Scott Polar Research Institute
University of Cambridge

Flipping the Switch:
The Transformation of Energy Markets
Submitted for a Ph.D.

S. A. Van Vactor
Scott Polar Research Institute
Darwin College
University of Cambridge

February 12, 2004

Table of Contents

Tables and Figures

Acknowledgements

I would like to thank my advisor Professor Peter Nolan and Dr. Piers Vitebsky for their many suggestions, help, and encouragement. Professors Richard Gordon and Arlon Tussing contributed immeasurably by providing scores of new research sources and enforcing rigorous thinking. Dr. Fred Pickel and David Ramberg provided helpful editing and suggestions that improved the research significantly. Professor David Newbery's seminars on regulation were an outstanding source from which to draw reliable information and research techniques. Any errors or omissions are, however, my responsibility.

1. *Introduction*

1.1 Goals of the Thesis

This thesis aims to describe and analyze the reasons for the remarkable transformation of a key sector of the world's economy. In the 1970s the Organization of Petroleum Exporting Countries (OPEC) emerged as a major force in world oil markets. The disturbances in the oil market and high prices set by the cartel caused one of the greatest economic dislocations of the twentieth century. OPEC's unchallenged control of the market lasted just under fourteen years, from 1973 to 1986. In that short period all of the important institutions of oil trading changed. The oil market transformation was followed by deregulation of the North American natural gas market and a worldwide movement to liberalize all energy markets, including electricity.

The approach taken to analyze these events is broadly captioned transaction cost economics. Institutional economists have used transaction cost analysis to understand the nature and scope of the firm. Financial economists have studied transaction costs to analyze the efficiency of trading mechanisms - "market microstructure" analysis. These schools of thought provide predictive tools for understanding firm size and scope and market structures. If it is costly to trade resources, firms will integrate and combine various activities into a single administrative unit. Likewise if trading volume is small and prices are volatile trading costs will be high and the market will be illiquid. The goal here is to gain a better understanding of these processes in the energy industries and explain why energy markets have changed so dramatically.

1.2 Energy Market Liberalization

In June 1979, heads-of-state from the six most powerful economies in the world met in Tokyo (General Accounting Office [GAO] 1980). This time the normal agenda of fiscal and monetary policy, exchange rates, trade barriers, and the like was smothered by concern about events in the spot oil market, where prices seemed to be out of control. A few years before, OPEC had asserted its long-latent market power during the Arab oil embargo and the price of crude oil had increased five-fold. Now the industrialized nations were faced with a new threat, stemming from their own rising demand for fuel. Companies that had been awash in Iranian oil before the revolution were suddenly desperately short and forced to join a long queue of buyers all shopping for the same scarce supply. By the time of the summit, spot oil prices had risen four-fold and it was only a matter of time until OPEC rubberstamped the market's voice as its own, in an attempt to sink the world into permanently high-cost energy.

Before 1979, the spot market for oil was little more than a balance wheel. Over 95% of international oil transfers took place within the integrated structure of Major oil companies. With rare exceptions the oil flowed smoothly; petroleum products were available on instant demand anywhere in the globe at reasonably stable prices. In contrast, critics asserted that OPEC was unsuited for the task it had seized. Instead of disciplined management with tight control of supply and marketing, the cartel careened from crisis to crisis and allowed volatile trade in a thin market to set the prices of the world's most important natural resource.

Despite its nefarious beginnings, this is the story of how a rogue market became the model for energy market liberalization around the world. Markets do not spring forth as finished works; they take time to mature. Efficient markets require a variety of institutions – information systems, trade associations and standards boards, government oversight, and knowledgeable participants in order to minimize operating costs and allow competition to flourish. The urge to compete may be potent, but the opportunity to do so is not. The barriers to effective competition are many - the consequence of poorly conceived regulations and government policies, misinformation, the blunders of poor managers, and the mindset of firms that seek to dominate a market. For many, the world oil market is now a model of efficiency, but it certainly had a rough start.

In the space of four years, from 1979 to 1983, crude oil acquisition shifted from a closed system of long-term contracts and internal company transfers to an open commodity market. The new market adopted a specific paradigm, one pioneered by the agricultural industry in the nineteenth century. Broadly, such markets are referred to as commodity markets and they include two distinct and interdependent components. The first and primary part of the market concerns the physical movement of the commodity. In this segment of activity, spot prices guide day-to-day decisions about immediate production, storage, and consumption. Trade occurs through bilateral negotiations and purchase contracts are adapted to the specifics of the buyer and seller. The second segment of the market establishes the relationship between today's spot prices and expectations about prices in the future through futures or forward trading. The stream of prices into the future provides longer-term signals governing investment decisions and allows producers and consumers to manage price risk. In agricultural, oil, gas, and many financial markets, exchange-based trading in futures contracts is the fundamental source of price discovery. Unlike bilateral trading, futures trading depends on establishing high-volume standardized contracts, because exchanges have high fixed costs. Successful futures trading in a commodity lowers the transaction costs of trading in all segments of the market by standardizing commodities, improving information, and lowering enforcement costs.

Writing before the transformation of energy markets, Houthakker (1959 pp. 156-157) lamented: "Viewed in this light, futures trading would seem to be one of those marvels that ought to be invented if they did not already exist. Yet the number of futures markets is surprisingly small…"

A crucial element of any successful commodity market is the presence of marketeers, or market makers, as they are known in securities markets. These are companies or individuals that act as brokers (connecting willing buyers and sellers) and that also buy for and sell from inventory. Their willingness to hold the commodity in the period of time between the arrival of an active buyer and seller substantially improves market liquidity and satisfies producers' and consumers' demand for immediacy. Even in the broader economy this is the most frequently occurring market structure; intermediaries buy from primary producers in wholesale markets, hold an inventory, and sell to consumers in retail markets. Although this structure is the most prevalent, it is commonly ignored in economic analysis, which focuses on demand and supply schedules. It is this intermediary segment of an industry that shoulders the bulk of transaction costs.

A study of transaction costs in the energy industry is particularly useful because energy prices are extremely volatile, which frequently results in regulation or price controls. Price volatility is high because energy commodities are **component products**. These are products in which both the buyer and the seller must invest in inflexible fixed assets for both use and production. Industry-agreed standards allow competition between producers, but nonetheless constrain consumer behavior in the short term. That is, owners of gasoline-powered cars cannot switch to coal, electricity, or compressed gas without replacing or modifying their fixed assets. Given the large investment consumers have already made and the cost of switching to an alternative, their demand schedule for a specific energy product is highly price inelastic in the short run.

1.3 Transaction Costs and the Study of Industries

Transaction costs and their study within the field of economics may be traced back to an early paper published by Ronald Coase (1937). Coase did not use the term "transaction costs," but he asked a question central to the study of economic institutions: If markets are the most efficient means to allocate resources, why do firms exist? This question cannot be evaded in any study of the energy industries, where many of the world's largest corporations operate. The elephantine size of oil, gas, and electric companies is not new; it has been a characteristic of the industry since its early foundation. Only two firms have appeared systematically in *Fortune's* list of the ten largest firms: Standard Oil of New Jersey (once the Standard Oil Trust and now ExxonMobil) and General Electric. Exxon, in its various guises, was founded by John D. Rockefeller and insiders believe it still reflects the personality of its founder. General Electric is now a conglomerate, but got its start (and still does much of its business) in electricity generators and related devices.

Conventional wisdom suggests that the reason for the large size of energy companies is economies of scale, particularly for energy supply projects, which can take years to develop. However, many energy companies are also vertically integrated, i.e., they operate in distinct segments of primary production, transportation, upgrading and retail marketing. This raises a key question: Are organizational structures better explained by scale economics or by transaction costs?

Since Coase's (1937) article, transaction costs have been categorized as costs associated with information, bargaining, measurement, and enforcement. One branch of transaction cost economics, identified with Oliver Williamson (1985) and Klein, Crawford, and Alchian (1978), has analyzed the impact of these costs, particularly bargaining and enforcement, on a company's organizational structure and contractual arrangements. Another branch, identified with Douglas North (1997), has studied changes in transaction costs over time, noting that a great deal of economic development can be explained by the advancement of institutions and markets that allow transaction costs to be reduced or eliminated.

Energy commodities are complex products compared to many other goods and services purchased by consumers. They are complex because they come in unusual forms – liquid, gas, or electrons – that often require unique packaging and handling. Energy products are frequently toxic and dangerous, which adds to the cost of changing title, particularly in a modern economy. Intuitively, a commodity's characteristics ought to impact transaction costs and therefore the industry's structure. This was the view of Paul Frankel (1969), who argued that the oil industry's natural order – vertical integration – was determined in large measure by the fact that key petroleum products were liquid and part of a continuous flow, which made "security of supply" of paramount importance to producers, refiners, and marketers.

When the concept of transaction cost was first developed it was immediately applied to the securities market. In his classic article, Harold Demsetz (1968) analyzed the cost of trading company shares. In this case, the definition of transaction cost was concise and easily quantified – the difference between the "bid" and "ask" price quoted by market makers on the floor of the exchange. Demsetz concluded that the level of transaction cost depended on the volume and frequency of trading. His analysis leads to the observation that transaction costs and liquidity are related concepts. That is, in markets with low liquidity, transaction costs are high and vice versa. A host of studies followed Demsetz's paper, but the analysis could not be extended to commodity markets, because futures exchanges do not keep bid and ask records.

Although there are no explicit records, transaction costs are highly relevant to futures trading. Futures exchanges, of course, prosper by offering futures contracts that attract traders. The majority of new contracts offered fail, i.e., they do not attract enough trading to cover the fixed and variable costs of trading the contract on an exchange. Deborah Black (1986) studied the launching of futures contracts and attempted to generalize why some succeeded and others failed. Some of the more important elements she discussed were the commodity's storability, its homogeneity, price volatility, the size of the market, and low delivery costs.

The studies by Frankel and Black suggest that different industrial structures might be explained, in part, by the physical characteristics of the commodity itself. That is, whether or not the commodity is homogeneous, divisible, transportable, durable, composed of hazardous material, has "non-rival" characteristics, or is a component requiring use with other products may affect industry configuration. Commodities vary significantly in their combination of characteristics. Thus, it should not be surprising that transaction costs and market institutions vary widely from one industry to another, and the more complex the commodity, the more complex the market.

1.4 Revolution and Counter-Revolution in the Oil Market

In many instances a distinction cannot be made between markets and the firms that operate them, as was the case with vertically integrated oil firms. At the peak of their popularity, the Major oil companies explored around the world for oil, discovered it, developed fields, built pipelines and tankers to move oil to refineries, refined products, branded them, and finally marketed the oil to consumers in their own retail outlets. All of this activity took place without revealing the market price of crude oil or intermediate products; the only visible prices were those charged to consumers in the retail market.

Speaking of works by Paul Frankel, Edith Penrose, Helmut Frank, Jack Hartshorn, and Morris Adelman, Robert Mabro (1992) lamented: "Thereafter, [post 1973] for reasons that are difficult to identify or comprehend, the rich stream of works on the industrial economy of oil dwindled to a trickle." The trickle is explained, in part, by the fact that oil firms themselves have become largely irrelevant in determining the price of the world's most important commodity. In 1973, the oil industry was little more than the firms themselves; a study of oil companies was a study of the oil market. That structure dissolved in the days surrounding the Arab oil embargo of 1973, as the power to set oil prices passed to the OPEC cartel.

OPEC's primary stewardship lasted thirteen years and ended in disarray when the crude oil market collapsed during the first few weeks of 1986. By the time of the cartel collapse, a wholly new structure for determining oil prices was firmly in place. The new structure had as its centerpiece a commodity exchange – a futures market where participants could hedge against price volatility and, as necessary, deliver the commodity. The New York Mercantile Exchange (NYMEX) started successfully trading heating oil futures in 1978 and extended futures trading to the crude oil market in 1983. The futures market was, however, only the most visible part of a much larger supporting market, where every type of crude oil all over the world was exchanged. The spot oil market, previously damned by consuming nations' heads-of-state, had become the instrument of OPEC's surrender. The spot market proved as effective in bringing prices down in 1986 as it had in driving them up in 1979.

By its nature the oil industry ought to have high transaction costs. Aside from the special problems of transportation and packaging identified by Frankel (1969), crude oil is highly heterogeneous. In nature, crude oil varies from thick tars to natural gasolines. In addition, most crude oils contain a variety of impurities: sulfur, waxes, heavy metals, etc. Each type of crude oil requires a slightly different type of refining process in order to produce the most profitable slate of petroleum products. Due to the diversity of crude oil, refineries are often designed to process specific crude-oil streams; substituting other types of crude oils results in lower volume runs or excessive production of low-valued products. This is the type of asset specificity that Williamson and Klein identified as having high transaction costs and, consequently, that provides a motive to integrate. Indeed, it may have been a factor explaining vertical integration through much of the twentieth century.

During the 1973-74 oil crisis, OPEC increased prices, but its members retained their relationships with oil refiners, replacing crude oil ownership rights with long-term contracts. They put in place a flexible pricing system that allowed each producing country to adjust prices in order to account for the diversity of crude oil quality. In 1979, however, myopia swamped their decision-making as most OPEC members took advantage of the shortage caused by the Iranian revolution by reneging on contracts and shifting supplies to the spot market where they could get much higher prices. Their opportunistic behavior upset the economics of vertical integration and resulted in a significant increase in the transaction costs of allocating crude oil to refiners.

Coase's theory implies that the sudden shock of high transaction costs would change the industrial structure. The first reaction of the oil industry was predictable; having lost assured access to OPEC oil reserves, the Majors set about replacing them with more secure sources in North America, the North Sea, etc. One consequence was a wave of consolidation in the industry: Standard Oil of California acquired Gulf; Texaco acquired Getty, etc. During this period, however, oil prices were declining, which sharply reduced the incentive to acquire high-cost oil reserves.

A simple return to vertical integration with non-OPEC resources was not economically feasible. Likewise, short of capital and understaffed, OPEC national oil companies did not integrate into the downstream sector. Refiners learned that they would have to depend on unstable OPEC supplies. The market adapted to the new reality and discovered ways to reduce the transaction costs of trade. This was achieved through modest changes in the size or structure of oil companies and through the introduction of new market institutions – better information systems, commodity exchanges, streamlined contacts and arbitration, reduced bargaining time, etc.

1.5 Evolution in the Gas Market

While there are many similarities in oil and gas trading in today's market, there are also differences. Fundamentally, the natural gas industry depends on a network of interconnected pipelines to connect producers and consumers and create a market. Unlike the oil industry, the North American gas industry was not vertically integrated; it was segmented into three wholly separate tiers – producers, transmission pipeline, and distribution companies. Because the producing sector is highly fragmented, it is naturally competitive, but under the regulatory system it had little or no access to a market. Prior to deregulation, transmission pipelines provided two fundamental services: the transportation of gas and the so-called "merchant function." Transmission pipelines purchased gas from producers and resold it to distribution companies, mostly under long-term contracts. Implicit transmission rates were calculated from the average price differentials. Deregulation of the North American gas industry amounted to eliminating the pipeline's merchant function and replacing it with regulated transportation rates guaranteeing third party access and, at the same time, allowing large customers to buy directly from producers.

During the period leading up to the oil market collapse in 1986, the North American natural gas industry underwent a radical change. The gas market, steeped in four decades of mummifying regulation, was grossly inefficient, with prices received by some producers ten to twenty times higher than those received by others. Inefficient pricing at the wellhead created a patchwork of simultaneous shortages and surpluses. At the same time, redundant capacity in transmission and distribution pipelines, along with cross-subsidization of residential consumers by large users, jacked up cost-based transportation rates. Gas producers and large consumers sought to free themselves from the quagmire by "buying direct," bypassing over-priced pipelines and distribution companies.

The process of gas market deregulation took over a decade, culminating in 1992 with Federal Energy Regulatory Commission (FERC) Order 636. At the close of their book on the U.S. natural gas industry Tussing and Tippee (1995) commented: "It is hard to think of an established industry at any time in any country whose market structure and dynamics, intellectual environment, and business culture changed as radically within a single generation as the natural gas industry of North America."

The success of the North American natural gas deregulation rested to some extent on the shoulders of the newly liberalized oil market. Gas and oil are connected through the producing sector of the industry. In exploring for oil, companies frequently find gas. In the early years, gas was a waste product, but as the marketability of gas expanded, exploration companies began looking for it in its own right. As the gas market matured, conventions and institutions used in trading for oil were adapted to the gas industry.

The complexity of the regulatory framework and the limited market experience of gas producers and distribution companies led to the rapid expansion of middlemen, known as marketers. These companies acted as brokers between buyers and sellers; in addition, they were willing to take title to the gas, hold inventory, and act as market makers. Natural Gas Clearinghouse (now Dynegy) started as a pure marketing company, but during its growth it acquired a variety of energy assets. Enron started as a pipeline company (some would call it a "pipe dream" company), but as it expanded it ignored its assets and, instead, focused on trading and the knowledge base of its managers, creating what they called "virtual" assets.

The maturity of North America's gas pipeline system helps explain why a competitive market could emerge despite the barriers created by pipeline economies of scale. In this system, pipelines do not necessarily run directly from a producing field to consuming centers. Over the years production from various regions has waxed and waned; at the same time, consuming centers have shifted as new customers were connected to the grid and others shut down. To accommodate shifting demand and supply, the pipeline system has been extensively interconnected. The consequence is a system that frequently has spare capacity. In most cases multiple pipeline routes can be used to ship gas. Even though transmission pipelines are regulated, setting rates and guaranteeing access, there is also a measure of competition between pipeline companies.

As the gas market developed, it centered at hubs and market centers – the points in the pipeline grid where there is significant interconnection. These became pricing points where blocks of gas were bought and sold by producers, consumers and middlemen. As trading at each of the hubs increased, transaction costs declined, liquidity increased, and a rational ordering of natural gas prices emerged. Typically producing areas such as Western Canada and the Gulf Coast have the lowest gas prices. Consuming areas remote from producing areas, such as New England, typically have the highest prices.

Enron and the many companies that attempted to emulate it brought something quite new to energy trading, something that had been largely unnecessary in oil trading. Enron provided a package of services – the purchase and sale of the commodity and the risk management tools necessary to moderate price volatility. Many of these tools were adapted from the financial industry, whose traders populated the early Enron trading floor. Unlike the oil market, natural gas had been commonly sold on long-term fixed price contracts (long-term contracts in the oil industry were based on posted prices that were flexible and shifted with market conditions; moreover, although OPEC had largely dropped out of the day-to-day management of oil prices, it did act to moderate market extremes; the cartel's key countries maintained sufficient spare producing capacity to prevent oil prices from spiking due to unexpected shifts in demand). Gas storage is more costly than oil or coal storage. Consequently, gas prices are more volatile and distribution companies (with regulated retail rates) are particularly vulnerable to such volatility. Thus, Enron's package of gas supplies, with reliable fixed prices, was particularly attractive.

The growth of the North American spot gas market was accompanied and cultivated by the development of futures trading. In 1990, NYMEX launched a contract in natural gas trading based at Henry Hub in Louisiana. The structure developed to serve the oil market was thus quickly adapted to natural gas. Henry Hub proved to be a brilliant choice. The extensive pipeline interconnection and proximity to Gulf gas deposits created a highly liquid market - one of the more successful commodity markets in the world.

Following the success of U.S. natural gas deregulation, the U.K. privatized its gas company, creating BG Transco, a national transmission company, with regulated access to the system for producers, distribution companies and large consumers. The International Petroleum Exchange (IPE) operates a successful futures exchange for natural gas. The U.K. gas market is connected to continental Europe by the Interconnector pipeline, which to some extent integrates the two markets.

Gas market liberalization in Europe is proceeding under the European Union (EU) Natural Gas Directive. Historically, natural gas prices were indexed to petroleum product prices. The experience in North America demonstrated, however, that the gas market can and should seek its own equilibrium level. On the one hand, Europe has good supply diversity that could lead to an efficient market. Gas supplies come from the extremities – the North Sea, Siberia and Northern Africa - external sources that are naturally diversified. On the other hand, these supply regions are unlike North America. Except for the North Sea, natural gas supplies are controlled by governments and/or national oil companies. Further, dependence on Russian gas is expected to increase. As a consequence of these gas supply sources, Europe may not be able to develop the same type of market as that in North America.

As noted, gas trading is connected to oil trading by companies that produce both oil and gas. Similarly, gas and electricity trading are cross-connected by gas-fired generators. Companies that operate such generators typically buy and sell both gas and power. Almost all marketing firms trade both commodities. Moreover, there is often a close connection between power and gas pricing, since the cost of converting natural gas to electricity is usually predictable.

1.6 Breakup of the Power Industry

Unlike the natural gas industry, the electricity industry has strongly favored vertical integration. Historically this was due in part to economies of scale in generation, which led to integrated transmission grids, which also have economies of scale. Transaction costs have also played a role in provoking vertical integration, and since electricity cannot be stored cheaply, the grid must be constantly balanced or it will collapse. This requires minute-to-minute centralized control over suppliers and/or buyers. As a consequence of these characteristics, most electricity supply systems around the world have been publicly owned or privately owned and regulated as natural monopolies.

During the energy crisis of the 1970s, a number of analysts (Lovins 1976) argued that the industry's economies of scale had reversed, meaning that smaller localized units could displace large central station generators. This led to a change in federal law in the U.S., encouraging the entry of independent power producers (IPPs). These generators were, as Lovins and others had predicted, smaller in scale, but more importantly they were independently owned. Intitially this made little or no difference to the industry's structure, because to sell power, IPPs had to contract to utilities. Over time, however, development of independent suppliers created the momentum for a market where power could be bought and sold by a larger number of participants.

In general, electricity pricing in regulated utilities was highly inefficient. Although precise regulatory systems varied, virtually all were cost based with little or no sensitivity to daily and seasonal demand peaks or supply emergencies. Reforms were proposed that would shift retail and wholesale power pricing to accommodate known swings in demand and supply (Turvey and Anderson 1977). Such pricing was not efficient, however, because the single largest cause of a shift in demand is weather, and it is not predictable far enough ahead to set regulated prices. Analysts (Schweppe et al 1980) began to argue that it was feasible to develop a spot power market that would accommodate hourly shifts in demand and supply and create efficient pricing.

The U.S. experimented with market-based pricing of electricity in the Western Systems Power Pool (WSPP), which FERC created in 1987 (Lehr and Van Vactor 1997). This was, however, a very limited wholesale market. Essentially, it was trade between integrated utilities, each of which controlled an operating area. The transmission system in the West, however, was diverse enough to allow the dozen or so companies to carry on active competitive trade. This market developed a sales contract, product definition, and other devices that lowered transaction costs and increased liquidity. The WSPP became the model for bilateral trading in North America and to some extent in Europe.

While a power market was evolving in the U.S., the U.K. chose to create a whole new one in a single stroke. The U.K. had been saddled with a highly inefficient industry, left over from post-war nationalization. As time wore on original agendas changed, and the administration of the electricity sector became little more than a vehicle for subsidizing the coal and nuclear industries. The Thatcher government sought to privatize the electricity sector, but unlike the earlier privatizations of British Telephone and the gas industry, they sought to create the semblance of a competitive market from the onset. Many models were considered, but it was finally determined that the industry should initially be divided into three sectors: power generation, transmission, and distribution. Transmission and distribution would require regulation, but a competitive market could set wholesale prices for generated power as sold into a common pool. Fear of coordinating difficulties caused the government to create an effective duopoly rather than a diversified supply sector. However, the door to competition was left open, because barriers to entry were more or less eliminated. This turned out to be quite significant, because new gas-fired generators could produce power at lower cost than many existing facilities.

Norway began with a step-by-step approach to deregulation in 1991, rather than a comprehensive restructuring of the industry as in the U.K. In 1993 Nord Pool opened, which has since broadened to include Sweden, Denmark, and Finland. Unlike the U.K., the Norwegian power sector is dominated by hydroelectric power, where capital costs are large, but variable costs are negligible. Because reservoir inventories must stretch over multiple seasons facing the risk of a drought, planning is very different from that of the U.K., where marginal costs are set by the cost of coal, oil, or natural gas in a relatively predictable fashion. One consequence was Nord Pool's early development of a system that integrated futures, forward, and spot pricing. In contrast, the system in the U.K. was built primarily around half-hourly spot pricing. Longer-term contracts were determined in the Over-the-Counter (OTC) market as contracts for differences against pool prices, and were thus not as transparent as long-term contracts in Nord Pool.

Experiences in the U.K., Nord Pool, and various power pools in North America were demonstrating the complexity of liberalizing electricity markets. The need to continuously balance the grid was difficult enough, but in the presence of congestion, efficient pricing required a complex melding of bidding and system operator adjustments. In an integrated utility, electrical engineers familiar with the system could solve these problems through centralized dispatch, based on their specialized knowledge of the "merit order" of the generating options. In a market setting, however, the role of the system operator became the focus of an intense debate with ideological overtones, especially in California. Advocates of the pool approach believed that the system operator should manage the transmission grid and operate a market, which ordered dispatch and set prices in each zone of the grid. Opponents viewed such a role as akin to central planning, at odds with the whole idea of decentralized decision-making.

In the mid-1990s, as California debated the future for its power industry, Daniel Fessler, the President of the California Public Utility Commission (CPUC), visited the U.K., toured the U.K. power pool, and was enthralled by its concept. Under the CPUC's leadership, California set about completely restructuring its industry along the lines of the U.K. and Norwegian model. Unlike the U.K., however, California's important power utilities were privately owned. The State had three such utilities, each with its own franchised operating area, generating units, supply contracts, etc. In addition to the three principal utilities, there was a menagerie of publicly owned utilities and a few small private utilities. Most importantly, California remained part of the integrated western transmission system and was dependent on power imports from adjoining states, particularly surplus hydroelectric generation from the Pacific Northwest. When the entire West suffered a severe drought in 2000 and 2001, California's restructured market collapsed in financial disarray.

Rightly or wrongly, the California experiment has caused electricity restructuring to stall in the U.S. and in other countries. It has also revealed the substantial transaction costs frequently hidden in the electricity industry. It cost nearly $1 billion to set up California's power market and grid management system. The collapse of the market cost California consumers over $10 billion, and it also resulted in the bankruptcy of Pacific Gas & Electric as well as a number of energy suppliers. On the other hand, the California crisis may have been partially responsible for developing futures trading and a viable forward market. Such developments are essential to the efficient workings of a commodity market.

1.7 Determinants of Transaction Costs in Energy Trading

Energy markets are an excellent basis by which to study the interaction of commodity characteristics, transaction costs, and market institutions. This is because key characteristics of energy commodities vary enormously. Oil is a liquid that can be packaged in everything from gallon jugs to super tankers at atmospheric pressure. Consequently, oil can be stored and it can be bought in large and small lots. Crude oil in its natural state, however, varies widely in density and the number and types of contaminants, thus there are measurement and quality control issues associated with its transport and marketing.

Natural gas, on the other hand, requires much more costly packaging. It is most commonly collected and distributed by a network system of pipelines or liquefied through a costly process for ocean transport. Although natural gas varies somewhat in quality in its natural state, it can be standardized at reasonable cost. High transaction costs in natural gas trading arise from the inflexibilities in the pipeline delivery system that constrain arbitrage, limiting liquidity at key nodes in the infrastructure.

So far, electricity markets have proven to be the most costly and difficult to establish. Electricity has distribution constraints like natural gas, but, as explained, faces additional complications. One consequence is that electricity markets remain relatively immature and are particularly prone to price instability.

A detailed study of the costs of bilateral spot trading in heating oil, gasoline, crude oil, and natural gas in the North American wholesale market is made in Chapter 7. Trading costs are measured at 49 market centers in various regions by subtracting "sell" prices from "buy" prices. This measure of transaction costs excludes elements such as information gathering and enforcement, but includes implicit brokerage fees and the value of "immediacy." A similar measure has been used in other studies to analyze the trading costs of buying and selling securities in stock exchanges.

The classical linear regression model is used to explain the cost difference between trading natural gas and oil. The estimation determines that there are significant cost differences between the two types of fuels and that "basis risk" is also a strong determinant of transaction cost. Basis is the difference between the price of a commodity that arises due to differences in location or quality and the "marker" price commonly used as a general index of price movements. If basis is predictable, transaction costs will be lower because price determination will be more transparent and liquidity will be higher, lowering information and search costs for the regional market in question. Low basis risk also allows futures or forward contracts (tied to the marker price) to be used to hedge, reducing the risk of holding positions and broadening market participation.

Comparable data on buy and sell prices are not available for electricity trading in North America. Other trading data, however, suggest that electricity has a much higher price volatility and basis risk then either oil or gas. In addition, electricity decays immediately and the grid must remain in constant equilibrium. These characteristics lead to the inference that the transaction costs of trading electricity are higher than those for either oil or gas. Data on bid and ask prices from the day-ahead market in the U.K. confirm this relationship. In the U.K., the cost of trading electricity is two and one-half times higher than natural gas.

1.8 Statement of Thesis and Brief Summary of Conclusions

In a period of just over two decades, the petroleum, natural gas, and electricity industries have undergone a revolutionary transformation as a consequence of changing economics and regulatory oversight. These industries have frequently been analyzed with respect to their cost structure, resources, competitiveness, and political impact. Transaction cost economics, however, has seldom been applied to the energy industries. When examined, it is a useful tool for understanding both why the energy industries have undergone such a radical transformation and why a traditional commodity market structure has been adopted in the oil and gas markets, but not, so far, in the electricity industry.

Ronald Coase (1937) opened a continuing debate on why firms exist and its corollary, what costs are associated with using the price mechanism. Since Coase's article, a large volume of economic literature on transaction costs has blossomed. Most of the academic research has been aimed at the costs *firms* avoid when they integrate in order to avoid trading resources at arm's length. In general, the literature on transaction costs has ignored an equally important question: what impact do transaction costs have on the type of *market institutions external to firms* that are adopted by the industry? This is an important question because markets come in a multitude of forms, from the traditional agricultural exchanges of the nineteenth century to modern commodity exchanges, bazaars, retail stores, etc. And, of course, it is a key issue underlying the transformation of energy markets.

With respect to the oil and gas industries, there is little evidence that the problem of specific assets is a significant determinant of firm size and organizational structure. Over the last two decades the oil industry has undergone significant changes in ownership rights and the cost of trading, but only modest changes in firm structure. The degree of vertical integration among the Major oil companies has been reduced and they have become more specialized, but the fundamental organization has changed little. Inertia, economies of scale and risk aversion, rather than transaction costs, appears to be the principal motives that explain why Major oil companies retain at least the outline of a vertically integrated structure, where the companies operate in each important sector but do not feel compelled to balance refinery capacity with crude oil production. Similar observations can be made about the gas industry. Economies of scope in oil and gas exploration can explain the integration of the oil- and gas- producing sectors. Despite a decade of full deregulation, gas producers in North America have not integrated with distribution utilities. In contrast, power producers and electricity distribution companies in the U.K. are merging, creating a vertically integrated structure. The motives for these mergers appear to be related the high cost of trading electricity, risk aversion, and the fact that retail prices in distribution companies have not fallen as far as wholesale prices. Not surprisingly, vertical integration has reduced the volume of trading and liquidity, so that the organizational structure becomes a self-fulfilling prophecy.

Recent mergers in the U.K. electricity sector, the failure of electricity futures trading to develop, and the huge cost borne by California in its power market restructuring demonstrate that, so far, the transaction costs of trading electricity are higher than either oil or natural gas. This is explained in part by the complexities of spot power trading and grid management that arise from the fact that electricity cannot be economically stored, the grid must be balanced, and that transmission costs are unpredictable.

While the impact of transaction costs on firm size and structure remains debatable, their impact on market institutions is clearly evident. The breakup of vertical integration in the oil industry by OPEC and collapse of long-term sales contracts in the North American gas industry precipitated the development of a whole new set of market institutions. Futures exchanges, marketers, risk management services, trading floors, standardized contracts, trade press price reporting, etc. are all part of the market transformation.

The fundamental characteristics of a commodity do matter in setting transaction costs and, thus, impact industry structure. In North America, both oil and gas commodity markets are mature, with comparable levels of trading and liquidity. Nonetheless, natural gas trading has systematically higher trading costs – the difference between bid and ask prices – than does oil. This is explained primarily by the fact that gas must be delivered in a fixed network infrastructure, while oil has a much more flexible transportation system. Defining transaction costs as the difference between bid and ask prices, as is done in the financial industry, has a distinct advantage in that it allows quantitative analysis aimed at identifying the causation of transaction costs and how these costs change over time. It also reveals an important parallelism in two concepts: liquidity and transaction costs. Although these concepts are frequently treated separately they are conceptually similar; low transaction costs are the hallmark of a liquid market, just as high transaction costs reflect an illiquid market.

2. Energy Market Liberalization

2.1 The Movement to Liberalize Energy Markets

The movement to liberalize energy markets developed primarily from two separate sets of experiences: one in the U.S. and one in the U.K. During the 1980s, the U.S. abandoned its heavy-handed regulation of the natural gas market and replaced it with a sophisticated commodity market. The regulatory shift prompted the industry to reorganize gas sales, transmission, and procurement into an open wholesale market that depended on competition rather than regulation to set prices. Buoyed by the successful liberalization of its gas market, the U.S. has since sought in fits and starts to restructure its electricity industry with less success. Contemporaneously, the U.K. embarked on a steady path to privatize its nationalized industries - notably rail, telecommunications, natural gas, and electricity - and implemented regulatory schemes that relied as much as possible on competition.

The policy shifts in the U.S. and U.K. cannot be characterized as unchallengeable successes, although on balance, most observers view the changes as improvements. The multiple options governing how privatization and/or market restructuring might proceed have created a policy dilemma for many governments. State-owned enterprises often have a substantial asset value – value that is enhanced if the asset is sold as a monopoly. Carving the industry up and creating a competitive market often reduces expected proceeds from privatization. In its initial restructuring plan for electricity, the U.K. sought a compromise approach and created an effective duopoly, which protected proceeds from privatization but may have resulted in inefficient electricity pricing. Likewise, for different reasons, the California experiment with electricity restructuring proved disastrous, both for the state and for industry. Elsewhere in North America, electricity restructuring has been more successful.

Along with the wave of restructuring and privatization has come a new utility model, fashioned primarily around the experience of the world oil market and the North American natural gas market. It has been generally accepted that the transmission infrastructure (pipelines and wires) should remain either publicly owned or privately owned and regulated, since transmission is a natural monopoly. On the other hand, gas producers or power generators can effectively compete. One reason for this change is a reversal of economies of scale in supply projects. A battery of small companies can produce onshore natural gas cost-effectively, just as power can now be cheaply generated in relatively small facilities. These changes make it feasible to foster competitive wholesale markets for gas and electricity, although the precise form these markets should assume remains debatable.

In developed economies, policy makers have taken steps to open markets and sell state-owned assets. In Europe, the EU leads the initiative aiming to build a continuous market across Europe; liberalizing policies are also being adopted by other advanced economies such as Japan. The policy initiatives are overseen and to some extent coordinated by the International Energy Agency (IEA) and the Organization for Economic Cooperation and Development (OECD). In reviewing the development of electricity market restructuring, the IEA (2001 p. 7) observed: "Reformed markets require regulators to perform new tasks, such as ensuring open access to the electricity network, protecting the ability of consumers to choose their supplier, and enforcing antitrust laws." The need for regulatory oversight to change (or to be developed for the first time) is generally recognized in the advanced economies, but is often ignored in less-developed countries.

In North America, market liberalization of the energy sector has been led by regulatory changes in the gas industry. In Europe it has been the reverse: gas market liberalization has followed electricity. In part, this reflected the pattern in the U.K., but it also reflected a greater degree of maturity on the part of the electricity industry. According to Jonathan Stern (1998 p. 91): "the electricity industry was prepared to concede that liberalization could bring efficiencies and advantages to the industry…By contrast, many of the established actors in European gas industries still regarded the introduction of liberalization as the equivalent of the end of civilization." Fundamentally, however, gas supplies are more concentrated in regions external to the EU – Russia and Northern Africa, making the outcome of liberalization risky for consumers in the EU, since competition cannot be assured. James Chalker (2003) reports that the EU expects its dependence on imported gas to rise from 40% to 70% over the next twenty to thirty years. Moreover, dependence on Russian gas will continue to grow, reaching 60% of EU imports. Russia's giant Gazprom controls the vast majority of Russia's production and all of its effective export capacity to Europe. Unless expectations change, these facts could further delay gas market liberalization.

The EU's Council adopted a directive on the internal market for electricity in December 1996 (IEA 2001 p. 36) and one on the gas market about a year later. The electricity directive gave member countries just over three years to implement the program, which mandated that electricity markets must be opened to competition for larger customers – 30% of the market by 2000 and 35% by 2003. In addition, third party access to the grid must be assured by establishing independent transmission service operators. The gas directive takes a similar course, mandating third party access for large consumers. However, Stern (1998) notes, "the resulting Directive is a relatively weak document in comparison with the hopes and aspirations of the late 1980s." He concludes that this is due to opposition from the gas companies, but the supply dependence on Russia must have been a factor as well.

Both the EU's Electricity and Gas Directives are loose knit documents, intended to provide guidance to member countries. This is necessary due to the diversity of the European energy infrastructure, its companies, and its regulatory procedures. The consequence has been a variety of gas and power market structures with some success and occasional failure. Nonetheless, energy market liberalization is proving beneficial and the trend is unlikely to be reversed, despite the setbacks arising from the collapse of Enron and other energy trading companies.

The consequence of energy privatization and market liberalization has been much more important outside the front-rank economies, because many less developed countries were desperate to raise cash. Here the dilemma between maximizing sales value and economic efficiency is really striking. According to the World Bank (2003 pp. 7-8):

> Between 1990 and 2001, 132 developing and transition economies (DTEs) have taken substantive steps to introduce private participation in infrastructure sectors. During that period, over U.S.$750 billion of investment in infrastructure projects with private participation took place in the DTEs—in the form of divestitures, green field projects, and management and operational contracts with major capital expenditures.

Included in these figures are telecommunications and other utility-style investments in non-energy infrastructure, but the energy share is still the substantial majority of asset value.

The World Bank (2003 p. 9) recognizes a tradeoff between the dangers of ineffective state management and the potential distortions arising from monopoly pricing, noting: "In many countries, inefficient public enterprise, especially in the infrastructure sectors, were draining the state budgets, diverting resources from other social priorities such as health and education, undermining the banking sector and impeding the development of the private sector." Nonetheless, the immediate motive for most of the privatization, particularly in South America, was simply to supplement government revenue. In many cases that resulted in inefficient redistribution.

Other questions need to be raised about the DTEs' uncritical adoption of regulatory schemes from North America and Europe, and South Africa is a good case in point.

In 1995 the World Bank sought to encourage the development of the Pande gas field in Mozambique. Mozambique is an exceptionally poor country that had been ravaged by civil war. While there was no demand for the gas in Mozambique, South Africa's mining and industrial core was only 400 km to the south. Gas from the Pande field could be used to offset the higher cost of South African coal gas to the advantage of both countries. Enron proposed to construct a pipeline from Mozambique to Johannesburg and, with the World Bank's support, lobbied the South African government to create third party access to existing gas pipelines with regulated rates, which would connect to about 600 industrial consumers. At the time the existing pipeline system was an unregulated monopoly owned by South Africa's national energy company, Sasol.

At first glance the proposal to liberalize the existing gas market appeared progressive, providing direct benefits to the two countries and improving economic efficiency. Closer examination, however, revealed some serious problems. First, coal gas and natural gas were mutually exclusive fuels. Coal gas has about half the thermal content of natural gas and cannot be mixed with methane. Most importantly, however, granting third party access to the pipeline would not, in this case, create a competitive market, because there was only a single supplier in Mozambique. The only impact of the regulatory shift would be to shift monopoly power from South Africa's national energy company to Mozambique's national energy company, Empresa Nacional de Hidrocarbonetos de Mozambique. This would certainly advantage Mozambique, but not necessarily South Africa. The core point hardly needs stating: regulatory and administrative changes to liberalize energy markets by opening access to transmission infrastructure do not create competitive markets if either supply or demand is concentrated in the hands of a few.

Thus, Europe may discover that it needs the buying power of Ruhr Gas to counter the selling power of Gazprom.

The calamity in California, and the contraction of trading volumes and liquidity in North American and European energy markets in 2002, has also raised uneasy questions about the stampede into privatization. The debate is healthy and resolution of the policy dilemmas will take years to unfold. Nonetheless, there is much to be learned from what has already transpired, which is the motive behind this study.

2.2 Commodities and Their Market Structures

This section sets out to provide the reader a brief sketch of key terms and concepts in commodity trading as it applies to the energy industry. For those already familiar with commodity markets, trading terms, and procedures, this section can be skipped.

Commodity-style trading in the oil market began during the Iranian Revolution, when the OPEC system of contracting broke down. The modern energy market evolved over two decades and as it matured, new institutions arose and the trading structure spread from oil to natural gas and then to electricity. The institutions and structure of energy trading grew out of existing trading arrangements, from other commodity markets (mainly agriculture and metals), and from the financial industry. Today's oil, gas, and electricity markets are highly sophisticated and complex, although most of the concepts have roots back to the nineteenth century.

Despite the extraordinary success of energy futures markets at NYMEX and IPE, energy contracts are not the most widely traded. Derivatives trading in financial markets now set the pace. Table 2.1 contains the latest data from the Futures Industry Association on the volume of futures and options contracts. 92% of the contracts traded are financial; 3% are energy, and the original commodity markets of metals and agricultural products total just over 3%.

Energy markets have long made a distinction between **spot** and **term** contracts. A spot deal is described as a short term, "one-off" transaction between a buyer and a seller. On the other hand, a term contract connotes a continuing relationship between the buyer and the seller. Long before OPEC asserted control over crude oil pricing, oil companies in North America posted prices for crude oil purchased at the wellhead. (OPEC attempted to emulate this system through the "official prices" that its members set for each of their crude oils). Posting in North America continues to this day, but prices are now revised multiple times a month based mainly on movements in the NYMEX price, adjusted for differences in quality and location. Posted prices are used principally as a basis for long-term contracts and for royalty and tax computations. The system allows an oil producer to enter into a term contract with a buyer and retain flexible prices (frequently necessary when pipelines or other inflexible infrastructure must be installed). Once a contract is entered into, there is not much, if any, price negotiation. Instead the counter party depends on the good faith of the posting company.

Table 2.1
Volume of Contracts (in Millions)

Global Futures and Options	2003	Percent
Equity Indices	3,845.8	50.3%
Interest Rate	1,912.7	25.0%
Individual Equities	1,275.9	16.7%
Energy Products	226.3	3.0%
Ag Commodities	168.8	2.2%
NonPrecious Metals	82.1	1.1%
Foreign Currency/Index	73.0	1.0%
Precious Metals	65.0	0.8%
Other	0.7	0.0%
TOTAL	**7,650.3**	**100.0%**

Source: Futures Industry Association
First six months data at annual rate

Buyers and sellers might also enter into term contracts where the price of the commodity is fixed over the entire length of the contract. Fixed price term contracts were rare in the oil market, but they were the standard in natural gas and electricity markets. This difference can be explained by the greater flexibility in oil sales and deliveries. Natural gas and electricity have fixed transportation systems, which lock specific buyers to specific sellers. Williams (1985 pp. 178-180) refers to these types of assets as "specific assets," the presence of which likely lead to long-term enforceable contracts or integration of the two firms. This concept will be developed and critiqued more fully in subsequent chapters.

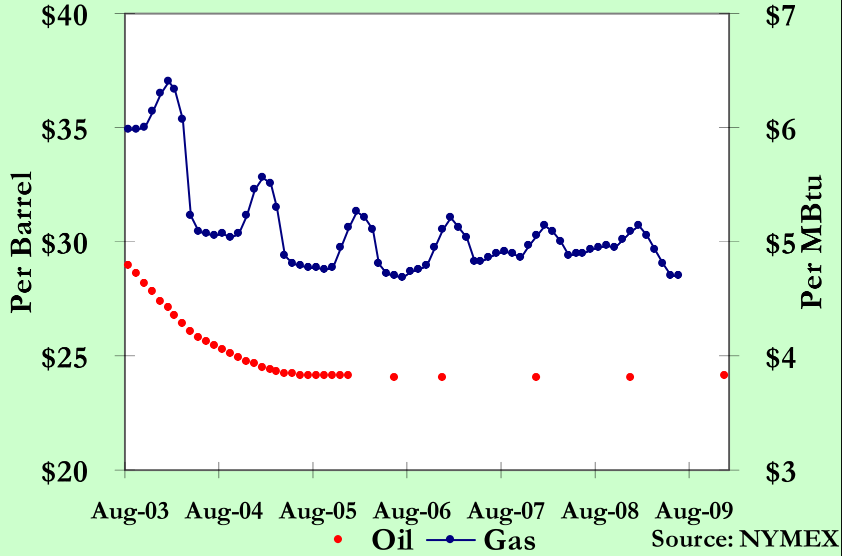
The notion of spot and term prices also underlies trading in a futures exchange, although the arrangement is more formal and there are crucial financial differences. Typically, the exchange offers trading in a series of forthcoming months. Unlike a term contract, however, prices are set individually for each month in daily trading. These prices are **forward** prices; that is, they are prices set at the time of negotiation for delivery at a future time. The main difference between forward prices and **futures** prices is standardization and liquidity.[1] The relationship between a series of forward prices is referred to as the **term structure**; traders also refer to a set of forward or futures prices as the **forward strip**. The price set for the first month being traded is referred to as the **prompt price**. Prompt prices correspond to published spot prices. Spot trading in crude oil was not very well defined before futures trading began. As the futures market matured, the industry adopted contractual conventions that matched delivery conditions of the futures contract. Spot oil prices now reported by the Trade Press reflect the price of oil agreed to today for delivery in the next (prompt) month. In contrast, spot power

[1] Carlton (1984 pp. 241) notes: "What is the main advantage of using a futures market instead of a forward market? The main advantage is liquidity." Along with that liquidity comes objective valuation; a significant issue for energy companies in the last few years.

prices are normally determined for next-day deliveries. The difference reflects the difference in commodity characteristics and transportation systems.

Figure 2.1

June 2003 NYMEX Gas and Oil Futures Prices



Determinants of the shape of a forward strip (whether prices rise or fall) are complex. Generally, forward prices reflect expected future spot prices (Stoft 2002 pp. 90). That is, when entering into a long-term agreement neither the buyer nor seller will agree to a price that is much different than their expectation of what they could, on average, buy or sell the commodity for each day. The cost of storing a commodity combined with expected production costs is what links the price in the first month to forward prices in the remaining strip. Gold has low storage costs, so spot prices are seldom much different from forward prices. Agricultural prices often have a seasonal cycle reflecting expectations about crop yields and the cost of storage. The natural gas forward price strip has a strong seasonal component, reflecting high storage costs, fixed production capacity, and strong winter demand.

When forward prices are progressively lower in future months, traders refer to such circumstances as **backwardization**. This pricing structure might, for example, arise before harvest time, when the forthcoming yield is known to be high and prices can be expected to decline. The alternative, **contango**, refers to market conditions where forward prices are progressively higher than spot prices. Figure 2.1 illustrates the forward price curves of oil and natural gas in June 2003. Crude oil prices exhibit strong backwardization, reflecting an expected price decline since Iraqi oil production was expected to come back on line. Natural gas prices reflect backwardization and a strong seasonality.

Commodity and securities markets operate what is known as **continuous trade,** i.e., during the hours the exchange is open, buyers and sellers may trade assets more or less instantaneously. With continuous trading in an exchange, prices change throughout the day, with special attention paid to the closing price. In contrast, many emerging electricity markets have been designed as once-a-day or once-an-hour **auctions**. The auction apparatus is usually employed by the grid system operator as a means to determine prices, balance the grid, and resolve transmission congestion.

One of the most important concepts in commodity markets is the distinction between **physicals** and **financials**. Futures contracts are designed as financial instruments and are not intended to be a primary source for procuring a commodity in its physical form. Typically, less than one percent of futures trading contracts are ever delivered. The primary demand for futures contracts is for **hedging**. For example, a buyer depending on spot purchases might reduce price risk by entering into companion futures contracts. If the spot price of the commodity rises, purchasing costs will rise, but this will be offset by a corresponding increase in the value of the futures contract. Likewise if spot prices fall and it costs less to buy the commodity, the gain will be offset by a parallel loss on the futures contract.

Derivatives are other types of financials. A derivative is a financial contract that derives its value from the price and availability of another commodity. The energy industry uses two types: standardized futures and option contracts purchased on a futures exchange such as NYMEX, and **over-the-counter (OTC)** derivatives tailored to the specific needs of the buyer and seller and offered by a third party - usually a financial institution. The OTC operates through **bilateral** negotiations - that is, through a web of personal contacts, telephone negotiations, and confirming faxes. In the OTC market there are a variety of derivative types; the most common are **options** and **swaps**. A swap is effectively the same as a futures contract but traded in the OTC; it is normally a temporal or locational **Contract for Difference** (**CfD**). For example, a local distribution company might contract for gas with prices indexed to their region as published by the Trade Press. In order to ensure stable prices for the next year, the company could enter into a separate CfD or swap with a financial company. If spot prices are above the marker in the contract, the distribution company would receive a payment equal the difference; if spot prices are below, they would pay the financial company. Options are traded on exchanges and in the OTC. Options provide the buyer the right (but not the obligation) to buy a commodity at a fixed strike price (a **call**), or the right (but not the obligation) to sell (a **put**).

OTC derivatives can be used to hedge either commodity prices or **basis**. Basis represents the difference between the price of one commodity and another. The price difference may arise due to geographic or quality differences. **Basis risk** arises when the commodity in question deviates in unpredictable ways from the **marker price** used by the futures exchange or in OTC derivative contracts. A marker price is the price of a commonly traded form of the commodity that is widely watched by the industry. When OPEC dominated prices, the marker was Arabian Light crude oil. When futures exchanges became the principal source of price information, the marker switched to West Texas Intermediate (WTI) in North America and Brent in Europe.

Generally, the OTC market refers to financial contracts necessary to manage price risk. The physical procurement of the commodity is supposed to take place in a separate transaction. This distinction is not always crystal clear, because many financial firms are also involved in the physical supply of the commodity. Likewise, some producers are willing to offer both physical and financial contracts or roll them all into one deal. There is also a third class of energy companies: **marketers** and **brokers** offering both physical and financial deals. Marketers take title to the commodity while brokers simply seek to match a buyer and seller. The role of energy marketers is paralleled in the securities market by a class of brokers active on the floor of an exchange, known as **market makers**. These companies are crucial in providing **liquidity** - the ease with which a commodity can be bought and sold - without discounting or paying a substantial premium.

The concept of liquidity is not as easily grasped as it might seem. Lee (1998 pp. 50-51) discusses all of its implications:

> One conception of a perfectly liquid market is that of a forum in where it is possible to buy and sell an infinite amount of the asset being traded, at the same time, without delay, and at the same price. As is self-evident, this definition is composed of several distinct components. The difference between the bid and the ask price for small-sized orders, or the '**width'** of the spread, is one measure of liquidity. Another is the '**depth**' of the market, a gauge of the manner in which the spread widens or narrows as the size of the transaction becomes larger. A further element of liquidity is the market's '**immediacy**', namely whether there will immediately be a price at which it is possible to execute all trades. Another aspect of a market related to liquidity is the speed at which orders can be executed, or analogously the expected time market participants need to wait before an order on the other side of the market appears. Yet a further element of the liquidity of a market relates to the dynamic properties of transaction prices, and in particular the extent to which transaction prices diverge from and revert to equilibrium prices, or the so-called '**resiliency**' of the market. [Emphasis added.]

As Lee's expansive description suggests, liquidity and trading costs are clearly related. In commodity and securities markets they are like opposite ends of a scale. In other words, an illiquid market has high trading costs and a liquid one has low costs.

The most popular form of contractual arrangement in liberalized energy markets is a pair of parallel contracts. The first contract is for physicals, to sell or buy the commodity. The buyer and seller may establish a long-term relationship by **indexing** the price of the commodity to prices published in the trade press, which reduces the transaction cost of constantly searching for a counter party and a good price. Indexing prices to the spot market, however, creates substantial price risk. This risk can be modified by a second contract: a futures contact if basis risk is low or a similar derivative from the OTC market. Separating the two reduces **counter party risk** - the risk of default by the exchange partner - because firms specializing in financial contracts are better prepared to minimize risk by diversifying their portfolio.

Despite the similarity of results in a mature market, important distinctions can be drawn between futures exchanges and forward prices determined in the OTC. Exchanges are public by design. When **open pit** trading was first devised, exchanges developed several important rules. First, the exchange was to be entirely neutral; it took no position in the commodities being traded and, thus, had no exposure to radical changes in the market. Second, ownership and control of the exchange rested in brokers who held seats in the exchange. Since these seats could be bought and sold, exchange members had a vested interest in the success of the exchange (the more successful the exchange, the more valuable the seat). Third, all bids and offers were to be called out so that all traders could hear them. Fourth, the results of all transactions were to be recorded on the exchange's books, the results compiled, and prices made public. Open pit trading is slowly being replaced by electronic systems, and although these rules may seem arcane given modern communications and computers, they succeeded because they were self-reinforcing. The biggest problem electronic exchanges have is that traders use them to obtain information and then make their deals in private, constraining the exchange's revenue.

Another important difference between a commodity exchange and OTC forward contracts concerns counter party risk and credit (Black 1986). Commodity exchanges operate on **margins**. That is, participants must post a portion (about 20%) of the market value of the commodity contract to the exchange. As the price of the commodity changes, participants either have to increase the amount posted or they receive a refund. This calculation is made daily using **mark-to-market** accounting. Some companies, particularly Enron, appear to have abused mark-to-market accounting. It is, however, the only way to keep track of value in commodity markets where prices can change radically in a few days. Firms use mark-to-market accounting to determine the value of their forward contracts negotiated in the OTC, but do not post daily margins. Instead, counter party risk is covered by the reputation of the firm or letters of credit. This gives well-financed firms a definite competitive advantage.[2]

Both the OTC and futures exchanges perform important functions and have broad benefits. The Executive Vice President of NYMEX (Wolkoff 2000) describes the role of a futures market in this way but his points could also apply to the bilateral trade in the OTC market:

> Futures markets provide two important economic functions: price transparency (price discovery) and risk shifting (risk management).
>
> Price transparency is the constant reporting to the world of the prices of actual trades being made at the exchange. With tens of thousands of energy contracts traded daily, each reflecting a binding commitment to make or take delivery of a specific commodity, price information is made available in real time, on a virtually continuous basis. Thus the true world reference price referred to earlier.
>
> Risk shifting, in the secure liquid markets that the New York Mercantile Exchange provides, allows commercial interests to "hedge" the risk of price fluctuations that could affect profitability and planning of their business operations. For the commercial participant, the result is a form of insurance against the financial adversity that can result from volatile energy prices.
>
> A futures contract is an agreement between two parties for delivery of a particular commodity at a specific time, place, and price. Once initiated, a futures contract obligation can be satisfied by either taking an equal and offsetting futures position or by going through the delivery process and taking possession or making delivery of the commodity.

[2] Ironically, the California Power Exchange did not require the state's three large utilities to post margins or other extensive credit backup, because no one could imagine that they would default. Two did and the dispute, concerning about $4 billion, took years to settle.

> The vast majority of market participants opt for the former, and as a result, futures contracts are primarily used as financial rather than physical management tools.
>
> An options contract bestows upon its owner the right, but not the obligation, to buy or sell the underlying futures contract at a specified time and "strike" price. The option to buy is called a call, and the option to sell is a put. A major appeal of options is their similarity to term insurance. The option is purchased for a one-time fee called an option premium. Depending on how the market moves, the option may be sold for a profit, exercised, or allowed to expire worthless, with the holder's loss limited to the premium paid.

Experience in all three energy markets – oil, gas, and power – suggests that futures exchanges and OTC trading are mutually dependent. Exchanges ensure accurate price transparency of key commodities, but cannot serve all of the needs of the market. The volume of trade in commodities with unusual qualities or in isolated geographic markets is too small to be economic in an exchange. At the same time, however, the immediate price transparency of exchange trading guides OTC trading, reducing the risk of inefficient contracting. Basis swaps directly link the two types of activity, ensuring greater trade volume and lower transaction costs for both.

A mature commodity market with well-established futures contracts and an OTC provides buyers and sellers a transparent portrait of expected prices in future periods. These prices can be used to determine the fair market value of a long-term contract. In immature markets without futures, the relationship between current spot and longer-term contract prices certainly exists, but it is not transparent and often appears random and unexplainable. The consequence is inefficient investment decisions and sometimes a barrier to entry. In early stages of development, the gas industry relied on complex and rigid contracts in order to obtain funding. No one would build a pipeline (and no bank would finance it) unless there was assurance of a supplier's willingness to supply gas and a buyer's willingness to consume it. While these arrangements are often necessary to underwrite the construction of gas infrastructure, they are not flexible and have difficulty in accommodating unexpected changes in the economic environment. Moreover, the large scale of the projects requires massive financing, which may exclude all but the largest companies.

2.3 Commoditization

Although policy makers would like to take credit for the benefits that liberalized energy markets have bestowed, they have had little role in the development of the key institutions that made it possible. Instead, those institutions have evolved primarily as means for to increase profitability by reducing the transaction costs of transferring resources.

In recent years, traders and market pundits have described the "commoditization of certain markets that have undergone technological or regulatory evolution" (Mango and Woodley 1994). In the marketplace, commoditization can mean many things. To some it simply means greater competition and lower profit margins. The term should be used more profoundly, however, to denote the fundamental change in the market structure for particular goods and services. In this sense, commoditization may be described as a transition from a "closed" to an "open" market. In the last few decades, such structural changes have occurred in the markets for foreign currency, air travel, telecommunications, personal computers, and the energy markets.[3]

Phil Verleger (1987 p. 163) was one of the first to recognize the significant structural shift in the oil market: "It is argued that oil has become more like other widely traded commodities, such as corn, wheat, copper, and silver. In a word, oil has been 'commoditized.'"

Policy initiatives aimed at deregulating or liberalizing markets have often provoked the transformation of industries. In particular, the North American gas market (discussed in more detail in Chapter 5) was heavily regulated. Federal deregulation of interstate gas transport created an open market for wholesale gas and it has generally led to lower prices. Deregulation also fostered competition in the long-distance telecommunications market at the retail level, and telephone rates fell dramatically due, in part, to increased competition. It is worthy to note that of the liberalized markets - telecommunications, trucking, airlines, and energy - only the gas and oil industries have taken on the commodity market structure described in Section 2.2. The electricity industry may never adopt the commodity paradigm for reasons to be explained in Chapter 6.

The process of market liberalization – the metamorphosis from a closed to an open market – includes five major elements:

- As markets are opened, a commodity typically becomes *unbundled* from other products and services normally associated with its sale;
- This unbundling accompanies price discovery: the collection and dissemination of pricing information;
- Price discovery necessitates some sort of product and contract standardization in order to compare prices reliably;
- Dissemination of prices that identify geographic and other quality differences creates the opportunity for quick profits, which, in turn, attract traders and new firms to the industry, increasing the liquidity of the marketplace;
- If prices prove to be volatile (as most energy prices are), inefficient companies will face financial peril and will have to transform themselves in order to survive. As a consequence, forward markets develop and futures trading becomes a necessity, allowing price risk to be separated from contracts for physical delivery (Tussing and Hatcher 1994).

The dramatic change in the computer market is a prime example of unbundling. The hardware underlying many computers is now considered a commodity – parts are interchangeable and practically anybody can get in the business. A couple of decades ago, computers were sold primarily as mainframes in a proprietary package, in which hardware, software, maintenance and training were bundled together under a single fee.

[3] The first successful large-scale futures market outside the agriculture industry was established by the Chicago Mercantile Exchange in 1972. This was prompted by the end of the gold standard and the extreme volatility that resulted in foreign exchange markets at that time (Millman 1993).

Economic rent in the industry is now earned from specific components (such as the processor chip) or from specific operating systems and applications software.[4]

During the period of natural gas deregulation, the term "unbundling" was mostly used to advocate the pricing of gas separately from transportation and inventory services. In the 1970s, industrial customers often paid a single price throughout a utility's service district. The regulated price sometimes reflected the cross-subsidization of the residential sector, and usually failed to account for the economies or diseconomies of particular customers. It seldom, if ever, matched the utility's marginal cost of delivery.

Separating the market value of a product from the price of associated services and other products in a closed market is a complex and often impossible task, as independent prices for the goods and services bundled together are seldom available. Moreover, the entire package often includes a complex term structure, where the sales price is an amalgam of current market conditions and expectations about the future. Prices in one contract (or regulated market) are not usually comparable with prices in another, which may leave buyers and sellers in a state of confusion and prices not satisfactorily performing their equilibrating role. The consequence of this may be an endless cycle of mismatch between demand and supply.

Efficient markets require reliable information about prices, since they facilitate trade and investment decisions. The bulk power market in the Western U.S. was exchanging power at average prices of $20 to $30 per megawatt-hour (MWh, or one hour of electricity consumption at a constant rate of one megawatt) in the late 1970s and early 1980s. At the same time, power utilities reckoned their avoided costs at more than $100 per MWh, while state and federal authorities were approving or coercing investments that would be viable at such prices. In short, there was no connection between price signals and investment decisions. This was partially responsible for the "stranded cost" regulatory problem and led California on its disastrous course to plan its restructured power markets exclusively around spot trading, as briefly discussed in Chapter 6.

Price discovery and unbundling go hand in hand. In the case of the energy markets (oil, gas, and power), mismatches in demand and supply necessitated a small, flexible and largely unbundled spot market to fill the void. As explained in more detail in subsequent chapters, the spot oil market was initially the Rotterdam market; in gas it was the Texas Intrastate market; and in power it was the WSPP in the U.S. and the creation of the England-Wales power pool in the U.K. Spot markets, however, do not operate effectively to smooth regulatory and contractual irregularities unless spot prices are transparent and contractual procedures relatively unconstrained. As trade increases and balancing becomes important, traders are willing to pay for the collection and publication of price information. Initially, participants in the existing market can ignore spot markets and spot pricing. Ultimately, however, spot prices begin to play a much larger role in the setting of day-to-day activities and in planning the industry's development.

---

[4] It is ironic that the two greater sources of IBM's market power, operating systems and processing chips, were inadvertently transferred to Microsoft and Intel.

Price discovery may be stillborn unless it can describe a standard product. Everybody has a general idea of the price of a hand-made oriental rug, but that does not help much with the bargaining; subtleties of size, fiber, dye, and density aid pricing only for those with a highly trained eye. Oriental carpets will always be highly diverse and, indeed, would lose their charm if they were not. However, such diversity is neither necessary nor desirable for energy products. Active trading requires common standards on volume, impurities, thermal content, timing of deliveries, credit, term structure, etc. Futures trading requires a further step: the identification of an actively traded product in the physical market to which futures contracts can be tied and the whole market effectively indexed.

As markets refashion themselves from closed to open, commonly traded standard products emerge. In the crude oil market, the standard is barrels (42 gallons) of light sweet crude traded in the mid-continental U.S. Natural gas prices are measured in millions of British thermal units (Btu) at Henry Hub in Louisiana. Electricity is standardized in bilateral trading in megawatt hours for on-peak and off-peak daily, weekly and monthly blocks of power, delivered at various hubs and control areas. Power exchanges and pool auctions typically price power on an hourly or half-hourly basis.

Personal computers have made it much easier for traders to reconcile the bewildering heterogeneity of transactions in the physical market with the need for a standardized index or market commodity with which to track general price movements. Individual sales contracts at different locations, with different terms and conditions (or with slightly different product specifications), can then be marked against general price movements.

Market liquidity rests on a variety of underlying characteristics such as: the number, heterogeneity and (de)-concentration of participants, the transparency of transactions, the quality of market information, and, ultimately, the proportion of the stock or flow of the relevant product that is negotiable (i.e. in play or potentially drawn into play by new information). These are, of course, many of the same characteristics that determine transaction costs. When trade volume is low and the number of traders constrained, it could prove extremely costly to close out a futures or forward contract. Successful futures trading and the commoditization of a market depend on substantial increases in liquidity and its counterpart, reduced transaction costs.

As spot trading in crude oil and petroleum products climbed in the 1980s, the number of trading companies grew exponentially, reaching several thousand. A similar pattern has been observed in North American gas and electricity markets, as these markets have been systematically opened up to broader trade. In the beginning of 1994, there were nine marketers registered to trade with the FERC; three years later there were 275. When markets open, they do so with great speed.

In the oil and gas industries, spot markets emerged as a balance wheel, bringing unexpected dislocations of demand and supply into equilibrium by serving as secondary markets for surplus volume. These volumes had originally been subject to long-term contracts or sale at regulated prices, and were resold at different prices. The greater the difference between the regulated or contract price and the potential resale price, the greater was the incentive for buyers to seek arbitrage gains in the spot market (and, concomitantly, the more rapidly these markets grew).

An open commodity market does not progress by stages – advancement occurs in physical and financial sectors simultaneously (although perhaps at different rates), including the evolution of futures and forward trading. As trading expands, commodities are standardized and unbundled, and reliable price information becomes available, allowing buyers and sellers to separate price and physical risk. Prices often seem stable in closed markets, although this may be illusory. Pricing in the commodity markets is actually erratic from day to day and week to week, although it may be less volatile over the long-term cycle.

In closed markets, buyers and sellers try to protect themselves from unexpected changes in market conditions by implementing long-term contracts that fix both volume and price. When markets shift unpredictably, as energy markets did between 1973 and 1986, the contract structure becomes untenable, and it becomes difficult for buyers and sellers to agree on long-term prices. Such circumstances create the demand for futures and forward markets, as described in section 2.2.

In the U.S., futures trading has contributed enormously to the commoditization of the oil and gas markets, with substantial benefits to both consumers and the industry. The contract allows producers, refiners, processors, and large energy consumers to manage energy price and supply risk separately. In addition, the development of the futures exchange has greatly increased liquidity and competition. It was not so long ago when energy industries were constrained by regulation; the price of world oil was shrouded in secrecy with an oil shock entailing long queues at service stations and a crippled economy. In contrast, during the Gulf War of 1990-91, when there were no price controls or cultural inhibitions to raising prices, a full-scale invasion and bombardment in the Persian Gulf prompted a relatively modest market response. Similarly, only two decades ago the demand and supply of natural gas was balanced by closing down schools and hospitals. Energy issues have since moved from the front page to the business page, where they should have been all along.

2.4 The Role of Market Makers in Other Markets

A review of the literature on financial markets (O'Hara 1995, Lee 1998, and Demsetz 1968) emphasizes the crucial role of market makers in commodity and securities trading. Market makers (also referred to in a slightly different role as specialists in stock markets) are the foundation of continuous trading and are essential for liquidity. These entities act as middlemen and are willing to buy or sell the commodity instantaneously. They are neither producers nor consumers; instead, as noted in section 2.2, they manage an inventory and fulfill two primary functions. First, they act as brokers coordinating demand and supply. Second, they fulfill a market participant's "demand for immediacy," the ability to buy or sell on a minute's notice. Market makers offer to buy and sell simultaneously and the spread between their offer prices is a measure of both transaction costs and liquidity. In financial markets the role of market makers is precisely defined.

Although these functions appear in other markets, the role is often hidden by other activities and the categorization is much less clear.

Most economic theory ignores the role of intermediate trade and concentrates on producers' supply and consumers' demand. Transaction cost economics obviously cannot avoid the study of middlemen and the institutions that facilitate exchange. The observations concerning financial markets are helpful in interpreting market institutions in the broader economy. Fundamentally, there are two types of marketing arrangements – brokers and marketers. Brokers, such as real estate agents, auctioneers, etc., do not take title to the goods and services they sell. Marketers (most retail outlets, for example) do take title and manage inventories; they organize the distribution of goods and services temporally and geographically. These two aspects of a market go beyond simple arbitrage, which is the usual function attributed to middlemen; they reduce transaction costs and provide benefits and services to producers and consumers. Convenience stores are exactly what the name implies; they are stores open twenty-four hours at key locations for the convenience of consumers. These stores act as a market maker for consumers and satisfy a demand for immediacy.

Gasoline marketing is a good example of the complex interrelationship of transaction costs, the theory of the firm, and marketing institutions. In the U.S. there are three basic tiers to marketing the product. First, the branded Major company may operate company owned and operated gasoline service stations with managers and employees on payroll. These are usually high-volume stations in close proximity to the company's infrastructure (refineries, pipelines, tank farms, etc.) and along major transportation corridors. These types of stations have the lowest administrative cost to manage. Second, the Major refiner may contract with independent individuals who will manage a company owned or personally owned station for profit. The oil company delivers the gasoline, sets its wholesale price and station rent but allows the contractor to maximize profits. Typically these stations are somewhat remote from key infrastructure, are slightly lower in volume, or have side activities (car repair, etc.) that are complex to manage. Finally, Major oil companies may contract with jobbers (marketers) that purchase the gasoline at wholesale terminals and have their own tank trucks and storage facilities; jobbers then distribute the product to independent dealers and their own stations. Usually these third-party marketers serve the most remote markets, which would present the highest administrative cost if a branded refiner would attempt to operate them as in the first category. In all cases, however, the purpose of the marketing arm is to deliver the product at times and places that are convenient to the consumer. Prices at various points of distribution reflect spatial competition, transaction costs, and transportation costs.

2.5 The Nature of Energy Demand and Supply

Williamson (1985), Klein, Crawford, and Alchian (1978), and others have identified and analyzed the problem of asset specificity as related to transaction costs. The essence of this problem is that exchange often involves *ex ante* investment in specific assets that, *ex post*, would not have an alternative use. Examples cited are a printing press that is built to publish a specific magazine; in such a case, the two entities depend entirely on each other. As each issue is published, either the magazine owner or the printing press owner may be tempted to renege on contractual obligations and demand additional payments or discounts. Accordingly, transaction costs are high; *ex ante* bargaining may be protracted and complex and *ex post*, enforcement could be a serious problem.

The problem of asset specificity can arise in an *ex ante* competitive market. That is, before the specific investment there may be a multitude of magazine publishers and printing presses. Competition in advance of investment leads to efficient price bidding between printer and publisher. After the investment, however, competition ceases to be an effective constraint and either the buyer or seller may be tempted into "opportunistic" behavior.

Inflexible assets certainly exist in the energy industries, particularly with regard to infrastructure investments that are constrained to a specific location – good examples would be oil and gas leases, pipelines, etc. However, the energy industries also have another complexity in that virtually all energy commodities are component products, i.e., their production and use requires substantial capital expense by both the buyer and the seller in equipment necessitating compatibility and, thus, continuing interdependence. Conceptually this is similar to, but not the same as, the problem of asset specificity analyzed in transaction cost literature. It is different from the problem of asset specificity because *ex post* the market may still be competitive, i.e., buyers and sellers are locked into a technology, but not to each other.

It has often been noted that the demand for key energy products is extremely price-inelastic in the short term. This is because consumers have made large investments in cars, appliances, and equipment that are designed to use a specific type of energy. For example, electrical appliances depend on the continuous delivery of electricity over a network system within a specific frequency range. Cars depend on standardized fuels to run efficiently, or sometimes to run at all. Once consumers have made their appliance choices, they become dependent on particular types of energy suppliers. Further, they have usually made their choice based on the expected price of the commodity. If energy prices change radically, consumers find it difficult to adjust quickly. Faced with rising prices and changing expectations, consumers will eventually buy more efficient cars, but such substitutions take time. In the short run, when gasoline prices double, consumption does not drop significantly. A variety of econometric studies of energy demand have demonstrated low energy price elasticity in the short term, but relatively high price elasticity in the long run.

Energy producers face a parallel problem to that of consumers. They have to design infrastructure to produce energy products that meet specific standards and meet consumers' expected demand. Often these investments are inflexible. When oil prices increased dramatically in 1979 the demand for heavy fuel oil dropped substantially, while the demand for gasoline did not. As the market adjusted, unsophisticated refineries (those not capable of producing a high proportion of gasoline and middle distillates) had to be upgraded or closed down.

Although many investments made by energy consumers and producers are designed to use one specific type of energy commodity, this need not cause the problem of asset specificity analyzed in the transaction cost literature. As noted, this is because there may be substantial competition both before and after capital investments are made. As to be described in Chapter 5, efforts made in the natural gas industry to unbundle the purchase and sale of gas from the transportation system transformed the industry, switching the economic relationship from one of asset specificity to component products. Although petroleum consumers may be locked into the need to purchase a certain type of commodity, they still have a choice of suppliers, because there are common industry standards. Even with choice, however, component products create highly inelastic demand and supply schedules, which lead to volatile prices. The tradeoff between price stability (achieved by high-cost regulation or inefficient market concentration) and competition is a common thread that has run through the history of all energy industries. Likewise, the proponents of gas and electricity regulation believe consumers should not be required to cope with the price volatility and suspected price gouging inherent in cyclical energy markets.

Wholesale energy markets are a complex interlacing of specific assets and component products. While wholesale gas and electricity are standardized products with multiple buyers and sellers, the delivery infrastructure is specific to producing facilities and consumption nodes. Creating adequate liquidity (or competition) at various trading points is the principal objective of market liberalization in jurisdictions seeking to substitute competition for franchised monopolies or state-owned enterprises. Because the oil industry's transportation sector is far more flexible, it has fewer problems with asset specificity than either gas or electricity. With the exception of the concentration of low cost oil in the Middle East (and the consequent market power of OPEC), the exercise of market power in the oil industry, when it occurs, has concerned control over transportation and logistics.

The linkage between natural gas production and the final consumer is sometimes referred to as the "gas chain." The same type of chain links electricity consumers to producers. Petroleum product consumers are not tied to producers through a fixed delivery infrastructure, but through product specification and standardization – component products.

All the energy industries have some form of interdependence between final buyers and sellers and interdependence linking the various stages of production. As a consequence, the organization of these industries has tended toward vertical integration and/or long-term contracts. Events in the last two decades have demonstrated that market concentration, vertical integration, and rigid contracting need not be a natural order for the oil and gas industries. Transportation systems can be unbundled from the commodity. Product standardization can be achieved through regulatory agencies or industry institutes and standards boards. Futures contracts and other derivative securities can moderate the inherent price volatility of energy commodities. The shift in regulatory approach and the spontaneous development of market institutions designed to lower the transaction costs of trading are the means by which energy markets have been transformed.

In North America and to some extent in Europe, natural gas has become the "fulcrum" of energy pricing. This is because the demand for gas is closely linked to the prices of both petroleum products and electricity. The interrelationship is complex, but it can be broken down into specific arenas of use.

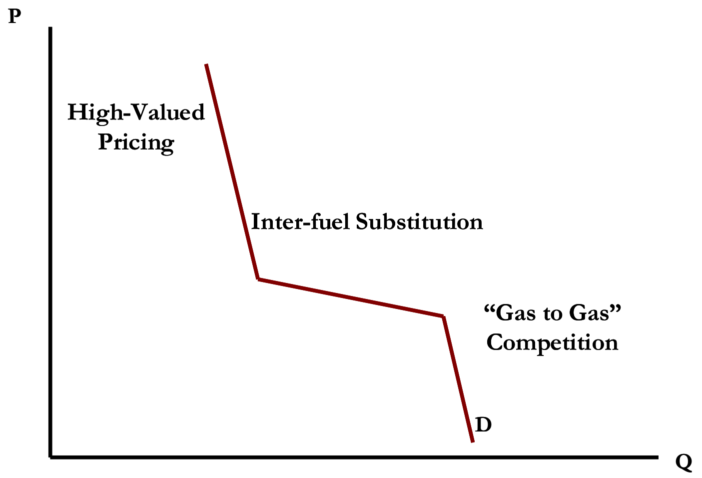
The emergence of low-cost gas-powered electricity generation directly links gas and electric prices, because the marginal cost of electricity is frequently determined by gas prices in much of the world. The tradeoff between the value of gas in direct sales, as compared to its use to generate electricity, is a precise calculation based on the "heat rate" of the unit. As the technology for combined-cycle (CC) and combustion turbine (CT) units has improved, the heat rate has declined. That is, fewer and fewer Btus of gas are required to generate a megawatt (MW) of electricity. Older CTs may require 10 to 14 million Btus to generate one MWh of electricity. New CCs, however, can generate the same amount of electricity with less than 7 million Btus. Generators can do a quick calculation of electricity economics by multiplying gas prices (in million Btu) times the heat rate and adding a small charge for operation and maintenance expenses (O&M), usually around $2 per MWh. In the above example, if gas prices are $3 per million Btu, the efficient plant will be able to generate electricity at a cost of $21, plus $2, per MWh.

Electricity prices can occasionally diverge from expected gas-power heat rate conversions if highly inefficient units are dispatched or if the system is capacity constrained, i.e., sufficient generation capacity is not available to meet peak demand and load must be shed (effective demand elasticity). The difference between observed electricity prices and generation cost is known as the "spark spread." To ensure long-term investment, the spark spread must average enough to cover the amortized cost of the capital invested in the plant and/or there must be side payments, commonly known in the industry as capacity payments. In the period leading up to the California energy crisis, the spark spread in Southern California averaged $10.50 per MWh, while the average in the largest power pool on the Atlantic Coast (PJM) averaged $14.78 per MWh (Van Vactor 2000). Except for relatively rare price spikes during periods of peak demand, gas prices are reasonably good predictors of electricity costs, establishing the link between the two markets.

To some extent electricity serves the high end of consumer energy demand (or as OPEC officials would cast it – "noble uses"). Electricity is used for lighting, refrigeration, entertainment, etc. - the elements of a modern economy that most consumers now consider "essentials." Gas-powered electricity generation also has a lower environmental impact than power produced with other fossil fuels. Consequently, gas used for electrical generation is not as price sensitive as in other uses. In contrast, when natural gas substitutes for oil, it is usually at the low end of the scale, in the boiler or black fuels market. In this segment of use, where gas, coal and fuel oils compete, demand is extremely price sensitive.

The functionality of these various market segments can be represented graphically. Figure 2.2 illustrates a demand schedule for natural gas, broken into three parts. The first segment is price-inelastic. These are the high-valued uses of natural gas: electricity generation, precision industrial processes, home appliances, etc. The demand schedule falls steeply through the first segment and then turns into a nearly flat line, where the demand for gas encounters substantial price sensitivity. In this segment, there is extensive competition between fuels and prices are driven to thermal parity, i.e., the price of fuel oil, coal, and natural gas per Btu come to rough equivalency. There will always be some price differences due to handling and procurement costs, etc.

Figure 2.2
Natural Gas Demand in Three Segments



The final segment of the gas demand schedule has been referred to as a market of "gas to gas" competition. In this segment of demand, gas thermal-equivalent prices fall below those of competing fuels, just as in the first segment gas prices are above those of alternative fuels. In the first and final segments of demand, the gas market effectively decouples from other energy markets. The best example of gas-to-gas competition was the North American gas market in the 1980s. During the period in which OPEC attempted to maintain excessively high oil prices, the heavy fuel oil market was nearly eliminated. By and large, natural gas filled the void.

To sum up, the investment in specialized equipment – component products – to produce and use specific forms of energy can create highly price-inelastic demand schedules in the short run. The consequence can be extremely volatile prices. Energy price volatility can be mitigated through fuel substitution if industrial energy consumers have designed flexible facilities, i.e., facilities that can utilize gas, oil or coal. In mature energy markets, such as Europe and North America, where a variety of energy sources are available, natural gas is the fulcrum of energy choice, playing a critical role in dampening price volatility for all forms of energy.

2.6 The Physical Attributes of Energy Commodities

In his influential book on the oil industry *Essentials of Petroleum*, Paul Frankel (1969 p. 11) observed:

> The *leitmotif* of any discussion about petroleum must be its liquid state.

> The problems involved will be considered first against their technical, or rather scientific background; the fact that most petroleum products are volatile liquids delimits their possibilities and determines their role among similar or competing materials.
>
> Starting with this knowledge, the specific features of the exploration of oil-fields and the exploitation of oil-wells will have to be investigated.
>
> The next stage is refining. Here the fact that a liquid cannot be "handled"—in the original sense of the word—fixes the pattern of the industry. Refining requires little labour but elaborate plant.
>
> Lastly, we shall see the consequences of the liquid state of petroleum in transport and marketing where it entails the use of specialized equipment which puts the oil trade into a category all of its own.

Frankel's book was written before the development of transaction cost economics (to be explained in more detail in Chapter 3). Nonetheless, it is interesting to note that he also recognized the significance of specialized equipment in explaining the structure of the petroleum industry. In addition, he emphasized an aspect of commodities that has not been investigated in a comprehensive way – their physical characteristics and the impact these characteristics have on marketing costs and industrial structure.

In the case of petroleum, Frankel emphasized two aspects: it is a liquid and it is volatile. In Frankel's view, these characteristics explained much of the industrial structure. Frankel tells an interesting anecdote about Rockefeller and the Standard Oil Company. He explains that Standard Oil was the first company to adapt its marketing to the nature of the commodity it was selling. Initially, the primary market for oil was kerosene for lighting. According to Frankel (1969 pp. 36-37):

> They [Standard Oil] appear to have been the first—not only in the States but in most European countries—to devise a method of storing kerosene immediately prior to its sale to the consumers. They provided shops with a little tank from which the retailer could draw the required quantity at the time of sale, and they replenished the stock in the tank from horse-drawn road tank wagons. The retailer was never faced with the necessity of handling packages, he was only concerned with the liquid itself. Such an arrangement offered considerable advantages, and retailers were easily persuaded to sign an agreement to the effect that they would not sell any kerosene except that supplied by the company providing the tank.

Standard Oil was highly successful with their scheme, except in Asia where they moved slowly, opening the door for Shell. According to Daniel Yergin (1991 p. 68), Marcus Samuel, the founder of Shell, was able to establish a similar kerosene distribution system in Asia ahead of Standard Oil.

Frankel's comments are obviously important in understanding the petroleum industry, but they also reflect the timing of his analysis. The fact that oil is a liquid was especially significant in the early and mid twentieth century, because it was rapidly displacing coal, wood, and charcoal as the primary fuel for commerce and industry. As a liquid, oil could be transported much more cheaply than solid fuels, but only if new types of delivery systems could be developed. Standard Oil was not only successful in marketing kerosene; it was the first company to fully exploit the advantages of rail transport and pipelines.

In modern terms, Frankel's analysis misses a step; a more general way to view the problem is to examine the impact of a commodity's physical characteristics on transaction cost – the cost of the marketplace. Variation in such costs can help explain why different industries have different market structures. The energy industries are particularly interesting to study in this respect because the principal energy sources – oil, gas, coal, and electricity – all have radically different physical characteristics. Oil is, of course, a liquid; natural gas is delivered as a gas; coal is a solid fuel, and electricity is the movement of electrons delivered over a wire. Chapter 3 will develop a more generalized approach to support this analysis.

3. Transaction Cost Economics and Energy Commodities

3.1 Defining Transaction Costs

By now two points should be clear. First, the transformation of energy markets was not a simple undertaking for either policy makers or the private entrepreneurs seeking to craft new market institutions to cope with OPEC's challenge. From the perspective of political leaders, their initial motive for managing the oil market was entirely practical; the supply and price of an essential commodity had spun out of control. The impact of high oil prices on employment and economic growth for industrialized economies was serious and the situation had to be corrected. Second, energy companies had a lot to learn. The oil industry had been vertically integrated and oil trading had not played an important role in profitability. Now, it was costly and time consuming to procure raw materials, where before it had been an internal company transfer. Put another way, the transaction costs of trading in the OPEC market were high. Companies would have to adapt to the new structure and find ways to reduce the costs if they were to compete with their rivals. That is, they would have to find ways to reduce transaction costs. But what exactly is meant by such costs? They must at least be categorized if not defined before they can be analyzed.

Despite the obvious presence of transaction costs and their frequent use in institutional economics, the concept is not well defined. In fact, the definition varies depending on the author and the analytical use to which it is put. Ronald Coase first introduced the idea of transaction costs, although he did not use the terminology at that time. Coase (1937) noted that using the price mechanism costs something and that firms integrate multiple operations internally as an alternative. According to Michael Dietrich (1994 p. 19) the first author to use the phrase "transaction cost" was Kenneth Arrow in 1969. However, in his article, "The Problem of Social Cost," Coase (1960) captioned his description of these costs as "the cost of market transactions taken into account." Since the work of Ronald Coase a large body of literature on transaction costs has evolved.

The popularity of the transaction cost concept lies in its ability to explain exceptions to more general models of economic behavior. But a theory of exceptions is by necessity ad hoc and the explanatory power is dulled by imprecise definitions and inconsistent application. Coase's (1960) definition is tantalizingly vague: "In order to carry out market transactions it is necessary to discover who it is that one wishes to deal with, to inform people that one wishes to deal and on what terms, to conduct negotiations leading up to the bargain, to draw up the contract, to undertake the inspection needed to make sure that the terms of the contract are being observed, and so on."

North (1997) defines transaction costs as primarily measurement and enforcement. This is more precise, and provides a useful starting point for explaining why transaction costs vary – not just temporally, but from one industry to another. North (1997 p. 157) elaborates on measurement in an advanced economy: "…increased resources are necessary to measure the quality of output. Sorting, grading, labeling, trade marks, warranties and licensing are all, albeit costly and imperfect, devices to measure the characteristics of goods and services." The problem of enforcement is far more complex involving employer rules, regulations, incentives, punishments, etc. and contractual failures preceding and following exchange. Although it is more precise, North's definition of transaction costs may be too narrow.

James Buchanan (1985 p. 97) is critical of the transaction cost concept, discussing rather than defining "…three broadly defined categories of problems that are placed in the transaction costs rubric…" He discusses information-communication constraints, free-rider constraints, and strategic behavior. In his view, ex-ante consensual rule making can resolve most of these apparent dilemmas and there need be no departure from neoclassical price theory. Buchanan also identifies the unexplained complexities of the approach, underscoring once again that for the Coase Theorem to hold, as Coase himself stated, transaction costs must be the same under a different allocation of rights.

Coase's (1960) observation about the significance of transaction costs in affecting the allocation of resources should have provided a strong incentive to explore the concept and define it precisely, because it significantly weakens the concept of economic efficiency. In an extreme view, (as Buchanan 1985 and others have pointed out) the Coase Theorem can be used to justify whatever exists. Among rational economic agents, the only reason why a potential gain from trade will not be realized is because the transaction cost of doing so exceeds its value. Thus, some economists have viewed the transaction costs concept as an ideology, rather than a tool, which once again has inhibited an objective development of its definition and inclusion in economic analysis.

Due to the elusive nature of the more general point about transaction costs, the active use of the concept has been considerably narrowed. Dietrich (1994 p. 15) defines transaction cost as the cost of discovering prices, negotiating, and concluding contracts. This approach would suggest that brokerage fees, search time, and certain types of information costs should also be included as a broad category in this definition, although these costs are often internalized and cannot be easily quantified. Cootner and Ulen (1996 p. 84) define "…transaction costs corresponding to three steps of an exchange: (1) search costs, (2) bargaining costs, and (3) enforcement costs."

Much of the confusion and academic debate derives from concern about unifying the concept of transaction costs with neoclassical price theory and the traditional applications of economic efficiency. Positive economic analysis ought to come first. Until transaction costs are consistently categorized (let alone defined), the activities affected by these costs cannot be clearly identified, measured, and compared. This inhibits a rigorous study of why transaction costs vary from one industry to another, which is a key step in understanding the issue. Consequently for this analysis the following typography and classification of transaction costs has been developed:

- Information or search costs:
  - Search time
  - Advertising
  - Price discovery

- Bargaining and negotiating:

- Brokerage fees
- Bids and offers
- Haggling
- Contract preparation and review
- Immediacy

- Measurement:
  - Weighing and sizing
  - Sorting and classifying
  - Grading and assaying
  - Packaging and labeling
  - Spoilage
  - Record keeping and accounting
  - Support of industry associations and institutes

- Enforcement:
  - Distribution design and control
  - Security
  - Clerking
  - Credit control
  - Registration of trademarks, copyrights, and patents
  - Contracting and licensing
  - Safety rules and responsibilities
  - Regulation
  - Legal fees and litigation
  - Insurance
  - Warranties and guarantees

Breaking down and detailing the various types of transaction costs invariably results in some dispute over classification or inclusion. Not everyone would agree that all of the items listed above are transaction costs. Some could be classified as part of a firm's production cost function or part of a consumer's preference function. In any case there is considerable overlap. For example, trademarks serve at least two purposes. They prevent other producers from co-opting the goodwill and reputation of a branded product. At the same time, they reduce information costs for consumers, when there is genuine heterogeneity in competing commodities. Likewise, regulatory agencies or voluntary industry associations frequently set standards and product specifications. These organizations may reduce measurement and enforcement costs as well as improve information for consumers.

The categorization of various types of transaction costs, however, avoids an important question. Why should these costs be treated separately rather than integrated as a part of neoclassical demand and supply theory? In this respect there is one important common characteristic that sets them apart: *Transaction costs are those costs associated with the exchange of a commodity; costs that might be borne by the supplier, the consumer, middlemen, any or all of the above.*

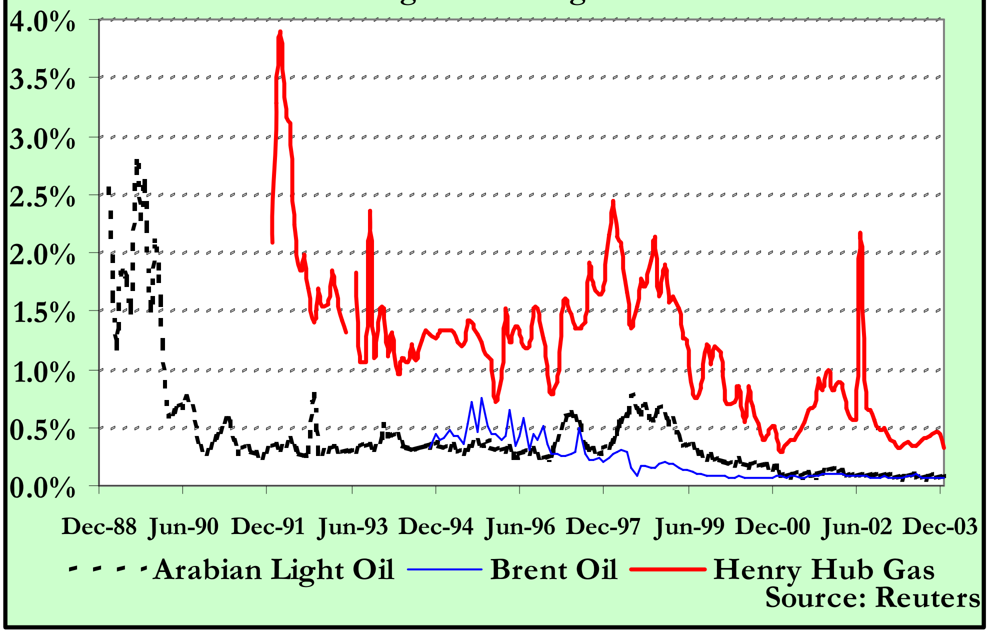
Put another way, transaction costs shape the institutions commonly referred to as the marketplace, and they are central to the functioning of a modern economy.

3.2 Transaction Cost Economics Applied to the Energy Industries

3.2.1 Falling Transaction Costs

All trade has some explicit or implicit transaction costs. In addition, such costs vary depending on the characteristics of the commodity and the industrial structure. This chapter explores the impact of commodity characteristics on transaction costs in a general framework. Chapter 7, in turn, analyzes in depth the differences in trading costs between the North American oil and gas markets, where these costs can be measured and compared.

Figure 3.1
Change in Trading Costs



It is important to note that trading costs, as a key element of transaction costs, have changed over time as energy markets have matured. Figure 3.1 illustrates the trend of trading costs, measured as the difference between buy and sell prices, for the purchase and sale of three key energy commodities. (This calculation may also be interpreted as a measure of liquidity.) Arabian Light crude oil was for many years the marker price set by the OPEC cartel. Principally from the Ghawar oil field in Saudi Arabia, it is the largest volume crude oil stream in the world. Brent crude oil is an important North Sea crude oil and is also the marker for the futures market in London. Henry Hub is the most important natural gas trading point in North America and is the marker for natural gas futures trading in New York.

As will be explained in more detail in section 3.6 and Chapter 7, this measure of transaction cost is the same as that used in other studies of financial markets. The difference between buy and sell prices is most easily understood as a broker's margin.

From the perspective of the discussion of transaction costs described in section 3.1, it is obviously incomplete. For example, traders would incur unmeasured information costs in deciding whether or not to enter into the trade. Further, large volume trading blocks can disturb the relationship between buy and sell prices. Nonetheless, there are two important conclusions to be drawn from these data. First, the cost of trading natural gas is consistently higher than that of trading for oil; the reasons for this are explained in Chapter 7. Second, the cost of trading all three commodities has dropped significantly. It is interesting to note that the trend of declining cost for Arabian Light continued through the Gulf War of 1990-91 and rose during the Asian financial crisis. This may be explained by the greater risk of default during the financial crisis than the disruption caused by the war.

What is significant is the relative change in these measurable costs; the trend is irrefutable. Clearly the decline in trading costs in the oil and gas markets has been dramatic, representing in the case of Arabian Light a cost reduction greater than an order of magnitude. How do these quantitative results mesh with observable behavior? What has the industry done to reduce trading costs? As energy markets have transformed themselves, the costs of many of the other components of transaction costs – information, bargaining, measurement, and enforcement – have declined. The following is a thumbnail sketch of these costs in the energy industries and some explanation as to why they have changed.

3.2.2 Information

As pointed out, component products have highly inelastic demand and supply functions in the short term. Moreover, demand for certain energy products fluctuate radically with the seasons and daily changes in the weather. Storage costs vary depending on the commodity and location. All of these factors combine to place an extraordinary premium on information in energy trading.

Before weather satellites, it was claimed that airline pilots could make good money in certain agricultural futures markets, because they were often the first to observe severe storms. For example, a pilot noting a violent storm moving towards Florida's orange groves could buy frozen orange futures by radioing his broker. Energy trading companies try to do the same thing. Enron was known to employ spotters around major coal and nuclear generating plants. When steam was emitted from the plant, the spotter would phone directly to Enron's trading floor with the news that the plant was ramping up. A few minutes extra notice gave Enron's traders just enough advantage to make the costly system of information procurement worthwhile.

Energy markets, like all markets, have taken advantage of computer technology. A modern energy-trading floor is interlaced with video and data displays. Traders are confronted with a series of overhead screens with continuous updates of financial information from world stock and bond markets, the latest weather forecasts including satellite and radar images, continuous updates of the price of every traded energy commodity, etc. Successful traders are up early reviewing events in markets around the world from electronic news services and the trade press. Each important energy commodity has at least a half-dozen successful newsletters and data sources.

3.2.3 Bargaining

Bargaining costs have declined significantly in energy trading, as a consequence of 1) the maturity of trading networks, including the development of formal exchanges, 2) greater price transparency, 3) standardization of commodity definitions and pricing terms, 4) identification of standard points of delivery, 5) development of common terms of art and legal language, 6) the use of standardized contracts, 7) the growth in trade associations, and 8) streamlined regulatory processes.

When the oil market exploded in 1979 there were a limited number of traders from the oil companies, joined by a few brokers and traders from the financial industry. The number of traders was limited by perceptions about how oil supply contracts should be secured. The various OPEC members had national supply companies, but government ministries often made the key decisions about supply contracts. In many cases these negotiations were little more than barter, which can be time consuming and costly because market values are not immediately apparent. With the development of futures trading in crude oil, however, a whole new class of speculative traders entered the market; traders whose aim was to trade for profit instead of commodity procurement. These traders focused on relative prices and substantially increased market liquidity by expanding the network of those willing to buy or sell at a particular time. The growth of North American gas trading further expanded this network. In general, bargaining costs are lower in a large trading network, because competition and the threat of competition sharpen offers.

Obviously, greater price transparency reduces bargaining costs. Reliable price information narrows the degree of separation between buyers and sellers. If the price information is sophisticated, containing, for example, a series of forward prices similar to what is available from futures markets, it allows traders to tailor complex patterns of future delivery volumes. In addition, reliable price transparency provides buyers and sellers the option to index contract prices to published values.

Another element in reducing bargaining costs is product definition and standardization. Petroleum products are good examples. In a modern market each product fits precise specifications as to thermal content, viscosity, number and types of contaminants, pour point, etc. The products are standardized regardless of which refinery produced them or from which crude-oil feedstock they were refined. The only variation is in some branded products, so that the seller can make claims of superiority. Standardized product definitions allow substantial wholesale trade and because they increase substitutability, they reduce bargaining costs.

Standardized delivery points, like standardized products, aid in reducing bargaining costs. Obviously if transportation costs or location differentials are well known, the range of bargaining outcomes is reduced. Over the last two decades, common trading points, market centers and hubs have emerged for all energy products. Details on these hubs are provided in this chapter and in Chapters 4 and 5.

It is often said that English is the language of commerce, and in the case of energy trading, virtually all oil contracts are written in it.[5] The language issue is slightly more complex than that, however, because as markets mature they develop their own terms of art and commonly understood phrases. Mutual understanding of these terms facilitates bargaining, in part, by reducing uncertainty and clarifying obligations.

The North American natural gas industry relies primarily on the Canadian Gas EDI short-term sales contract. This contract is twelve pages in length (small type) and covers issues like: definitions, performance and obligations, transportation, nominations, imbalances, quality and measurement, taxes, billing, payment and audit, title, warranty and indemnity, financial responsibility, defaults and remedies, force majeure, and term. The electricity market relies on the Edison Electric Institute (EEI) contract prototype and the WSPP contract format.

The most relevant part of each of these contracts, and the only part that changes from one trade to another, is the last page, which contains the "confirmation" of key trading elements – commodity, price, volume, time and location of delivery. Some companies have their own variation on these contracts, particularly for long-term deals. By and large, however, the use of standardized contracts allows the traders to focus on the key terms of the deal and not get bogged down in technicalities and legal language. The use of standardized contracts has contributed to an explosion of trading. In hearings before FERC on the western power market controversy, it was estimated that over an eight-month period there were in excess of 400,000 bilateral trades.

Trade associations have also played a big role in reducing bargaining costs. These include the American Petroleum Institute (API), the Gas Research Institute (GRI), etc. As is explained throughout this work, trading in energy commodities has been greatly facilitated by the accompanying development of derivative contracts for hedging and risk management. The International Swaps and Derivatives Association (ISDA) has been particularly important in standardizing the approach. They, too, have developed a standardized contract and generic guidance for members.

Regulatory agencies and review processes have been greatly streamlined in order to facilitate trade. In the United States, FERC opened up gas and power markets by staging the introduction of "market-based" pricing. Prior to the authorization of such trading, any gas or power sales contract had to be approved by the Commission. As explained in Chapter 5, the increasing complexity of the number and types of sales in the gas market made this a virtual impossibility. As a consequence, FERC focused on the conditions of trade (the degree of competitiveness) rather than the specific pricing terms of the contract.

[5] Due to its Anglo-Saxon origins, the petroleum industry operates almost exclusively in English throughout the world. On the other hand, as gas and power trading has expanded to European markets contract language has diversified.

3.2.4 *Measurement*

Measurement is not a serious problem for the exchange of most energy commodities and costs have not changed much, nor do they appear to depend on firm structure. Retail meter reading for gas and power is costly, but the high-volume delivery systems of wholesale trading significantly reduce per-unit costs. Moreover, containment vessels for delivering the product have long lives and low maintenance costs. Crude oil is the most costly of wholesale energy products to measure, particularly in small lots. The quality characteristics of crude oil can be approximated with a hydrometer, indicating its density or gravity. Once a crude oil's density is known, key refining characteristics can be estimated. Details on refining characteristics can be obtained through a laboratory assay of a sample of the crude oil. Assaying is not, however, normally required if the origin of the crude oil is known; crude oil is typically sold by its field (or blend) name and its measured density. Companies are careful about sampling crude oil and petroleum products from marine tankers to be sure the crude oil or petroleum products have not been contaminated by water or other impurities. Pipelines may ship in batch mode or blend various crude oils, trying to maintain constant proportions.

It is worth noting that many of the gas and power delivery systems of the Former Soviet Union (FSU) were not metered. This had two consequences, unnecessarily high demand and pipeline leakages. As the FSU has slowly reformed energy consumption has dropped significantly and the pipeline leakage problem has been largely resolved. According to BP (2003) energy demand has dropped 31% in the Russian Federation, due to better conservation and economic decline.

3.2.5 *Enforcement*

Energy industries are plagued by the problem of contract disputes. The enforcement problem arises for three fundamental reasons. First, many primary energy resources, prolific oil and gas fields, hydroelectric sites, etc., have large economic rents. That is, their production costs are substantially below the costs of competing alternatives that set prices in an open market. Companies that own the rights to these resources are under constant pressure from public organizations and other companies to share the associated economic rent of their resource base. Second, as already explained, most energy supply projects are capital intensive, that is, they have large fixed costs relative to variable costs. Once these projects are in place, they too harbor huge economic rents, with great pressure to dissipate them. Third, the high elastic demand and supply schedules are prone to accidents. Relatively modest changes in demand or supply can cause profound changes in price. California's power prices increased more than tenfold compared to equivalent periods in 1999 and 2000. At the least, such radical price changes seem unfair; they usually provoke a public outcry and litigation.

Aside from accidents, enforcement costs for wholesale energy trading have declined. The primary reasons are much better credit and risk management controls, standardized contracts often containing mandatory arbitration, greater liquidity, and significantly improved price transparency. These developments reduce the odds of a dispute and if a dispute occurs, contested elements are narrowed. Nonetheless, it would be unfair to conclude that there are fewer lawyers - there are undoubtedly more. That, however, reflects the far greater volume of trade, rather than an increase in the percentage of contracts in dispute.

It is worth commenting on the growth of credit and risk management in energy trading. Traditionally, economic theory has treated the analysis of risk separate from transaction costs. In a classic example, the internal rate of return of a risky investment project would be expected to be higher than the rate of return of a low risk project, even though the two might have the same transaction costs. This stark division is not as easily applied to energy trading. Arguably, the management of credit and risk are transaction costs, and they are not insignificant. Some of the cost may be attributed to information costs; keeping tabs on whether or not your counter party has credit problems. Other costs are the price paid by a firm to reduce risk. Properly organized, the use of derivatives and other financial instruments reduce a firm's risk and stabilize profitability. The management of energy trading activities is not, however, without its own set of risks and oversight costs: embezzlement and unauthorized speculation, to cite two examples. If the cost of trading (using the market) is equal to the transaction costs avoided by the company that chooses instead to integrate, then much of risk management is clearly one of those costs.

3.3 Transaction Cost and Liquidity

Harold Demsetz (1968) was the first to study the transaction costs of trading securities. He studied the spread between bid and ask prices in the New York Stock Exchange. As noted, the spread between buying and selling is a narrow definition of transaction costs, since (among other things) it excludes the time and cost involved in deciding to buy or sell. It is often associated with brokerage fees, although the bulk of these financial studies make clear that comprehensive explanations are far more complex. Most studies of the spread between bid and ask prices have extended Demsetz's approach while staying close to the original methodology.

The concept of liquidity links institutional economics and studies of financial markets. High transaction costs are associated with illiquid markets. Securities that turn over rapidly generally have a smaller spread in bid-ask prices than do securities that trade infrequently. Likewise, the study of transaction costs in institutional economics (Williamson 1985) has focused on "bounded rationality, opportunism, and asset specificity." These problems arise mainly in illiquid markets, where information is costly, trading is limited, search and negotiations periods are long, and there are only a few active buyers and sellers.[6]

Liquidity is also the link between commodity characteristics and transaction costs. Complex commodities, e.g., those that are immobile, heterogeneous, indivisible, etc., are seldom part of a liquid market. This is because commodity complexity segments markets. A pipeline network is the best example of this phenomenon. Even when

[6] Often the alternative explanation for such market imperfections is the exercise of market power. Determining whether transaction costs are unavoidable or a reflection of economic inefficiency is at the heart of the debate about the validity and analytical value of the transaction cost approach.

third-party access is granted, buyers and sellers are constrained by the delivery infrastructure. The limited numbers of buyers and sellers at key hubs reduce liquidity and raise transaction costs, even though prices throughout the system may respond to the same broad market movements.

3.4 The Narrow Tradition of Institutional Economics

In his original work, Coase sought to explain why firms exist and why they take on a given size and structure: "…the operation of a market costs something and by forming an organization and allowing some authority (an 'entrepreneur') to direct the resources, certain marketing costs are saved" (1937 p. 392). This tradition has been extended over the years and recent work is typified by Oliver Williamson (1985 p. 1) who comments: "Contrary to earlier conceptions – where the economic institutions of capitalism are explained by reference to class interests, technology, and/or monopoly power – the transaction cost approach maintains that these institutions have the main purpose and effect of economizing on transaction costs." North also echoes this point of view (1997). Economic development arises from improvements in both technology and in institutions that govern the efficiency of markets. Technology and capital accumulation by themselves do not explain why many less developed countries severely lag in economic development. Steven Cheung (1998) notes that: "like any other institution, the market was created to reduce transaction costs, subject to other constraints." Although the literature on the institutions of capitalism is fascinating and provides many insights it is frequently criticized because the subject matter and methodology do not allow for easy quantification, making it difficult to form a testable hypothesis.

Institutional economics has focused on enforcement, which is a narrow aspect of transaction costs to the detriment of its intellectual development. This analysis would be better described as the economics of contracts, rather than transaction cost analysis. Both Williamson and Klein have emphasized the problem of opportunistic behavior, where assets are fixed and buyer and seller remain mutually dependent after the investment has been made. This debate has focused on an important example, the acquisition of the Fisher Body by General Motors. Klein, Crawford, and Alchian (1978) had identified this event as an example of vertical integration in order to prevent opportunistic behavior. Other researchers have refuted the observation, to which Klein (2000) replied, detailing the facts and reasoning behind the original observation. Coase (2000 p. 30) himself weighed into the debate:

> The erroneous statement of the facts in the Fisher Body – General Motors case has misdirected the attention of economists and has stood in the way of the development of a more solidly based treatment of the problem of asset specificity. The view that I formed in 1932 and discussed in my Yale lectures was that the asset specificity problem was normally best handled by a long-term contract rather than by vertical integration and that "propensity for opportunistic behavior is usually effectively checked by the need to take account of the effect of the firms actions on future business."

The narrow focus of transaction cost analysis in institutional economics is even greater than Coase's recent critique. Originally, Coase (1937 p. 390) made a second important point: "Up to now it has been assumed that the exchange transactions which take place through the price mechanism are homogenous. In fact, nothing could be more diverse than the actual transactions which take place in our modern world. This would seem to imply that the cost of carrying out exchange transactions through the price mechanism will vary considerably as will also the cost of organizing these transactions within the firm."

Despite Coase's observation on the variability of transaction costs, the body of economic literature on the subject is slender and seldom systematic; it is focused on specific industries, rather than a comparison across industries. The neglect is remarkable because, since then, advanced economies have reallocated substantial resources from manufacturing to marketing. Not unexpectedly, the complexity of marketing mechanisms and price structures has also expanded dramatically. Thus a systematic understanding of why the cost of the price mechanism varies would seem essential to modern economic theory.

Two polar examples help identify the wide variation in transaction costs and hint at causality: the allocation of traffic at an intersection and the marketing of gumballs. The allocation of traffic at an intersection is almost always determined on the basis of rules – stop signs, roundabouts, traffic lights, policemen, etc. The notion of drivers negotiating and pricing the order of transit is somewhat ridiculous. From this example, it is tempting to conclude that a market for transit rights would produce prices so low that it wouldn't be worth implementing, which is why a system of rules prevails. That, however, can't be true for some intersections during peak hours. More than a few drivers would be willing to pay a lot to jump the queue. The obvious problem is the high cost of monitoring the intersection, enforcing order, price discovery, and the small size of the market. High transaction costs, beyond the value of the commodity, are the primary reasons why markets for intersection rights do not exist.

In contrast to traffic routing, consider the lowly gumball machine. Gumballs are quite cheap, but not free. The price of a gumball is certainly a lot less than what some commuters would pay to get home earlier. Decades ago somebody figured out how to market gumballs cheaply. The dispensing machine is simple and cost-effective. It can be placed practically anywhere; it doesn't require clerking time or costly overhead. Gumballs are hardy and easily last through a prolonged replacement cycle, while the right to cross an intersection is a momentary thing. Far and away the most interesting features of gumballs are their size, hardcover, and shape. The commodity itself is designed to be appealing to consumers and at the same time to fit the dispensing machine – the least cost means to bring it to market.

3.5 The Transaction Costs of Trading in Financial Markets

In contrast to research on transaction costs and institutions, the analysis of trading costs for financial securities has been highly quantitative. This is because exchanges keep records of various kinds of costs and because the efficiency of the market structure is of concern to competing exchanges and to regulators. As Locke and Venkatesh (1997 p. 229) expressed it: "Financial market transaction costs are of general interest to academic researchers, as well as being of particular interest with regard to the optimal design and regulation of market microstructure."

Harold Demsetz (1968) demonstrated that there are two types of transaction costs in trading securities. First, there is the explicit brokerage fee for the purchase or sale of company shares. Typically these fees decline as the value of the transaction increases. Secondarily, there is the implicit cost of trading – the difference between bid and ask prices that represents the profit margin of market makers (floor traders who are willing to hold an inventory of each listed share). Demsetz found that the spread between bid and ask prices varied with the volume of trading in the shares. That is, if the market maker's inventory turned over slowly, they had to charge more than if there were a rapid turnover. The task of market makers is to provide liquidity. Even in robust markets it is unlikely that orders to buy and sell are going to arrive simultaneously. In order to facilitate trade someone must be prepared to buy and sell from inventory so that the customer will be able to complete an immediate transaction.

Since Demsetz's article there have been a large number of studies of the transaction costs of trading securities. Most have concluded that in addition to the volume of trading, the price volatility of the security impacts the spread between bid and ask prices. This makes sense, because the greater the price volatility, the greater the risk to the market maker to hold an inventory.

Grossman and Miller (1988 p. 620) compared highly liquid securities markets to residential housing, "where virtually none of the transactions pass through a dealer's temporary inventory." They concluded that in illiquid markets "intermediaries provide brokerage or search services, not immediacy." Similarly, Neal (1992 p. 319) noted:

> The Grossman and Miller model predicts that when the intrinsic risk of an asset is low (real estate is a good example), a broker market will emerge. As the intrinsic risk rises, other things constant, so does the demand for liquidity, and centralized markets emerge with a specialist or batch call structure. Assets with a high level of risk, such as futures contracts on commodities and Treasury bonds, create extreme demand for liquidity and are traded in a competitive market maker structure.

Several observations concerning commodity characteristics and transaction costs can be drawn from these comments. First, all things are not constant in the above example. The physical characteristics of real estate assets are radically different than Treasury bonds. Low liquidity in the real estate market is caused partly by low turnover – most people will buy only one or two houses in their lifetime – and partly by the diversity in quality and size of the housing stock, a house's indivisibility, and its immobility. Furthermore, there are interesting counter examples. For example, dealers, rather than brokers, dominate car sales, even though the market is not especially liquid.

Nonetheless, the general observation is correct, if incomplete. Liquidity does play an important role in determining transaction costs, because it is linked to commodity characteristics. Consider the most obvious problem - product heterogeneity - that requires sorting, grading, and search time to determine relative prices. Two effects result from this problem. The one most commonly addressed in transaction cost analysis is information: buyers and sellers have to determine quality gradations and this introduces all sorts of problems in asymmetrical information and/or the drudgery of inspecting the product (Akerlof, 1970). Equally important, however, product heterogeneity segments the market. Buyers seeking a specific variation of a generic product must find a seller of like quality. Since there is potential substitution, there is competition in the market, but it can also lead to one-on-one bargaining and often a price or cost penalty if the transaction must be immediate. In addition, certain aspects of commodity characteristics – particularly geographic segmentation and spoilage (or high storage costs) – lead to price volatility. Stigler (1951) characterizes such small-number problems as the "extent of the market." The implications of this analytical approach with respect to the oil industry will be summarized in Chapter 4.

The foregoing point leads to the key assumption of ceteris paribus in analyzing transaction costs. All things equal, homogenous commodities are likely to have lower transaction costs, just as high-volume markets do. But it is as important to recognize that low-volume homogenous commodities may have higher transaction costs than high-volume heterogeneous ones.

3.6 Studies from Commodity Markets

There is a small volume of literature dealing with the reasons for failure of new products on commodity exchanges. This is a useful starting point for an analysis of why transaction costs vary: because product failures are usually associated with high transaction costs, and because commodity markets are highly competitive. In a commodity exchange, there are usually a diverse number of buyers and sellers, prices are transparent, and price information is made public. Most participants are price takers, and exchanges constantly monitor behavior in an attempt to prevent traders from cornering the market. In short, commodity markets operate nearly like perfectly competitive markets, as envisioned by neoclassical economic theory. Despite the presence of intense competition, transaction costs remain significant. These commonly include: brokerage fees, a "liquidity penalty" (the difference between bid and ask prices), and the risk of a default or dispute arising from miscommunication or non-performance. Moreover, participants frequently undertake research or other investigations to improve information and reduce risk prior to the transaction.

In a modern economy, most commodity markets are a combination of bilateral trade in the commodity (physical or cash market) and exchange-based trade in futures contracts (financials). As a derivative tool, a futures contract is intended to be closed out with an associated financial gain or loss. If the motive for the transaction is hedging, then the financial gain or loss from futures trading should offset, or nearly offset, the loss or gain from trading in the physical market. For example, a petroleum refiner can lock in a guaranteed profit margin by trading futures in both petroleum product and crude oil derivatives to coincide with physical obligations. Although futures contracts are intended to be financial instruments, they must be directly linked to the physical market. If necessary, a participant can insist on delivery of the commodity. The terms and conditions of delivery are itemized in the underlying contract. Futures contracts are standardized, which is in sharp contrast to the physical market where there may be distinct differences in location delivery, product quality, and other terms and conditions. It is standardization that allows high turnover, leading to a liquid market.

Futures trading is more than a century old, and there are a surprisingly large number of commodity exchanges that compete to offer futures contracts for a wide array of commodities. Until recently, most futures contracts were associated with agricultural products. Since the 1970s, however, a whole new series of contracts have been launched dealing with stocks, bonds, currency, and energy. The competition to develop a successful contract is strong and over half of the commodities offered fail (Silber 1981 p. 128), i.e., volume is insufficient to cover cost and sooner or later the contract disappears.

Dennis Carlton (1984) studied the evolution of futures markets noting that until 1970 these markets were limited to metals, industrial products, oilseed, livestock, food, and grain. From 1921 until 1970, when financial products were introduced, the number of contracts increased slightly from 50 to 78. The volume of total contracts traded increased from 1 to 150 million from 1954 through 1984 (Carlton 1984 p 249). The volume of trading has continued to increase and in 2003 the total will be about 8 billion contracts (Futures Industry Association).

According to Carlton (1984 p. 242) successful contracts have five features: 1) there is a high level of uncertainty about price and availability of the underlying commodity, 2) there are price correlations among similar products, i.e., basis risk is predictable, 3) there are a large number of participants and an appropriate industrial structure, 4) transactions have large value, and 5) prices are freely determined without regulation. The third point concerning industrial structure is particularly significant to the energy industries, because Carlton claims that vertical integration can stifle the development of a futures market.

Black (1986 p. 6) surveyed the literature on new futures contracts and identified six requirements for product success, but her analysis focused more on the characteristics of the commodity:

1. The commodity must be durable and storable;
2. Units must be homogenous;
3. The commodity must be subject to frequent price fluctuations with wide amplitude;
4. Supply and demand must be large;
5. Supply must flow naturally to a market [(1) competitive cash market and (2) low delivery costs];
6. There must be breakdowns in an existing pattern of forward contracting.

Black discusses each of the above points and two issues merit elaboration. First, the storability of a commodity is no longer essential. This is because, as Black (1986 p. 7) recognized, "…futures contracts had a forward pricing function separate from their inventory guiding role…Goods for futures delivery did not have to currently be in inventory as long as they would become available through production". This opened up the potential market for futures contracts to a wide variety of commodities, including crude oil and petroleum products, where the flexible surface inventory is only a few days' supply. Second, the issue of product homogeneity is crucially important to the success of a futures contract. That is because contracts are standardized around a specific product and location. It is not necessary that the commodity be homogenous for a futures contract to succeed, as long as the value between various grades can be objectively determined and location cost differentials are known and stable.

According to Black (1986), success or failure in designing new futures contracts rested on two issues – the choice of the commodity and contract design. Her comments on the commodity itself were particularly interesting. She developed a generalized theory of contract success and used statistical analysis to support it. Her statistical analysis suggested that success or failure of a futures contract rested on the opportunities for "cross hedging." If a new contract concerned a commodity that had a close substitute, hedgers might choose to trade the substitute, rather than the new contract. In this instance, hedgers faced a trade-off between two transaction costs. The new contract might be specific for the product they wished to hedge, but if trading volume was low, the spread between bid and ask prices could be very high and shift unpredictably, increasing the liquidity penalty. Alternatively, prices in the established contract might not be perfectly correlated with the commodity for which they sought to reduce risk. The most cost-effective choice for reducing risk would determine success or failure of the new futures contract.

Success or failure of futures contracts demonstrates that differences in transaction costs have concrete effects and that they vary depending on the characteristics of the commodity. The next step is to generalize this approach in order to gain deeper insight into its role in determining market structures.

3.7 Causation

Like Coase, Cooter and Ulen (1996) note that transaction costs are sometimes low and sometimes high. They (1996, p. 84) observe: "Consider this question by looking at the three elements of the costs of an exchange. Search costs tend to be high for unique goods or services, and low for standardized goods or services. To illustrate, finding someone to sell a 1957 Chevrolet is harder than finding someone to sell a soft drink." Bargaining costs depend on the degree of "public" knowledge of the parties, the clarity of private property rights, and the number of parties that must agree. Cooter and Ulen (1996 p 86) contend: "Enforcement costs, the third and final element of transaction costs, arise when agreement takes time to fulfill." While these observations are helpful, they may be too narrow and frequently depend on an implicit assumption of imperfect competition. Even casual observation suggests that the transaction costs associated with the exchange of standardized goods and services in highly competitive markets vary enormously.

In order to focus on issues of competitiveness, neoclassical economic theory necessarily simplifies the nature of commodities and assumes homogeneity. But to help explain why transaction costs vary from one industry to another, it is necessary to relax the latter assumption. The approach suggested here begins with an analysis of commodity *characteristics*, summarized in Table 3.1. These are a set of generalized properties that categorize commodities and impact transaction costs. There is extensive literature on the impact of some of these characteristics on markets, while others have been neglected. What often matters, however, is how the characteristics interact with each other to increase or decrease transaction costs because, in the real world, commodities have multi-dimensional properties. There is another problem too. The set of transaction costs identified earlier define a broad set of activities that are often unrelated, even allowing some characteristics to impact the components of transaction costs in opposite directions.

Table 3.1 -- Commodity Characteristics and Associated Costs

Homogeneity or Heterogeneity – Many groups or classes of commodities have variations in quality, location, and other characteristics that require grading, sorting, etc. and may segment the market. In general, the more heterogeneous a commodity is, the greater the measurement and negotiating costs.

Divisibility – Many products that are considered classic "commodities" are easily and cheaply divisible, e.g., agricultural and natural resource products can be purchased in small or large lots. Generally, transaction costs per unit increase the smaller the volume purchased due to measuring, packaging, and handling costs. On the other hand, indivisible products that are durable can be made divisible through a secondary rental market, but with higher search time and enforcement costs.

Mobility – Real estate, roads, pipelines, and amusement parks are goods and services that are specific to a particular location, while cars, precious stones, small consumer goods, and the like are cheaply transportable from one location to another. The mobility of a commodity impacts search time, enforcement, and negotiating costs.

Durability – There are two dimensions to durability. Some commodities decay whether used or not, others depreciate with use. If a product or service decays without use it may impact transaction costs through wastage or the necessity of costly preservation. Uncertainty about depreciation can increase information and measurement costs.

Hazard – Some commodities are dangerous if mishandled. Explosives, fireworks, poisons, energy products, tools, etc., can damage persons and property. Such products impact enforcement costs, requiring special packaging and warnings, and the assumption of liability risks.

Rivalry – A "nonrival" commodity is one that can be consumed contemporaneously without degradation by multiple consumers or reproduced by buyers at low marginal cost. Enforcement costs arise in excluding free riders and by prohibiting or limiting secondary markets. Rival commodities are by nature exclusive to one consumer.

Components – If component products are produced and marketed separately, it increases information and measurement costs to ensure compatibility. Setting industry standards can reduce such costs.

3.8 Discussion

3.8.1 Homogeneity or Heterogeneity

Commodities often vary in quality, which requires grading and sorting and, in general, transaction costs are inversely proportional to the degree of heterogeneity. Within a product class, the price structure must capture the general shift in demand and supply (relative to all other goods and services) as well as the commodity's diversity. Diamonds are an excellent example of the pricing problem. The "four 'Cs'" – carats (size), color, cut, and clarity – determine the relative value of one diamond to another. In order to track how the price of diamonds compares to all other commodities and precious stones, the industry follows the price of "Diamond D," which is a flawless, clear white, one-carat stone, cut in traditional style. The price of a stone of different size and quality is then estimated from the Diamond D price by trained appraisers. Crude oil, which in its natural state ranges from sulfur-laden tar to "natural gasoline," is priced against a "marker" crude oil. As noted earlier, when OPEC dominated pricing the marker was Saudi Arabian Light oil. Following the success of futures trading in crude oil, the marker shifted to West Texas Intermediate in North America and Brent North Sea oil in Europe. Prices of the marker crude oil are determined by its availability (or that of a close substitute) in the daily spot or cash market, in tandem with the cost of storage and trading in futures exchanges and in forward trading. Crude oils at different locations or with different quality characteristics are priced from the marker with appropriate discounts or bonuses.

Because it can be costly to determine quality and associated relative values, there are a wide variety of privately- and publicly-funded institutions, independent from both the buyer and the seller, that set standards, grade the product, and otherwise ensure minimum levels of quality. To return to the earlier example of diamonds, grading or assay of stones for retail sale is most often done independently by non-profit organizations like the Gemological Institute of America. Assay of crude oil is a combination of government and private initiative. In the United States, a database of the assays of over 1,500 crude oils is maintained by the Department of Energy at the government's expense. In addition, refiners maintain their own proprietary assays of crude oil streams they are likely to purchase.

The precise market response to the challenge of commodity diversity varies throughout the world, reflecting differences in taste and social customs as well as different opportunities. In the U.K, the Tea Tasters Council certifies the quality of tearooms around the country, but the grading of tealeaves remains the responsibility of the brand-name packager, where quality is assured through reputation and tradition. In Canada, independent tasters grade tea and the quality of each package is certified. The U.S., as its custom, has a hodgepodge of public and private institutions concerned with the grading (and promotion) of tea. In many industries with heterogeneous products, standards are set by the reputation of the firm, which relies on trademarks and similar devices to preserve unique product features and to reassure consumers as to quality. Establishing independent institutions that grade products and set standards frequently reduces information costs. Competing firms may sponsor these institutions in a combined effort to promote the product and to standardize quality differences if the benefits exceed the alternative of higher transaction costs.

3.8.2 Divisibility

There are two dimensions to divisibility. In the first, a commodity might be divisible by dividing it into smaller parts. In the second, divisibility can be achieved by creating a market structure that allows different consumers to use the commodity at different times.

In the first type of divisibility, the principal transaction costs concern measurement, packaging, and handling costs. Most agricultural and natural resource products can be purchased in small or large lots. Wholesale markets typically develop measurement standards based on the commodity's physical characteristics, historical transportation, and containment facilities and are priced accordingly. Lumber is measured in board feet, crude oil in barrels (based on the historic 42 gallon wooden barrel, transported in wagons), grain in bushels, etc. In almost every case, there are independent organizations or government agencies that set and enforce measurement standards.

It is no secret to most consumers that in retail markets per-unit prices are cheaper for large volume purchases. Although the point is obvious, it is not trivial. Retail prices reflect packaging and handling costs as well as manufacturing costs. Packaging and handling per-unit costs normally increase the smaller the lot sold. Put another way, over a certain range, there are economies of scale in transaction costs related to packaging and handling. On the other hand, the ability to sell a commodity in small lots increases the potential number of buyers and sellers, increasing liquidity and reducing the bid-ask spread.

The transaction costs associated with dividing a commodity's use over time are especially interesting. Laundromats, hotel rooms, car rental agencies, and time-share condominiums are examples of various secondary products that have evolved to divide the use of a commodity that would not normally be divisible. Typically transaction costs in rental markets are proportionally higher than for outright purchase. Rental properties tend to depreciate faster because renters have less incentive to prevent damage. Rental companies have high administrative costs due to the resulting enforcement problems. They cannot determine in advance which customers are likely to steal or cause damage and thus require credit references, deposits, insurance, etc. Information costs are higher, because consumers are often confronted with a complex array of pricing and terms that vary considerably from one rental company to another.

There is another wrinkle too; divisibility may cause the market to be segmented into various groups of consumers. The time profile of demand (including or excluding weekends) often creates a natural segregation of consumers. Travel, for example, is easily split into business and leisure categories, resulting in complex pricing structures for airline tickets, rental cars, hotels, etc. Market segmentation, however, increases transaction costs due to higher information, administrative, and enforcement costs.[7]

3.8.3 Mobility

Some commodities, such as real estate, are absolutely immobile. Others, like cars, are designed to be mobile. Typically, increased mobility raises one type of transaction cost (enforcement) and may decrease another (negotiation).

[7] Airlines frequently employ elaborate schemes to prevent business travelers from escaping the Saturday-night stay rule. Recently in the U.S., they prohibited travel companies from issuing a set of overlapping tickets, which would allow two trips at tourist rates.

Goods like precious stones and metals, fancy cars, etc., require special care to protect against theft or loss. In the case of jewelry, the value is large compared to its size, which makes theft easier and the cost of prevention higher. A car, of course, is by nature transportable. The risk of theft and the risk that uncontrolled cars will do great damage to other people and property have resulted in special laws and regulations governing the licensing of cars and drivers.

On the other hand, immobility may segment markets and reduce liquidity, thus raising negotiating costs. The wholesale markets for natural gas and electricity are constrained by fixed distribution systems, which limit the volume of trading and number of traders at each delivery point. Trading activity throughout the grid comes to bear on price determination, even though only a few buyers and sellers are active at each node, increasing the scope for one-on-one negotiations and the cost of making a deal.

3.8.4 Durability

Durability, like divisibility, has two dimensions. The first is temporal. Some commodities can be stored indefinitely without consequence, others with some-to-substantial decrease in quantity or quality, and still others are absolutely non-storable. The second is the extent of deprecation; a commodity may not decrease in quantity or quality with use, decrease to some degree in quantity or quality with use, or decrease completely in quantity or quality with a single usage.

When a commodity decays rapidly there are trade-offs between spoilage, costly preservation, and complex pricing. Fresh fruit and vegetable sellers, restaurants, transportation systems, etc. all plan for a percentage of waste or unused capacity. One alternative to waste is a much more complex and costly pricing structure. In the face of spoilage, prices will often be discounted, requiring search time by both buyers and sellers. Peak load pricing in electricity and telephones can make production more efficient, but it is more costly for consumers to deal with the complex structure and more costly for the supplier to administer pricing, in that it requires special meters and complex billing. In addition, the life expectancy of many commodities that spoil can be extended at some cost. The most obvious example is refrigeration.

Transaction costs are also associated with depreciation. Consumers have to take into account expected depreciation when comparing the prices of various durable commodities. Likewise, manufacturers have to build in consumer preferences of the tradeoff between durability and price. Both of these activities require market research, or disappointment will result.

The difference in pricing structures between hotel rooms and rental cars demonstrates the difference between decay and depreciation and provides evidence of the impact of transaction costs. As a general rule in the U.S., consumers can reserve cars without an advanced payment. On the other hand, normally at least one night must be paid in advance to reserve a hotel room. If an innkeeper loses a nightly room fee, it is gone for good. The cost of daily depreciation of an occupied room is small relative to the loss stemming from a vacancy. In contrast, cars (particularly rental cars) depreciate more rapidly from use than from time. Thus, car rental agencies are far less concerned that a reservation be kept.

3.8.5 Hazard

Most hazardous material has special safety and related regulations that govern its production, distribution, and sale. Obviously there are substantial liability issues associated with the markets for hazardous goods. Like many commodity characteristics, the degree of hazard or toxicity varies and transaction costs are proportional. Common household products, like ammonia, are at one end of the scale, while plutonium and uranium are at the other. As noted earlier, information costs arising from product heterogeneity (quality variation) can often be reduced or eliminated by the use of semi-private institutions, branding, etc. that ensure standardization and quality control. Such institutions are seldom used to govern the trading of hazardous material, because the consequences of misuse or mislabeling can be so serious. Indeed, in most modern economies the penalties for misuse exceed the normal civil sanctions and violators can often be prosecuted under criminal codes.

3.8.6 Rivalry

When a commodity has nonrival attributes and nonpaying consumers cannot be excluded, it is defined as a "public good." There are, however, many nonrival commodities in which a means has been found to exclude (or at least inhibit) free riders. Examples are theatre, sports events, museums, and most intellectual property including computer software.

There are two types of transaction costs arising from the various types of nonrival commodities. The first is the cost of exclusion. This problem arises because multiple consumers can simultaneously share the commodity in question. Theatres are often designed both for the presentation of entertainment and discriminatory pricing. Obviously in a properly designed theatre, those that don't pay can be excluded by a system of ticketing and ushers, but ticket prices can also be tiered to seating arrangements. A system of reserve seats, however, likely increases enforcement costs. Enforcement and control at rock concerts is much more difficult than at the ballet. As a consequence, the ballet tends to have reserved seats, while rock concerts tend toward general admission.

The second type of cost concerns measures to prevent the development of secondary markets outside the control of the producing firm. This is the central problem of intellectual property rights. The effectiveness of copyrights and patents varies enormously from one country to another. Even with copyright protection, vendors take all kinds of measures to prevent secondary markets. Software publishers invent elaborate schemes to prevent copying. Frequently this actually detracts from the quality of the software or burdens the user with extra cost. Likewise software publishers tie in the sale of the software with instruction manuals and continuing support. Book publishers create elaborate binding, print parts in blue ink, and take other measures to make it very difficult to reproduce the publication. The cheaper it is for consumers to reproduce the commodity, the higher the transaction cost of enforcement.

3.8.7 Components

There are many commodities that are component products, i.e., one is completely useless without the other. Computers and software, telephones and connection networks, guns and bullets are just a few examples. This type of component product may create substantial search and information costs for both producers and consumers to ensure compatibility, if the components are sold separately.

The transaction costs of dealing with these types of products can be substantially reduced by mutual agreement on industry standards. The computer industry frequently cites the evolution of guns and muskets as an analogy for understanding the economic significance of industrial standards. Until the invention of the bullet, rifles and muskets were hand crafted and owners frequently cast their own shot. The use of bullets, however, required that rifle and pistol bores be standardized. Once standards were set by the industry, the manufacturing of guns could be separated from bullets, allowing specialized production and greater competition.

In a modern economy the list of component products requiring standardization is huge. Sometimes the standards are arrived at by historical accidents, but they often arise due to a dominant manufacturer. The Colt 45 Peacemaker, for example, firmly established itself as the leading standard for pistols. There are many other examples, particularly in the energy and computer industries. Petroleum product standardization is a complex process of government and private interaction. The process involves automobile manufacturers, environmental regulators, safety experts, refiners, and private institutions. The elimination of lead additives, which were used to enhance the performance of internal combustion engines, took years to implement at great cost.[8] The business strategy of just about every company in the high tech industry is aimed at developing and controlling a key standard. Microsoft is hated by its competitors, because its operating system is the dominant standard. As a consequence, other software and hardware companies must depend on Microsoft's cooperation to ensure their products are compatible.

If an industry standard is the exclusive property (by copyright or patent) of a single company, it can bestow huge windfall profits on the lucky owner. The enhanced value arises from the reduction in transaction costs that would be associated with a variety of confusing standards and incompatible products.

Williamson, Klein and others have used the term "specific assets" to describe a special type of component product. These are assets that depend on each other to have economic value and have little or no alternative uses. Ex ante competition may create a variety of options for either the buyer or seller. But once a commitment is made, the two

[8] There were some direct environmental concerns about the lead additive. The main reason it had to be phased out, however, was that catalytic converters necessary to reduce polluting gases were quickly damaged by lead. In the United States the process of eliminating gasoline with lead additives took nearly two decades.

are locked together, and as a consequence the buyer and seller are likely to enter into a long-term enforceable contract or vertically integrate into a single firm.

3.9 Combining Characteristics

As suggested earlier, it is the interaction of the characteristics in Table 3.1 that impact the level of transaction costs. This can be illustrated by returning to the problem of ordering traffic through an intersection. Assuming the commodity is the right to transit, the following characteristics likely add to the transaction cost of establishing a market. The value of that right is highly heterogeneous; there are multiple intersections, each would have a different price at different times. Moreover, the commodity is not divisible, except by sequential ordering. There is risk in transiting an intersection if someone else jumps the queue, which requires enforcement. The right to transit is immediately perishable. Transit rights are a complementary product to the ownership of motor vehicles. Of the seven characteristics identified in Table 3.1 only two of them – transportability and rivalry – would suggest lower, rather than higher transaction costs. In most circumstances ordering traffic is best done by a system of rules rather than a market.

Contrast the transaction cost of managing traffic to the purchase and sale of common shares of a listed company. Although there may be some diversity in share terms and conditions, there is normally one dominant issue. Shares are divisible to a single unit, with brokerage fees scaled to the volume purchased. Shares are easily transported and stored. They do not depreciate and there are no complementary products; there are no nonrival characteristics. Share ownership carries no risk, other than a possible decrease in value. The characteristics of common stock shares suggest that transaction costs of trading this commodity will be relatively low.

Gold has a unique history among commodities, because for centuries it played an extraordinary role – the dominant medium of exchange and store of value. Houthaker (1959 p. 147) notes that for a participant in a barter economy: "The introduction of money makes it possible for him to avoid a large part of these transaction costs and risks." A modern economy needs a low-cost medium of exchange, and it is only recently that government-backed currencies have replaced gold. There are a wide variety of precious stones and metals, so why did so many cultures settle on gold as the central medium of exchange? In all likelihood it was to economize on transaction costs, and Table 3.1 can explain why.

Gold is homogenous and its purity and weight can be determined with simple tests. The metal is malleable and easy to shape into jewelry, etc. Gold is easily divisible; it is valuable relative to its weight, which makes it easily transportable; it is durable and does not decay or erode, although there were depreciation techniques, known as "sweating," in which coins could be debased. Gold does not require complex or expensive packaging. It is not toxic or dangerous. Gold (except when cast as art) has no nonrival characteristics and is not a complement to other commodities. In contrast, precious stones are highly heterogeneous and not easily divisible. Other metals are too scarce or too abundant.

Transaction costs (rather than Gresham's Law) may also explain why checks, bankcards and government-backed currency have largely replaced gold as the primary medium of exchange. They are less costly to transport and much easier to divide than precious metals although not as durable. On the other hand, preventing the development of a secondary market (counterfeiting) is a far more serious problem with currency and checks than with gold coins. The dominant form of currency has changed over time as technology has improved and society has discovered a variety of means to reduce transaction costs.

3.10 Limitations and Comments

Commodity characteristics are important, but they are not the only explanation as to why transaction costs vary. Market power, social conventions, the nature of demand and supply functions, private property rights, cultural differences, history, the size of the market, etc. all may have an impact on market structures and transaction costs. Nonetheless, the physical characteristics of the commodity do have an important impact on transaction costs no matter what the cultural circumstances, and these features have been under-emphasized in the study of industrial organization.

One assumption of a perfectly competitive market is the absence of transaction costs, or at least the presumption that such costs are sufficiently neutral that they can be neglected. As a consequence, the study of transaction costs has usually proceeded by relaxing the assumptions of a perfectly competitive market. Cootner and Ulen (1996, p. 86), for example, identify ten factors affecting transaction costs: standardization, property rights, number of parties, friendliness of the parties, familiarity of the parties, reasonable behavior, speed of exchange, contingencies, monitoring costs, and the cost of punishment. While all of these items have an undeniable impact on transaction costs, they may arise due to a variety of unexplained causes, many of which can be linked to imperfect competition. They do not explain or predict why one industry has higher transaction costs than another.

On the other hand, the approach suggested here – linking transaction costs to the physical complexity of the commodity – provides a predictive tool. Such analysis can begin with the obvious: dangerous commodities will have a higher transaction cost than those with negligible risk; heterogeneous commodities have higher marketing costs than standardized products, etc. But this approach also allows a more subtle analysis. It can also explain why wholesale trading in electricity is bound to have higher transaction costs than trading in natural gas or oil. Such observations are predictive and can be empirically verified; thus this approach provides a testable hypothesis.

4. Revolution and Counter-Revolution in the Oil Market

4.1 Stabilizing Force or Monopoly?

For most consumers, petroleum markets appear relatively unchanged. Gasoline is still purchased from local service stations, heating oil is delivered on demand and most, if not all, of the familiar brand names dominate sales. The modest changes in retail marketing, however, mask much more substantive changes in industrial structure and wholesale marketing. The petroleum industry has been transformed as it has adapted to OPEC's control over resources, two spectacular supply shocks, new marketing institutions, new technologies, depleting resources, and regulatory metamorphosis.

The most obvious change to both retail and wholesale markets concerned price stability and perceptions about the industry's market power. "For decades before World War II, the world price was the U.S. price plus freight," (Adelman 1995 p. 4). Following the war prices fluctuated within a narrow range (except for the Suez Crisis) until the Arab Oil Embargo of 1973. Price stability during this period is frequently credited to an industrial structure that had endured since the breakup of the Standard Oil Trust. Critics of the industry maintained that world oil markets were dominated by a handful of Major vertically integrated companies. In the view of such critics, the Major oil companies could control the flow of oil from their own fields, by their own tankers to their own privately held refineries, and on to strictly-controlled marketing outlets. The object of such control was to keep resource prices low and retail prices high, ensuring above-normal rates of return on the vast capital investments required to supply oil and run the world's machinery.

John Blair (1978 pp. 27-28), a voracious critic of the oil industry, put it this way:

> The nature of the industry is such that stability of price requires almost complete control over markets, since as has been repeatedly demonstrated, it takes only a relatively small amount of "uncontrolled" supply to disrupt the market...When independents secured concessions in Venezuela and later in Libya, it did not take long for their supplies to depress world prices. Control over supply through jointly owned operating companies and restrictive long-term contracts has had to be supplemented in a variety of ways. In these efforts, the major international oil companies have received the enthusiastic cooperation of governments in both the producing and the consuming countries. By means of a web of cartel arrangements set up in most of the world's consuming countries, they secured control over most of the world markets.

In contrast, Professor Adelman (1995 p. 5) wryly comments: "By 1970, the oil companies were in the saddle only as well-paid jockeys; the oil exporter nations owned the horses and collected the winnings." Although it lasted decades, the period of price stability ended abruptly. In 1974 and 1979 (when faced with high prices and long queues), motorists would fondly recall the days before OPEC when there was full service and enticing side benefits. In the view of OPEC officials, however, the oil industry's vertically integrated structure had been nothing more than a means to transfer vast natural resource wealth from the poor to the rich.

Surprisingly, in the rapidly growing market of the 1960s, profits in the petroleum industry were not extraordinary. The petroleum industry's return on invested capital was middle-of-the-road for the manufacturing industries, and below drugs, tobacco, chemicals, etc. (Johnson et al 1976 p. 101). Blair's comments are not, however, totally unfounded. He was reflecting a deeply held belief on the part of many independent oilmen and industry critics since the days of John D. Rockefeller.

Paul Frankel (1969 p. 76) writing well before John Blair, offered a strong counter-opinion to the allegations, simultaneously explaining and defending the oil industry's structure:

> As pipe-line economics hinge mainly on constant flow, i.e., on steady and concentrated supply and demand, nobody was to benefit more by the advent of pipeline transport than the biggest operator. A pipeline made sense only as part and parcel of an adequate and balanced organization.
>
> LEVIATHIAN
>
> For the first time we see the advantage of the complete organization, of what we now call the integrated firm.

Frankel (1969 p. 76) went on to quote W.S. Farish, a former President of Exxon, who in 1969 described the petroleum industrial structure:

> The conditions under which integration are desirable are: (1) large volume of business in a single commodity group; (2) highly specialized production, manufacturing, transportation, and distribution techniques; and (3) substantial advantages (at some stages) in large-scale operation. These conditions characterize the petroleum industry, and it follows therefore that the relations between any one of the stages of the industry and the other next to it are peculiarly close. The refiner needs to be assured of his market. The marketer needs to be assured of his supply.

Professor Adelman (1995 p. 44) provides an updated view of the motives for vertical integration, tying it to small markets and high transaction costs:

> Production was too risky without an assured outlet, known as 'finding a home for the crude.' Refining was too risky without an assured supply of crude…The obvious solution was vertical integration. Refining-marketing joined with production by merger or by branching upstream into production or downstream into refining. Vertical integration saved the transaction costs of incessant search and negotiations. Buyer-seller contracts do it today, and could have done it then, if there had been an appreciable number of buyers and sellers. But the need to enter both production and refining-marketing together made entry more difficult. Thus inertia preserved a concentrated integrated structure even as the original cause – economies of scale in small markets – slowly disappeared.

Petroleum producers and refiners can be paranoid about markets and market access. This is because for most of the century there was a surplus of oil and gas supplies. Surplus supply in a capital-intensive industry is a painful experience for firms, since competition can drive prices to variable costs, well below the level necessary to recover capital. Once a company loses market share many believe it will be forced to "buy it back" through fierce price competition.[9] The seller's markets of 1973-74 and 1979-80

[9] For example, in 1983 ARCO had a surplus of Alaska crude oil. In order to save the transportation cost of shipping it to the Gulf Coast, they expanded refinery capacity and steeply discounted gasoline prices.

were short-lived. Bradley (1989 p. 3) described the broader experience as follows: "For the industry as a whole, a buyer's market for crude oil persisted from the mid-1920s to the early 1970s, interrupted only by World War II and its aftermath. The experience that began in 1973 was a sharp break from over four decades of peacetime experience."

The petroleum industry is a cyclical industry. Oil supplies are dependent on a number of variables, some of which are accidental and most of which are unpredictable. Great discoveries are seldom anticipated and if new supply is large enough it can send prices to unexpected lows. During the boom and bust periods of discovery in Texas and Oklahoma early in the century, oil was often said to have run loose in the streets, not worth the cost to prevent spillage. Likewise, new technologies can open up completely new frontiers. Most recently, the development of deepwater drilling and production has led to extensive oil development in the Gulf Coast, Brazil, and West Africa.

For most of the twentieth century natural gas was a waste product. Travelers recall driving through Texas at night in the 1950s with the sky lit up from natural gas "flaring." Similar images still abound in the Middle East. Gas produced in association with oil has often been a waste product to be burnt on site. The development of high-pressure gas transmission lines, combined with the incremental expansion of local distribution companies, provided an outlet for surplus gas in North America, but the capital investment to complete the connections was substantial.

When faced with high capital costs, economies of scale, and chronic (and unpredictable) surplus, an industry is likely to erect defenses. There are three classic responses to such circumstances. First is a natural tendency for consolidation, frequently blocked by anti-trust laws, particularly in the United States. Second, over the years, companies will lobby for government protection – protection from cheap foreign imports and from their own competitive instincts. Third, companies may seek to moderate the impact of unpredictable prices and markets through diversification. Initially, the most popular form of diversification in the oil industry was vertical integration – the combining of all segments of the industry, exploration and production, transportation, refining, and marketing - into a single company.

Statistically, the oil industry is relatively unconcentrated compared to many industries. This has often been the consequence of government intervention.[10] And, since the rise of OPEC the entry of national oil companies such as Saudi Aramco (originally named the Arabian American Company but now the national oil company of Saudi Arabia) have further diversified the industry. In the consuming nations, particularly the U.S., anti-trust officials have sought to prevent excessive market share of any one company by carefully monitoring mergers and acquisitions. However, much of the reason for public

---

The consequence was a West Coast price war in which ARCO doubled its market share, mostly at the expense of Chevron. It took years for Chevron to rebuild its position in the market.

[10] It is interesting to note that the market for some types of pet foods, breakfast cereals, and other retail products frequently have market concentration ratios five times that of gasoline or heating oil; this is why critics of the industry such as Blair have focused on more nebulous indices related to "market control" and the like.

apprehension about the practices of oil companies arises from their history and consumers' perceived dependence rather than from conventional techniques of analysis.

Table 4.1

Super Majors Oil Production and Refining Capacity in 2002

	Million B/D Oil Production	Million B/D Refinery Capactiy	Percentage Production	Percentage Refining	Self-Sufficiency
ExxonMobil	2.5	6.3	3.4%	7.5%	39.7%
Royal Dutch Shell	2.4	4.4	3.2%	5.2%	54.1%
BP	2.0	3.5	2.7%	4.2%	57.1%
Chevron/Texaco	1.9	2.3	2.6%	2.7%	82.6%
Saudi Aramco	9.5		12.8%		
World	73.9	83.9	100.0%	100.0%	

Source: Annual Reports
BP Energy Statistics

Table 4.1 illustrates the point in stark terms. The most recent wave of consolidation reduced the number of Major oil companies. The four largest are now sometimes referred to as "super" Majors and yet their relative positions in the oil market are almost trivial as compared to the 1960s. The largest, ExxonMobil, has only 3.4% of world oil production. As a group, the four super Majors have less oil production capacity than Saudi Aramco. ExxonMobil is the world's largest refiner, but controls only 7.5% of the total capacity. These market shares fall far short of the level that usually concerns anti-trust officials, although shares do vary from one region to another.

The extent of market power exercised by the Standard Oil Trust remains a highly debated topic, despite its antiquity. As Yergin (1991 p.101) chronicles, Ida Tarbell's series of exposés on John D. Rockefeller and his Trust began in *McClure's* in November 1902. It was a broad-based attack on the company, and whether intended or not, it tainted the entire industry. Tarbell's articles launched a variety of private and public investigations, culminating in the formal breakup of the Standard Oil Trust in 1911. The issue was thought to be resolved until research by McGee (1958) argued that the historical record of grievances by Standard Oil had been vastly exaggerated. In particular, he sought to refute the claim of predatory pricing, concluding that Standard Oil did not drive out weaker competitors. Instead, it consolidated the industry through reasonably priced mergers and acquisitions. More recently, Granitz and Klein (1996) have counter-argued the McGee point, demonstrating that the Trust did exercise extensive market power. In any case, it is claimed that Standard's legendary monopolistic abuses arose from coordinating and manipulating railroad shipping interests, rather than direct control over refining or oil production.

The debate over the culpability of Standard Oil is an interesting one, because the breakup has had such an impact on the implementation of American anti-trust laws and a resounding impact on the worldwide structure of the petroleum industry. The Trust was broken into 38 separate companies, many of which went on to become larger than the original parent (Sampson 1975 p. 39). Since then, the U.S. Department of Justice (DOJ) and Federal Trade Commission (FTC) have closely monitored market shares of each segment of the industry. When mergers are proposed, these agencies review the proposal and frequently force the parties to divest assets that would result in undue concentration. This has been an ongoing process for almost a century and has an obvious impact on the industry. If a U.S. company seeks to grow in size it must do so outside the United States or integrate into non-competing economic activities.

In the United States, protection of the domestic oil industry took two forms. The Texas Railroad Commission implemented control over oil production in Texas, and nearby oil-producing states implemented close imitations. The Commission developed a system of "pro-rationing," based on the production capabilities of each field within the Commission's jurisdiction. A total for the oil producing regions was calculated and compared to estimates of market size. Each producer was then allocated a share of the market, a ration of production. The state-based pro-rationing system constrained supply, but left it to federal regulation to stabilize price. In March 1959, President Eisenhower implemented the Mandatory Oil Import Program (MOIP). The purpose of the program was to constrain foreign imports by allocating import rights. The effect of the program was to prop up domestic prices somewhat higher than prevailing world levels. MOIP was dismantled in 1973, when world oil prices surpassed domestic oil prices due to President Nixon's price stabilization program.

Companies faced with significant constraints on their activities – anti-trust laws that prevent horizontal expansion and/or regulatory practices that, intended or not, limit growth – must find other means to cope with market volatility. Diversification, which can take many forms, is the obvious answer.

The oil crises of the 1970s prompted a number of unusual acquisitions. U.S. Steel acquired the Marathon Oil Company. Dupont acquired Conoco. In these examples, large energy consumers were faced with the threat of rising oil prices and sought to hedge their future by acquiring energy companies. Other companies, such as Tenneco, were diversified as conglomerates by design. Tenneco began as a gas pipeline company, but its corporate head office managed a variety of unrelated divisions and subsidiaries, including the Newport News shipyard. Other oil companies, enjoying the sudden windfall of higher oil prices, acquired a host of unrelated businesses. The combination of energy companies as part of corporate conglomerates has not, however, been common. Energy companies are capital intensive and highly specialized; the benefits obtained from diversification can be more than offset by inept management. Most of the 1970s experiments failed and the industry has become increasingly specialized.

Following the breakup of the Standard Oil Trust, most petroleum companies chose to vertically integrate. To many it seemed, and still seems, a natural order. In the words of oilmen, oil and gas are "continuous flow industries." Crude oil flows to the surface with each well connected to gathering lines, and the gathering lines are connected to trunk pipelines and storage facilities (necessary to even out the flow and segment the crude oil streams by their quality characteristics). The trunk pipelines connect to marine tanker terminals or directly to refiners. Once crude oil arrives at a refinery, it is blended with other streams to optimize the refinery output of products. The products are shipped by pipeline, marine tankers, barges, and trucks to "tank farms," from where they are delivered to service stations, factories, and homes. Virtually all the oil stored in this process is working inventory: volume necessary to keep pipelines and refineries full. Any interruption in the process can have devastating financial and engineering consequences. "Shutting-in" crude oils in many fields can mean the permanent loss of the reserves; operating a refinery at less than full capacity quickly draws red ink. The economics of these operations can be very precarious. During the soft market of the 1980s, Tosco was forced to close down a reasonably efficient refinery in order to raise cash by selling the working inventories of crude oil and petroleum products.

In addition to presenting a natural way to expand, vertical integration was also perceived as an effective tool for economic diversification. In general, a sudden rise in crude oil prices pushes up the cost of petroleum products. As petroleum product prices rise, demand slackens, which is likely to cause surplus refinery capacity, pulling down refinery margins. A company that owns both crude oil production and refining capacity can offset the loss in one sector with the profits from another. To some extent vertical integration in the petroleum industry is self-stabilizing.[11] Nonetheless, the usual anxiety over market outlets, the normal stabilizing features of vertical integration, and the delicate fabric of protection were simply overwhelmed by the extraordinary rise in crude oil prices from 1973 through 1980 and the consequent reaction of oil producers and consumers. These events created a shock so profound that it completely transformed the industry.

4.2 OPEC Asserts Control

No study of the petroleum industry can escape a discussion of OPEC, but the focus here is narrowed to a key issue – the transition in the ownership and control of the members' natural resources. In the 1970s, OPEC evolved into a price-fixing cartel, but that was never its sole purpose. When the organization was first formed it aimed to aid its members in the restoration of oil and gas ownership rights. Sampson (1975), Blair (1978), Yergin (1991), and others have detailed the manner and methods by which oil companies acquired concession rights in the important oil producing regions of the world.[12] The founding members of OPEC set about to correct what they perceived as a grossly unfair relationship. As of 1960, oil-exporting countries had little or no say over production decisions or pricing; their leaders were placated by gifts and ritual payments, often conveyed directly in the form of gold.

Over a period of two decades, all OPEC's members effectively eliminated the oil companies' property rights over resources. In some cases the assets were simply expropriated. In other cases, the transfer was accomplished by negotiation and "buy

---

[11] This intuitive explanation is hard to demonstrate statistically. *Pacific West Oil Data* tracks West Coast crude oil prices and the estimated profitability of five categories of refineries. The correlation coefficient between the monthly change in crude oil prices and these five estimates of refinery profits varied from –0.14 to –0.32 for the period 1992 through June 2003.

[12] Much of the primary information on the history and operation of the major multinational oil companies was gathered through a set of hearings held by the Subcommittee on Multinational Corporations of the Committee on Foreign Relations of the U.S. Senate during 1974. John Blair was staff aid to these hearings, which were chaired by Senator Church of Idaho. It is worthy to note that Senator Church lost his next election, in part because the citizens of Idaho were more concerned about the local economy than the exploits and chicanery of geologists and engineers in the Middle East.

out." The last major transfer was the transition of Aramco to Saudi Aramco. One side consequence to the ownership transfer has been a significant drop in the quality and availability of basic petroleum data on exploration, development, production and reserves. Campbell (1997 p. 77) has correctly pointed out that important oil reserve estimates from OPEC members are complete fictions. Nonetheless, these data are published uncritically by the *Oil and Gas Journal* and BP, lending them an aura of reliability that can be horribly misleading.

In 1973, OPEC produced 53% of the world's crude oil (British Petroleum, 2002). So when the cartel asserted control over production and pricing it drove a wedge between resources and markets, quickly undermining the companies' comfortable notion of vertical integration. Worries about replacing lost reserves compounded as the Iranian crisis unfolded and most companies discovered that their historic supply sources were at best unreliable.

4.3 Panic over Resources

From the perspective of political and industrial leaders in the developed world, the second great oil crisis following the Iranian revolution was a cardinal event. Virtually no one had expected bloody revolution in the heart of the Persian Gulf. Indeed, the Shah of Iran had been propped up by successive American presidents, including the newly elected Jimmy Carter who, in a rushed trip to Tehran in 1977, toasted the Shah with haunting words: "'Iran is an island of stability in one of the most troubled areas of the world,' he said. 'This is a great tribute to you, Your Majesty, and to your leadership, and to the respect and the admiration and love which your people give you'" (Yergin 1991 p. 672).

In the midst of the Iranian Revolution, cheap oil seemed permanently and irreplaceably gone. Moreover, the source of cheap oil, the Middle East from Iraq to Yemen, was shrouded in political and social chaos. Even if orderly supply could be restored, there was no guarantee it could be sustained beyond the next wave of terror or revolutionary fervor. For political leaders of the time, resolution of the problem seemed to require a massive change in industrial structure and life style, changes that the general public would not willingly adopt and markets would find difficult to accommodate.

Figure 4.1 illustrates the trend of oil prices since 1972. In inflation-adjusted dollars, crude oil prices reached over $70 per barrel at their peak in 1980. This change in circumstances was noteworthy not only due to the peak levels that oil prices reached, but also for the speed at which they increased and the prolonged period at which they remained high. In 1969, four years before the Arab oil embargo, oil prices were deeply depressed. Spot oil prices in the Persian Gulf were frequently less than $1 per barrel. Low quality heavy oils, such as those produced in Kuwait, could find no market at all and the oil was simply shut in. From 1969 to 1972, however, the world economy grew at a sweltering pace. Before OPEC suddenly discovered its market power in 1973, prices of petroleum products in Europe increased substantially (Van Vactor and Tussing 1987). The gasoline and heating oil markets in the United States were also strained. The U.S., however, had implemented a price control program, so excess demand created spot shortages, rather than increasing prices. During the summers of 1972 and 1973, there were occasional gasoline queues in the United States, particularly in tourist and high-growth areas, and many service stations operated at reduced hours.

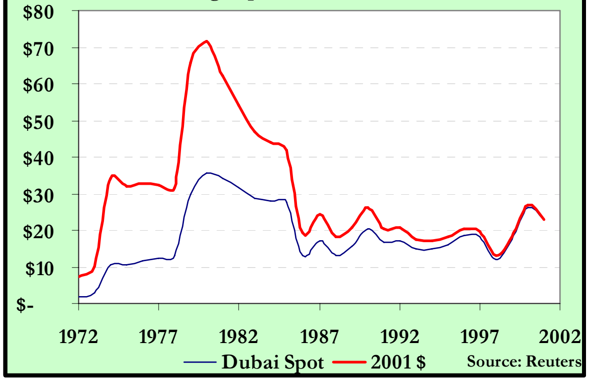


Figure 4.1
Average Spot Crude Oil Prices

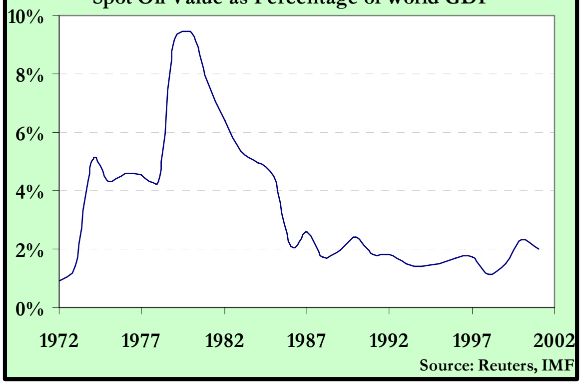


Figure 4.2
Spot Oil Value as Percentage of World GDP

In retrospect, the energy crises of the 1970s seem exaggerated. Nonetheless, they had a remarkable impact, primarily because the period of high oil prices lasted a dozen years, from 1973 through 1985. (In contrast, oil price disruptions during the Gulf War lasted only a few months and are barely visible in Figure 4.1). Visions of a return to pre-industrial pastoral life have now faded (at least as a consequence of "running out of oil.") It is all too easy to gloss over the economic impact of the 1970s shift in the oil market. The sudden shock of oil shortages had a serious impact on natural gas and coal production, power generation, car manufacturing, marine tankers, airlines, etc. Overall, the impact was profound and contributed to the worst economic slump for industrialized nations since the Great Depression. Figure 4.2 provides a thumbnail sketch of the impact by focusing on crude-oil prices as a percentage of world Gross Domestic Product (GDP). Incredibly, the value of crude-oil production (as measured by spot prices) rose from less than 1% in 1972 to over 9% in 1979 and 1980. Such a powerful shift in the relative value of a key commodity cannot help but to cause significant macro- and micro-economic impacts.

The meaning of oil politics and their impact is suitably vague. No one knows for sure what the hundreds of ministerial meetings between and among producers and consumers actually accomplished. The economic impact of a ten-fold increase in the cost of "consuming" oil, however, was much clearer. This was a signal that could not be ignored and although day-to-day movement was imperceptible, the functions of demand and supply went to work. From 1960 through 1973, oil consumption increased 7.6% per year (OECD 1976). From 1973 through 1979, annual growth slowed to 2.1%, and from 1980 through 1985 consumption actually declined by a rate of 1.2% per annum (British Petroleum [BP] 2003). In contrast, the supply of alternative energy, particularly natural gas, gathered speed.

Along with profound changes in oil consumption, the relationship between oil companies and the supplying host countries changed permanently after 1973. Oil exploration in the most prolific basins of the world was quite different from its historical evolution in North America and Europe. Outside the industrialized world, sovereign governments, rather than private individuals, controlled the ownership and exploitation of natural resources. Even within North America and Europe the greatest potential for new discoveries was offshore, in the hands of the public. This created a whole new set of hurdles for development. Obviously, projects would have to meet narrow economic criteria, i.e., the expected value of production would have to exceed expected cost. In addition, however, projects would have to provide direct benefits to sovereign governments in the form of revenue and indirect side benefits of jobs, community action, public works, etc. The new relationship might not have evolved at a revolutionary pace, were it not for the rapidity with which the world suffered oil crises: two in the space of only six years.

The 1973 oil crisis consisted of two parts. First, the Arab members of OPEC declared an embargo against the United States and the Netherlands. These nations were targeted for their support of Israel. Second, along with the embargo, Saudi Arabia dropped crude-oil production by about 5%. This is a relatively insignificant drop in supply, but demand had been growing at nearly double digits and Saudi Arabia was the main source of incremental oil. As a consequence of the supply constraint, crude oil prices lunged from an average of $2.48 per barrel in 1972 to $11.58 per barrel in 1974 (BP 2003.) Very significantly, however, this disruption in oil supply did not target any particular company. Thus the basic structure of the market was not stressed.

The Iranian crisis had the opposite impact. Within a few short weeks, Iranian oil production plunged from about 5.5 million to 410 thousand barrels per day (Energy Information Administration [EIA] website). The drop in supply had an impact on price levels predictably similar to the Arab oil embargo of six years before. During the course of the crisis, average crude oil prices rose from $13.60 per barrel in 1978 to $35.69 in 1980 (BP 1999). The Iranian oil crisis followed on the heels of U.S. natural gas shortages during the winter of 1976-77, which had been accompanied by much hand wringing by President Carter and his entourage. U.S. policy makers set about convincing themselves and the rest of the world that oil supplies were near their peak of production and would steadily decline in future years. [13]

Most significant for the development of energy markets, however, was the shortfall in crude oil that did not fall evenly across all companies in the second crisis, as it had in 1973. BP was one of the companies heavily dependent on Iranian crude oil. They, like any integrated Major, had a host of refineries, marketing outlets and long-term sales contracts that they were obligated to maintain. The integrated structure was not designed to accommodate such a shock. The loss of Iranian crude oil meant that BP had to replace it with other sources of oil and the sources available were from OPEC countries. But if BP and other companies dependent on Iranian oil were to lay claim to these alternative supplies, they would have to bid the oil away from other companies. In turn,

[13] Facts that surfaced after the crisis cast a very different light on events. Although Iranian oil production did collapse, the shortfall was quickly made up by other OPEC producers. However, the unexpected drop in oil consumption growth following the Arab oil embargo had created a huge surplus of tankers. These tankers were drawn out of mothballs and used as temporary storage. As anxiety over oil supplies worsened, more and more oil was added to inventory. During 1979 and 1980, world oil production averaged about 2 million barrels per day - 3% greater than consumption. Following the crisis OPEC, and particularly Saudi Arabia, then had to deal with a massive inventory draw down.

the various OPEC members would have to break long-term contracts with existing buyers. The staid market of the last seven decades was suddenly awash with brokers, merchants, marketers, and buccaneers all seeking to make one big deal.

There was another unusual twist to the Iranian oil crisis. As Iran's oil production declined, Saudi Arabia's increased. The Saudis however, had no desire to find new buyers. They depended on Aramco to produce and market their oil. Aramco was the exclusive province of four American oil companies, Chevron, Exxon, Mobil, and Texaco. Other companies without the Aramco connection scrambled to replace lost supply sources. As a consequence, the oil market devolved into chaos, with secretive prices and a prolonged disequilibrium.

4.4 OPEC's Pricing Schemes

The Iranian crisis did much more than just disrupt supply relationships between OPEC and its buyers; it also effectively undercut the cartel's oil pricing scheme. The primary purpose of the OPEC cartel was, of course, to set and maintain high crude oil prices. This is a difficult task for any cartel, but it is particularly difficult to accomplish for a commodity like crude oil, given its uneven quality and wide geographic dispersion.

Crude oil in its natural state varies from tar to natural gasoline. The "heaviest" oils do not flow at room temperature and must be heated to ship by pipeline. Very light oils, naphtha and natural gasoline, are volatile; mishandling can result in explosions or other potentially fatal fires or accidents. API has devised a method for measuring the density of crude oil. Their measure, API gravity (APIG), is based on a formula arising from specific gravity.[14] Typically, heavy crude oils range from 8 to 20° APIG. Medium crude oils range from 21 to 30° and light crude oils have an APIG of 31° or higher.

Another important element of crude oil quality is the number and types of contaminants mixed in with the crude oil. The most common of these contaminants is sulfur, but some crude oils also contain heavy metals, etc. The industry refers to high sulfur oils as "sour" and low sulfur oils as "sweet."

Most refiners prefer to run light-sweet crude oils, rather than their heavy-sour counterparts, for two reasons. First, light-sweet crude oils cause less wear and tear on refining machinery. Sulfur and other heavy metals are highly corrosive, increasing refinery downtime for maintenance and repair. Second, it is less costly to refine light-sweet crude oils into gasoline, jet fuel, and diesel - petroleum products that have the highest market value.

Refining consists of a series of steps, beginning with simple distillation. As the temperature of crude oil rises, lighter products boil off. The first products to evaporate are naphtha and gasoline, followed by middle distillates. Material left (usually at temperatures over 600° F) is referred to as residual oil and may be sold as heavy fuel oil, or if additional equipment is added to the refinery, it may be upgraded to higher-value

[14] Conversion of specific gravity (SG) to APIG (Schmidt 1951) is given by the formula:
APIG = (141.5 / SG at 60° F) –131.5

products through processes like catalytic cracking or coking. These more complex processing methods, however, require substantial capital investment and have a higher variable cost. In the most sophisticated refineries, the production of residual oil can be completely eliminated. Contaminants in the crude oils typically bond to heavier molecules, which means that during distillation sulfur becomes concentrated in the residual oil. With additional capital investment, sulfur can be removed from residual oil or eliminated entirely in the upgrading process. Since most countries have implemented environmental regulations that limit or prohibit the use of sulfur-laden fuels, low sulfur fuel oils command much higher prices than their sulfurous alternatives.

As the previous points hint, there is a complex relationship between crude-oil values, refining costs, and petroleum product prices. In general, a rise in the price of gasoline and distillates compared to heavy fuels results in a rise in the relative value of light-sweet crude oil. The relationship is often murky, however, because refiners have considerable latitude in the "slate" of products they can produce and in the combinations of crude oils they can run. In fact, refiners have to change the relative proportion of products produced with the season – increasing gasoline production in the spring and summer and heating oil production in the fall and winter.

The complexity of refining different qualities of crude oil in order to optimize a slate of petroleum products is made all the more difficult by the geographic dispersion of crude-oil sources, refineries, and markets. Most important crude oil sources are connected to refineries through the marine transport system. This results in transportation costs that vary from $1 to $3 per barrel, except during periods of extreme market stress. Far more important than the cost of moving crude oil, however, is the timing of shipments. Long-haul crude oils (such as those in the Persian Gulf) can take up to sixty days to move to a market. During such a period of time the relative value of various petroleum products can change dramatically.

Imagine the problem for the OPEC cartel as it tried to sort out and set prices for each of its members' crude oils. Algerian and Libyan oils, for example, are short-haul crude oils, only a few days from the Italian refineries. On the other hand, Oriente, the crude oil from Ecuador, travels by pipeline over the Andes, by tanker to Panama, over the isthmus by pipeline, and by a different tanker to Gulf Coast refineries. The North African crude oils are very light and low in sulfur, while Oriente is of medium density and sulfur content. Somehow all of these variables must be accounted for. Originally, OPEC tried to fix quality and location differentials for each of its members' crude oils. It was rumored that a computer program was designed to take into account all of the variables. Running on the highest speed mainframe computers at the time, it took 24 hours to compute. Cynics observed, however, that by the time program had arrived at a solution, the market had already changed and the results were no longer valid.

With or without input from the fabled computer program, OPEC managed to devise a much simpler and elegant system of price fixing. Robert Mabro (1975-76) first described how the OPEC algorithm worked.

In the first step OPEC set a "marker" price for crude oil. They chose "Arabian Light" as the marker. This is the crude oil from the Ghawar field, the largest oil field in the world, producing on the order of 4.5 million barrels per day. Once the marker price was set, Saudi Arabia was allowed to adjust production up or down as demand for petroleum waxed and waned. The second step was equally simple. Other members of OPEC were to accept fixed production quotas, but could allow prices to vary. In other words, an expected surge in oil demand would have two types of consequences for cartel members. Algeria, for example, could respond to the increased call on its crude oil by increasing price, holding production constant. Saudi Arabia would hold its price constant but allow production to increase. The cartel would meet periodically to assess production rates from each of the countries with respect to the overall price level they set.

As has been pointed out by Mabro (1975-76), Adelman (1995), and other economists, the power of the OPEC cartel rested with its ability to stabilize and maintain price levels after a price rise. In every case an unexpected drop in supply occurred during a period of rising demand. The excess demand sparked an increase in spot prices providing a signal to the Cartel that they could raise official, or term contract, prices. Once all of the Cartel's prices had risen to a new level, demand would slacken. At that point, OPEC depended on Saudi Arabia to cut back production sufficiently to preserve the price level.

The OPEC pricing scheme was clever and from the cartel's point of view it worked admirably well during the period between the Arab oil embargo and the Iranian revolution. It broke apart for the classic reasons: price chiseling among cartel members and new supplies brought on by high prices.

4.5 The Struggle to Fix Prices

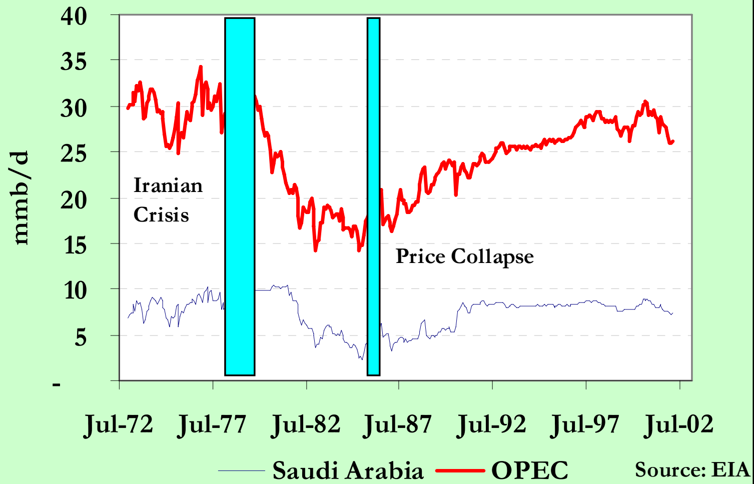
Although the Iranian revolution and subsequent war with Iraq created extraordinary uncertainty and apprehension, it did not directly affect the oil producing states in the southern portion of the Persian Gulf. Kuwait and, particularly, Saudi Arabia maintained their position as dominant producers. In fact, the downturn in Iranian and Iraqi oil further enhanced the importance of Saudi Arabia, who embraced its role as the cartel's swing producer. Higher production combined with record level oil prices swamped the kingdom with revenue, while the notion of "backward bending" supply curves gained sway over the minds of policy makers. According to this theory, the greater the surplus revenue, the less incentive the Saudis had to produce oil; the less they produced, the higher the price and the greater the revenue, and so on, and so forth.

Higher oil prices, however, had two substantive impacts on the world economy. The first was to remove purchasing power from consumers. For the U.S., Japan, and European countries this meant less to spend on other goods and services; instead, the cash flowed to OPEC, whose members did not know how to spend it. Second, higher oil prices had an immediate impact on the calculation of the consumer price index (CPI) and perceptions about the rate of inflation. In the U.S. the rate of increase in the CPI peaked at nearly 12% per annum. In response, the Federal Reserve Board tightened the money supply, driving interest rates to record highs and severely curtailing investment.

As a consequence of lower consumer spending and lower investment, the industrialized world experienced its lowest period of growth since the Great Depression.

Although it took several years to unfold, negligible economic growth and high oil prices had a predictable impact on the demand for OPEC crude oil. This surprised everyone, because the price increases of 1973-74 did not seem to have had much of an effect (this was, in fact, not true; oil demand continued to grow after the price rise, but at a much slower rate, approximately half the rate of the previous growth). Figure 4.3 illustrates the trend. After 1980 the demand for OPEC oil declined. Energy demand growth continued, but was focused on other fuels, particularly natural gas. As the swing producer for the OPEC cartel, Saudi Arabia took the brunt of the drop in demand.

Figure 4.3
OPEC and Saudi Arabia Oil Production



From 1975 through 1978, the Saudis had experienced a drop in oil demand and had demonstrated their willingness to stabilize the market by cutting production. Just before the Arab oil embargo, the Saudis were producing about 8.5 million barrels per day and were scheduled to ramp up production as demand warranted. In a 1971 forecast of the world oil market, Shell Oil projected Saudi production to reach 20 million barrels per day by 1980. If the investment in infrastructure had been made, such a rate of production would have been feasible. The Ghawar oil field alone had estimated reserves of 78 billion barrels, approximately ten times those estimated for Prudhoe Bay and Alaska's North Slope. At its peak, North Slope oil production reached 2.1 million barrels per day. Despite its potential production in 1975, Saudi Arabia appears to have been more than willing to cut back output and revenue from its August 1973 level.

The demand for crude oil is seasonal due to the surge in demand for space heating during the northern hemisphere's winter. As a consequence, the market picks up in the late summer and fall and sags each spring. Following the end of the Arab oil embargo Saudi production rose, peaking at 9 million barrels per day in October 1974; it then declined in order to balance the market, bottoming out in April 1975 at just under 6 million barrels per day. At the time, various economists, including Milton Friedman, predicted that OPEC would not maintain discipline among its members and prices would collapse.

Oil demand proved to be more resilient that expected, however, and Saudi production quickly rose, reaching nearly 10 million barrels per day in the winter of 1977. There were multiple reasons for the resurgence of demand. First was the lag between investment decisions and forthcoming production. The embargo and rise in prices was a jolt to companies dependent on OPEC oil. Their first reaction was to look elsewhere for new reserves. However, the cycle for exploration and development is much longer than two years. So, even with successful projects in alternative provinces, incremental oil supplies would have to come from OPEC. Second, the five-fold increase in wholesale crude oil prices did not initially have that much impact on demand. As explained in Chapter 2, in the short-run energy demand is highly price inelastic. Moreover, the net increase in the prices of many petroleum products (laden with excise taxes) was not substantial. Third, the demand for heavy fuel oil was not significantly impacted by the price increases, because prices were still below the threshold at which natural gas and coal were obvious substitutes. Fourth, the winter of 1976-77 was exceptionally cold and, as to be explained in Chapter 5, North America experienced serious shortages of natural gas brought on by U.S. price regulations. Thus the only available substitute was imported oil.

Table 4.2
Significant Price Changes

Date	Initator	Marker for Saudis	Prices for Others
Oct-73	Gulf Six	$3.65	
Dec-73	Gulf Six	$11.65	
Jan-74	OPEC	$10.95	
Nov-74	OPEC	$11.46	
Oct-75	OPEC	$11.51	
Jan-77	OPEC	$12.09	$12.70
Jul-77	OPEC	$12.70	$13.66
Jan-79	OPEC	$13.34	$14.10
Apr-79	OPEC	$14.55	What market will bear
Jun-79	OPEC	$18.00	What market will bear
Dec-79	Saudi Arabia	$24.00	What market will bear
Jan-80	Saudi Arabia	$26.00	What market will bear
May-80	Saudi Arabia	$28.00	What market will bear
Aug-80	Saudi Arabia	$30.00	What market will bear
Nov-80	OPEC	$32.00	$36.00
Nov-81	OPEC	$34.00	What market will bear
Mar-83	OPEC	$29.00	$29.00
Jan-85	OPEC	$28.00	Heavy/Light Adjsutment
Aug-85	Saudi Arabia	Net back Pricing	What market will bear
Jan-86	NYMEX	Market Collapse	What market will bear

Source: Deutschbank, EIA, OPEC Bulletin, Middle East Economic Survey

The reason for the 1973-74 oil price increases had less to do with the embargo than with the accompanying production cuts from Saudi Arabia and other Gulf states. During the Iranian crisis (with the exception of a brief three month period), the Saudis did the opposite. Saudi Arabia increased production and attempted to moderate price increases. This is something of a controversial point. Morris Adelman (1995) and others have argued that the one million barrel per day cut in April 1979 was aimed at engineering substantive price increases (as cuts in 1973 had done). However, the Saudis claimed otherwise. Saudi Arabia raised oil production to its maximum in July 1979 and, as summarized in Table 4.2, it was not until December 1979 (retroactive to November) that they made a significant increase in official prices, which then reached $24.00 per barrel. That price increase occurred only after the spot market had sustained price levels in excess of $30 per barrel over a prolonged period.

Table 4.2 also illustrates another crucial point. During most of the period when OPEC appeared to control oil prices, there was substantial dissention among the cartel's members. The second column of the table presents the apparent decision maker responsible for the key price change. At times it was clearly an OPEC decision, at other times it seemed to be the sole responsibility of Saudi Arabia, with the possible cooperation of a few close-by Arab states. As explained earlier, many OPEC members took immediate advantage of the spot market when prices were rising. Nonetheless, they and the rest of the world treated each OPEC meeting as a stellar event – ministerial delegates were frequently media stars. The media attention focused on the cartel's official price structure, because the price set for Saudi oil ended up fixing prices for about half of the Cartel's output and this was enough to set world oil price levels in most circumstances.

In the early days of the Iranian crisis, Saudi Arabia was cautious about raising prices. As the swing producer they had endured the slide in the oil market in 1975 and 1976 and were well aware that if the cartel set prices that were too high it would mean a drastic loss of market share following the crisis. Virtually all of the other cartel members pressed the Saudis (and the few Gulf States that coordinated production and pricing decisions with them) to raise prices. They did so because higher official prices allowed them to demand higher prices with little or no fear of losing their market.

As the Iranian crisis compounded, Saudi reluctance to increase crude oil prices wilted. During 1980 the price was increased multiple times, until at OPEC's 59th meeting in Bali, Indonesia the cartel, according to its press release, agreed: "To fix the official price of the marker crude (Arabian Light 34° API ex-Ras Tanura) at a level of U.S. $32 per barrel." Further, "prices of OPEC crudes may be set on the basis of an oil price ceiling for a deemed marker crude of up to U.S. $36/barrel." The use of different "official" and a "deemed" marker prices was OPEC's tacit recognition of a market reality. With a few exceptions, its members were simply charging what the market would bear.

When the spot market turned soft in 1981, motives changed and the stated object of the cartel became "price unification." The notion among members was to unify the deemed and official markers. Again, those members less likely to reduce production in a soft market pressed the Saudis to increase prices, i.e., to unify the price structure upward instead of downward.

The history of OPEC's internal contortions over price has normally been viewed as a conflict between members with low oil reserves and large populations (hawks) as opposed to members with high oil reserves and low populations (doves). The doves, of course, had far less need for immediate revenue than the hawks. This view, however, ignores a very important aspect of OPEC's behavior – the relationship of its members to the Major oil companies that discovered and developed the resources. In many cases these companies were still present and had substantial control over the production and marketing of the crude-oil streams. This was especially the case in Saudi Arabia, which had been slow to divorce itself from the four Major companies that produced oil in the Kingdom.

The Saudis had good reason to maintain a strong relationship with Aramco, because it was the instrument that Saudi policy makers used in the 1970s to stabilize the market. The arrangement was based on a complex system of mutual interdependence, an extension of the North American system of posted prices.

Recall that the object of a posting company was to set an acceptable price to the counter party - one, however, that smoothed out random fluctuations - a long-term deal rather than a spot relationship. At times this meant a better deal for one party than the other, but if the posting company set prices correctly, benefits and costs would cancel out in the longer term. Moreover, the smoothing function had some benefit to both buyers and sellers; it did moderate random fluctuations and reduce transaction costs associated with constantly changing prices. Given the broad diversity of crude oils, the determination of relative prices was not a trivial matter and constant fluctuations in the marker compounded the problem.

In any case, when OPEC assumed responsibility for setting world oil price levels, something like the posting system was retained; by design or luck, the Saudis seized on the arrangement as a means to manage their role as the swing producer. During market shortages, the Saudis would set the official price of their crude oil below the spot market. The Aramco partners would receive the net benefit, since they were integrated companies. As refiners, Exxon, Mobil, Texaco, and Gulf were competing with other refiners forced to buy much more expensive crude oil, but all received similar prices for the products they refined. During periods of shortage, the industry referred to these circumstances as the "Aramco advantage."

Saudi Arabia was careful about ensuring that the benefits of low-cost crude oil accrued to the four companies. During the Iranian crisis, the Kingdom's oil minister, Sheikh Zaki Yamani, issued what became known as the Yamani edict. The edict prohibited the Aramco partners from reselling Saudi crude oils at anything but the official price. This meant that these companies would refine Saudi crude oil before purchasing alternative feedstocks. Moreover, given their dependence on Saudi crude oil, the Aramco partners would continue to design refining and logistic facilities specific to Arabian Light and would not seek other sources of supply (refineries are usually most profitable when running the crude oils for which they are designed; thus, if necessary, they will pay a premium in the short-term to obtain the right feedstock). The Yamani edict was, in part, aimed at locking in Aramco purchases. When the market reversed, the Aramco partners would prefer to continue to buy Saudi crude oil due to the specialized investments they had made in refineries and logistical facilities.

Saudi Arabia was also buying goodwill. Obviously if a buyer receives special advantages for a prolonged period, the seller will expect something in return. More than that, however, Aramco and Saudi Arabia were locked in a perceived long-term relationship in which the balance of benefits and costs would swing back and forth. The Aramco partners could husband the high profit levels and be prepared to pay a premium over other supplies when the oil market grew weak. Even when taking a loss an Aramco company would be reluctant to walk away from the Saudi relationship, because in future shortages, they would once again benefit.

As the oil market softened in 1981, both Saudi Arabia and Aramco had reason to believe that prices could be stabilized and the market protected as it had been in 1975-76. Neither was prepared for the length of the downturn or the amount of price discounting other OPEC members would pursue. They were also misled by the apparent price inelasticity of petroleum product demand. Despite an approximate doubling of retail prices the demand for gasoline, heating oil, jet fuel, and other premium products remained strong. Only the demand for heavy fuel oil had shown weakness.

In the weakened market, the Aramco partners were given "lifting" requirements. That is, they had a minimum volume of Saudi crude oil they were required to purchase at official prices or they would jeopardize their relationship with the Kingdom. Saudi Arabia was willing and able to reduce production to within certain limits, so the first months of the weak market passed largely without incident. They were also willing to press fellow members of OPEC to lower official prices in order to stay the rapid decline in the demand for the cartel's oil.

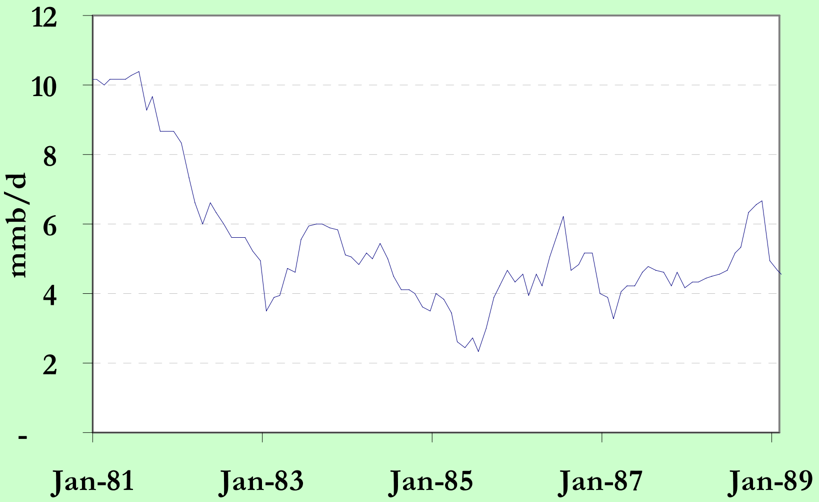
The first big test of OPEC and the Saudi-Aramco relationship came in the spring of 1982 as consuming nations drew down the massive inventories of crude oil and petroleum products they had built up during the crisis. By May, Saudi production had plunged over 40%. Through much of 1981 the industry had been waiting patiently for the cartel to make a final realignment of prices – to raise official Saudi prices and lower the deemed marker. In November 1981, OPEC finally made its move and the official price of Arabian Light crude oil was raised from $32 to $34 per barrel. At $34 per barrel, official prices now stood three times higher than they had before the Iranian Crisis began; they were over thirty times higher than spot prices in the Persian Gulf a decade before. Virtually no one expected another increase and this meant it was time to cut costs by reducing oil inventories to the minimum necessary. The resulting run on the market surprised everyone: from October 1981 to April 1982 the demand for OPEC oil plummeted another 20%, after having fallen by about one third over the previous two years.

OPEC and Saudi Arabia were in for further surprises, as the industrialized nations suffered their worst economic downturn since the Great Depression. The Saudis were forced to retreat from $34 crude oil and in March 1983 after weeks of negotiation, OPEC dropped the official marker price to $29 per barrel (Adelman 1995 p 213).

In 1984 Saudi Arabia shifted to a complicated scheme of selling a market basket of its crude oils, fixing proportions of heavy and light oils. They claimed to be seeking increased production of heavy crude oils (high oil prices had caused a substantial drop in heavy fuel oil demand, resulting in a higher demand by refiners for light crude oils so they could manufacture a greater proportion of gasoline, jet fuel, and diesel). One purpose of the arrangement, however, was to obscure the price being charged for the marker crude oil. In short, it was a complex way for the Saudis to discount prices from other OPEC producers and moderate the impact of their role as the swing producer.

Along with obscure maneuvering over types of crude oils sold and associated prices, the Saudis continued to scold the Aramco partners for under lifting, and the Aramco partners continued to lecture the Saudis about getting prices down. The Aramco advantage of a few years before had quickly turned into a disadvantage as the market grew weaker and weaker. The Major oil companies' ownership of crude oil reserves outside OPEC was limited, so the advantage of high crude-oil prices was more than offset by the disadvantages of compressed or negative refinery margins. Moreover, OPEC was not the only entity pressing for a greater share of the oil wealth. Norway, the U.K., and Canada all raised taxes and associated government take. Even the U.S. implemented a windfall profits tax on domestic crude oil. Saudi Arabia was confronted by the classic problems of managing cartel pricing. The Aramco partners may have had no motive to see the oil market collapse, but they had every motive to pay no more for crude oil than what other refiners had to pay. The dilemma was that if Saudi Arabia reduced prices, other cartel members followed suit, often with even larger discounts.

Figure 4.4
Saudi Arabia Oil Production



Source: EIA

Figure 4.4 is a microscopic view of Figure 4.3. As the significance of the Iranian revolution diminished, Saudi production remained high - over 10 million barrels per day through August 1981. Once the Saudis realigned prices at $34 per barrel, however, production began to drop (as it had done in 1975). This time, however, the plunge in demand was much greater. By February 1983, the call on Saudi oil had fallen to 3.5 million barrels per day, over a 65% decline. Price adjustments and rising winter demand temporarily restored production, but by January 1984 it was declining once again, despite the Kingdom's machinations aimed at restoring market share.

As the summer of 1985 approached, Saudi Arabia was faced with demand for their crude oil at less than 2.3 million barrels per day; the Kingdom would have to do something dramatic if it were to retain any presence in the market. Since over 1 million barrels of oil per day were required for Saudi refineries, it was evident that the Aramco partners were not living up to the implicit bargain. If the trend were to continue, in a few years Saudi Arabia would be forced to buy oil in the world market in order to stabilize prices, an unhappy prospect for a nation with more crude-oil reserves than any other country in the world.

Saudi Arabia had no choice but to abandon its role as the swing producer and by giving up the role, any pretense of controlling the level of oil prices. The Saudis did not, however, wish to confront other OPEC members with its policy change. Instead, they chose to work through the Aramco partners. Together, Saudi and company officials developed a "netback" pricing scheme. Formulae were devised that based the price the Saudis received for their crude oil on prices of various petroleum products that were expected to be refined from it. The scheme gave the Saudis two immediate advantages. First, it obscured the price they were receiving for their oil, making an immediate price war with other OPEC members less likely. Second, it guaranteed the Saudis a minimum market share. The Aramco partners retained a huge refining capacity and a large market share of the petroleum product market. With netback pricing formulae they had no fear of being undercut by lower-priced producers; thus it made sense to substitute Saudi oil for spot purchases.

The Saudi shift to netback pricing was well timed - just before the buildup of winter demand. Competitors were also slow to react. In the fall of 1985 spot prices actually rose slightly. But a more careful look at price patterns revealed a disturbing signal – the newly-established crude-oil futures market at NYMEX was in serious backwardization: prices for oil delivered in January, February, and March 1986 were substantially below December 1985 spot prices. In short, traders foresaw a substantial price decline, but no one expected a wholesale market collapse.

When prices began to decline in January 1986, the descent accelerated quickly. Netback pricing had an advantage in that it obscured the price of Saudi crude oil; it had a disadvantage in that it tied Saudi prices (and thus OPEC prices) directly to the market. Although precise information on the Saudi Arabia netback contracts has not been made public, it is understood that pricing was based on spot petroleum product prices, rather than actual Aramco retail prices. This made crude oil pricing highly volatile and tied to daily markets.

The collapse of crude oil prices had a major impact on the OPEC cartel and for a while many questioned its survival. For years, most of its members had been unwilling to share the burden of reduced production. From 1986 on, OPEC abandoned any pretext of setting a marker price, instead focusing on setting production quotas for its members and targeting a range of acceptable prices.

In its official review of market conditions of 1985, OPEC (1986 p. 12) commented:

> As subsequent events were to show, December marked the beginning of a period of severe pressure on spot crude prices. The main factors were already apparent: increasing world production at a time of replete inventories, mild winter weather and reduced incentives for refiners to purchase crude on a spot basis. A time of reckoning was at hand for those who had too long presumed upon OPEC's goodwill in shoring up oil prices without help from most other producers.

There was indeed a shortage of goodwill in 1986 as traders swarmed to the new crude-oil contract established at NYMEX. In the view of many OPEC analysts, the new exchange was responsible for driving prices down. Put more correctly, however, traders at NYMEX facilitated a quick drop in oil prices caused by too much supply and too little demand: the consequence of the cartel's own behavior. The market now reacted immediately to news about market fundamentals – weather, inventory levels, production rates, economic performance, etc. This was a substantial institutional change in the way the oil market worked and it had come about in only three short years. Before turning to the role of the NYMEX futures market in determining oil prices, however, it is necessary to describe the development of the spot oil market. Without the growth of active daily trading, a successful futures market would not have been possible.

4.6 The Development of Spot Markets

Until a few years after the Iranian revolution and accompanying energy crisis, most refining and marketing companies viewed wholesale crude oil and petroleum product spot markets with deep suspicion. Spot trading was frequently referred to as the "Rotterdam market," since much of the trading concerned shipments of crude oil and petroleum products in and out of the Netherlands refinery complex. Until 1979 volumes in this market were small, from 3% to 5% of the total flow (GAO 1980 p. 2). In contrast, by 1986 the total volume of spot trades had grown to 85% of total physical trade (Fesharaki and Rzavi 1986).[15] Originally, the market's sole purpose was to balance refinery flows; it was not intended to be a primary source of supply. When Iran's oil production fell to a trickle, however, suddenly companies like BP had to make large volumes of spot purchases in order to keep their refineries working and contracts fulfilled. This was not just disruptive to companies now dependent on spot oil supplies, it was also disruptive to OPEC's system of price administration. As noted, all of OPEC's members except Saudi Arabia were allowed to set prices at whatever the market would bear. The Saudis had to wait for an official OPEC meeting and this, alone, created extraordinary tension in the cartel.

The incentive for higher or lower oil prices did not fall evenly across OPEC's members. The cartel's hawks sought high prices; they needed the money now and saw little or no benefit from oil production in the next century. On the other hand, the doves, countries like Saudi Arabia and Kuwait, had vast oil reserves and sparse populations. They took the long view and feared that high oil prices would push the industrialized world into developing alternatives to petroleum, thus destroying the future value of their reserves.

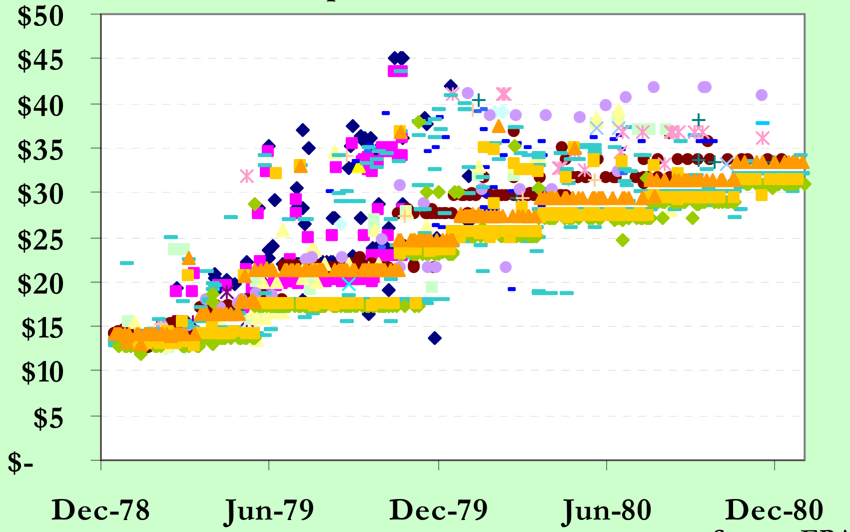
---

[15] This figure undoubtedly reflects long-term contracts indexed to spot crude oil or petroleum product prices, such as the "net-back" contracts that Saudi Aramco entered into in 1985.

As the 1979 oil crisis worsened, the hawks took every opportunity to raise prices. They increased "official" prices to standing customers and diverted cargoes to the spot market, where frequently they could achieve price premiums of 50% or more. Saudi Arabia, on the other hand, generally sought to moderate oil prices. They increased production and argued vehemently against excessive increases in the marker price.[16]

OPEC's behavior from 1979 through 1980 had unintended consequences. The first and most obvious was market chaos. As noted earlier, crude oil was an extremely important commodity to the world's economy. With such an important role on world economic output, it is hard to imagine how this market could descend into disequilibrium for a period of over twenty-four months. Data now available make clear just how befuddled the market really was. Figure 4.5 presents a scatter diagram of the prices of 21 OPEC crude oils. Some of the price variation is due to market fundamentals; relative price differentials are necessary to pull any market back to equilibrium. The price differences presented in the diagram are much greater than those in other periods and are largely unexplained.

Figure 4.5 -- Prices of Various Crude Oils Imported to the U.S.



Source: ERA

The data in Figure 4.5 represent prices paid by U.S. refiners for cargoes intended for import.[17] Most of the purchases were directly from the OPEC countries, but many were purchased on a secondary market from other oil companies and brokers. A market such

[16] Interpretations of Saudi behavior during this period vary, as the Saudis themselves were frequently inconsistent. The Kingdom, for example, decreased production by 1 million barrels per day from April through June 1979. Realizing that it may have erred, production was then increased to its maximum in July 1979 and remained high through the remainder of the crisis. Professor Adelman (1995) argues that the maneuvers were purposely aimed at increasing prices.

[17] U.S. oil companies importing crude oil during this period were required to report prices, volumes, and other pertinent information to the Energy Regulatory Administration (ERA). These data were reported on form ERA-51 and obtained years later from the federal government under the Freedom of Information Act.

as that depicted in Figure 4.5 opens up enormous opportunities for arbitrage. In addition to a substantial increase in trading by existing oil companies, a variety of commodity brokers entered the market. These included firms like Phillips Brothers (Phibro), Mark Rich and Company, etc. Many of these companies would go on to expand energy trading to natural gas and electricity as these markets opened. Intended or not, OPEC provoked the structure that would undo their system of administered pricing.

Trading records on the crude-oil market from the 1979-80 period are limited, which makes comparison to today's market difficult. Despite the limitations there are some comparisons that reveal striking differences, suggesting the depth of inefficiency in the oil market during the Iranian disruption. Recall that OPEC had a benchmark or marker price, Arabian Light. Prices of other crude oils would be expected to fluctuate around the marker – to vary with the seasons, changes in transportation costs, and changes in relative petroleum product prices. Energy futures markets also depend on the concept of a marker price. Traders can use futures contracts to hedge as long as the "basis risk" is low. Typically, when a market is mature and competitive, a price rise in the marker will be matched by similar price increases of the commodity in other regions etc., i.e., basis risk is low. On the other had, if there is little or no correlation between price movements, basis risk is high.

Table 4.3
Comparison of Basis Risk in Two-Time Periods

	Sahara Blend	Forcados	Bonny Light	Ekofisk	Minas	Cinta
Number of Observations	182	106	105	84	51	21
1979-80 Average Price	$27.73	$28.22	$30.97	$27.88	$21.28	$25.68
1979-80 Average Basis Differ	$5.10	$4.53	$7.42	$4.28	$0.44	$2.71
1979-80 Standard Dev. Basis	$4.11	$5.71	$6.58	$3.79	$3.47	$3.53
2000-01 Average Price	$26.63	$26.50	$26.02	$25.68	$25.65	$25.37
2000-01 Average Basis Differ	-$0.25	$0.15	-$0.06	$0.04	-$0.05	$0.79
2000-01 Standard Dev. Basis	$0.26	$0.30	$0.22	$0.38	$1.43	$1.24
Ratio 1979-2000	16	19	30	10	2	3
2000-01 Transaction Costs	$0.02	$0.02	$0.02	$0.02	$0.20	$0.20

Source: ERA, Reuters

Table 4.3 compares basis risk between selected crude oil prices and their relevant marker price to the present market. In the table, basis risk is measured as the standard deviation of the difference between the crude oil in question and the marker. The table identifies six different crude oils commonly traded during the Iranian crisis that are still in active trading today. Sahara Blend is a North African high quality crude oil relatively close to European refineries. Forcados and Bonny Light are Nigerian crude oils, produced in West Africa. Ekofisk is a North Sea crude oil (in fact, the first North Sea discovery) connected to U.K. refineries by the Norpipe pipeline. Minas and Cinta are both Indonesian crude oils, close to Singapore refineries and frequently imported to the U.S. West Coast. These crude oils, however, have a high wax content, and require special equipment to move and process.

The number of observations in this instance refers to the number of cargos purchased for U.S. import during the two-year period (data as described earlier). In the 2000-2001 time period the methodology is quite different. These prices are gathered by telephone survey and posted for each business day. The average price during the two periods is almost identical. Moreover, there was similar price volatility during the two periods. 1979-1980 dealt with rising prices due to the Iranian Revolution. 2000-01 also dealt with rising prices, because the Asian financial crisis of 1998-99 had caused an unexpected decline in crude oil and natural gas prices.

During the 1979-80 period, prices of the four crude oils normally serving European oil refineries averaged from \$4.28 to \$7.42 per barrel more than Arabian Light crude oil. This reflected the lag that was built into OPEC's procedures for increasing prices. It also reflected the frequently stated desire of Saudi Arabia and a few Gulf States aligned with them to keep oil prices low enough to preserve future markets. Nigeria, Algeria, and other OPEC members were, however, unable to maintain a consistent premium over the price of Arabia Light crude oil, causing a high standard deviation of the basis differential.

For the years 2000-2001, marker prices underlying the table were changed to reflect markers used by crude-oil futures exchanges. Thus, the prices of Sahara Blend, Forcados, Bonny Light, and Ekofisk were compared to the North Sea crude oil, Brent. Minas and Cinta were compared to the Dubai blend crude oil, commonly used as the marker in Asia. In today's market basis differentials are much lower and, more importantly, the standard deviation of the basis is only a few pennies per barrel. Put another way, if the price of Brent crude oil is known, prices of other crude oils that vary slightly in quality and location can be predicted with considerable accuracy. In contrast, during 1979-80 the price of Arabian Light crude oil was not a very good predictor of the price of other crude oils. In the extreme case, the variability of Bonny Light, the principal Nigerian crude oil, was thirty times higher in 1979-80 than in 2000-01. This relationship corresponds to the scatter data presented in Figure 4.5.

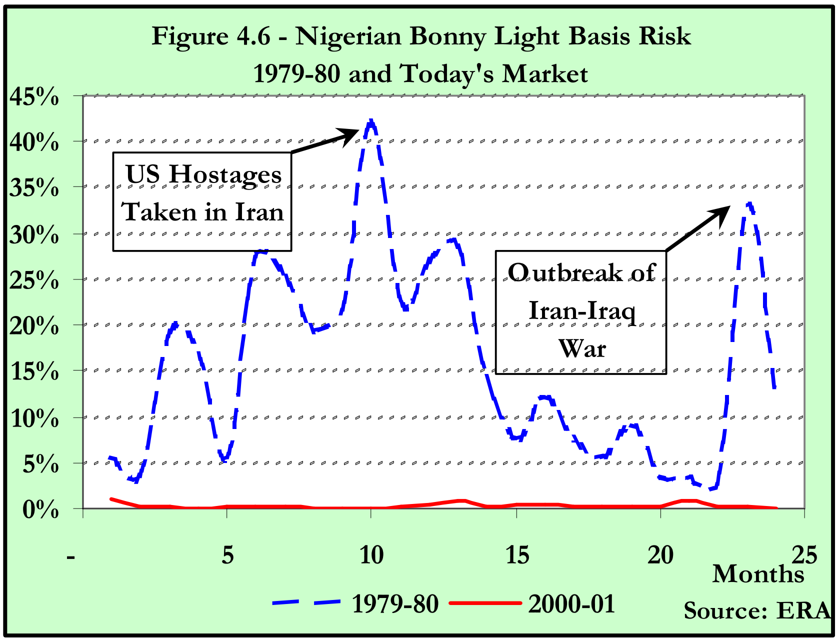


Figure 4.6 above provides a dramatic picture of the difference in basis risk during the Iranian crisis and today's market. Present day price fluctuations between Brent North Sea crude oil and Nigerian crude oils are negligible. Put another way, refiners or producers dependent on Nigerian crude oil can safely hedge prices using the Brent futures market. Oil prices may go up and down with world events and OPEC meetings, but there is little risk that the price of Nigerian crude oil will diverge from other key oil prices. This was certainly not the case in 1979 and 1980. During that period, prices varied in highly unpredictable ways, reflecting barriers to trade, poor information, and a host of associated transaction costs.

Asian crude oils are something of an exception in today's oil market. There are several reasons for the higher variation in price. Asian refineries are more spread out than their European counterparts. In addition, Asia does not have a mature crude-oil futures market. The Singapore Futures exchange tried to establish such a market, but it failed. Recently, the Tokyo Commodities Exchange (TOCOM) established a crude-oil futures contract based on a market basket of crude oils and denoted in Yen. This contract appears to have a better chance of success and if it succeeds, Asian crude oil markets should come into line with their European and American counterparts. The final line in Table 4.3 reflects an estimate of the current market's transaction or trading costs. This estimate is based on the difference of surveyed buy and sell prices. It is worthwhile to note that energy-trading margins in Europe are low compared to Asia. Comparable trading costs in the global oil market from 1979-80 are not available, but the data in Table 4.2 suggests that they were high, reflecting poor information, uncertainty, and confusion.

High transaction costs are not necessarily an incentive to trade less, because such costs also reflect opportunities for arbitrage. Those opportunities took on a colorful character during 1979-80. As OPEC countries began diverting sales from long-time buyers to the spot market, a new class of brokers and middlemen quickly entered the fray. At times the apparent range of crude oil prices grew as large as $20 per barrel. Since OPEC oil production averaged around 30 million barrels per day, even a dollar or two difference represented the opportunities for extraordinary profits. In fact, the new breed of brokers was rumored to be making millions of dollars on every trade. The most aggressive trader, Mark Rich, may have accumulated over $1 billion during the period. Much of this activity was tainted, because U.S. price control regulations created significant incentive for illegal arbitrage. Through January 1981 a large volume of U.S. crude oil, identified under price control regulations as "lower tier" oil, was required to be sold at about $6.50 per barrel. Black market traders began to "re-certify" greater and greater volumes of this oil to free market prices. Much of the re-certification was arranged through elaborate swaps of domestic and foreign crude oils. In the final month of the price control program, the administrative machinery completely broke down and enforcement became impossible. To no one's surprise, President Reagan decontrolled the oil market on his first day in office, January 21, 1981.

Table 4.4 (Treat 2002 p. 6) illustrates the build up in trading volume of spot crude oil cargoes from 1973 through 1985. (These data were originally compiled by Phil Verleger from the *Petroleum Argus* database and are only a partial listing). Other data compiled by

Table 4.4
Growth of the Spot Market

	Physical Cargoes	**Forward Cargoes**	**Total**
1973	54	-	54
1974	96	-	96
1975	154	-	154
1976	330	-	330
1977	327	-	327
1978	291	-	291
1979	315	-	315
1980	290	-	290
1981	405	-	405
1982	573	122	695
1983	1,014	806	1,820
1984	3,149	2,254	5,403
1985	5,628	3,058	8,686

Sources: Treat, Verlager, *Petroleum Argus*

the GAO and the ERA database used to illustrate prices in Figure 4.5 suggest that the volume of spot trading increased substantially during 1979 and 1980, approximately quadrupling during the disruption. In any case, during and following the disruption, spot trading increased dramatically and ultimately dominated the market.

According to traders active at the time (GAO 1980), many of OPEC's members sought to take advantage of panic buying during the Iranian crisis, frequently ignoring longstanding relationships with companies that had been instrumental in discovering and developing the host country's oil reserves in the first place. While such behavior garnered higher prices in the short term, it also undermined long-term relationships. When the oil market turned from a seller's to a buyer's market in 1981, refiners were all too willing to shop around for the best-priced crude oil. Moreover, in the declining market there was excess refinery capacity and low margins, creating a highly competitive environment. The secretive and semi-illicit nature of spot sales also suited many OPEC members as they sought to increase market share through price chiseling and discounting, but did not want to disclose their behavior to other members of the cartel. A favorite anecdote of the time was the Nigerian "slop" oil market. Slop oil is, as the name implies, small lots of oil from refinery and producing operations, which vary widely in quality and the number and types of impurities; thus, it is deeply discounted when sold. As competition in international crude-oil sales intensified, the volume of Nigerian slop oil (and its quality) greatly increased.

The furor over rising oil prices during the crisis caused a critical development in 1980. High transaction costs created a demand for information. The trade press began to collect and publish information on spot crude oil prices. *Platt's Oilgram Price Report* and *Petroleum Intelligence Weekly* were the two most widely read journals on the petroleum market. Until the Iranian disruption, however, they reported only anecdotal information on crude oil prices, instead of providing information on spot petroleum product prices and official or posted crude oil prices. As the spot market became increasingly important, reporting increased. In mid-1980 *Platt's* began systematically reporting crude oil prices.

Table 4.4 makes two important points. First, the volume of spot trade greatly increased from 1981 through 1985, when oil demand as a whole declined and the market was most competitive. Second, there was a new type of trade in "forward" contracts, which could be settled out for cash rather than the commodity. The development of a forward market was a first step in the radical changes that would follow.

The disarray of OPEC sales and the collapse of U.S. oil price regulations created a wholly different world for the oil industry. Before the Arab oil embargo, international crude oil production was, as Blair (1978), Frankel (1969), Adelman (1972), and many others have described, controlled by a few companies. When oil changed hands outside the integrated system it did so through secretive long-term contracts, barter exchanges, or in small volume lots necessary to balance transportation and refining incongruities. There was virtually no public information available on price. International crude oil pricing was often opaque on purpose in order to facilitate a seemingly medieval system of lords and vassals. In the absence of supply disruptions the system worked splendidly, but it was far too rigid to accommodate the 1970s oil shocks. OPEC's system of administered prices might have survived, if the cartel could have been more disciplined, but it was an extraordinary mix of incongruous nations.[18] The institutions of oil trading and pricing would have to change, and when the change came it did so with remarkable speed.

4.7 The Birth of Futures Trading

John Blair and other critics of the oil industry believed that the Major oil companies had put an organized system of control in place, which made it easy for OPEC to set and maintain excessively high crude oil prices. In this view the system was built around market rigidity, achieved largely through vertical integration. Under this system OPEC

[18] In one OPEC session negotiations on price had to be shelved as members attempted to reconcile differences between Iran and Iraq, which were at war.

could control crude oil prices by capturing the Majors' vertically integrated structure. The cartel would fix prices and the Majors would simply add on refinery, transportation, and marketing margins and pass the costs along to consumers. In such a tight structure, the absence of competition for market share locked everyone into a high-cost energy environment. This critique of the market was, however, short-lived because the critics did not foresee that the high transaction costs of trading with OPEC members would provoke a change in the institutions that govern the operation of the oil market. Nowhere was this more evident than in the rapid growth and success of energy futures markets in tandem with the spot market. Another analyst during this period (Safer 1979 p. 87) commented: "The development of an organized exchange market for oil products would help make the pricing process more competitive." Dr. Safer was pinning his hopes on the newly opened oil futures market in New York City. In 1978 NYMEX was a struggling commodity exchange, overshadowed by its larger and more successful competitors. In desperation, NYMEX decided to launch heating oil and heavy fuel oil contracts, based in New York harbor. Oil futures had been tried before – in the 1930s in California, in 1973 by the Cotton Exchange, and in 1974 by NYMEX – however, these contracts failed (Treat 2000 pp. 3-13). Trading in futures contracts depends on hedging demand and this, in turn, depends on price volatility and active trade in the physical commodity. During the middle of the twentieth century, the oil market was stable, prices seldom changed, and when they did the change appeared to be in the hands of a few decision makers. (Commodity exchanges are always concerned about someone cornering the market).

This time around NYMEX had good luck. As the winter of 1978-79 approached, the world grew more apprehensive about events in Iran. These events had an especially significant impact for New England homeowners, since the region suffers cold winters and is heavily dependent on heating oil. Suddenly, there was substantial interest in trading oil and hedging forward prices; thus, trading volumes grew dramatically. (The heavy fuel oil contract, however, failed perhaps in part because higher oil prices rapidly diminished that market). The nascent futures market did much more than open up hedging for heating oil; it had other benefits to bestow.

The success of the heating oil contract encouraged NYMEX to open other contracts, and in 1981 they launched a New York-based leaded gasoline contract (ultimately replaced by unleaded gasoline). During the same year London's IPE launched a successful gasoil (heating oil) contract based in Rotterdam. The year was also notable for the number of failed energy contracts. NYMEX failed in its attempt to establish heating oil and leaded gasoline contracts based in the Gulf Coast, the Chicago Board of Trade (CBOT) failed to establish an unleaded gasoline contract in the Gulf Coast, and the New York Commodity Exchange (NYCE) propane contract failed.

The real prize in energy trading fell to NYMEX, however, in 1983 when it launched a crude-oil contract. It was this contract that was to change so dramatically the face of the energy market. Again, NYMEX had extraordinary luck with its timing. In 1983 oil demand was in decline, due to high prices and the worst recession since the 1930s. The burgeoning spot market had produced a new class of oilmen who made profits through trading. For the first time since the nineteenth century, the crude-oil market was wide open to negotiation and trade.

NYMEX's crude oil contract was based on WTI, a mid-quality crude oil, gathered by pipeline from a variety of oil fields in western Texas. Much of the U.S. crude-oil pipeline system converged on Cushing, Oklahoma, where there was active trading and access. In 1983, crude oil purchased in Cushing could be transported north to Chicago and the Midwest, or it could be transported south to Houston and Gulf Coast refineries, where domestic crude oils competed with foreign imports. Cushing was an ideal geographic location for the new contract. NYMEX's contract volume now averages over 150 million barrels per trading day – just about double actual daily production throughout the world.

As volume in the NYMEX crude-oil contract increased, prices determined on the exchange began to have an impact on posted prices of crude oil. The industry referred to the impact as the "merc jerk." Traders also made a distinction between "wet" barrels (the actual flow of crude oil) and "paper" barrels (futures contracts for purchase and sale of crude oil).

4.8 Description of Today's World Oil Market

Ultimately, crude oil prices are determined by the demand for petroleum products and the supply of crude oil available to meet that demand (although supply is frequently constrained by OPEC). In order to track the broad shifts in crude oil value, the industry pays particular attention to the two crude oil streams that underlie the two largest futures markers. At the IPE in London, Brent crude oil produced in the North Sea is used as the marker. At NYMEX in New York, a market basket of medium quality crude oils produced in Texas and Oklahoma delivered to Cushing, Oklahoma are used as the marker. As noted, the daily volume traded in these commodity exchanges is enormous and the resulting prices are broadcast throughout the world.

Prices quoted at IPE and NYMEX track the general movement in oil prices, but they do not set relative values among the diverse number of crude oils produced all over the world. As discussed earlier, an adjustment has to be made for location. Brent crude oil is delivered by pipeline to Sullom Voe in the Shetland Islands. From its landing point it can be shipped to virtually any tidewater refinery by marine tanker. Most Texas and Oklahoma crude oils are "landlocked," but can be delivered to a variety of refineries in the U.S. mid-continent by pipeline. The prices of Brent and other crude oils are linked by the cost differentials of shipping the oil to refineries. These costs are largely determined by the cost of marine transportation, which is remarkably cheap.

Crude oil is concentrated in a few major geological provinces and tends to be located near principal marine waterways. This, of course, is because it is found in sedimentary basins and has its origin in primeval deposits of algae and other biological material. Likewise, refineries tend to be located near population centers, because generally it is cheaper to ship crude oil than refined products. The consequence of having both major markets and major producing areas near the seacoast allows the oil market to be knit together by a seamless transportation system.

As noted earlier, in their natural state, crude oils vary enormously in quality. The sorting out of the value differences among the various crude oils is done by telephone survey. Publishing companies, such as Reuters and McGraw-Hill's *Platt's Oilgram Price Report (Platt's)*, track daily prices of about 50 crude oils throughout the world. The various crude oils are often quoted as differentials to prices determined in the IPE and NYMEX exchanges. In essence, these companies track five separate markets:

- The European market, with crude oils from Russia, the North Sea, West Africa, and North Africa;
- Persian Gulf crude oils, which are mainly "sour" (high sulfur) oils from Saudi Arabia, Kuwait, Iran, Iraq, and Abu Dhabi;
- The U.S. Mid-Continent market, with WTI, West Texas Sour, and a variety of new oils produced offshore in the Gulf of Mexico;
- The U.S. West Coast market, with production of heavy California crude oils and ANS from Alaska;
- The Asian market, mainly in Singapore, with crude oils from Malaysia, Indonesia and other parts of Southeast Asia.

4.9 Vertical Integration Revisited

This chapter began with the historical view of the structure of the oil industry. Since the breakup of the Standard Oil Trust both the industry's critics and apologists have viewed vertical integration as a natural order to the industry, particularly integration between refining and crude oil production, where "security of supply" has dominated thinking since the nineteenth century. Following the Arab oil embargo, there was considerable political pressure in the U.S. to break up the companies. The impetus dissipated, but left in its wake a number of studies.

David Teece (1976) used transaction cost economics to explain the presence of vertical integration and provide an "affirmative rationale" for its structure. He made four main points: 1) vertical integration allows specialized decision making "economizing on communications expense," 2) it reduces uncertainty by allowing coordinated responses to unforeseen contingencies, 3) vertical integration reduces the risk of opportunistic behavior, and 4) it reduces information costs. In a more recent study Fernando Barrera-Rey (1995) commented: "the most popular explanation of vertical integration is that derived from the cost of transactions." Barrera-Rey ties this to the problems of asset specificity identified by Williamson and others as discussed in Chapter 3. As mentioned earlier, however, with the exception of certain geographic synergies, asset specificity is not a significant problem for the oil industry. Component products create highly inelastic demand and supply schedules, causing price volatility and increased risk. Price risk is normally not considered part of transaction costs.

Although Barrera-Rey's study (1995 p. 48) was directed more at the consequences of integration than its causes, he made an interesting conclusion: "On the one hand, our results point to a detrimental effect of integration on economic efficiency perhaps as a result of the diseconomies of size and diversification. On the other hand a greater degree of integration in these two stages is associated with a reduction in the variability of economic efficiency." Put another way, integrated companies are not necessarily more profitable, but their profits are more stable. Barrera-Rey's findings are consistent with what would be expected following the transformation of the oil market. Trading costs have shrunk to practically nothing, but wholesale prices for intermediate products are far more volatile. As pointed out earlier, the combination of refining and crude oil production are thought to be a natural hedge. Thus, companies that operate in both the refining and crude oil production sectors are to some extent automatically stabilized.

However, today's oil market is far more complicated than it was before the OPEC revolution, when there were seven companies dominating world trade. OPEC's members not only asserted control over resources, they also created national oil companies and gave them preferential access to resources. Many consuming nations followed the same practice, creating or enhancing national oil companies with preferential access to the market. Among other things, the creation of national oil companies obfuscates financial data on the industry, because these companies do not report a clear statement of profit and loss.

Table 4.5
Revenue by Category of the 185 Largest Oil and Gas Companies
(in Billions US $)

	Integrated Market	Integrated None	Integrated Resources	Specialized Market	Specialized None	Specialized Resources	Total
No.Companies	12	17	11	12	130	3	185
% Companies	6.5%	9.2%	5.9%	6.5%	70.3%	1.6%	100.0%
Average Revenue	$14.6	$55.3	$24.3	$8.8	$4.1	$22.1	$11.3
Group Revenue	$175.5	$940.9	$267.7	$105.4	$538.9	$66.3	$2,094.6
% Total Revenue	8.4%	44.9%	12.8%	5.0%	25.7%	3.2%	100.0%

Notes:
Source: Hoovers
Data are for 2001 or most recent year available
A few conglomerate corporations are difficult to characterize
"Integrated" refers to vertically integrated. "Market" preferential access to market. "Resources" refer to preferential access to resources.

Table 4.5 provides a thumbnail sketch of the 185 largest oil and gas companies (excluding gas utilities). 40 of these companies are classified by Hoovers as integrated, i.e., they operate in more than one sector of the industry. The remaining 145 are primarily specialized in one of the key sectors: exploration and production, transportation, refining, or marketing. Here integrated and specialized firms are further broken down into those with preferential access to the market, those with preferential access to resources, and those without a specific tie to a producing or consuming country. Those without preferential access to markets or resources remain the largest segment of the business, representing three-quarters of total turnover (revenue), which in 2001 was over $2 trillion. Integrated firms without preferential access remain the dominant force. Although in 2001 there were only 17 such companies, they represented nearly half of the industry's revenue. The second largest segment of the industry contains specialized companies without specific ties to a country; these companies are mainly refiners, exploration and production companies, and the oil services.

Because some of these companies are not publicly traded, the number reporting net revenue is somewhat less; the statistical data are displayed in Table 4.6. It should come as no surprise that companies granted preferential access to either markets or resources appear to be more profitable, i.e., have a higher net percentage of total revenue. This, of course, does not mean that they are more efficient. In all categories, integrated companies were slightly more profitable than specialized companies. They are, however, much larger in scale, which muddies the water. More conclusively, the coefficient of variation (even on cross-sectional data) was consistently higher for specialized firms than it was for the integrated ones.

Some oil industry analysts have observed that the oil industry is no longer vertically integrated, at least not in the same way as many electrical utilities, which control generation, transmission, and distribution in a specified grid. Instead, these companies are better characterized as conglomerates; holding an array of assets at various locations aimed to achieve diversity, stabilize net revenue, and provide the financial scale necessary to invest in the most profitable exploration plays. Investing in the most profitable sector of the industry (thought to be exploration and production) requires massive capital investments and patience, since some projects take a decade to bring to full fruition. Unfortunately, profitability in this sector is also the most volatile. Companies with erratic profitability find it more difficult to finance large-scale development projects, either through equity or borrowing. Holding a broad array of assets helps the Major oil companies manage their investments and gain entry into the best large-scale projects.

Table 4.6
Revenue by Category of 164 Oil and Gas Companies
(in Billions US $)

	Integrated Market	Integrated None	Integrated Resources	Specialized Market	Specialized None	Specialized Resources	Total
No. Companies	11	17	10	10	113	3	164
Turnover	$173.2	$940.9	$209.7	$95.1	$485.0	$66.2	$1,970.2
Net	$18.6	$55.0	$19.0	$6.4	$23.0	$5.5	$127.3
Profit Margin	10.7%	5.8%	9.0%	6.7%	4.7%	8.3%	6.5%
Coef. of Variation	8.0%	6.7%	12.6%	13.7%	10.1%	24.2%	10.3%

Notes:
Source: Hoovers
Data are for 2001 or most recent year available
A few conglomerate corporations are difficult to characterize
"Integrated" refers to vertically integrated. "Market" preferential access to market. "Resources" refer to preferential access to resources.

Teece's data (1976) can also be used to demonstrate that vertical integration was not necessarily the dominating paradigm even before the OPEC revolution. Table 4.7 displays the self-sufficiency ratios of U.S. companies in 1972. A company with perfect balance between crude-oil production and refining would have a ratio of 100%. The average ratio of 84% is not far off the expectation, but the standard deviation is 52%.

There are only 7 companies out of 25 that fall within the range of 80 to 120%. Three of these companies -- Gulf, Exxon, and Mobil – were considered true international Majors, part of the seven sisters. The two other U.S. sisters – Standard Oil of California and Texaco – had crude-oil producing capacity substantially in excess of refining capacity. (All of these figures are biased by the Aramco subsidiary owned at the time by Exxon, Mobil, Texaco, and Standard Oil of California.) If the transaction costs of buying or selling crude-oil feedstocks in 1972 had been prohibitively high, self-sufficiency ratios should have been closer to 100% for both large and small companies. Recall also that the original Standard Oil companies concentrated on transportation and refining, not crude oil production, before the Trust was dissolved. Thus, it is difficult to conclude that vertical integration is a natural order for the petroleum industry, either before or after the OPEC revolution.

The history of the acquisition, operation, and disposal of specific assets in a localized market can help shed light on the motives for vertical integration vis-à-vis the benefits of specialization. The U.S. oil market west of the Rocky Mountains is geographically isolated, and because the indigenous crude oil is much heavier than oils in other markets, it has a unique refinery sector. Within this self-contained market there are numerous examples and counter-examples to the vertical integration model.

Table 4.7
1972 Self-Sufficieny Ratios of U.S. Companies

Amerada Hess	52%
American Petrofina	14%
Apco Oil	62%
Ashland Oil	13%
Atlantic Richfield	87%
Cities Service	84%
Clark Oil	2%
Continental Oil	115%
Diamond Shamrock	32%
Exxon	97%
Getty Oil	188%
Gulf Oil	165%
Kerr-McGee	111%
Marathon	159%
Mobil	86%
Murphy Oil	40%
Phillips	60%
Shell Oil	65%
Skelly	132%
Standard Oil of California	154%
Standard Oil of Indiana	79%
Standard Oil (Ohio)	13%
Sun Oil	80%
Texaco	136%
Union Oil	77%
Average	84%
Standard Deviation	52%

Source: Teece (1976)
Includes domestic and international operations

In the 1960s the largest crude oil producer in California was Getty, which owned the Kern River oil field. This field was in the San Joaquin Valley, north of Bakersfield. Getty had two refineries, a large one in the Bay Area and a smaller one in Bakersfield. The company also had crude-oil pipelines, which ran from the center of the valley to the Bay Area and to Bakersfield. Getty decided to specialize in crude-oil production and the refineries were sold (along with a crude-oil sales contract) to Tosco. Tosco's primary interests were in developing oil shale in the Rocky Mountains, but when its development there collapsed, it had no choice but to concentrate on refining. At first, Tosco floundered: one refinery had to be shut down, its crude-oil purchase contracts expired, and it narrowly escaped bankruptcy. In the late 1980s, new management invigorated the company. In the 1990s, it focused very effectively on marketing and began to profitably expand without a guaranteed source of supply.

Vertical integration may sometimes seem to be more of an accident than a plan. In any case, profitability, not integration, is the primary goal. For years, the West Coast's largest oil producer, BP, had no refineries in the region. Instead, they sold the oil outright or exchanged it for crude oil delivered in the Midwest, where they did have refineries. BP finally acquired Mobil's refinery in the Pacific Northwest and began to market gasoline under the BP logo. The effort proved unprofitable and BP sold its refinery, service station and right to the BP logo to Tosco. Despite having no crude oil production, Tosco was able to operate the assets profitably (unlike BP) and in 1996 they expanded again by purchasing refineries, service stations, and the "76" logo from Union Oil of California. In 1999 BP acquired ARCO, but was required by regulatory authorities to divest ARCO's Alaskan oil-producing assets. They were sold to Phillips, which had no petroleum refining or marketing on the West Coast. As a consequence, Phillips acquired Tosco and a few months later merged with Conoco. BP ended up with ARCO's West Coast service stations and refineries. ConocoPhillips ended up with ARCO's Alaska oil production and Tosco's refineries and services stations, many of which were once owned by BP.

John Mitchell (1999 p. 9) has argued that: "… vertical integration was not an objective of the Exxon and BP mergers. Their attractions come from horizontal integration: *complementarity of properties, and superfluity of management,* in each segment separately." This is almost certainly correct. In the case of BP's acquisition of ARCO, one of the original strategies was to rationalize the cost of maintaining production in the Prudhoe Bay oil field and to better coordinate the development of satellite fields on the North Slope. Instead, the FTC forced BP to divest ARCO's Alaska assets, which in turn provoked a vertically integrated structure for both BP and ConocoPhillips on the West Coast.

The simplest of statistics confirm waning interest in vertical integration. The self-sufficiency ratios of all four Super Majors following the wave of mergers are less than one, and only Chevron-Texaco resembles the traditional vertically integrated company (see Table 4.1). Moreover, with only a few exceptions, the national oil companies of the oil exporting nations have not integrated into downstream markets. These observations are somewhat baffling. Vertical integration is frequently the stated goal of oil companies, particularly those that have ambitions to become Majors in the industry. The practice, however, falls far short of the theory. In the last three decades the industry has become more competitive, more fragmented, and more specialized.

Although OPEC can be blamed for some of the fragmentation, it is useful to return to the standard insights of neoclassical economic theory that stretches back to Adam Smith. In the *Wealth of Nations*, Smith (1776) noted that benefits from the division of labor were limited by the "extent of the market." George Stigler (1951 p. 189) picked up on this idea to conclude that: "…vertical disintegration is the typical development in growing industries, vertical integration in declining industry." When economies of scale to particular activities are limited by the size of the market, then firms find it economic to perform the function internally, rather than specialize. Put another way, companies integrate vertically when markets are small and illiquid and trading costs are high. As markets expand, however, trading costs decline and firms find it profitable to specialize. Overall, since 1973 the volume of petroleum consumption has increased at a slow pace. The "market" for crude oil and petroleum products, however, has expanded by many orders of magnitude. This is due primarily to the introduction of futures trading, which has vastly expanded the volume of turnover in product sales. Greater liquidity and its counterpart, lower transaction costs, provoke specialization and reduce the incentive for firms to integrate.

Regarding the impact of transaction costs on the oil industry, Nick Antill (1999 p. 3) summarized the current state of affairs: "Textbook theory suggests that there are two strong reasons for vertical integration. The first is that it may reduce transaction costs. If an operation is not integrated, then it will have to buy or sell to or from a market. This requires finding out what a market clearing price is. There may be some costs associated with the transaction itself." Further, the fixed infrastructure creates a second reason to integrate: "Investors are exposed to the risk of what I believe economists call 'ex-post opportunism.'"

Neither reason to integrate is significant for today's oil industry; it has been transformed. The growth in the size and reliability of the spot market has accompanied and provoked a substantial drop in the transaction cost of buying refinery feed stocks. Reliable price information is cheaply available and low brokerage fees have reduced trading costs to the bare minimum. In addition, the development of a robust forward market mitigates the price risks inherent in energy markets that arise from the problem of component products and specific assets. It is unlikely that either of these institutional arrangements would have developed for the oil market in the absence of futures trading, which enhances trading volume and price transparency. Whatever the cause, the presence of a mature commodity market for wholesale crude oil trading significantly reduces the incentive for vertical integration. Thus it should be no surprise that the self-sufficiency ratios of Major companies have changed as they have focused on greater profits through greater specialization. What is surprising is that the industry clings to the myth of vertical integration.

5. The Evolution of the Natural Gas Market

5.1 Network Delivery Systems

In its early stages of development the natural gas industry has significant economies of scale, arising from an inflexible network of transmission and delivery pipelines. The consequence is frequently government ownership or extensive economic regulation of privately owned companies. This has often resulted in a fossilized industry in which limited quantities of gas are provided to a few high-valued users at high prices. In the last two decades, however, the North American gas industry has evolved far beyond this organizational structure and is often held out as model for other countries and regions. The efficient pricing that flows from the deregulated wholesale gas market has bestowed many benefits to U.S. energy consumers; for over a decade it kept prices low and now that gas is less abundant it is allocating resources efficiently, minimizing the disruption of supplies.

The path of transformation in the U.S. gas industry was costly and protracted. It is a thorny path that others may avoid by observing what has worked and what has not. Europe is now in a staged transition to a market that will inevitably resemble something like the North American structure. While it will not be a copy of the U.S. and Canadian market, it will have features in common, including:

- A gas market built around a system of interconnected pipelines;
- Pluralistic ownership of the infrastructure, including all sectors of the industry: natural gas resources, production facilities, transmission pipeline, storage, and distribution systems;
- An enlightened system of regulation, with third party access to pipelines, i.e., no discrimination with regard to transport rates or access rules and transparency of transport rates, storage fees, gas prices, and all other economic variables that impact the efficiency of operation;
- An active spot market;
- Liquid and efficient forward and futures markets, with standardized products, participation of both speculators and hedgers, and a transparent term structure.

In the natural gas industry, consumers are connected to producers by a pipeline network, divided into two interdependent segments: long distance high-pressure transmission pipelines which connect producing with consuming regions and distribution systems which aggregate a multitude of isolated users into a viable market. In this regard, the industry has much in common with transportation, electricity and telecommunications networks. There is, however, a subtle difference between the various types of systems. With the telecommunications industry, interconnection is essential; that is, there must be two-way communication within the system. The value of a telephone (and now Internet) system is the ability of one node to connect independently to another node and to send information in both directions. There is no such complication in the gas industry: transmission need only be in a single direction, from producer to consumer. Interconnection is, however, important even if its role is different. As will be explained later, it is all but impossible to have an open gas market with active trading unless the system is interconnected.

The natural gas industry frequently exhibits economies-of-scale and is usually considered a natural monopoly. That is, per-unit costs decline as volume increases. The inflexible distribution system creates the need for a natural monopoly structure in the gas industry; it would be economically wasteful to build a series of duplicative and competing pipelines. The cost of the right-of-way for a pipeline system usually does not vary with respect to the volume of flow. Moreover, there are often economies-of-scale in the construction of high-pressure transmission pipelines. Up to a point, the larger the diameter of the pipeline, the lower the per-unit cost to transport gas.

The industry's status as a natural monopoly has been used to justify state ownership, extensive regulation, or both. The deregulation of the industry in North America has revealed one central point: many segments of the natural gas industry are not natural monopolies and competition, rather than regulation, can ensure fair prices for consumers. For example, gas exploration and production, marketing, balancing, and storage do not necessarily have economies of scale, so competition can be an effective substitute for regulation. The resolution of this dilemma in North America has been to "unbundle" the industry, that is, to separate, by regulation if not ownership, the industry's various segments. The segmentation of the industry, however, requires a shift in regulatory focus to ensure that competitive sectors of the industry have access to key elements of the transmission and distribution infrastructure.

City gas distribution systems are still extensively regulated in the United States by local Public Utility Commissions (PUCs). FERC regulates interstate transmission pipelines that connect producers to consumption centers. Regulation has dramatically changed however, with emphasis now aimed at ensuring non-discriminatory access to transportation, rather than setting gas prices. In North America, the economies of scale in long-distance transmission pipelines may not inhibit competition between them, because the system is extensively interconnected. That is, producers often have multiple choices with respect to shipping gas. Even though FERC sets transmission rates for pipeline owners, there is an active secondary market in spot transportation, with rates often below regulated levels.

One feature of the gas industry is frequently misunderstood outside of regions with a mature gas industry: an established gas market is almost always more valuable than the resource itself. The distribution systems that support gas markets are very costly and require time to construct. Not only does laying the pipelines require initial investment, but consumers have to be convinced to connect to that pipeline and use the gas as well. The aggregation of thousands of consumers into a market large enough to pay for the construction of long-distance transportation systems (Liquid Natural Gas (LNG) or transmission pipelines) may take years. The value of gas market aggregation has often been misunderstood because of the oil supply disruptions of the 1970s. These events focused attention on oil and gas as depleting resources. Gas is, however, much more abundant than oil. There are huge resources in the Middle East, Arctic and other remote corners of the globe that are untapped because of the high cost of transporting them to a market.

Although a fixed pipeline delivery system may bestow a natural monopoly on its owner, there are more substitutes for natural gas than either petroleum products or electricity. Liquefied petroleum gas (LPG) can substitute for natural gas in many applications. In addition, as pointed out in Chapter 2, many industrial consumers can switch between heavy fuel oils and coal with relative ease. In addition, it is the distribution utility, not consumers, that has invested the vast bulk of capital in the infrastructure; consumers simply pay per-unit rates. These features limit the ability of gas suppliers to extract monopoly rents and, again, underscore the importance of inter-fuel competition. This feature of the industry also helps explain why privatization of the gas industry is more likely to be successful than privatization of the electric industry, where consumers cannot easily substitute an alternative commodity.

It is instructive to note how the natural gas industry developed in both Europe and North America. At the turn of the last century, gas was used for lighting in urban areas. By and large, the gas was manufactured from coal. Electric lighting replaced gas lighting, but the important consequence was an existing gas pipeline distribution network, which could be used for other things, such as cooking and space heating, if the commodity could be supplied economically. Thus, when long-distance transmission pipelines were developed in North America, and when Europe discovered North Sea gas, the basic distribution infrastructure was already in place. This allowed the gas industry to develop much more rapidly than it might otherwise have. It also explains the difficulty in developing the gas market in Asia and other parts of the world, which do not have the foundation of town gas upon which to build.

5.2 Characteristics of Gas as a Commodity

Compared to crude oil and electricity, natural gas is a fairly simple commodity. In its natural state, it often contains impurities and other gases that must be removed. In most cases, however, it is easily and cheaply standardized. In contrast, as discussed in Chapter 4, crude oil ranges widely in quality and in degree of contamination. The consequence of this variation in crude-oil quality is a vast array of crude-oil prices and often a difficulty in determining the fair market value of a specific crude oil at a specific location. Natural gas seldom has this problem. Quality differences, if they exist, are normally measured in heat content and these are easily determined.

Natural gas storage is also not a serious technical problem, although it is sometimes expensive. Peak-shaving facilities can be constructed close to market. Field production can, of course, be varied with the seasons. Gas can be re-injected into depleted gas fields and withdrawn during periods of peak demand. Often these fields are close to the market, which helps balance seasonal demand against the capacity constraints of transmission pipelines.[19] The pipeline system itself can be used for

[19] For example, Oregon's only gas field, the Mist Field, is now depleted but used to store gas in order to meet seasonal peaks.

limited amounts of storage by adjusting pressure in order to meet variations in demand and supply. If gas production were abruptly terminated, consumers could still withdraw gas for up to three days while the pressure dropped. In contrast, the electrical grid must be balanced within a very close tolerance minute to minute.

As noted in Chapter 4, natural gas has always been the second child of the petroleum family. Companies that have explored for oil have been disappointed to find only gas. If the field is remote from a market, it is usually uneconomic to develop. For years, many natural gas discoveries were not even recorded. One consequence was that producers tended to market natural gas at whatever price it would fetch. The widespread availability of cheap natural gas in petroleum-producing states and the use of high-valued coal gas in cities and towns for lighting and cooking in other regions and states provided the incentive to develop high-pressure steel pipelines that could be used to transmit gas over long distances. However, U.S. interstate commerce is governed by federal, not state authorities, which led to the regulation of transmission pipelines by the federal government.

5.3 From Regulation to Markets[20]

The Natural Gas Act (NGA) of 1938 opened a Pandora's box for federal regulation of the gas industry in the U.S. and vested the regulatory authority in an existing agency, the Federal Power Commission (FPC).[21] In North America, large gas fields are separated from major consumption centers by vast distances and normally cross one or more state borders and/or national boundaries. The NGA established federal jurisdiction over the interstate transport of natural gas. In 1954, the Supreme Court's "Phillips Decision," in combination with a series of FPC regulatory rulings, dramatically extended the power and authority of federal regulators. Any gas moving in interstate commerce was subject to price regulation, extended to individual wellheads, within state boundaries.

Gas price regulation at the wellhead was implemented at a time when there were huge surpluses of gas. The overhang of North American gas discoveries in the 1950s and 1960s produced exceptionally low prices; regulations combined with long-term contracts locked low prices in, often for as long as the fields were expected to produce. As the energy crisis in the 1970s unfolded, however, the industry and regulators began to think about gas in quite different ways. Gas, particularly in incremental supplies, had been used as a boiler fuel—to heat water and generate electricity. But, low regulated prices, combined with growing oil scarcity, were creating shortages, and using gas as a substitute for coal or heavy fuel oil seemed wasteful.

Gas supply problems came to a head during the winter of 1976-77. Both Presidents Nixon and Ford had attempted to use the oil shortages of 1973-74 and the growing dependence on imported oil as a reason to produce more domestic energy. They were thwarted in this attempt, however, by low regulated oil prices imposed by the

[20] The following discussion is a very brief review of a highly complex subject. For details on the history of U.S. gas regulation see Tussing and Tippee (1995).
[21] The FPC became the FERC in 1978.

Democratically-controlled Congress, the inertia of the regulatory system, and the opposition of environmentalists. President Carter took office in mid-winter, during serious natural gas shortages, and he had a different idea. He believed that oil and gas were depleting resources that should be husbanded for future generations. As a consequence, he chose to emphasize energy conservation. One result was the federal Fuel Use Act of 1978, which prohibited the use of gas for electrical generation and other low grade uses (Gordon 1979). It was a peculiar dichotomy; one branch of the federal government was actively stimulating demand by holding prices low, while another authority was seeking to restrain gas use.

Earlier events in the United States had also had an important impact on industry structure. A series of financial scandals rocked the electric utility industry during the Great Depression and resulted in the Public Utility Holding Company Act (PUHCA) of 1935. Although this act was aimed primarily at electric utilities, it also had a substantial impact on the natural gas industry. PUHCA broke up the control of multiple energy assets by a holding company. The breakup included assets owned by Standard Oil of New Jersey, which held a substantial amount of the gas transmission capacity in the U.S. (Tussing and Tippee 1995 p. 85). The Act essentially precluded gas producers from integrating downstream into transportation and distribution; likewise, it inhibited distribution companies from integrating upstream. The net result was the independent ownership and operation of the transmission pipelines as separate companies. This is quite a different structure than in either the petroleum or electricity industries, where the largest companies are vertically integrated. The independent ownership of the pipeline companies was to be an important consideration in the style of gas market deregulation that took place in the U.S. because it made it much easier to unbundle products and services and regulate only selected activities. In contrast, the deregulation of the U.S. electricity industry has been much more complex due to the existence of hundreds of vertically integrated utilities.

The period from 1973 through 1985 was marked by two significant events in the U.S. gas industry that underscored the regulatory gridlock. By the time President Carter took office in January 1977 there was a full-blown energy crisis. It was an extremely cold winter and the curtailments of natural gas resulted in the necessity to close schools, hospitals, and factories. The consequence of this crisis was the Natural Gas Policy Act (NGPA), passed in 1978. Before the Act, federal regulators had been experimenting with a multi-tier price structure. The price for each category of gas depended on the vintage of its production. This provided an incentive for the industry to explore for and develop new gas deposits. The NGPA locked in the procedure with thirty categories of gas prices tied to gradual deregulation (Tussing and Tippee 1995 p. 152). Consumers continued to enjoy relatively low prices because the new high-cost gas was rolled in with the historic low-priced gas. In 1980, a great deal of gas was sold for less than $1.00 per million Btu at the wellhead. At the same time, the industry planned to construct the Alaska Natural Gas Transportation System (ANGTS), which would have delivered gas from Alaska to California at a cost of around $12.00 per million Btu.

Paul MacAvoy (2000) has estimated the gains and losses to producers and consumers from the two eras. During the period leading up to the shortages, from 1968 to 1977, consumers had a large net gain from controlled low prices of $38 billion; producers, however, lost that amount and more for a net loss of $6 billion (MacAvoy 2000 p. 55). The seesaw of offsetting losses and gains continued from 1978 through 1985 under the NPGA, but in the new era both groups lost. According to MacAvoy (2000 p. 72) consumers lost $4 billion and producers lost $45 billion.

Federal price regulation sometimes had positive, if unintended, consequences. In the 1970s a robust intrastate gas market developed in Texas. As long as the gas did not cross the Texas border, FERC had no authority to implement its complex pricing schemes, and local authorities had no stomach to regulate such a thriving market. In fact, the wide-scale availability of gas at reasonable prices was creating a booming economy in petrochemical and associated industrial products.

It was the Texas intrastate gas market that laid the foundation for a continental gas market. It was not an accident that Houston was where the energy marketing industry took root in the mid-1980s, and that it remains the energy trading capital of the world. For the 1970s, however, the free market was short-lived. The NGPA extended federal price regulations to the intrastate market. In return, gas producers were given gradual escalations in gas prices, leading to deregulation for all categories of gas.

An important consequence of the multi-price regulatory scheme passed by Congress was an unrealistic expansion of the industry. The high prices for new gas did, indeed, encourage additional supply. On the other hand, low rolled-in-prices stimulated demand both for the commodity and for pipeline capacity. Pipeline companies, which bought gas from producers and sold it to distribution companies, entered into a series of high-priced take-or-pay contracts, which were to haunt them a few years later. And, because they had an "obligation to serve," it led to the construction of storage facilities and connections to other pipelines to ensure energy security.[22] In retrospect, the result ought to have been obvious—substantial overbuilding. From 1980 to 1985, regulators and politicians fretted about the impact of price deregulation on consumers, anticipating a huge price spike or "fly up." One of the few analysts to recognize that prices were more likely to go down than up once controls were lifted was Dr. Arlon Tussing. In 1981, he received a great deal of notoriety by claiming that the theme song of the late 1980s would be: "I won't take and I won't pay, so sue me." He was right.

The excesses and mistakes of natural gas regulation in the U.S. are more than a story of historical interest. It was certainly not intended, but overcapacity, the Texas intrastate market, and interconnection provided the right framework for the evolution to a competitive market. By 1985 the "gas bubble" dominated the industry's thinking, and pipeline companies finally recognized that they might be stuck with a large

[22] The obligation to serve was considered a crucial part of the "utility compact." If a private company was to be granted an exclusive monopoly in a specific region it must submit to regulation. Regulators would set price levels adequate to ensure a reasonable rate of return on investment, but in turn would expect the company to serve all who requested service at regulated prices.

number of high-priced supplies that they could not pass on to the distribution companies.

The mid-eighties gas bubble also reflected the abundance of petroleum products. As noted in Chapter 3, the crude-oil market collapsed in 1986. Low oil prices increased competition for the boiler fuel market. Most importantly, however, the institutional changes in oil trading became a permanent feature of the industry. Refiners were initially disadvantaged when OPEC seized control of oil supplies and increased prices. In 1986, the tables turned and refiners had their choice of multiple oil suppliers. Most refiners were, however, part of integrated companies who produced both oil and gas. Unlike crude oil, however, natural gas had to be marketed to third parties. Thus, the institutional structure put in place to buy crude oil was adapted to sell gas.

FERC Order 451 in 1985, a series of intervening orders, and Order 636 in 1992 created the modern U.S. gas market. As mentioned earlier, the key to understanding the North American gas market structure was the independence of the gas pipeline companies. Since they were independently owned and operated, they could be transformed from merchants (buying and selling gas) to transportation companies, where they carried gas for other buyers and sellers.[23] The transformation occurred in a series of steps between 1985 and 1992.

The merchant function vacated by the pipeline companies was transferred to marketers. Gas marketers practiced classic arbitrage; they sought out producers wishing to sell gas, purchased it, and resold to consumers. In the early years, great profits could be made due to the many inefficiencies in the market. Although complex regulations and barriers to a gas market were unwinding in the 1980s, there were still many obstacles and roadblocks. These arose because a few companies owned key pieces of the infrastructure, which they had no obligation to share, and because state regulatory authorities were often reluctant to follow the federal lead. Thus marketers, producers, and large gas consumers conceived the notion of "bypass," where a wholly new pipeline segment would be built as an alternative to existing facilities.

Regulators quickly recognized that some of the pipeline proposals made economic sense; gas could be shipped more cheaply in the new facility, without impacting the economics of existing pipelines. This was referred to as economic bypass. Most of the time, however, the proposals were uneconomic. It would make sense for the project sponsors to build a new pipeline, but the reduced volumes in existing pipelines would increase their per-unit costs and the overall result would be net loss, rather than a net benefit. This was referred to as uneconomic bypass. Even if a project was uneconomic in a global sense, the threat of bypass was a powerful means to break down barriers and open up the gas market.

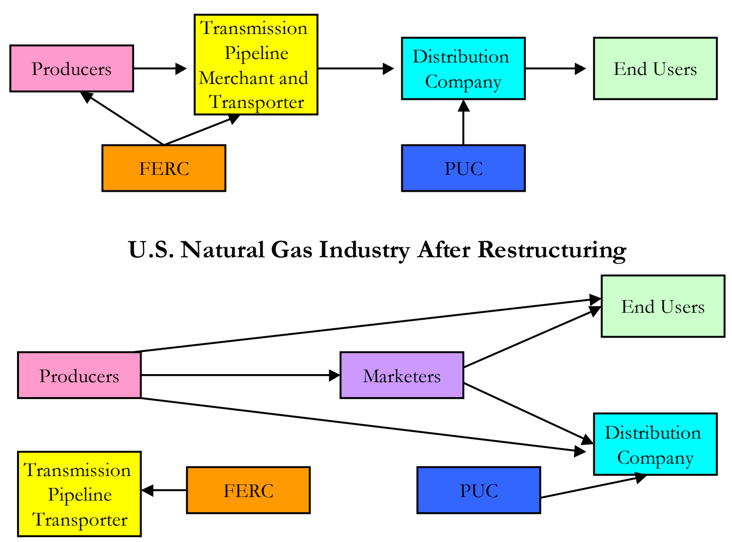
---

[23] Much of the regulatory upheaval at FERC concerned the means to effectively transform the pipeline companies to "common carriers," without using the name, which was prohibited by law. It is easy to become confused by all the different terms of art. The point is simple; the pipeline companies were transformed from merchants to carriers. (See Tussing and Tippee 1995).

The best way to explain how the threat of bypass opened up U.S. gas markets is to provide a specific example. As mentioned earlier, the federal government regulates interstate gas pipelines and the local governments regulate the distribution systems. In the 1980s a number of large industrial consumers were situated a few miles from the Pacific Gas Transmission (PGT) trunk pipeline that connects Canada to California. They were, however, required to purchase gas from distribution companies: Northwest Natural Gas and Washington Natural Gas Company. At that time PGT could buy gas in Canada for about $1 per million Btu; the company added about $1.50 for transportation and export fees and sold the gas to Northwest Natural Gas. The distribution companies added about $2.00 per million Btu for distribution charges and sold the gas to industrial customers. The distribution companies calculated rates (retail prices) based on the average cost of service for all customers. They also practiced cross-subsidization. That is, the high rates charged to large industrial and commercial customers were used to offset the higher cost of serving residential customers.

Industrial gas consumers in Washington and Oregon joined in a consortium to sponsor an alternative pipeline from British Columbia to Portland. They formed a company, hired attorneys, engineers, and investment bankers, and put together a prospectus for the new pipeline. The consortium spent enough money to convince PGT and distribution companies that they were serious. They were also able to determine the expected cost of delivering gas in the alternative pipeline. The new pipeline was not, of course, built. Instead, PGT and the distribution utilities conceded and allowed the industrial gas consumers to buy directly from gas producers in Canada (or from marketers) and pay a reasonable fee for gas transportation. Initially, the PUCs of the two states were reluctant to approve the concessions, but they too had to recognize the hard economic reality.

Figure 5.1
U.S. Natural Gas Industry Before Restructuring



Adapted from: Tussing and Tippee, *The Natural Gas Industry*

Figure 5.1 illustrates a summary of the gas industry's structure before and after deregulation. Before deregulation, FERC regulated producer gas prices and interstate transmission pipelines. The pipeline companies served two roles: merchant and transporter. As a merchant, the pipeline companies bought gas from the producers and resold it to distribution companies (and occasionally large end users). FERC regulated the pipeline companies, allowing them to pass through the cost of acquiring gas plus a cost-based fee for transportation. Pipeline companies had an obligation to serve distribution companies, who in turn were regulated by state Public Utility Commissions. They purchased gas from the pipeline companies, added a cost-based distribution fee and resold the gas to end-users. Importantly, however, FERC also restricted entry and limited competition from producers, because before a pipeline could be built it had to be backed by contracts for gas supplies.

After FERC Order 636 in 1992, the U.S. gas market evolved to an open market structure. Producers were free to sell gas at any price to any buyer. Marketers acted as middlemen between the producers, large end users, and distribution companies. FERC still regulates the pipeline companies, but the regulation is aimed at ensuring non-discriminatory access and fair transmission rates. FERC does not regulate gas prices. Marketers are certified by FERC and are required to submit annual reports, but otherwise are left to make profits or take losses. Large end users can buy from whomever they please and pay transportation fees regulated by FERC or the local PUC. Distribution companies are still regulated by local PUCs, but often have the ability to offer reduced rates to large end users. The residential and small commercial sector remains regulated by PUCs. Retail prices are based on the average cost of acquiring gas. To date this has not been a problem, with one major exception, because wholesale gas prices have fluctuated within a reasonable range.[24]

5.4 Gas Market Liberalization in Europe

Before turning to gas trading and the institutions that support the modern market, it is worthwhile to compare and contrast the restructured North American gas market to the status of deregulation in the EU. While many of the concepts of an open gas market are inherent in EU plans, the terms of art and specific description of activities are quite different. For example, in Europe the term Third Party Access (TPA) is used to connote the term "open access" in North America. The concept is the same and, of course, it is the key element in establishing a market-based industry.

Multiple languages and the complex political structure of the EU have made it very difficult to agree on a common plan to restructure the gas industry. In general, Europe's gas market is one or two decades behind North America in its development. Industrial structure in the two continents is, however, similar. With the principal exception of Gaz de France, Europe's gas industry is divided into three sectors—producers, transmission companies, and distribution companies. Thus an American-

[24] The obvious exception to this generalization is the California Energy Crisis of 2000-2001. Retail gas rates have risen for customers of various utilities as a result. Proceedings are currently underway in the courts and among the regulatory bodies to determine the cause(s) of the price spikes, and if mitigation measures should be imposed.

style open market has been feasible for some time. In 1992 the EU attempted to elicit such a development with a draft directive for common rules on unbundling and TPA (Stern 1998 p. 96). The initiative failed. Finally in February 1997, Common Rules for Electricity became effective, and natural gas followed in August 1998. The EU Directives do not, however, necessarily ensure a rapid and identical response in each country. Indeed some analysts believe that in the best circumstances, it will take the gas industry ten years to evolve into a competitive market (van Oostvoorn and Boots 1999). The California gas and power disruption and collapse of energy trader Enron has further slowed market liberalization in the EU as politicians question the benefits of rushing into unbridled competition. Development of a competitive market has been further hindered by the decrease in trading companies in Europe – Enron was an important pillar of the EU gas markets before its downfall. In electricity, market structures vary from the U.K., which started with a relatively formal wholesale power pool, to Germany, where there is competition in retail markets but no mandatory participation in an exchange or common power pool and the industry remains vertically integrated.

Europe's gas producers have not been as diversified as those in North America. For decades, investment has been the result of the political aspirations of the host countries and business decisions of private companies. Moreover, there has been a large number of new supply and pipeline projects scheduled for completion at about the same time (Stern 1998 pp. 33-60). Although a number of these projects were postponed due to normal market pressures linked to abundant supply, net European pipeline capacity has expanded over the last few years. In short, Europe is in the throes of a boom in gas supply at the time that efforts are being made to liberalize the market. One consequence will be moderation in European gas prices, just as North America's gas prices rise. Canada's reserve-production ratio has declined from over 30 to around 10 in the last decades. This is combined with extremely high depletion rates in mature U.S. gas fields (EIA 2001), suggesting that it may be some time before U.S. gas prices moderate.

The gas market in the U.K. is the most advanced in Europe. BG began as state-owned enterprise. In 1986 it was privatized and in 1993 it was split into two companies by the U.K.'s Monopolies and Mergers Commission: a transportation company and a marketing company (Stern 1998 pp. 33-60). The transportation company is now known as BG Transco. As the transportation company evolved, the market in the U.K. was slowly liberalized. In 1990 large industrial customers were allowed to buy gas directly from producers and pay standardized transportation fees. In 1992 smaller industrial customers were allowed the same privileges, and in 1998 the right was extended to all gas consumers in the U.K. Today 19.5 million customers are able to choose their own supplier. BG Transco transports gas, but does not have a merchant function; the firm is responsible for ensuring that gas is delivered to consumers safely and efficiently (Transco web page).

BG Transco has six reception points for gas and 22 strategically placed compressor stations. The high-pressure transmission grid has more than 130 offtake points. Delivery of the natural gas futures contract takes place at the "National Balancing

Point" which is the notional point in the grid where the system is in balance between input and withdrawal. Onshore delivery of the gas from producers takes place in Scotland and Eastern England. BG Transco manages the balance of gas in the system through its National Control Centre; Area Control Centres step the gas down to each of the regions.

In the U.K. system, BG Transco itself is the "hub." The key concept is not the point where the commodity is exchanged for cash, but the "network code." This code is "…a legal and contractual framework to supply and transport gas. It has a common set of rules for all industry players, which ensure that competition can be facilitated on level terms. It came into effect in March 1996 after two years of negotiation between Transco and the shippers." (Transco web page). In the U.S., pipeline balance is achieved through a system of penalties for buyers and sellers that fail to meet their nominations and schedules. Although the systems are similar, BG Transco is more active in balancing the system by buying and selling gas and using its storage facilities.

BG Transco's activities are described as an "information business," which is an interesting idea for a pipeline company. They explain that to manage a large number of transactions by independent buyers and sellers, large amounts of information need to be rapidly exchanged and processed. BG Transco manages the market input and output using a computer system known as "UK Link."

Despite an earlier start, the European gas market lags the U.K. in its development. But clearly the new Interconnector pipeline will play a large role in the natural gas market's formation. The Interconnector connects the U.K.'s gas industry to the continent at Zeebrugge, Belgium, and gas flows in either direction. On the main continent, there are substantial interconnections around the giant Groningen gas field in the Netherlands and in central Germany. The Interconnector has a capacity of 20 bcm from the U.K. to the continent and 8.5 bcm in reverse. It is anticipated that it will have 15 bcm of capacity under long-term contract and 5 bcm will be reserved for spot contract carriage. According to Oostvoorn and Boots: "Before the U.K.-Interconnector has started operating, plans have been advanced to make Zeebrugge in a European hub, comprising both physical and paper trade." (van Oostvoorn and Boots 1999).

Although the Interconnector has been in service for several years, it has not had a substantial impact on the European gas market. According to analysts at the IPE, the European market remains dominated by a few big players, with long-term contract pricing as the principal form of business. A significant spot market has not yet developed in continental Europe. If anything the impact of the Interconnector has been the reverse. Mulcare (2002) maintains that it has increased U.K. spot prices, particularly during the winter heating season.

Europe may be lagging, but the U.K. has developed a viable daily spot market and natural gas futures. The IPE in London launched a natural gas futures contract in January 1997. The gas contract is based on a month-long delivery schedule with five hundred million Btu delivered each calendar day. Delivery takes place in the U.K. natural gas grid at the National Balancing Point.

Given that the U.K. gas system is now connected to the European continent, prices determined in this market will, sooner or later, have a major impact on prices throughout Europe, replacing the system of indexing gas to petroleum products. It will be very interesting to see how the market evolves in Europe and which region will emerge as the major market center or hub. The country that develops a successful gas market, with futures exchanges and high volume trading, will almost certainly play a dominant role in the industry's development in neighboring countries and, at the moment, that appears to be the U.K.

5.5 The Natural Gas Spot Market, Interconnected Pipelines, Hubs, and Market Centers

Although Europe and North American gas markets have many differences, they also have similarities. As gas markets mature on the two continents, common attributes will increase.

In the early stages of market development, gas pipelines usually have a single direction of flow from a producing province to a consumption center. But as more pipelines and storage facilities are built and, as the industry develops, the direction of flow in the grid is less obvious. Northwest Pipeline, which stretches from British Columbia to the Southern Rocky Mountains in the U.S., connects two producing areas. Gas is drawn off the pipeline along its entire length and the null point in the pipeline moves back and forth as levels of production and consumption change in the various spurs. A network develops when multiple pipelines are interconnected and it is the interconnection that compels the creation of a market. Complex gas grids do not necessarily have a distinct physical flow of the gas from producing wells to consumption nodes. In any case, the flow of gas usually doesn't look anything like the contractual commitments that arise from trading.

The natural gas industry is very capital intensive. Developing a gas field, constructing pipelines, or even installing a gas furnace also requires a substantial investment in a "component" product – one that depends on another to be viable. (Component products are explained in more detail in Chapter 3). The investment is worthless unless the gas is continuously available over the life of the equipment. For that reason, buyers and sellers in the formative period of the industry normally enter into long-term contracts. However, production problems, inadequate storage, pipeline bottlenecks, weather, etc. are largely unpredictable. Once pipelines are interconnected, consumers and producers find it beneficial to adjust contractual obligations through trade. For example, a factory may have to shut down for a month due to a labor dispute. If the owners have a take-or-pay contract for gas supplies, they are stuck with the bill unless they can resell the gas. An interconnected pipeline grid provides a much broader base in which to find a willing buyer.

The above example also demonstrates the role of arbitrage in the gas market. It is likely that a reseller of gas does not have the expertise to find a buyer and make the best deal, thus he will seek out experts in the market - marketers and brokers - who know potential buyers of the surplus gas. The first players in the new market are likely to be brokers—middlemen that simply match buyers and sellers. In early market development, the spot market is not reliable because of low volumes. No one wants to get stuck with a commodity that cannot be resold. As the spot market gathers volume, however, the most active middlemen are likely to be marketers. They take short-term positions in the commodity, which greatly enhances a market's liquidity.

To a large extent market maturity is measured by the size and reliability of the spot market. Once a robust liquid spot market is established, long-term contracts are less important. Having a diverse number of buyers and sellers, rather than linking a specific buyer and seller together with a long-term contract, achieves security of supply. An active spot market eliminates the risk of physical shortages or overflowing stockpiles. Instead, buyers and sellers face price risk, which gives rise to the development of futures markets. As might be imagined, however, establishing spot markets in a rigid distribution system, or when there is a lasting imbalance of demand and supply, is foolhardy.

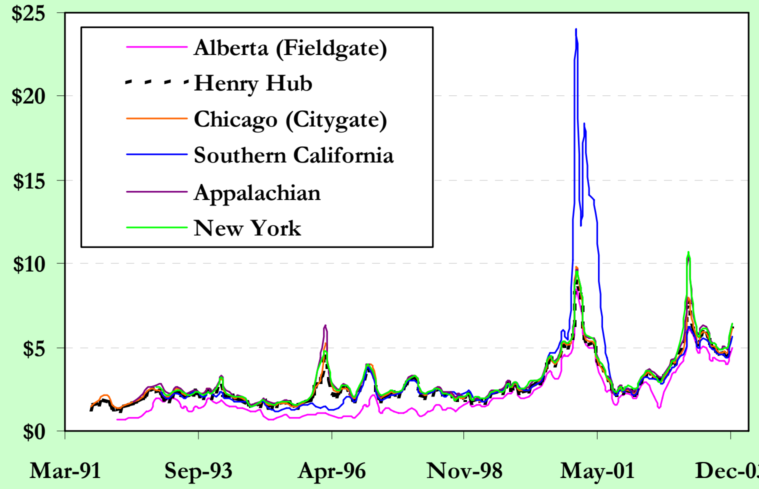
A quick glance at today's North American and European gas pipeline maps provides a clear impression. They form a spider's web of interconnected nodes, particularly in North America. In certain parts of the web, the lines are more thickly concentrated and interconnect a more diverse group of buyers and sellers. Any junction between two major pipelines has the potential to be a market center. Market centers are, as the name suggests, a point in the infrastructure in which there are frequent changes in ownership. A market center normally has an identifiable transportation differential between gas fields and consumption nodes as well as ownership arrangements that allow for trading.

Over time, market centers develop in which there is a great deal of physical interconnection and highly active trading. These heavily used market centers are called hubs. They are specifically configured to facilitate arm's length purchases and sales. In North America the single most important hub is Henry Hub, which is owned by the Sabine Pipe Line Company (a subsidiary of Chevron-Texaco). Henry Hub is interconnected to thirteen separate pipelines. It operates two compressor stations that allow the Hub to make deliveries to any of the high-pressure pipelines to which it is connected. Henry Hub is located near Lafayette, Louisiana in close proximity to the giant gas fields on shore and in the Gulf of Mexico. Pipeline hubs have also developed in Alberta, Canada at the confluence of pipelines that lead from Canada's producing fields to markets in Eastern Canada and the U.S. The most important of these hubs is AECO-C.

When active trading began in the North American gas market in the 1980s there were huge profits to be made, because the market was so inefficient. It is usually the case that the first companies to enter a new market and establish themselves grow rapidly and are very profitable. Companies like Enron and Dynegy (formerly Natural Gas Clearinghouse) became legends because they understood how the market was changing and how to position themselves in the new structure.[25] In the early days it was not uncommon to earn huge profit margins on quite simple transactions. This was because there were huge price differentials between regions and companies. All that had to be done was to follow the advice of Adam Smith (1776 p. 431): "Buy as cheap and sell as dear as possible."

The lure of high profits is what made the gas market change so dramatically. Companies that were trying to "buy cheap and sell dear" had an unintended impact: price differences across the continent began to disappear. From Mexico to Canada and from the Atlantic to the Pacific, regional differences moderated and the gas industry finally had a continent-wide market. Figure 5.2 illustrates recent prices and three disturbances that have caused larger than normal basis differentials to emerge. During the winter of 1995-96 there was a huge price spike in most of the United States. Prices in California, the Pacific Northwest and Canada were, however, unaffected because of shipping bottlenecks. Anyone that could ship gas from Alberta to Chicago could make a huge profit. This did not escape the notice of gas producers and others in Alberta, and they sought to expand pipeline capacity to eastern markets. They were successful and as Figure 5.2 illustrates, for a time Alberta enjoyed some of the highest prices in the continent. During the winter of 1996-97 all prices rose together in order to accommodate seasonal peaks. The winter of 1997-98, however, returned the Canadian market to its previous pattern. Shipping bottlenecks prevented adequate supplies of Alberta gas to flow the U.S. Midwest.

Figure 5.2 Monthly Average North American Natural Gas Prices



The most dramatic event in the chart concerns the California electricity and gas crisis of 2000-01. The disruption, discussed in more detail in Chapters 6 and 7, is a major, if short-lived, exception to the integration of the North American gas grid. California gas prices diverged from the Henry Hub price due to inadequate pipeline capacity between the State and producing regions. The constraint arose due to the radical

[25] Although Enron is now also legendary for one of the biggest corporate bankruptcies in U.S. history and allegations of illegal activities, the company was fundamental in the development and function of a liberalized market for energy in the U.S. (See Section 5.7).

increase in the demand for gas-fired generation in combination with constraints on intra-state "take away" capacity.

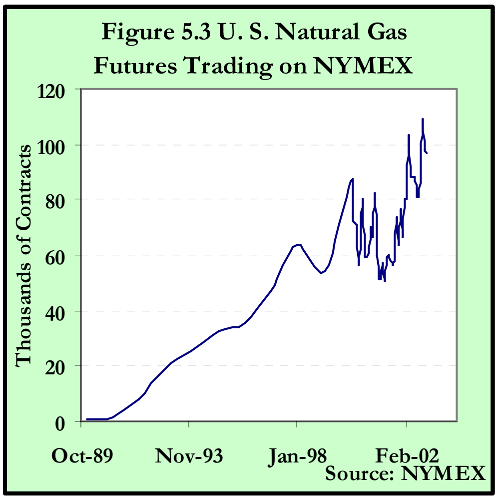
Following the California Energy Crisis gas prices declined; high prices had provoked a supply response and a recession lowered demand. As recovery has proceeded, however, gas prices have risen, reflecting reduced on-shore supply opportunities. One consequence has been a shut down of North American fertilizer and petrochemical industries that depend on cheap natural gas as a feedstock.

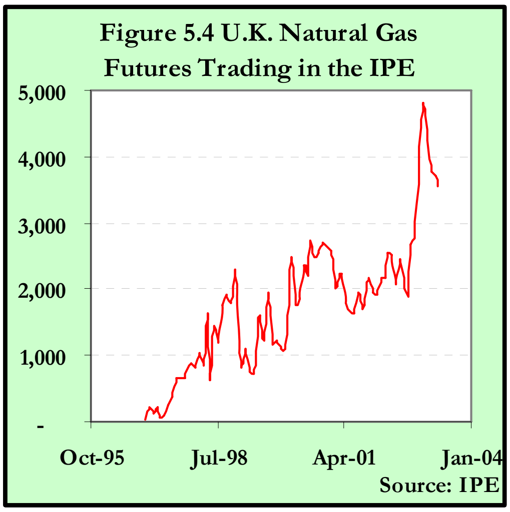
5.6 Natural Gas Futures

A robust spot market is an essential prerequisite for a successful futures contract. NYMEX studied the evolving natural gas market for several years before deciding on the specification of the product and the delivery location. In November 1989 Henry Hub was chosen as the delivery point because of both the extensive interconnection and Sabine's performance record, and in April 1990 the contract began trading. For the small number of futures contracts that are swapped for physicals, Sabine handles the deliveries just as it would any cash transaction.

NYMEX's gas futures contract was an immediate success with steady growth, as illustrated by Figure 5.3, which illustrates average daily volume by year and month. Slightly over 5,000 contracts were traded in the first month of trading. One year later, volume had reached 20 thousand contracts, and the year after that the total for April was nearly 100 thousand contracts. Monthly volume now hovers around two million contracts, putting gas on a near par with the crude-oil contract. Notably, despite all the recent controversy in gas trading and the manipulation of reported prices, the volume of gas futures trading has remained near peak levels.[26] To some extent, this demonstrates the significance of price transparency and the efficiency with which NYMEX, as an exchange, provides this service.

[26] The accounting discrepancies that followed the Enron scandal also uncovered instances in which traders had reported false data to the trade press in order to manipulate prices in their contracts, which were indexed to published prices. Reforms undertaken by FERC, the CFTC, and the industry itself have tightened the reporting process and a number of the traders that reported false data have been sentenced to prison terms.





Although gas futures trading did not begin in the U.K. until 1997, the contract there appears to be successful too. Natural gas futures trading volume is now averaging over three to four thousand lots per day, which is about one-half U.K. summer gas consumption and one-quarter of winter consumption. Figure 5.4 illustrates the growth in trading. Initially, the principal traders were gas marketers and producers. Once the volume grew, investment banks took an interest. Before the gas futures contract was launched, there had been over-the-counter trading in gas contracts for about three years.

As is evident in Figure 5.2, when gas prices go up in one region, they normally go up in other regions in parallel fashion. Thus, a gas consumer in Chicago can hedge expected purchases of gas by buying futures contracts that are tied to deliveries in Louisiana. Since futures contracts can be swapped for the product (at a future time), they are all part of the same market. When the gas market opened in the mid-1980s, there was little or no price consistency. The substantial variation in price at that time was explained by a number of factors: regional differences, variation in contract terms and other conditions, contract vintages, etc. As futures and physical markets matured, arbitrage ensured that price variation moderated and, with a few exceptions, it was mostly eliminated.

There is a feedback effect between volume growth in a futures contract and the physical market. Price signals determine how the gas will flow, and it does not take much of a price difference to resolve any market irregularities. As price differences at various geographic points diminish, the futures contract becomes more and more useful as a hedge, with a resulting growth in volume. The increase in futures trading enhances the market's liquidity, diminishing transaction costs. Lower trading costs in turn result in greater trading and contract turnover. The determinants of transaction costs in gas and oil trading are analyzed in more detail in Chapter 7.

The price difference between the various regional hubs and market centers (basis or basis differential) can be substantial in the gas market. The inflexibility of the pipeline delivery system significantly increases the risk that prices in one region will diverge substantially from other regions for prolonged periods of time. During the 2000-2001 period, gas prices in New England and California unexpectedly rose to levels much higher than in the central U.S., increasing basis risk and reducing the effectiveness of the NYMEX futures contract as a hedging device.

It is important to grasp that basis differentials are not the same thing as transportation costs. In a pipeline delivery system the molecules that the buyer actually receives may bear little or no relationship to the ultimate source of supply; it is like putting water in a reservoir and then drawing it out elsewhere. With a single producing and consuming region, price differences might be explained simply by the cost of moving the commodity from one region to another (to a large extent, the system of oil pricing prior to the oil shocks was based almost completely on transportation cost differentials). In a complex patchwork of pipeline intersections and multiple producing and consuming regions, such as that illustrated in Figure 5.2, price differentials are set by a host of interactive factors. These can include differences in production costs, variable pipeline costs, bottlenecks and capacity constraints, exchanges and displacement, etc. The interaction of these costs and demand is extremely complex. The market sorts out basis differentials through price bidding for gas production, pipeline capacity, and storage. While the results are clear, the inner workings of the process are largely unknowable.

The North American natural gas pipeline network has produced a system of pricing and distribution very different from its regulated predecessor. Arlon Tussing (2004) describes the determination of relative prices in the North American pipeline systems in this manner:

> The North American pipeline network is composed of literally tens of thousands of discrete segments, and literally hundreds of thousands of intake or delivery points. Each segment, intake or delivery point, has a unique minimum and maximum capacity whose determinants may be physical, contractual or regulatory. Cost functions for each of these components are non-linear and often discontinuous.
>
> With respect to the economics of backhauls, some segments, intake/delivery points, and interconnections are physically, economically, contractually, and legally reversible; others are not. "Backhaul rates" still exist in many filed pipeline tariffs. But market participants viewing the arbitrage opportunities inherent in locational price differentials, do not have to be expressly cognizant of the physical direction of flows in the various pipeline segments.
>
> Given such a framework, a clear explanation of the determinants of basis is not really attainable, beyond an enumeration and description of the factors and forces involved. Locational price differentials in large networks are in principle not solvable by multivariate analysis, nor can they be usefully simulated with sufficient precision to guide sales and purchase or dispatch decisions, even with immense data input and computational power. The market reveals basis by empirical observation, and reacts to it by arbitrage ---involving trial and error/successive-approximation. As a result, participation in natural-gas commodity markets is exceedingly imprecise and riddled with error. Successful trading more often than not depends on an intuitive grasp of the system as a whole.

Under FERC oversight, pipelines make firm capacity available to shippers at fixed rates based on cost in the primary transportation market. Shippers have the right to resell unused capacity in a capacity release (secondary) market. Rates set in the secondary market are akin to spot prices for gas, i.e., they vary with the daily market and provide transportation flexibility. At times the secondary rate is above regulated rates and at times it is below. The decision to expand pipeline capacity is made based an open season of bids. If sufficient numbers of prospective shippers are willing to pay for the expansion, the pipeline will likely proceed with the investment. Otherwise, the expansion is shelved. The substantial price increase in the delivered price of gas in California during 2000-2001 provoked an aggregate 20% increase in capacity from Canada and the Rocky Mountains to the state in a series of pipeline expansions.

Consider the circumstances of a gas buyer in Chicago. A variety of marketing and energy firms will offer to sell gas at an array of prices. The buyer might decide to purchase gas bundled with transportation and other services. Alternatively, the buyer might decide to acquire the gas at a producing region and contract for delivery directly to the plant in Chicago. If the buyer elects to buy gas in the field combined with spot transportation, the cost of gas will likely parallel delivered prices in the Chicago area. If, on the other hand, the buyer elects to secure firm transportation, the cost of gas could depart from delivered prices in Chicago. The buyer would likely be protected from price spikes in the local market but could not take advantage of a local glut. A buyer that purchases unbundled gas may also have to deal with balancing services and other delivery issues if actual gas consumption varies from contract volumes.

The volatility and complexity of the gas market necessitates risk management for most participants. So far the discussion of derivatives to manage price volatility has concerned futures and options contracts traded on an exchange. Given, however, that there is considerable risk in gas trading arising from unexpected movements in the basis differential, the NYMEX contract frequently proves inadequate for buyers and sellers in remote regions. There have been two solutions to this problem. The first has been the development a smaller volume exchange in Alberta, the Natural Gas Exchange (NGX). The second is the rich tapestry of financial and physical products available through bilateral trading in the OTC market.

The second most important gas-producing region in North America is the western sedimentary basin of British Columbia and Alberta. U.S. West Coast states and the upper Midwest have long been dependent on Canadian natural gas. In response to the two oil shocks of the 1970s, the Canadian government briefly constrained sales through export levies at the border to price natural gas at the thermal equivalent of heating oil. The gas bubble that emerged in the 1980s, however, forced Canada to reconsider its gas export policy. At the time, the gas surplus in Canada was substantially greater than in the U.S., with the reserves to production ratio peaking at 30, three or more times higher than that which is necessary to develop and produce gas economically. Once export restrictions were relaxed, Canadian producers began active marketing to U.S. buyers and the volume of exports grew by 389% from 1985 to 2001 (BP). Canada now supplies 18% of U.S. gas demand, exporting a greater volume south than it consumes itself.

Alberta's natural gas exchange began operations in 1994, sponsored by Westcoast Energy Inc., Canada's largest gas pipeline company (NGX). It has subsequently been acquired by OM, a Scandinavian firm specializing in electronic exchange platforms. NGX provides not only a trading platform, it includes clearing and settlement for both physical and financial products. In essence there are three types of products offered. First, there is the purchase and sale of natural gas at various hubs and market centers in Canada on the U.S. border. Second are basis swaps which allow Canadian gas producers and U.S. buyers to hedge transportation differentials between the field and U.S. delivery points. Finally, there is a basis swap between Alberta and Henry Hub, which allows market participants tied to Canadian supplies to hedge against broad movements in gas prices using the NYMEX natural gas contracts.

5.7 The Significance of Enron

It is not infrequent in the development of markets that a particularly aggressive and innovative company ends up defining the shape of the entire industry. In the oil industry John D. Rockefeller founded Standard Oil Company and defined an industrial structure that has lasted for a century. Bill Gates and Microsoft have done the same thing for the software industry. Kenneth Lay was not nearly as well known before the historic bankruptcy of the now-infamous energy giant, but under his leadership, Enron evolved from a modestly successful domestic gas transmission company to a Major international energy company.

Enron once described itself as the "the world's first Major international gas company" (Enron Corporation 1997). It would probably be more accurate, however, to describe Enron as the world's first global energy trading company, because it was trading and the use of modern risk management tools that built the company.

In the 1980s, energy industry executives were caught in the crossfire between emerging spot markets and the tradition of securing energy supplies at any cost. Most electricity and gas companies were regulated as franchise monopolies. They had the exclusive rights to market in a particular region, but with this right came the "obligation to serve." That is, their prices were regulated on a cost-plus basis, but, in turn, they could not refuse to serve a customer in their service territory. Electricity companies that had to purchase fuel for generation would place great emphasis on security of supply and be less concerned about prices paid. Gas distribution companies also would not be price-sensitive in their wholesale purchases. To this regulatory mix was added the experience of the 1970s, when there were multiple shortages of oil and gas. Even though "gas-to-gas" competition was producing low spot prices, the market was still described as a gas bubble and the common expectation was that prices would soon be on the rise.

In the 1980s, environment, gas and electricity companies were willing to pay a substantial premium over natural gas spot prices in order to secure fixed price long-term supplies. Many gas producers, on the other hand, were reluctant to enter into long-term contracts of any sort. In general, producers were receiving a fraction of the value of the high-priced take-or-pay contracts they had agreed to during the price run-ups of the 1970s. In their experience, they were required to supply gas if their contract prices were below spot prices, but buyers seemed to be able to abrogate contracts when the price structure reversed.

Enron seized on the arbitrage opportunity inherent in the gas market as a means to leverage their pipeline and producing assets. Anyone willing to buy gas short and cheap could sell it long at very high prices, if they were willing to take the risk. Enron could have made extraordinary profits pursuing that strategy, but that would have been risky and unnecessary. Instead, they developed an arsenal of risk management tools that reduced their risk and still generated enormous profits, because their competitors were way behind. Neither gas producers nor buyers yet realized the full potential of derivates trading. In the simplest arrangement, Enron would sell gas to a utility at a high fixed price and buy gas from producers on the spot market at much lower prices. They would then cover their position by buying or selling gas in futures and options contracts, which locked in a guaranteed profit and eliminated risk.

Enron's early growth might have been fueled by risk-taking and perhaps it was. The "gas bubble" lasted much longer than anyone anticipated, which meant that spot prices were below long-term contract prices for over a decade. However, the recent market for both electricity and gas has had a number of price run-ups, especially during the California Energy Crisis of 2000-2001. In every case, Enron appears to have made hefty profits, while many of its competitors took substantial losses. Enron's sophisticated procedures, market monitoring, and control of its traders appeared to have protected the company from highly volatile markets, if not from its own accountants.

As electricity deregulation began, Enron expanded and became the largest marketer of power in the U.S. until their bankruptcy. It is interesting to note that since the wholesale electricity market opened up in the U. S. gas marketing companies have dominated it, because they had the necessary skills to trade successfully.

Enron also leaped ahead of its competitors for on-line Internet trading. Enron-on-Line was one of the most visible and heavily used of the new services. Enron had announced that total trade at their site for the first five months of 2000 was over $50 billion. This was still small in comparison to the annual trade of $2 trillion for energy at NYMEX, but it was an impressive start.

In addition to new domestic activities, Enron International was highly active in the liberalization of European energy markets. They acquired gas pipelines and electricity assets in South America and Asia. They were well on their way to being a Major company, although probably not the one they had envisioned only a few years before.

On December 3, 2001, Enron declared bankruptcy. Its strategy all along had been to acquire hard assets only long enough to learn as much information as possible, hire on the best and brightest employees gained in the acquisition, use them to trade in the commodity, and then sell off the underlying asset (such as a pipeline or generator). Its focus on trade and risk management was much more lucrative, and Enron was more agile without the encumbrance of hard assets. However, when the run on the banks started and some of their trading positions became liabilities, Enron had nothing solid to back them up. In retrospect, it is not difficult to see that banks, investment firms, and other financial institutions, with their well-balanced blend of investments and assets across multiple industries, are better equipped to handle trade and risk management in industries as volatile and risky as electricity and natural gas trading.

5.8 Market Impact on Organizational Structure

Since the breakup of the Standard Oil Trust, the preferred organizational structure of the oil industry has been vertical integration. As pointed out in Chapter 4, however, the facts about vertical integration do not always match the legend. Nonetheless, refiners often seem convinced that they need to control their own supply sources and markets if they are to be profitable. There ought to be a similar and even stronger logic for the natural gas industry: gas producers could easily become stranded by pipeline owners that refuse to complete connections or shut down, unless price discounts are offered. In the language of transaction cost economics, oil and gas pipelines are classic "specific assets;" they connect to individual fields or markets and have virtually no alternative value. Thus, gas producers should be motivated to acquire pipelines and distribution companies. However, despite the strong incentives to combine, there are few examples of vertically integrated gas firms.

In Europe, geography and constrained private property rights have heavily influenced the separation of the ownership of natural gas fields and distribution systems. Many of the region's largest gas producing countries – Russia, Norway, Algeria, and Tunisia – are, of course, outside the core of consuming states – France, Germany, Belgium, and Italy. Indeed Europe's gas transmission and distribution companies have been described as "gatekeepers" (Stern 1998 p. 10). As the term implies, the gatekeeping companies are, as often as not, instruments of national policy. Similarly, gas development and production decisions from Europe's periphery are largely in the hands of governments, not private companies. Thus, the opportunities for vertical integration are significantly reduced or wholly eliminated. The U.K. provides a counterexample to the European dilemma; North Sea gas producers might have integrated into pipelines and distribution companies but for regulatory barriers.

Vertical integration of the gas industry in North America has also been constrained. As noted earlier, most natural gas crosses state lines, bringing it under the regulatory umbrella of the federal government, which mandated cost-based pricing until the mid-1980s. Prior to the gas industry's deregulation, a transmission pipeline buying gas from an unaffiliated company was allowed to pass the full price along. If, however, the pipeline supplied itself, only historic costs plus a reasonable margin were allowed. Such a cost-based system is unworkable in oil and gas exploration, where profits from a prolific field have to offset losses from an undefined number of dry holes. According to Mulherin (1986): "It is apparent, therefore, that the rate regulation of the Federal Power Commission presented a strong disincentive for affiliation between producers and interstate pipelines. This Act together with the Public Utility Holding

Company Act led to a significant reduction in the amount of vertical integration in the natural gas industry."

Economies of scope also help explain the segmented market. As noted, natural gas discoveries were often the side benefit of oil exploration. The specialized skills necessary to find oil match those necessary to find gas and have little value outside the sector. Since the preferred structure of the oil industry was vertical integration, further amalgamation into gas pipelines and distribution companies would create an unwieldy corporate structure; one where the advantages of specialization might be overwhelmed by the problems of coordination and communication. The corporate culture of a franchised distribution monopoly is fundamentally different from the risk-taking outlook of prospectors.

Mulherin (1986) also suggests that federal regulatory oversight provided a strict framework for contract enforcement, which significantly reduced the risk of opportunistic behavior. He noted that gas contract structures, with "take or pay" provisions and other complexities, were aimed at protecting ownership rights associated with specific assets. The regulatory and contract structure may indeed have inhibited opportunistic behavior, but it had a downside: it created an inflexible market, where prices were not allowed to balance demand and supply. This inflexibility forced regulators, gas suppliers, pipeline companies, and distribution companies to forge a wholly different type of structure, one based on the successful oil commodity market. When the North American gas market was deregulated, policy makers had a viable market model, an alternative to vertical integration or rigid long-term contracts – the oil commodity market.

6. *Breakup of the Power Industry*

6.1 Historical Industrial Structure

In the early twentieth century a variety of high-octane entrepreneurial firms made a painful transition from rapid growth and innovation to apathy and stagnation. It is striking to look at early photographs of the U.S. electricity industry run amok, with competing power lines clogging streets. Duplication and waste were obvious, as well as the public hazard arising from the confluence of lethal power lines a few feet above urban crowds. Leapfrogging demand, uncertainty over standards (the basis on which to homogenize the delivery characteristics of electricity), and the value-added of obtaining rights of way through public and private land led to voracious competition. Obviously the companies that obtained key rights of way and established voltage and other technical standards would achieve nearly unassailable market dominance. There is, of course, a modern parallel in the struggle over control of operating systems for personal computers. New technologies create the opportunity for the accumulation of great wealth through secret proprietary systems, patents, and uncertain property rights, and thus are attractive, if risky, investments. Inevitably, the early markets that support these new technologies have high transaction costs.

The rapid expansion and chaotic development of the industry was followed by the Great Depression, which in the U.S. precipitated the collapse of a number of privately owned electric utility holding companies, including Samuel Insull's empire.[27] The resulting bankruptcies and financial disorder gave rise to the PUHCA in 1935, administered by the Securities and Exchange Commission (SEC). This Act broke up multi-state holding companies and constrained the development of utilities, essentially limiting them to local distribution areas. Richard Gordon (1992 p. 59) summarized its impact: "The act requires the companies to operate as coordinated, integrated systems, confined to a single area or region." Gordon (2001) also estimated that in 1998 there were 3,170 utilities operating the U.S. (many, however, are small distribution companies than own no generating resources). Instead of a national grid, North America ended up with scores of separate operating areas and pools with very limited coordination. In contrast, most European nations opted for a national grid, frequently dominated by a single distribution company.

The rapid electrification of Europe and North America culminated in different types of ownership rights and regulatory oversight, but a frequently occurring industrial structure: vertically integrated utilities. The industry centered around three key elements: power generation, high voltage transmission in an interconnected grid, and distribution. The need to integrate generation and transmission was largely unquestioned, because a power grid links buyers and sellers in simultaneous generation and consumption. Traditionally

[27] Samuel Insull was general manger for Thomas Edison's electric manufacturing companies. He moved to Chicago and over several decades built an empire of 60 gas and electric utilities. In 1932, in the midst of the depression, the over-leveraged empire collapsed, destroying the savings of thousands of investors. Insull was indicted for fraud, but a jury failed to convict him, in part, because he too was wiped out by the financial collapse. Insull is credited with the wide-scale promotion and development of the electricity industry.

dispatchers have managed a centralized grid, with the authority to order generators to increase or decrease output and load to curtail consumption as necessary. For most commodities, there need not be simultaneous production and consumption, since inventories may be used to smooth out the differences or consumers can simply be turned away. In an electrical grid, however, matching electricity demand and supply is not just desirable; it is a necessity. In this respect electricity as a commodity is nearly unique: if minute-to-minute demand exceeds supply, voltage drops, transformers pop, and the grid collapses in a blackout.

During the period of rapid expansion, it was also discovered that power generation had extensive economies of scale; i.e., building larger and larger facilities could reduce per-unit costs. Thus, there were thought to be strong elements of natural monopoly in all segments of the electricity industry. By the 1970s, however, scale economies of generation had reached their limit at a capacity for coal and nuclear of around 1,000 MW. In addition, new technologies in fuel cells, cogeneration, solar cells, and distributed generation promised to reverse the advantages of large-scale generation. This gave rise to the notion that the generation segment of the power industry could be separated from transmission and distribution and that generators might compete, rather than be consolidated into a single regulated monopoly.

Electric utilities throughout the world are frequently a complex combination of public and private ownership. In almost all cases privately owned utilities are regulated with respect to price and customer obligations. A nearly universal standard is applied when utilities are privately owned, and it is known as the "utility compact." That is, when companies are granted an exclusive monopoly or franchised service territory, they must, in return, grant service to any and all customers at regulated rates. The extent of regulation varies from country to country, but is most complex in the U.S. where there are many overlapping regulatory agencies and rates have been based on cost with allowance for a normal return on capital and adjustments based on the "prudence" of the utility's activities.

As explained in Chapter 4, the transformation of the oil market from OPEC's system of contract sales at cartel prices to an open commodity market was an evolutionary process initiated by participants who sought to avoid constraints, lower transaction costs, and reduce risk. In contrast, Chapter 5 observed that the transformation of the North American natural gas industry was the product of unworkable economic conditions, piecemeal regulatory changes, and the migration of oil trading institutions to the gas industry. The transformation of the power industry, on the other hand, is remarkably different, because policy makers in the U.K., the U.S., and other countries have often set out to revolutionize the industry in a single stroke.

6.2 Independent Power Producers

One little-noticed change in U.S. regulatory procedures in the late 1970s had more to do with the fundamental structural change of the power industry than any other policy measures taken in response to the energy crises. In the face of the natural gas shortages of 1976-77, Congress passed a comprehensive legislative package known as the National

Energy Act (NEA) of 1978.[28] The various programs and incentives reflected a shift from President Nixon's and Ford's polices which were aimed at promoting conventional energy supplies of nuclear, oil, and gas. Instead, President Carter emphasized energy conservation and the development of "renewable" energy. The new package of legislation contained the Public Utility Regulatory Policy Act (PURPA), which became the key framework for the development of IPPs.[29]

Although it was not its apparent purpose, PURPA undermined the exclusive rights of integrated electric utilities to construct and own generating facilities. Essentially, the new law mandated that utilities must contract with "qualifying facilities" (QFs) to purchase power at their "avoided cost," i.e., the cost to the utility of constructing new generating facilities. The interpretation of the law and its application varied within the U.S., but it was particularly relevant in California, New York, and New England, both of which faced high development costs. PURPA's advocates had in mind that it would encourage the development of wind, solar, cogeneration[30], and other types of energy resources that had low environmental impacts. In practice, however, PURPA became most significant in encouraging the development of conventional energy resources, particularly gas-fired generation. Such a development was all the more ironic because the NEA also harbored the Power Plant and Industrial Fuel Use Act (Fuel Use Act), which prohibited the use of natural gas for new power generating facilities, with the important exception of cogeneration.

California was an especially attractive arena for the development of QFs due its strong demand growth, high-cost utility generation, and the opportunity for large-scale cogeneration projects in its heavy oil fields. The State's major utilities were determined to construct additional nuclear reactors at per-unit costs in excess of $100 per MWh. The consequence was a set of "standard offers" developed through a regulatory process. These offers ensured that developers would obtain long-term contract prices matching the utilities' avoided costs.

Because the owners of the State's heavy oil fields were companies like Texaco, Chevron, Shell, etc., California's prospective QFs did not have the usual financing barriers, but there were multiple regulatory hurdles. Production in these fields required the injection of large amounts of steam in order to pump the oil to the surface. Typically about one out of five barrels of oil was burned in order to generate the necessary thermal recovery. Proposals were made to build a natural gas pipeline from the Northern Rocky Mountains to the San Joaquin Valley, allowing additional natural gas to be substituted for oil in

---

[28] The NEA was the outgrowth of President Carter's 1977 National Energy Plan. It contained: The National Energy Conservation Policy Act, The Fuel Use Act, The Public Utility Regulatory Policy Act, The Natural Gas Act (see Chapter 5), and the Energy Tax Act.

[29] The IEA's *1978 Review of Energy Policies and Programmes of IEA Countries* describes the shift in regulatory policy as follows: "The Public Utility Regulatory Policy Act of 1978: State regulatory agencies are required to review rate structures to encourage conservation. FERC is empowered to demand integration of electricity grids and wheeling, and to arrange power exchanges between utilities and industry." No mention was made of what became of the key provisions that created independent power producers. At the time, electricity trade and pricing were understood to be part of the regulatory paradigm. No one envisioned market-based pricing.

[30] Cogeneration is known as the combined production of heat and power (CHP) in Europe.

Thermally Enhanced Oil Recovery (TEOR). Gas was used for the combined production of steam and power, but the economics hinged, of course, on being able to sell the surplus electricity. After extensive debate at FERC, the CPUC, and the California Energy Commission (CEC), the Kern River pipeline was approved and construction was completed in 1992. Throughout the early 1980s, the CPUC offered a series of Standard Offer Prices for power purchases from QFs. Long-term contracts were offered by the State's utilities and included both capacity and energy payments.[31] California's utilities also contracted for wind power and other types of generation that qualified under PURPA. By the time the state restructured its power market, there were over 600 PURPA contracts, representing about one-fifth of supply.

New England is at the end of the energy supply line in North America, with virtually no indigenous oil, gas, or coal production. More than any other region, it is tied directly to the international oil market. The dependence on imported oil caused the region's electric utilities to embrace nuclear power, but these projects were much more costly than anticipated. According to Fred Pickel of Tabors, Caramanis, and Associates, one of the region's largest utilities, New England Electric, recognized the advantages inherent in contracting with QFs and other IPPs rather than constructing high-cost risky plants. They opened their grid to other potential generators in order to further diversify supply.

As noted, PURPA's advocates had meant to encourage small-scale renewable energy supplies as an alternative to nuclear power. At the time no one envisioned the dramatic improvement in the thermal efficiency of gas-fired CT and CC plants.[32] Most utilities treated these options as suitable to meet peak demand, but not for base load and, of course, the Fuel Use Act effectively blocked any planning for larger-scale gas generation. Given relatively low gas prices arising from the gas bubble (see Chapter 4), however, there were enormous incentives for the development of large-scale gas generation to meet base load demand.

The attractive PURPA offer prices in California and New England gave QFs substantial incentive to develop large-scale gas-fired generation, if exceptions could be found to the Fuel Use Act. Even though it was large-scale, the TEOR development in California could qualify for an exemption as a QF, because technically it was cogeneration. According to Fred Pickel, New England did not have that advantage, but it did have aggressive IPPs who recognized the opportunity. In 1984 a developer, J. Makowski & Associates, set about to construct a stand-alone large-scale gas-fired generation plant in New England by seeking an exception to the Fuel Use Act. Simultaneously arguing for "no market power," they succeeded in getting released from the strict cost of service regulation, thus creating the first IPP. They were successful, paving the way for other entrepreneurs. The group also participated in the development of the Iroquois gas

---

[31] Long-term power supply contracts frequently contain a capacity and energy payment. The capacity payment reflects the amortized cost of the capital investment, while the energy price reflects the expected cost of fuel and other operational and maintenance expenses.

[32] Traditionally power generation from oil or gas was thought to "lose" two-thirds of the primary thermal energy. A kWh of electricity has a heat content of 3,412 Btu. Before the recent technological innovations, it took an average of around 10,000 Btu of gas, oil, or coal to produce a kWh of electricity. Modern CC facilities now require as few as 6,000 Btu of gas to produce one kWh of electricity.

pipeline. Soon there was growing pressure to repeal the gas use prohibition in the Fuel Use Act, and it was amended in 1987.

Nonetheless, the development of gas-fired power generation was constrained by fears of potential gas scarcity even after the Fuel Use Act was amended. The gas shortages of 1976-77 and oil shortages of 1979-80 were still fresh in everyone's mind, along with repeated warnings from analysts and policy makers of future shortages. The Fuel Use Act was, at best, a loose-knit prohibition on new gas-fired generation. It was, however, accompanied by an onslaught of government policy pronouncements, studies, and proposals that were aimed at reducing dependence on oil and gas, which on the margin had to be imported. Instead, the nation's utilities were encouraged to develop coal and nuclear generation, and in most regions this meant adopting the nuclear option.

In retrospect it is clear the focus on nuclear power was a costly diversion. Proposals to build nuclear power stations provoked a surprising amount of public antipathy and aroused the combative instincts of environmental coalitions. While the federal government advocated the development of nuclear power, it had little control over the pace of construction; site selection and approval was strictly a local regulatory decision. Site selection invariably involved the drafting of an environmental impact statement (EIS) and accompanying environmental and public safety hearings. This created a forum for opposition and objection, raising a broad array of issues and significantly lengthening planning and construction timetables. The staunch environmental opposition had an unintended impact: it obscured and degraded economic and financial analysis. Pressed on a variety of safety and environmental issues, electric utilities were extremely slow to recognize or acknowledge that nuclear power was simply not cost-effective in many instances.

When nuclear generators were constructed, the high cost had a substantial impact on retail rates. As the controversies over site selection died down, new battles were waged over allowed cost recovery. In addition, escalating power prices curbed demand growth and compounded planning uncertainties. U.S. electric utilities were caught in a bind; the utility compact required that they serve all customers, but there was increasing resistance to their generation proposals. PURPA offered an attractive solution. Utilities could contract with QFs at avoided costs. Regulators would approve the contract in advance, eliminating the risk of not recovering capital. Further, the cost and hassles of site approval would be shifted to another company, one not easily identified by local opposition groups. In addition, utilities that wished to remain in the generation business could create separate subsidiaries of their parent companies. As long as these companies were independent of the utility, they could be excluded from cost-based regulation, promising much higher profitability.

Maturation of the gas market helped lower the cost of gas-fired IPPs. Building a merchant power plant in the 1980s was an extraordinary feat. According to Fred Pickel, developers described the feat as holding "seven greased pigs." The developer had to 1) secure a power contract with the local utility, 2) contract with a supplier for fuel, 3) complete and manage a construction contract, 4) obtain a site certificate from state and local regulators, 5) obtain all construction, operational, environmental, and safety permits,

6) for cogeneration a steam sales contract would be required, and finally, 7) complete financing. Even though there was a gas bubble and spot prices were low, long-term gas purchase contracts could only be acquired by paying a substantial price premium, reflecting an expectation of future shortages and the absence of a liquid forward market. All in all, the task was a massive one, where the transaction costs involved formed a substantial barrier to entry.

Although they proceeded in fits and starts, PURPA projects slowly drove a wedge between power generation and the integrated utility structure. To be sure, the generator and the utility were locked together in a long-term contract. The contract committed the capacity of the generator as well as gave dispatch control to the utility – the physical linkage between generation and transmission was not changed, but property rights were. This meant that PURPA generators and grid owners had diverging economic interests.

6.3 Spot Pricing and Power Pools

Following World War II, many European countries had inadequate electricity generation capacity. One solution, particularly in France and the U.K., was the adoption of peak-load or time-of-use pricing, instead of a flat price. Because customers draw power directly from the grid at their discretion, utilities are forced to maintain substantial reserves to meet demand at its highest peak. If customers are offered discounts for power consumed in off-peak periods (usually weekends, late evening and early morning hours), they may be willing to shift their consumption away from the peak hours. A drop in peak demand, relative to off-peak demand, can reduce the amount of unused generation capacity, allowing existing capacity to be used more efficiently.

Regulated electricity prices in the U.S. have been based on actual costs incurred, and the higher cost of a new project cannot be included until the plant is operating. One consequence is an acceleration of demand growth in periods of rapid inflation, straining generation capacity. In the midst of the 1973-74 oil crisis, many U.S. utilities' capacities were stretched to their limits. One consequence was increased reliance on oil-fired generation, because it could be installed quickly, was more flexible, and was sometimes a carryover from decisions made when oil was cheap. Federal policy makers saw marginal cost or time-of-use pricing as an energy conservation tool and a means to improve the economics of base-load facilities - coal and nuclear generators - reducing the use of imported oil.

The interest shown by federal energy policy makers in marginal cost pricing was paralleled by interest at progressive utilities and state regulatory commissions. Experiments and studies in marginal cost pricing were conducted at Wisconsin Power and Light, Sacramento Municipal Utility District (SMUD), and the Los Angeles Department of Water and Power (LADWP) (Cicchetti, Gillen, and Smolnesky 1977). Much of this experimental work was based on research conducted for the World Bank by Ralph Turvey and Dennis Anderson. Earlier Boiteux (1949) had published the rationale for the approach. The Electric Power Research Institute (EPRI) also studied time-of-use pricing and developed programs for its implementation.

As interest in marginal cost pricing grew, however, it became increasingly clear that in a modern power system the idea was incomplete. Peak American electricity demand was concentrated on summer air conditioning load. Peaks were determined mainly by hot spells, which were not predictable. Rate schedules set months (or years) in advance could not be sensitive to unforeseen events. In addition, capacity shortages (when marginal costs were highest) did not necessarily occur during periods of peak demand – it also depended on maintenance schedules and forced outages. This led to the inevitable idea that real-time pricing, sensitive to the variation in marginal costs and demand, was required to optimize the system.

Early advocates of electricity rate reform recognized the difficulty in capturing marginal cost with pre-packaged rates. "If marginal cost can be described so briefly, why is it so difficult, technically speaking, to reflect it accurately in tariffs? We have touched on the answers earlier: high variability and unpredictability in demand, and cost and bother in metering" (Turvey and Anderson 1977 p. 15). The problem of demand variability is, of course, heightened by the inability to cost-effectively store electricity. In addition to seasonal variation, electricity demand varies depending on the time of day, day-of-the week, etc. In theory, real-time pricing would require consumers as well as producers to monitor price changes throughout the day, substantially increasing transaction costs. Even if consumers failed to watch power switches, metering, like all measurement, is a transaction cost, a cost that for small-volume consumers might be considerably higher than the benefits. Despite the constraints, the CEGB in the U.K., and EDF in France had developed reasonably sophisticated peak load pricing tariffs, which required real-time metering of large customers.

Despite the practical and cost-related problems of developing real-time electricity pricing for residential customers, it certainly made sense to consider it for large-scale commercial and industrial users. The practice of load shedding (cutting power to specified large customers in return for lower rates) was a procedure that recognized the implicit high value and cost of electricity at periods of peak demand. Marginal cost pricing, PURPA supplies, and load shedding contracts were studied by academics, particularly at the Massachusetts Institute of Technology (MIT), where in the late 1970s a theory of spot pricing began to evolve based initially on industrial load modeling. Summarizing a decade of experiments and study, Schweppe, Caramanis, Tabors, and Bohn (1988 p. xvii) noted: "There is a need for fundamental changes in the ways society views electric energy. Electric energy must be treated as a commodity which can be bought, sold, and traded, taking into account its time- and space-varying values and costs."

6.4 Experiment in the West

Utilities in the U.S. had recognized for some time that without well-coordinated regional grids, operating costs would be higher than necessary. On the one hand, managing a small operating area allowed dispatchers to achieve intimate knowledge of the system in order to efficiently match generation and load. On the other hand, inadequate integration meant that each utility faced greater risk from forced outages and was required to maintain higher than necessary reserve margins. In addition, utilities operating in a single region could not take advantage of seasonal differences, hydroelectric surpluses, etc. The solution was to form power pools.

In the eastern U.S., high population density meant that operating areas converged on one another, with a dense interlacing of transmission and distribution lines. Utilities in this region banded together to form "tight" power pools, including the Pennsylvania- Jersey-Maryland (PJM) pool, the New York Power Pool (NYPP), and the New England Power Pool (NEPOOL). The essence of a tight pool was an agreement among participants to centralize operations and operate generating units in an economically rational sequence – "economic dispatch." Pool operators would dispatch units based on their marginal cost, taking into account the impact on grid congestion. FERC approved the formation of power pools and allowed trading between utilities on the basis of "shared savings." That is, the difference in marginal cost between the unit being brought up and the one being shut down would be shared equally between the two participants.

The western power grid could not be organized in the same manner as pools in the east. This is because the region's principal population centers and resource bases are widely dispersed. Operating areas are mainly centered around the region's cities and each city is served by a different utility. Long distance transmission lines connect the coal-bearing regions of the Rocky Mountains to the region's cities. They also connect the Pacific Northwest and Canada to the Pacific Southwest and Northern Mexico in a set of transmission lines known as the "North-South Intertie." The combined capacity of the Intertie is about 8,000 MW, roughly 5% of the region's generation capacity. The line allows the Pacific Northwest to move power to California during the summer, the latter's period of peak demand, and to receive power during the winter. Beginning in 1971, FERC approved a series of seasonal exchange agreements between utilities in the north and south.

In 1987 FERC approved the formation of the WSPP. This pool was quite different from the ones established in the east in that it allowed "market-based pricing." Unlike the share-the-savings schemes that were based on cost, utilities in the west could charge whatever the market would bear for both transmission and energy. This established a genuine wholesale market, but it was limited, because there were significant barriers to entry. Also, the WSPP market was not transparent. Participants knew about power prices, but the general public had no access to them.

Although the WSPP market was opaque and not closely watched by either regulators or observers outside the industry, a number of remarkable institutions and trading conventions developed under its umbrella. The trading system was entirely bilateral. Unlike the tight power pools of the east, there was no central authority. Trade was between scheduling utilities, with each controlling its own operating area, thus minimizing reliability issues. No one was obligated to trade, but if power could be obtained from a neighboring utility cheaper than the cost of local generation, there was an obvious incentive to exchange.

Trade quickly zeroed in on day-ahead negotiations, referred to as the "pre-scheduled" market. This was a natural evolution from the scheduling of resources and loads.

Dispatchers within a utility's operating area are responsible for matching resources to loads. In a conventional utility the dispatcher directly controls the ramping up and down of generating units. The dispatcher does not, of course, have control over output and load from a neighboring utility with which there is an interconnection. In order to ensure stability between the two systems, the net delivery or receipt of power expected to flow between the two operating areas must be scheduled in advance and the change in expected flows updated as circumstances change. This fundamental concept was extended to embrace the concept of a market. As noted, IPPs were under direct contract to utilities that controlled their output. Sometimes, however, they had spare capacity that could be offered as spot power. This incremental power would have to be scheduled. Further, if the seller intended to wheel the power across a utility to a buyer in another operating area, the seller would have to negotiate transmission fees. Thus scheduling and the market became inexorably intertwined.

The WSPP divided daily power deliveries into two segments – on-peak and off-peak. The on-peak period was sixteen hours from 0600 to 2200; the off-peak period was the remaining eight hours. The tight power pools of the East divided deliveries much more precisely, with hourly increments of cost (Blackmon 1985 pp. 61-65). This simplification was made easier in the west by the preponderance of hydroelectric power, which reduced the cost of following load. In retrospect, this simplification identified a crucial problem in the operating methods and pricing of tight power pools. Procedures were extremely complex, with hourly prices and hundreds of pricing nodes, which meant that the transaction cost of trading in such a system would be high, reducing liquidity and making it easier to exercise market power. In contrast, bilateral trading (when unconstrained by regulatory oversight) has sought to simplify product definitions and trading arenas in order to ensure that contracts may be quickly bought and sold.

Trading in the WSPP was by and large competitive. The region was highly diversified and no single utility dominated generation. The largest share of generation (29%) was owned and controlled by the federal government through its agencies, primarily the Bonneville Power Administration (BPA) (Lehr and Van Vactor 1997 p. 246). The next largest generator was Southern California Edison (SCE) with 12% of supplies, but its fundamental interests were more as a buyer than as a seller. Most utilities in the region were vertically integrated. However, the extreme variability in hydroelectric generation created an actively traded market.

One of the most important contributions of the WSPP was a standardized contract for purchase and sale that substantially reduced transaction costs. The WSPP contract standardized all of the formalities: legal obligations, circumstances of delivery, resolution of disputes, etc. Once this contract was accepted, traders could make deals in a matter of minutes with a single page confirmation by fax. In contrast, negotiating a QF or IPP sales contract took years and involved the oversight of accountants, lawyers, and administrators.

Despite the substantial efficiencies of WSPP trade, there were many obstacles. Initially, only scheduling utilities could trade. QFs, most of which were in California, were locked into long-term contracts and dispatch was entirely at the control of the utility controlling the service territory. No pricing information was publicly available until 1994. Marketers were effectively barred from participating until 1993. Long before the WSPP could mature into a modern commodity market other nations embarked on electricity industry restructuring, which had very different institutional arrangements and procedures. Particularly, the development of power markets in the U.K. and Norway had an enormous impact on the pace of deregulation in the U.S. and the form it would take.

6.5 Privatization in the U.K.

While the U.S. was struggling with primitive power trading arrangements in the western market and its tight power pools, the Thatcher government made plans to turn the industry on its head. Since the Conservative Party victory in 1979, the Prime Minister and her cabinet had sought to privatize many of the nationalized industries in the U.K. In 1984 they completed the privatization of British Telephone (BT), but in so doing they chose a traditional utility model. BT was configured as a nation-wide monopoly with an exclusive franchise and appropriate regulatory oversight, in part because that structure maximized the value of the government asset. The privatization of BT was followed by British Gas in 1986; again, floated as a regulated monopoly. There were multiple goals in privatization, but the two most important were to garner funds for the Treasury and to improve customer service and efficiency in the various industries. The flotation of BT raised £3.9 billion and BG raised £5.6 billion (Thomas 1996 p. 44). In the case of the electricity industry, the government's objectives were somewhat inconsistent and not easily fulfilled. John Cheshire (1996 p. 37) summarized the situation as follows:

> The nationalization of the industry following the Second World War was widely seen as essential given the need to restore and expand electricity supply capacity as a precondition for economic recovery. Whilst successive efforts were made to reform the structure of the nationalized industry and to specify more fully its economic and financial objectives, for nearly forty years there was little debate between the major political parties about the basic form of ownership. On this there was broad consensus until successive Conservative Governments launched their privatization programme in the late 1980s. Even then, the privatization of the electricity industry was seen as likely to prove perhaps the most politically contentious and technically challenging of them all.

The technical challenge began with the most basic of problems. Joseph Chamberlain observed in 1881 (Surrey 1996 p. xv) that gas, water, and electric services: "involve interference with the streets, and with the rights and privileges of individuals. They cannot, therefore, be thrown open to free competition, but must be committed, under stringent conditions and regulations, to the fewest hands." Cheshire was right about the political challenges too. The U.K.'s electricity industry was not a nationalized monolith, which might have made restructuring easy. It consisted of three physically distinct grids: The England and Wales national grid managed by the Central Electricity Generating Board (CEGB); Scotland, divided into the North of Scotland Hydro-Electric Board (NSHEB) and the South of Scotland Electricity Board (SSEB), with the two regions coordinated as a single grid; and Northern Ireland Electricity, with a small interconnection to Ireland.

The CEGB was responsible for managing the grid and generation facilities in England and Wales, while electricity distribution was managed by 12 regional boards. In 1987 the

U.K. electricity supply system was dependent on three primary energy sources: nuclear generators with 6,519 MW of capacity, fossil fuel generators with 50,263 MW of capacity, and hydro generators with 4,085 MW of capacity (Cheshire 1996 p. 16). The bulk of the fossil fuel generation capacity relied on domestic coal, which had made the system particularly vulnerable to the coal miners' strike of 1985.[33] A principal motivation of the Thatcher Government was to decentralize the power supply system in order to avoid dependence on the miners' union.

Publicly, the Prime Minister was committed to two key goals: the introduction of competition (rather than the BT model of a regulated monopoly) and the preservation of nuclear power as an active supply source of power. Moreover, the U.K. had a centuries-old commitment to the coal industry. These competing objectives would prove inconsistent time and again as the U.K.'s industry leaders and civil servants struggled to design an industry structure that would mollify all of its participants and meet the government's goals.

It was unquestioned that the wires – the distribution and transmission system – were a natural monopoly, but U.K.'s analysts saw an opportunity to create a competitive supply by decentralizing the ownership of generation facilities. Unlike those in the U.S., British designers were not constrained to protect the private property rights of vertically integrated utilities. The England and Wales grid and generation resources were in the hands of a single public organization, the CEGB. If, however, ownership of the grid were to be separated from generation, it could create serious managerial and coordination problems.

Evolving energy markets in the U.S. offered two distinct models; the western electrical and gas market system of bilateral trading or the tight power pools of the eastern seaboard. In the western U.S., grid management was conducted by integrated utilities in distinct operating areas. The same utilities traded power, so the roles of grid manager and trader were inseparable. In theory, the U.K.'s regional electric boards might have acquired the generation resources in a structure similar to the U.S. western power market. However, the generators that served England and Wales were designed to meet national demand; they and the grid could not be split up efficiently in a manner that would serve the needs of twelve separate integrated utilities. Alex Henney (1994 p. 55) noted that CEGB argued strongly against breaking up the integrated structure in any form, pointing out that it provided a smoothly operating merit order and central planning for investment expansions. The government was, however, committed to a competitive model of some kind, and under strong pressure to decide on the new architecture and implement it before the next election. According to Henney (1994 p. 53), he and Bob Peddie were able to explain that the CEGB's concerns about fragmented supply and the loss of a merit order were misplaced. Instead, "the concepts of spot pricing and a U.S. style 'tight' power pool with centralized dispatch (such as the New York Power Pool) could be modified and combined to create a real time spot market which priced electricity at the marginal cost price." Thus, the earlier ideas of Ralph Turvey and the MIT scholars on

[33] The CEGB also had considerable oil-fired capacity, but except during the coal-miners strike of 1984-85, coal generation exceeded oil generation by about 7 to 1, (Parker 1996, p. 121).

marginal cost and the spot pricing of electricity were woven into the plan for the new power market.

In February 1988 the government released a white paper that outlined the expected structure. It segmented the industry into three key components. The twelve regional boards were to be converted to regional electric companies (RECs). Generation and the grid were separated, with the grid to be operated by the National Grid Company (NGC) as a common carrier. Existing generation was divided into two companies, National Power (including the nuclear stations) and a smaller company, PowerGen. IPPs would be allowed to enter the market as required.

Steve Thomas (1996 p. 55) saw a power pool as the centerpiece of the new industrial structure. It "was to be the main price-setting arena into which buyers and sellers would place bids. Contracts of limited term outside the Pool were anticipated for power purchasers that needed greater predictability in their costs, but pricing of these would tend to use Pool Prices as their benchmark." Conceptually, it was the inverse of existing commodity markets – inside out. Typically, a commodity is traded in bilateral physical markets – one-on-one arrangements between a buyer and seller. Futures contracts are traded on exchanges, with the exchange as the counter party. In the new U.K. system, the physical market would trade through the Pool, with the Pool as the counter party, while forward contracts would be bilateral.

The new market was designed and the design was implemented in a remarkably short period of time; active planning began following the general elections of 1987 and the market opened in April 1990. It was not, however, without flaws and incomplete components. The Pool market was planned as a price-setting double auction, with load (demand) bid in as one function and generation (supply) bid in as another. Instead, CEGB's dispatch software had to be adopted to mimic the demand schedule removing "the possibility of buyers placing bids for the price at which they were prepared to buy power, and made it a market-clearing mechanism for generators" (Thomas 1996 p. 58). Regulatory oversight was important from the beginning. The decision to create only two active generating companies limited competition in the new market and set the prelude for market reform. Henney (1994 p. 63), however, pointed out that this was a legitimate problem: "Namely the government was concerned about the managerial complexity of splitting the CEGB into many pieces in a short period of time because of the perceived difficulties of finding suitable directors and senior managers, allocating staff, and being able to get the parties to negotiate contracts with the area distribution companies and British Coal."

The focus of the U.K. market was a daily auction, which "used a version of the CEGB's cost-minimizing software to draw up an operating schedule and to calculate the System Marginal Price (SMP) for each half-hour" (Evans and Green 2003 p. 1). However, neither suppliers nor buyers wanted to base revenues and costs solely on the SMP, due to its expected volatility. Consequently, an OTC derivatives market in "contracts for differences" (CFDs) developed. According to Newbery and McDaniel (2003 p. 3), more than 90% of all electricity sold was actually priced under CFDs and not by the half-hourly system price. In addition to the SMP (plus or minus the adjustments by their CFDs), generators were paid a capacity payment for making capacity available, based on calculated loss-of-load probabilities times the associated value of lost load (Henney 1994).

Initially, there were too few active suppliers in the original design: National Power and PowerGen. IPPs were, however, allowed to enter the market and new gas-fired generation began to displace existing coal-fired generation, even though their cost was sometimes above the avoided cost of existing facilities (Evans and Green 2003 p. 1). The paucity of competitors provided ammunition for critics of the new market, who concluded that if market power was being exercised, then prices were higher than necessary. Likewise, there was an eager stable of critics who perceived that prices were too low and that the market was undercutting the coal industry, which was a political problem.

In 1998 the U.K power regulator, Offer, described how the U.K. power pool worked and analyzed its shortcomings. According to Evans and Green (2003 p. 3), many of the arguments were controversial but were ultimately accepted by the government. Consequently, a plan for new electricity trading arrangements (NETA) was drawn up and implemented. The new market opened in March 2001.

The design for NETA was radically different than the former Pool. Virtually all trading was shifted to the bilateral market, with voluntary rather than mandatory participation. The NGC operates only a balancing market, which was purposely designed to be small. Load serving entities (LSEs) are expected to own or contract for resources in the bilateral market that closely match their forecasted load. Discipline is achieved by maintaining a lower purchase price than sales price in the NGC balancing market, creating a penalty for scheduling errors and discrepancies. In other words, correcting load-forecasting errors in the NGC balancing market was intended to be more expensive than contracting for the same volume of power in the bilateral market. It was anticipated that LSEs and generators would enter into long-term purchase contracts. As the date of delivery approaches, adjustments to these contracts (due to unusual weather or forced outages) are made through day-ahead and day-of short-term sales. In addition to bilateral trading, NETA encouraged the development of privately funded power exchanges, such as the Automated Power Exchange (APX), that were expected to provide such adjustment services.

NETA allows regional distribution companies to construct and own generating facilities. In addition, before NETA there had been mergers and a realignment of assets between the generating and distribution companies. As a consequence there has been a shift back to vertical integration, with the exception of the transmission grid, which remains a regulated monopoly under the NGC. Ironically, the U.K.'s new power market structure is closer to the original WSPP design in the Western U.S. than it is to the initial structure chosen by the Thatcher government in 1989. It is especially ironic because before the NETA reform, Daniel Fessler, the President of the CPUC, visited the U.K., observed Pool bidding, and was inspired to model California's power market restructuring on a design the U.K. abandoned a few years later; elements of the U.K. design were also adopted by Australia, New Zealand, Norway and other countries.

The U.K. electricity privatization and market restructuring was revolutionary and the consequence of government-mandated market design. In contrast, market reforms for oil and North American natural gas have been largely evolutionary and, more often than not, adopted spontaneously by participants who were simply looking for lower-cost alternatives as the means to maximize profits. To put it another way, market designs thrust from the top down may be successful in meeting broad economic and political policy objectives, but they often ignore transaction costs where much of the perceived efficiency gains can be lost. As Surrey (1996 p. 10) noted: "Although competition may be effective as a means to lower costs in generation, the potentially very high transaction costs (e.g., in trading and settlement systems, new metering, and arrangements to guarantee security of supply) may outweigh efficiency gains that may be available at the smaller end of the supply market." This is an especially difficult dilemma for policy makers concerned with the electricity industry, because it has higher transaction costs than either the natural gas or oil industries, and it is easy to omit these costs from restructuring plans.

6.6 Restructuring in Norway

Eight months after the U.K. opened its new electricity market, Norway approved the Norwegian Energy Act of 1991 (Nord Pool 1997 p 1). Two years later Nord Pool ASA was established as a subsidiary of Statnett SF, the manager of the national grid. Initially Nord Pool established a market for the physical delivery of power, Elspot, but it quickly followed with a forward market, allowing blocks of power to be traded months in advance, and a futures market, Elterim. Nord Pool was expanded to include Sweden in 1996 and by 2000 Denmark and Finland had joined.

Unlike commodity market structures in the U.K. and U.S., the Scandinavian electricity futures exchange is not separated from physical trading. That is, futures, forward, and day-ahead physical trading are all part of a seamless whole. The Scandinavian system also allows bilateral trading, which may be wholly independent of Nord Pool, or it may be integrated with the exchange's scheduling and settlement activities. The ease by which the component parts of electricity trading fit together explains much of the reason for Nord Pool's success.

Another key feature of Nord Pool is its flexibility and adaptability. The trading structure has evolved (like North American oil and gas markets) over a decade, increasing in sophistication, while preserving its customers and adjusting institutions to meet their needs. A brief history of the evolution of the Pool's financial markets (Nord Pool 2003) illustrates the nature of this evolution:

- In 1994, on-peak and off-peak forward products were simplified to a single base load product.
- The same year, the weekly auction was replaced by a continuous trading system, using telephones and a "white board" to record bids.
- Physical settlement was shifted to financial settlement, using Elspot prices, for forward products.

- In late 1996, telephone bidding was replaced by an electronic trading system, developed by Sweden's OM Group.
- In 1997, Nord Pool standardized its forward products to conform to those being traded in the bilateral market.
- Options contracts were added in 1999.
- In 2000, CFDs were adopted to accommodate price differentials between Elspot and individual area prices in the four countries.

Grid management in Nord Pool, as in the U.K., is simpler than in the U.S. This is because there are fewer transmission bottlenecks and fewer entities managing the grid. Norway, itself, has two primary bidding zones, north and south. Denmark has an east and west zone. Sweden and Finland are each treated as single zones but use various techniques to offset congestion.[34]

Power prices in Nord Pool are constructed around the day-ahead hourly prices set in Elspot. Trading takes place in a double auction in which participants submit demand and supply schedules, relating prices and quantities to be delivered in each separately identified region, for each hour. Submissions take place up to the time designated for the auction. Demand and supply schedules are aggregated across the whole Pool and this determines the hourly system price. As long as all the schedules are feasible - that is, there are no transmission bottlenecks - then the hourly system price is the same across the entire grid.

In the event there is congestion and aggregated schedules are not feasible, the grid operator uses the individual demand and supply bids to identify which generators to back down in the surplus regions and which ones to increase output where there is a deficit. These relative adjustments calculate individual outputs that make the best possible use of the transmission grid and determine relative prices. Price differentials are the effective transmission tariff to transmit power from a region of surplus to a region in deficit.

As noted, the distinctive feature of Nord Pool is the smooth integration of physical and financial trading, all handled by the same institution. In addition to managing the Elspot day-ahead market for physical delivery, the exchange offers standardized futures and forward derivative products.[35] All derivative products are settled for cash, the value determined by the difference between day-ahead hourly Elspot prices and the contract price. The basic unit of a derivative contract is one week; longer-term contracts are organized as four-week units and the weekly block breaks into seven days, each day consistent with the Elspot market. All of Nord Pool's derivative products are designed to match product types traded in the OTC market. In addition, the Pool handles

[34] The methodology for eliminating congestion in each zone varies. In Norway, Statnett SF uses a forecast of load and generation to identify likely bottlenecks. It then uses the Elspot bids to adapt generation to the constraints, establishing multiple zones and prices. Svenska Kraftnat, the grid operator in Sweden, uses the regulation (balancing) market to create counter flows, so the entire Swedish grid can be treated as a single zone.

[35] Forward products are similar to futures products, except that they are not marked to market and, therefore, Nord Pool does not require adjustment in margin payments as the market prices of products vary.

scheduling and settlements for many of the OTC trades. All of the above features stress the importance of product compatibility, ensuring the highest possible liquidity for trading.

The Elspot market is a once-a-day auction. Originally Nord Pool designed futures and forward trading along similar lines. "In 1994 the financial market's weekly auction trading system was replaced by a continuous trading system, with bids written on a white board on the trading floor. All bidding, price quotation and execution of trades were carried out by phone between market participants and the trading floor at Nord Pool" (Nord Pool 2003). According to Per Horth, the CEO of Nord Pool in the 1990s, the reason for the shift to continuous bidding was an illiquid auction. Auctions enhance liquidity in the sense that they draw all market participants together at a specified time and place. On the other hand, the auction structure – which sets a single price, often restricts reselling, and may have an unpredictable outcome – does not encourage market makers, which reduces the market's liquidity. Historically, market makers have been essential for the smooth operation of commodity and financial markets. (See Chapters 2 and 7 for explanation and elaboration on these terms.)

In the early years, Nord Pool's futures and forward markets were little more than formalized OTC markets, but since they were conducted in the exchange they had greater transparency. The reason for the informality was because Nord Pool actively sought to mimic the type of products that were being exchanged in bilateral trading and because it takes time for more formal mechanisms to mature. In 1996, Nord Pool introduced PowerCLICK, an electronic system of continuous trading. Bids may now be placed directly to the exchange by telephone or through PowerCLICK.

Elspot and forward trading prices throughout the various zones of Nord Pool are reconciled by CFDs. For example, a power producer in Sweden that wanted to lock in forward contract prices with a "perfect hedge" would have to go through the following steps: first would be the sale of power into Nord Pool or the OTC forward market. Such contracts are written as CFDs, settled against the Elspot hourly price. (If the generator's contract price is higher than the Elspot price, the generator will receive the Elspot price plus a differential paid by the counter party or Nord Pool. If the contract price is lower, then the generator would receive the lower price and the differential would go to the buyer offsetting the higher Elspot price). Second, the Swedish power producer would have to enter into an additional CFD, which offsets the differences between Elspot and Swedish zonal prices.

Although the perfect hedge described above sounds complicated, it is not too different from the arrangements a gas producer in North America would have to cobble together to ensure a fixed forward price. For example, an Alberta gas producer could sell gas forward on the NYMEX futures market for delivery at Henry Hub, Louisiana. Then the producer could enter into a swap arrangement, which would lock in the differential between Henry Hub and Alberta prices. Such instruments are available in the OTC or at the Alberta NGX gas exchange. In the process described above, the first transaction removes price risk from the basic commodity (stabilizes the marker price) and the second removes basis risk. The basis risk market for Alberta gas has developed because

Canadian gas prices often move differently from Henry Hub prices and because the Canadian market is large.

By all accounts, Nord Pool is a successful commodity market for electricity. One important reason why it has succeeded when other power markets have failed is because Norway has a large hydroelectric supply, allowing water inventories to cheaply substitute for power storage (hydroelectric supplies may also have been one of the reasons for the earlier success of the WSPP). In addition, Nord Pool's products (the contractual form in which the commodity is bought and sold) were designed to integrate smoothly with one another and to subdivide as the date of generation approaches. Unlike commodity markets in the U.S. and U.K, physical and financial trading are integrated into the same institution, which helps ensure compatibility and streamlines procedures. Further, the transmission grid in Nord Pool is relatively unconstrained and an effort is made to integrate the market as efficiently as possible when congestion does occur. Nord Pool, like the oil market and the North American gas market, has been allowed to evolve step by step over a decade. It did not start with a grand inflexible design and the market has been adjusted as necessary to accommodate traders, producers and distribution utilities. Finally, unlike California, Nord Pool has not been confronted with a massive supply shock, which has made it acceptable to both the industry and power consumers.

6.7 The Special Problems of Power Markets

Both the U.K. and Norwegian power markets succeeded in part because their transmission systems were relatively unconstrained and generation capacity was more than adequate. When power is in surplus, the market structure can be kept straightforward and simple, but as the industry approaches either generation or transmission capacity limits, special complexities emerge that question the efficiency of traditional forms of bilateral trading that dominate other commodity markets.

When generation capacity is constrained, it often requires very high prices to clear the market. As explained in Chapter 2, the demand for energy products, particularly electricity, is extremely price-inelastic in the short run. This feature allows generators who believe that they are the final margin of supply to exercise market power. Although the impact of such behavior is easily exaggerated, generators have occasionally acted in this manner. Importantly, the opportunities to exercise market power are most likely to arise during periods of system emergency. Prior to restructuring, a guiding principal of all regulated utilities was to ensure system reliability. In a market setting, however, a system that is perfectly reliable, i.e. always has spare capacity, is overbuilt. If prices alone are intended to bring forth adequate investment, the system must occasionally reach its capacity limits. Prices must then rise periodically to levels adequate to ration scarcity and recover capital costs. The goal that regulators have set for themselves – to discern the difference between the price spikes necessary to encourage efficient investment and those that simply transfer wealth – is an extremely difficult objective.

When the transmission system is congested, "loop flow" causes a complex interaction between all paths in a transmission grid. That is, transmitting an excessive volume of electricity in one segment of the grid can cause congestion elsewhere. This externality is distinct from the problem of economies of scale and has no parallel in oil and gas pipeline networks. This can make the iterative process of a bilateral market inefficient when generation and transmission are privately owned.[36] The externality can only be resolved by a central pool mechanism that internalizes transmission externalities. This led to a ferocious debate in the U.S. on the extent to which the grid operator should be a passive manager of schedules or an active participant in the market. Clearly the grid manager must manage the system in emergencies, during unexpected forced outages or demand surges. The role of the system operator in determining pre-scheduled prices in the presence of congestion was, however, extensively debated during planning of the restructured California market. SCE was the principal advocate of an active role for the system operator, while Enron and NYMEX spearheaded the opposition. Unfortunately, the entire debate was obfuscated by a profusion of special interests, and despite the brainpower devoted to it, an obscure compromise resulted. (California decided to mandate the use of an auction-based power exchange, but separate it from the grid operator).

The problem of pricing in a congested grid is straightforward once the nature of electricity transmission is understood. That is, congestion is not a simple bottleneck between the source of supply and the point of consumption, as in a gas or oil pipeline. Since electricity flows on the path of least resistance, congestion costs impact every point of generation and load if the network is interconnected. Integrated utilities solve the congestion problem through the specialized knowledge of their dispatchers. Power pools, which meld the assets of various utilities, resolve the problem through the "economic dispatch" of generating units by the pool operator based on merit order. The order in which various generators were called took into account their impact on congestion throughout the grid. Naturally, the engineers who had successfully managed power grids for decades were deeply suspicious of any system in which no one was in charge. Further, the U.S. grid, unlike ones in the U.K. or Scandinavia, was never designed as an integrated system or even as a series of self-contained regional grids. Rather, they reflected the operating areas of integrated utilities with varying degrees of interchange capacity. In the U.S. this severely compounded the problem. Paul Joskow (1997 p. 122) summarized the issue at the time: "The key technical challenge is to expand decentralized competition in the supply of generation services in a way that preserves the operating and investment efficiencies that are associated with vertical and horizontal integration, while mitigating the significant costs that the institution of regulated monopoly has created."

When an electricity grid is congested, it is partitioned into a series of sub-markets. If each of these zones is treated separately, there may be too few buyers and sellers to

---

[36] Bilateral trading, as the name implies, is conducted by one-on-one negotiations between a buyer and seller. If the purchase involves the shipment from one region to another, transportation costs are obviously a key component of the transaction. For most commodities, such costs can be estimated in advance with reasonable certainty. In bilateral trading the traders have knowledge of the bids and offers of other parties, but do not know, or need to know, details of other transactions, i.e., how much they are producing or consuming, or when and where. Optimizing a power system, however, requires the coordination of generation and transmission to a level of detail not necessary for other markets, because transmission costs depend on the decisions of all other actors in the grid. In effect, the market needs to know the solution before bargaining can begin.

establish effective competition. This will allow a generator to take advantage of its location to extract monopoly rents. By underbidding capacity, the generator can exercise market power and set inefficient prices, exacerbating what might already be a strained market.

In theory, the difficulties of congestion management can be reconciled by the simultaneous auction of energy and transmission capacity. This is achieved in Nord Pool by the submission of demand and supply schedules tied to specific points of generation and load. Transmission is not auctioned separately; rather, the demand and supply bids, which relate various combinations of price and volume for each participant to their point of connection, are used by the system operator to balance the grid in the most economic way. If there is surplus power in one zone and a deficit in another, relative prices are adjusted in each zone until local demand and supply functions balance. The process of increasing and decreasing zonal prices creates a price differential across zones. This is the effective transmission fee to move power from one zone to another. However, in grids with hundreds of nodes in each zone impacted by congestion, such a system requires complex management tools. In these pools, such as PJM in the U.S., "locational marginal prices" (LMPs) are calculated for individual nodes within the grid using linear programming models.

According to Steven Stoft (2002 p. 398): "Depending on the algorithm, central computation usually computes prices and optimal dispatch simultaneously. There is no choice as to the definition of the least cost dispatch (except for rare ambiguities). It is best to think of the central computation as first optimizing the dispatch and then finding the prices." The computation of optimal dispatch depends, of course, on the quality of input data. Tight power pools traditionally based such computations on the heat rate of the various generating units combined with fuel costs. In a bidding system, such as Nord Pool, supply schedules would be inclusive of both generating and transaction costs. More importantly, unless the system is closed (no bilateral trading is allowed and the auction is compulsory) competitive bids will be based on alternative costs, which are the prices for which the power could be sold elsewhere. These bids may be unrelated to heat rates or the marginal cost of producing power from specific units.

The main problem of power market pricing is that its complexity leads to high transaction costs, whichever system is adopted. Optimizing grid usage in real time often requires hundreds of different price signals, calculated every ten minutes. If the prices are determined by decentralized bidding, it requires active participation by suppliers and consumers twenty-four hours per day. Imagine, every power consumer hovering by their electricity meter watching real-time prices in order to decide when to do the wash. Obviously, the transaction costs of operating such a market are prohibitive; corners have to be cut. The question is how and where. How perfect does the system need to be? In particular, if the grid management and real-time pricing become so complex that they inhibit the development of forward markets, the cost may greatly exceed the benefits. Schweppe and his collaborators believed that sufficient diversity of generators and a number of large industrial and commercial customers along with market makers would be adequate to establish an efficient spot market. For these entities transaction costs would be low enough to trade wholesale electricity.

The tension between efficient pricing and a successful market is captured in the following comment from Nord Pool (2003):

> The ideal tariff system for trade conducted on OTC and bilateral markets, or via a power-exchange operated spot market, should be characterized by the following principles:
>
> - Market participants should know the transmission costs at their location (grid connection points) without having to enter into negotiations with grid owners or system operators.
> - Transmission costs to a market participant should not depend on the location of a trade counterpart.

Nord Pool's principals for a successful market are often at variance with the principals of efficient grid management, unless there is substantial surplus capacity in transmission. In formal terms, successful forward and futures markets require predictable basis risk; otherwise they are a series of unconnected illiquid markets. Such an organizational structure runs the risk of inhibiting market growth if there is frequent congestion. Moreover, constrained forward trading enhances the risk that market power will be exercised by generators during periods of capacity shortages.

6.8 California's Frankenstein

In 1995 California's electricity rates for its three largest private utilities were the highest among western states and among the highest in the nation. The reasons were complex. Pacific Gas and Electric (PG&E), SCE, and San Diego Gas & Electric (SDG&E) had all invested in nuclear power during the period when its costs were rising out of control. They also had ambitious plans for new projects, but the CPUC and CEC blocked this development path. Instead, the State's regulators chose to stimulate the development of private generators under PURPA, the federal law that encouraged the development of IPPs. Since the offers were based on each utility's avoided costs, prices were high – exceptionally high, in many cases over $100 per MWh.

California faced a very different restructuring problem from that of the U.K. The three big utilities served eighty percent of the state's electricity load. The utilities, however, were vertically integrated, controlling distribution and transmission grids in their service territories. Northern California was under the operational control of PG&E and Southern California was split between SCE and SDG&E. Aside from the privately owned utilities, there were a number of publicly owned utilities in the middle of these service districts, including two large ones, the LADWP and SMUD. Thus, restructuring involved three difficult preliminary steps: 1) find a means to divest adequate generation from the utilities in order to create a meaningful market; 2) create an independently operated transmission grid, and 3) somehow knit together the piecemeal operating areas of the three utilities into a statewide grid. Resolving these issues required the active participation of the utilities, their supporters, and their critics.

The central problem for the California utilities was their "stranded costs." At the onset of restructuring, average generating costs were about $70 per MWh, roughly twice the price of power traded in the wholesale market. The utilities' costs included high-cost nuclear power, QF contracts at a range of prices, low-cost coal, and a number of infrequently used gas-fired generators. Under the regulatory compact, these costs were passed through to consumers, along with the cost of transmission, distribution, marketing, and allowed profits. The total to residential customers was over $100 per MWh. Industrial and large commercial customer rates were somewhat lower, reflecting higher volume deliveries and discounts for interruptible service. Within a given service area, consumers had no choice but to buy from the designated utility. If, however, choice were to be allowed, consumers would flock to lower-cost suppliers. Given high fixed cost, the stampede by their customers to alternative suppliers would exacerbate the difficulties and provoke a "death spiral" for the incumbent utilities. Thus, it was crucial for PG&E, SCE, and SDG&E that a mechanism be included which allowed them to pay off the sunk costs of their investments.

In 1994 the CPUC released a blueprint for reform in order to stir up debate: *Order Instituting Rulemaking on the Commission's Proposed Policies Governing Restructuring California's Electric Service Industry and Reforming Regulation,* otherwise known as the "Blue Book." James Bushnell and Carl Blumstein (1994 p. 3) provided a contemporary critique of the initial plan that captured the schizophrenic nature of California's policy objectives:

> The Blue Book embraces two competing strategies for the future structure of the California electric power industry. One is a bilateral retail wheeling structure, and the other is a U.K.-style transmission pool. The Blue Book describes a framework for implementing the former, but articulates a long-term vision more aligned with the latter. Under a retail wheeling structure, utilities would continue to be vertically integrated but would have to compete for the demand of their native load customers. Utilities would still provide transmission network services and be obligated to carry the transmission load of competitors at regulated prices. In the United Kingdom (U.K.), the creation of the pool was accompanied by the complete separation of ownership of generation, transmission and distribution. The transmission owners play the role of "market makers."

Bushnell and Blumstein (1994 p. 3) went on to say: "If the vertically integrated utilities remain largely intact in a bilateral retail wheeling structure, the coordination abilities would enhance reliability and reduce transaction costs. However, the utilities would also have a correspondingly large capacity for the exercise of market power. They also noted (1994 p. 6) that: "Any serious consideration of a U.K.-style pool in California will also have to account for the fact that California utilities are only a subgroup of the relevant western power market, which spans multiple states and three countries."

Indeed, while California was debating the options for restructuring, the existing wholesale power market itself was rapidly changing. The leading energy futures exchange, NYMEX, had its eye on a promising new market – electricity. The market was not only huge (on the order of $100 billion per year) but prices were extremely volatile. It seemed to be an ideal target for expansion. As noted in Chapter 3, however, a successful futures contract depends on a well-developed physical market for the commodity. In 1994 North American electricity trading was in its infancy. The best-developed market was in the West, but trading and supporting institutions fell far short of a mature commodity market. If NYMEX's managers were to successfully launch an electricity futures contract, they would have to prod the industry into accelerated change. At the least, three vital components were necessary: public information on bilateral prices, the development of hubs and market centers for scheduling and settlements, and the identification of a widely-traded standardized contract that in scale and timing was compatible with the futures trading paradigm.

As noted, part of the reason for Nord Pool's success lies with the manner in which physical and financial contracts are integrated into a seamless cascade of macro to microscopic deliveries. This is an institutional structure that is very difficult to achieve in the U.S. due its market development histories. In the nineteenth and early twentieth century there were a variety of commodity exchanges for physical markets.[37] As the cost of communication, enforcement, etc. declined, however, they were no longer cost effective and were replaced by bilateral trading. By 1990 the only remaining commodity exchanges in the U.S. were devoted exclusively to futures trading, where standardized contracts and high volumes could justify the large fixed cost. For energy markets, the clear distinction in the U.S. between physicals and financials allowed separate regulation with separate regulatory bodies; FERC covered interstate gas pipelines, electricity transmission, and related pricing, while the Commodity Futures Trading Commission (CFTC) covered the integrity of brokers, exchange settlements, margin rules, compliance, etc. NYMEX and other futures exchanges have elaborate rules and procedures for closing out financials for physicals, but their trading is focused exclusively on standardized financial contracts.

In 1994 NYMEX's officials knew that the electricity market was immature and was unlikely to support the rapid growth of an electricity futures contract. Nonetheless, they sought to develop the product early into the process of deregulation, before other competitors beat them to it. Much of the West Coast electricity industry was receptive to the idea, particularly utilities in the Pacific Northwest that faced significant risk arising from the variability in hydroelectric generation. The California utilities were split: PG&E supported the concept of a futures contract, but SCE was hostile. To a large extent, positions were based on notions of how the California market ought to be restructured. Those that supported NYMEX's efforts favored the bilateral retail wheeling structure described by Bushnell and Blumstein. Those that sought a more structured approach, a power pool along the lines of the U.K. model, viewed the NYMEX initiative as counter productive. Without resolution, the western power market developed along two tracks, evolving into two diametrically opposed models. It was, however, several years before problems associated with the incongruities of the two systems would be fully understood. In the meantime, California proceeded with its plans to remake the industry.

The final decision on California's new electricity market rested with the Governor and the state legislature. In 1996, Governor Wilson signed Assembly Bill (AB) 1890, making

[37] The last formal exchange for physical commodities in the U.S. was the Portland Grain Exchange. The West produces "white" wheat, rather than "red" wheat, which is the standard for the Midwest, thus the Chicago commodity exchanges were not suitable for the western product. Moreover, the Columbia River and its tributaries provided an inexpensive link between inland granaries and the Pacific Rim market. In short, distinctive location and quality differences necessitated separate pricing for Western wheat. The exchange closed in the 1980s and was replaced by telephonic bilateral trading, with centralized accounting and settlements.

California the first state to radically change its electricity industry. The final structure was a political compromise. The main concern of the utilities was met; they would (in theory) receive a full and complete payment for their stranded costs. Consumers got something too. Existing debt was to be restructured with lower interest and an extended payment period. As a consequence, retail rates could be cut by 10% and frozen during the transition period (four years or less in which stranded costs were expected to be recovered).[38] In addition, the utilities would be required to divest a good share of their generating assets in order to ensure a competitive market.

The debate over the technicalities of the structure also ended in compromise. Instead of a "Poolco," advocated by SCE, or bilateral trading advocated by PG&E, Enron, and other potential entrants, the state decided to split grid management and the market into two organizations: an independent system operator (ISO) and a power exchange (PX).[39] For the transition period, the utilities would be required to bid all of their resources into the PX and purchase their entire load through the same organization. PX prices would set payments to suppliers and the difference between them and frozen retail rates would determine the pace of stranded cost recovery. Independent suppliers who wished to enter this portion of the California market would have to bid their generation into the PX. Crucially, however, the utilities' mandatory participation in the PX was to last only four years (or less). After that time, the market would be allowed to set its own institutional arrangements. Thus, if bilateral trading proved to be more cost-effective than an exchange (that is, had lower transaction costs), the PX would close and the final market structure would not be too different than other states in the western market, except that the utilities would no longer be vertically integrated.

It is impossible to say which way California's transitory market structure might have evolved, because two and one-half years into its transition, California experienced an extraordinary energy crisis. Much has been written about the causes of the crisis, which may be summarized in three categories: 1) the crisis was caused by a "perfect storm" of adverse market fundamentals: unexpected demand growth, a catastrophic drop in hydroelectric output due to the second worst drought in history, and natural gas supply problems due to reserve depletion and pipeline bottlenecks; 2) the crisis was caused, or seriously exaggerated, by market manipulation and the exercise of market power, and 3) the crisis was caused by poor market design and regulatory mistakes. Obviously all three had some impact, and the debate centers around which one of these factors was the principal cause. Irrespective of the cause, the experiment in restructuring is blamed for rate increases, which were implemented a year after the crisis began. As a consequence, California has retrenched, and many of its politicians now seem determined to return the

---

[38] The transition period was planned for four years. If, after this period, stranded costs were not fully recovered, the utilities would absorb the remaining cost. On the other hand, if stranded costs were recovered earlier, the transition period would end before its four-year schedule. When the crisis began in the summer of 2000, SDG&E had fully recovered its stranded costs, allowing retail rates to fluctuate with the market. This caused a political furor, paving the way for more serious events as the crisis unfolded.

[39] In the U.S. the ISO concept has generally been replaced by Regional Transmission Organizations (RTOs.) RTOs are less complex, less costly, and less integrated than ISOs, but the function is largely the same. In Europe, grid managers are frequently referred to as Transmission System Operators (TSOs).

state to cost-based regulation. The state, however, remains constrained by a high-cost environment and little is likely to change.

The California experiment has revealed a substantial amount about transaction costs and many of the flaws can be traced to the fact that these costs were largely ignored when the market was designed. High transaction costs arose from the California experiment for at least four reasons. First, the dramatic break in market procedures thrust substantial administrative costs on new and old organizations: the total estimated cost to set up the new system was about $1 billion. Second, the substantive changes required all participants to learn wholly new procedures. In many cases the procedures were so complex that they were improperly performed, causing significant inefficiencies, particularly during periods of market stress. Third, the complexities of grid management combined with the separation of the power exchange from the system operator opened up unusual opportunities for gaming the system, adding compliance and oversight costs. Fourth, the transmission infrastructure was not designed to accommodate a statewide market. Rather than upgrade the infrastructure, market designers chose to obscure the inefficiencies with special market rules and de facto subsidies and side payments, which in turn led to further opportunities to game.[40]

6.9 Vertical Integration in Power Markets

The spectacle of a number of vertically integrated utilities competing in wholesale power trade seems like a massive contradiction in terms. And yet this was the predominant organization of the modestly successful WSPP, is the dominant form in Germany, and represents the new structure of the England and Wales power market. In the case of the WSPP and other parts of North America, markets were opened in an established industrial structure dominated by vertically integrated companies. In contrast, the development of a vertically integrated structure in the U.K. has been a gradual development provoked by the market structure adopted in 1990 and revised a decade later. There is a neat circular logic. In the U.S. vertical integration has inhibited market development, while in the U.K. the market structure seems to have created the need for vertical integration.

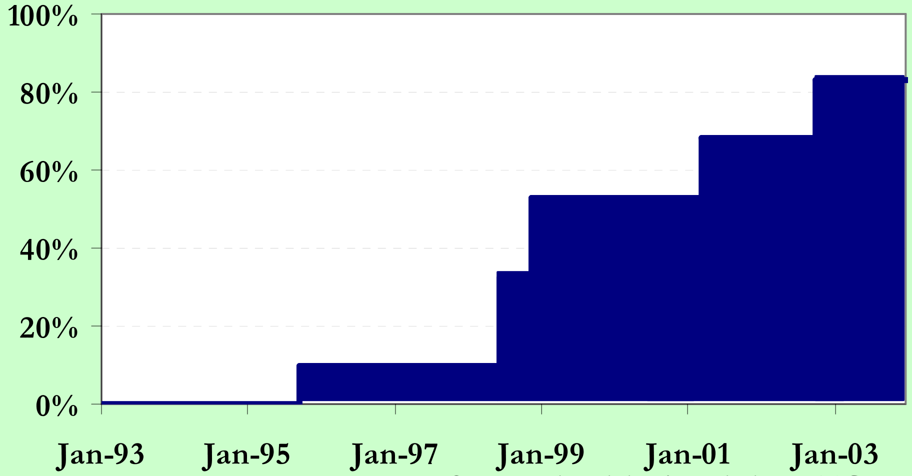
Figure 6.1 illustrates the trend in the U.K. Following privatization, the England and Wales pool had twelve distribution utilities and effectively two generating companies. Over time the two large suppliers, National Power and PowerGen, divested facilities, and new generators entered the market. The consequence was vertical integration (missing, of course, the transmission system that interconnects generators and load).

The motive for vertical integration of the power industry in the U.K. has little to do with the problem of specific assets for the obvious reason that the grid is independently

[40] The most obvious of example of the ad-hoc market design was the "dead zone" for San Francisco. The city had inadequate transmission capacity into the peninsula and high-cost generators inside. To keep San Francisco in the political compromise that restructured the industry, pricing in the city zone was excluded from the market-based paradigm and essentially subsidized. The Bay Area also turned out to be the region of highest demand growth for electricity. It should not have been a surprise that it was the region that suffered the most severe shortages during the crisis.

owned and operated with regulated access – a fundamental structure that no one expects to change. Thus opportunistic behavior can be ruled out. Instead, there are two primary motivating factors, price volatility and illiquid markets.

Figure 6.1 Proportion of Distribution Utiltities Owned by U.K. Generators by Number of Customers

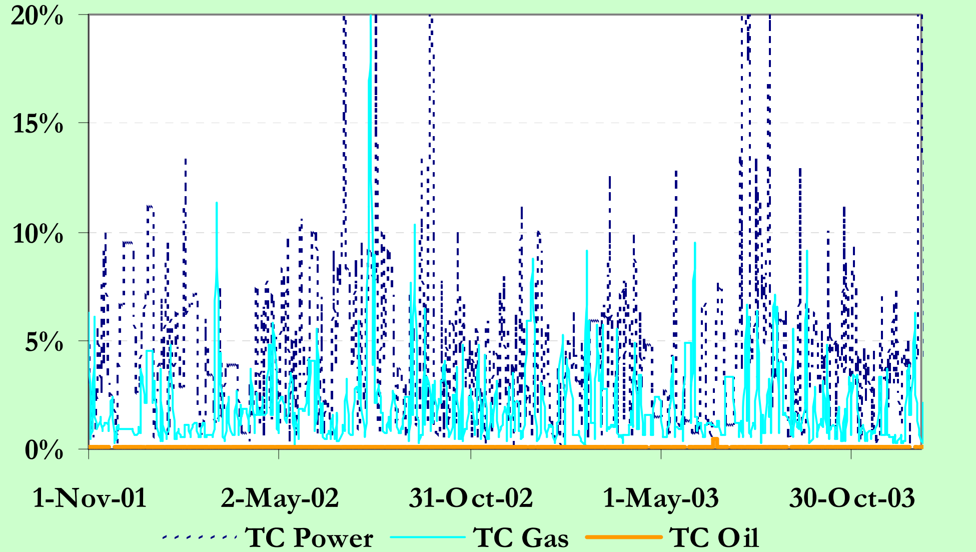


Source: Electricity Association, UK Companies

In Chapter 3 the characteristics of component products were explored, noting that large capital investments made by both buyers and sellers create highly inelastic demand and supply functions in the short term. Of all energy commodities, electricity is the most extreme in this respect. In the very short term, electricity consumers will pay extraordinarily high prices to avoid power curtailments. Likewise, when a generating system reaches capacity, it typically reaches an absolute limit. During peak emergencies electricity prices have soared to levels one hundred times greater than normal. Except for the California crisis, high prices have not persisted, since they usually reflect short-term forced outages or unusual weather. Nonetheless, such extreme prices in a single day can undo a year's worth of profits or push companies into bankruptcy.

Unlike well-developed oil and gas commodity markets, a transparent, liquid forward market for electricity in the U.K. failed to develop. According to John Bower (2002 p. 10): "Despite the volume of electricity that was covered by contracts, forward markets were highly illiquid and largely limited to semi-annual contracting rounds, held in April and October." Bower (2002 p. 21) further commented that liquidity did improve as ownership of generation fragmented and as direct access broadened. Nonetheless, for this and other reasons, as Bower (2002 p. 21) explained, the U.K. adopted NETA with the aim of "adopting trading arrangements mimicking those in traditional commodity markets." So far, however, NETA has failed to foster such a market structure. An electricity futures market was opened on the IPE but failed, in part because it was ill timed. Day-ahead trading is limited, and as Figure 6.2 demonstrates, the cost of trading electricity was two and one half times higher than natural gas and sixty times higher than Brent crude oil during the period between November 2001 and December 2004. The cost of trading gas in the U.K. has averaged just under 2% of its total price, slightly greater than that in North America (see Chapter 7). The cost of trading electricity, by far the highest of any of the energy markets, has declined from 4.9% in 2003 to 4.4%, but remains volatile.

Figure 6.2 Comparison of the Difference of Bid and Ask Prices in the U.K.



As the U.K. market evolved, generators were faced with three threats – declining wholesale prices, highly volatile prices, and an illiquid forward market. In such an environment it is not surprising that generators would seek to acquire distribution utilities with their built-in markets. So far these acquisitions have proven beneficial for the generators. According to David Newbery (2003), retail prices have not declined as rapidly as wholesale prices, and prices charged by existing RECs have remained above those offered by new entrants. As the original duopoly in the U.K. power market broke down, competition lowered prices, but the generators traded "horizontal for vertical integration" (Newbery 2003).

The successful oil and gas commodity markets described in Chapters 4 and 5 have an important common feature that is frequently overlooked when policy makers seek to liberalize energy markets. Both oil and gas markets began with independent groups of producers and consumers, i.e., vertical integration did not dominate the market structure. In the case of the oil market, OPEC had broken up the traditional structure. Regulation of the North American gas industry had blocked vertical integration and, instead, encouraged inflexible long-term contracts. Deregulation forced the breakup of these contracts, opened third party access to the pipeline network, and encouraged the creation of a flexible spot market. In both cases, once historic barriers were removed there were large numbers of willing buyers and sellers, the prerequisite for competition and commodity markets.

Australia adopted an electricity market design not too different from the original one in the U.K., i.e., a day-ahead auction for physicals with the OTC offering contracts for differences. In Australia, however, there is a futures market. According to press reports by the Sydney Futures Exchange volume is increasing and the contract is expected to be a success. Recent volumes have not lived up the hype and other observers dispute this claim.[41] If there is a difference between market developments in the U.K. and Australia, it may be explained by the degree of vertical integration and the rules that govern trading activity. According to Hugh Outhred (2000 p. 3) the vertical and horizontal integration of the former state-owned monopolies was purposely broken up in order to introduce competition. Australia has also had considerable price volatility and in sharp contrast to California, distribution utilities are required to hedge their spot market purchases. These features created a demand for forward trading and if the market is not too fragmented a robust futures market may develop.

The WSPP demonstrated that some degree of vertical integration is not inconsistent with a workable day-ahead spot market. German electricity distribution companies are vertically integrated and own generation units. So far, this has not inhibited the development of the spot market, mainly because there is substantial surplus capacity and little or no transmission constraints in the German grid. The structure may, however, be constraining the development of forward and futures trading.

Germany allows both bilateral and exchange trading. Germany's two electricity exchanges merged in 2001 to form the European Energy Exchange (EEX). In general, the EEX's product design and operating methods are similar to Nord Pool with two exceptions. The EEX does not manage the transmission grid or administer a real-time market for grid balancing; that is done by three German TSOs. Further, the EEX does not attempt "market splitting," the separation of the grid into zonal or nodal prices to account for transmission constraints. According to data provided by the EEX on their website, the volume of trading in the day-ahead market, for hourly and for on-peak and off-peak blocks, has steadily increased. Volume growth in forward contracts has not, however, been as substantial.

NYMEX's electricity contracts for the western power market failed; volume peaked in January 1998, three months before the restructured California market opened. By January 2000, futures trading volume had dropped to less than 10% of its peak. These results might be accounted for by bad luck and timing; California's restructuring divided the market in half and blocked three of the region's largest utilities from participating in bilateral trade or futures trading. On the other hand, NYMEX's month-long power contracts for the Midwest have also failed. The industry structure in the Midwest is nearly identical to that in the WSPP before California's restructuring, yet futures trading did not take hold there either.

Regulatory barriers and product design, rather than industry structure, may also help explain the failure of electricity contracts at NYMEX. In the U.S., Australia, and the U.K., there is a sharp division between physical and financial trading. In the case of the energy industry in the U.S., FERC regulates energy trading and the CFTC regulates futures markets. For the electricity market, this division is artificial, because any contract to deliver electricity is a forward contract – no one holds the electrons. Among other things, this means settlements or clearing in the electricity market are much more

[41] According to Professor Ronald Ripple of Macquarie University, and former research director of Australian Derivatives Exchange, the press release by the Sydney Futures Exchange is misleading, because it accounts for open interest rather than trading volume and aggregates eight separate contracts.

complex than any other commodity. It is best to think of the market as a cascade, an unwinding of contracts as the moment of delivery approaches.

The cascade goes something like this: Two parties to a trade might start with a multi-year contract: for example, the delivery of 100 MW per hour during all peak hours at a fixed price. The generator might, however, choose to do maintenance for one month and contract with another generator to cover deliveries during that period at some other level of price. During the month of maintenance the substitute generator may have a forced outage or unexpected economic opportunity and end up covering the obligation with yet another purchase for a week or a day; again, at a different price. During the period of dispatch there can be all kinds of adjustments. The generator may produce too much or too little, as the load may be less or more than expected. The grid operator will cover the variation with its ancillary services market, often with prices that vary every ten minutes. When meters are finally read and delivery volumes are compared to contract prices the ledger entries are intensely detailed. The complex inter-relationship between multiple parties also substantially increases counter-party risk. When the California electricity market collapsed, many parties did not fully comprehend their exposure. However they are done, settlements in the electricity market are costly and in some instances risky.

The regulatory division in Anglo-Saxon countries between financial and physical trading requires two separate settlement procedures. When commodities are simple it is not too costly to monitor and settle contracts; at the same time it provides greater diversity and flexibility. Initially, NYMEX developed a simple month-long electricity contract. The immaturity of the market and unpredictable transmission constraints resulted in high basis risk and the products were not very useful for hedging, despite low settlement costs and low exchange fees.

NYMEX is adapting to the unusual characteristics of the electricity market by completely revamping its electricity products. Robert Collins (2003 p. 2) describes the new structure: "The new contracts are based on monthly, weekly, and daily LMPs determined in the PJM Interconnection – hub, one of the most competitive power markets in the country. It is the first time that the Exchange has divided a commodity market into varying blocks of time." The monthly contract is traded in NYMEX's traditional open-pit, while the weekly and daily contracts are offered in an electronic framework – ClearPort. The electronic system has another unusual twist; traders can strike a bilateral deal, then list it and clear it through the exchange. For some traders this reduces transaction cost by reducing credit costs and counter party risk. However, the trading is limited to those hubs and contracts supported by NYMEX. Further, for those involved in the physical purchase and sale of power it requires dual settlements, unlike Nord Pool. When the veneer of NYMEX's new arrangement is stripped away, it bears a remarkable similarity to the one abandoned in the U.K.

Nord Pool, as the principal example of a successful electricity market, is instructive regarding the incompatibility of vertical integration and commodity markets. In January 1996 Sweden and Norway opened a joint trading exchange and the market evolved as described in Section 6.6. At that time, the Swedish and Norwegian electricity industry was a hodgepodge of independent generators, some vertically integrated utilities, and distribution companies. Four of the largest Norwegian generators supplied fifty percent of the market and only one (the second largest) was vertically integrated (London Economics 1997). Sweden had a similar situation, and in both of the countries small distribution companies also owned some generating resources. Significantly, a large number of generators and distribution companies were publicly owned, which reduced the likelihood that any of them would seek to vertically integrate through a merger or acquisition. Further, combining financial and physical trading in a single organization reduces settlement costs and counter party risk. Given these circumstances, it is understandable why a successful forward (futures) market would develop in Nord Pool and why similar markets have not matured in other jurisdictions.

When the experience in North America is compared to that of the U.K. and to the historical experience of the oil industry, it leads to the conclusion that excessive vertical integration is inconsistent with the traditional commodity market structure.[42] Put another way, either vertical integration or commodity markets are solutions to the problem of high transaction costs and price volatility, but they are unlikely to coexist since one is the alternative to the other.

As a commodity, electricity is more complex than either oil or gas, which, as explained in Chapters 3 and 7, increases transaction costs and lowers liquidity. These problems, combined with the inherent problem of specific assets, provoked vertical integration in the industry. Such barriers to market development created by high transaction costs associated with specific assets can be overcome by regulatory regimes that mandate third party access to the transportation infrastructure. For a true commodity market to evolve, however, there must be independence between buyers and sellers in sufficient numbers to support liquid trading. Otherwise, the incentive to integrate is strengthened by the extreme price volatility of a small balancing market. The higher the cost of storage, the more volatile is the balancing mechanism. Smaller volumes of trade and greater volatility decrease liquidity, increase transaction costs and drive participants to integrate.

[42] It is important to emphasize that these are alternative market structures, neither of which determines the degree of competitiveness. OPEC's interventions demonstrate that market power is exercised in the crude-oil market, which is now dominated by futures and forward trading. Likewise, the operations of the WSPP prior to 1998 demonstrate that a vertically integrated structure can be largely competitive when there are a sufficient number of buyers and sellers (Lehr and Van Vactor 1997).

7. Transaction or Trading Costs in Wholesale Energy Markets

Wholesale energy markets are a useful microcosm for the study of transaction costs. Oil and gas markets are mature and, in North America, continent-wide. Crude oil and natural gas are now the largest traded non-financial commodities in the world, dwarfing agricultural produce and most metals markets.

Central to the development of an efficient commodity market is the establishment of a marker price. Most commodities have variation in quality and/or regional distribution. This can fragment the market. In order to reduce information costs and improve liquidity, the industry typically focuses on a specific location and product where trading is most active. A marker price will emerge, acting as an index to which all other prices are compared. If the commodity in question has a successful futures market, the marker price is normally the underlying product for futures trading. Even without a futures market, however, marker prices frequently emerge as a trading index. For example, Saudi Arabian Light was used as the marker price for crude oil before the establishment of futures trading in New York and London (based on WTI and Brent crude oils).

In the United States, NYMEX has established high-volume futures markets for crude oil, gasoline, heating oil, and natural gas. In London, the IPE operates the Brent and gasoil futures markets and has expanded into natural gas and electricity. TOCOM trades gasoline and kerosene contracts and recently introduced a crude-oil contract. The growth in energy futures trading has accompanied blossoming spot trade in daily "physicals" that covers every conceivable energy product by region, delivery schedule, and volume. In addition to established futures exchanges, there has been a substantial increase in OTC derivative products and a rapid growth of electronic trading. The most notable success was Enron-On-Line (EOL), which, before Enron's collapse in December 2001, was said to handle from one-quarter to one-third of all natural gas and electricity trading. Since the collapse of EOL, trade has migrated to other platforms, including the Intercontinental Exchange (ICE).[43]

Market information is readily available to market participants through a vast array of systems providing the latest data on weather, infrastructure bottlenecks and facility outages, prices, risk management tools, etc. Most energy-trading companies trade in multiple energy markets and institutional arrangements are remarkably similar. Successful innovations in one market are quickly copied to another. As explained earlier, these systems have been adopted to reduce the transaction costs of trading and, in general, such costs have declined over the last decade.

Despite similarities in the three principal energy sources of oil, gas, and power, there are also differences. Electricity was the last market to be deregulated and successful futures markets have yet to be established. The existing high-voltage transmission grid is not suited for a national market in the U.S. since it was designed to support the operations of vertically integrated utilities, with price and access regulations enforced somewhat

[43] EOL was a "one to many," exchange in which Enron was always the counter party. In contrast, ICE is a "many to many" exchange in which the organization acts as a broker matching buyers to sellers.

inconsistently by both local and national authorities. In that structure efficient trade was a secondary motive.

With one regional difference, oil prices move in tandem across the continent. The exception is the West Coast petroleum product market, which is isolated by regulation and geography from the remainder of North America. (There are no major pipelines connecting the Gulf Coast to the West Coast, trade between the regions is limited, and tanker traffic from West to East is costly).

In the 1990s regional natural gas prices converged in North America. The period of the gas and power market disruption of 2000-2001 was a major exception. In that disruption, spot natural gas prices in California diverged dramatically from other regions for over one year. Power prices have yet to converge to a national norm and may never do so. This is because the width of the continent and sparse population in the center prevent the easy synchronization of the grid. To avoid cyclical interference, the eastern and western systems were designed independently and are operated separately.

7.1 Review of Commodity Characteristics and Pricing Conventions

The primary commodities studied – crude oil, heating oil, gasoline, natural gas, and electricity – have significantly different physical characteristics. Crude oil is highly heterogeneous, varying in quality from tar to natural gasoline, and has differing amounts of impurities that impact the cost of refining. As of December 2001, there were 34,969 oil fields in the United States (*Oil and Gas Journal* 2001 p. 155), with each field possessing unique quality characteristics and logistical considerations. Crude oil is priced by field name (or a blend from multiple fields) and the price is normally based at the wellhead. In contrast, natural gas is homogenous by design once it enters the pipeline transmission and distribution system (impure gas is normally cleaned and standardized at or close to the field) and is priced at market centers and hubs.

Petroleum products - gasoline and heating oil - are standardized but there are quality variations. As with crude oil, there is constant grading and assaying of products to ensure they meet the correct specifications. Generally, petroleum products come in smaller lots than crude oil. They also require "clean" pipelines, tankers and storage tanks, all of which add to handling costs. Petroleum product prices in specific locations account for both differing transportation costs and local market conditions. Both crude oil and petroleum products have flexible transportation systems: marine tanker, pipeline, tanker truck, rail, or barge can move the commodities. Transportation bottlenecks emerge, but they usually do not last long, given the flexibility.

Natural gas, as a gas, does not have a flexible transportation system and is more costly to store than oil. The various pricing points represent nodes in the infrastructure where there is ample storage and/or multiple pipelines converge. The most important of these hubs is Henry Hub in Louisiana. Henry Hub is the key hub for Gulf Coast natural gas and it interconnects major pipelines in the region. It is also the basing point for NYMEX futures trading. Because of its location and interconnections, Henry Hub has the highest liquidity (highest volume, frequency of trading, and number of independent traders) of any gas market center in the continent. Another hub of interest concerns a large storage facility in Alberta, operated by the Alberta Energy Company (AECO), which helps tie Canadian gas producers to the U.S. market.

As a commodity, electricity is even more complex than natural gas. It is standardized, but differentiations are made between Direct Current (DC) transmission and Alternating Current (AC) with high-voltage transmission, distribution, and home use. Like natural gas, the distribution of the commodity takes place through a fixed infrastructure. Unlike oil and gas, electricity decays immediately; generally if it is not used when it is produced, it is wasted.[44] Grid balancing is also a complex problem for a market to resolve. Electric load and generation must be balanced to maintain delivery system reliability. If the grid does collapse, service may not be restored for hours or days. A natural gas grid can collapse also, but the tolerances are much greater – on the order of two to three days of pressure decline, rather than a few minutes. Electricity, like gas, is priced at various hubs, but the geographic point may be less distinct, encompassing an operating area or service territory, rather than just an interconnection of transmission lines.

7.2 The Impact of Commodity Characteristics on Transaction Costs and Liquidity

Complex commodity characteristics segment markets, which reduces liquidity and raises transaction costs. Market segmentation occurs through geographic or spatial constraints, trading hours and venues, or through product differentiation.[45]

Liquidity may be associated with the degree of competitiveness of a market, but it need not be. For example, the market for housing is highly competitive: for most urban areas there are normally a large number of sellers and buyers. Nonetheless, it is not a liquid market. Buyers are seeking housing of a certain style, assured quality, size, and neighborhood. The commodity is indivisible, heterogeneous, and immobile. Sellers want to avoid high counter-party risk with regard to financing, commitment, and liability. And, of course, the buyer wants a low price and the seller a high price. The consequence of this complexity is both high search and information costs for the transacting parties and high fees to the brokers who make their living by reducing these costs.

Similar issues permeate energy markets. There are problems in dealing with the diversity of crude oil. For example, at any one time a refiner may need to increase or decrease the volume of light petroleum products it produces. This can be cost-effectively implemented by changing the proportion of light and heavy crude oils in the refinery feed stock. For example, a refiner seeking to produce more gasoline will search for light crude oils. However, such a shift segments the market by effectively excluding heavy crude-oil producers. Of course, if this action prompts a rise in light crude oil prices and a drop in heavy crude-oil prices, the refiner may abandon the plan. In all cases, however, it requires greater search time and information costs to sort out the optimum mix.

---

[44] There are a variety of storage devices utilized by the industry, but relative to oil and gas they are very costly and only economic under strict conditions.

[45] Obviously, much product differentiation in retail markets is driven by consumer demand. Wholesale markets, on the other hand, often must cope with heterogeneous raw materials (such as crude oil). Here the lack of standardization may be an impediment to efficient trading.

The geographic segmentation of the natural gas market also increases transaction costs. Although the North American pipeline system is highly interconnected, directional flow frequently limits the number of customers a producer can directly serve, constraining the commodity's mobility. Demand and supply must be equilibrated through a series of transactions, where one supplier displaces another. For example, higher demand in Seattle diverts gas from British Columbia and Alberta to the West Coast. The resulting reduction in deliveries to Chicago causes prices to rise, so producers in Louisiana divert gas that would have been sent to the Mid-Atlantic to the upper Midwest. Although displacement integrates the whole market, it is neither instantaneous nor costless. New spot purchases and sales have to be made and longer-term contracts have to be adjusted. In the short-term, the market is segmented into specific delivery nodes, where relative prices fluctuate and transaction costs vary as the market shifts.

7.3 Measuring Transaction Costs

As noted in earlier chapters, the spread between buy and sell (or bid and ask) prices can be used as a proxy for transaction costs. This figure will not capture the hidden costs associated with using the price mechanism but includes a premium for "immediacy," which may or may not be treated as an issue separate from transaction costs.[46] Despite the measurement's shortcomings, however, relative transaction costs between products should be captured by the difference between buy and sell prices, if market institutions are similar. Such data can be used to identify the key variables impacting transaction costs, measure their impact, and test whether or not commodity characteristics are relevant.

Demsetz (1968) studied transaction costs in the NYSE and found that bid-ask spreads were inversely proportional to the frequency of trading. Demsetz noted that the bid-ask spread was something more than a brokerage fee, i.e. a fee for matching buyers and sellers. In securities markets, specialists (a type of market maker) buy and sell from their own inventory in order to provide immediacy to buyers and sellers. Demsetz (1968 p. 36) commented: "The ask-bid spread is the markup that is paid for predictable immediacy of exchange in organized markets, it is the inventory markup of retailer or wholesaler." Ho and Stoll (1980) also studied the market structure of equity exchanges and identified an additional key variable impacting bid-ask spreads: price volatility. This is a key point in analyzing and interpreting the difference between buy and sell prices. If these transaction costs were only brokerage fees, price volatility should have a negligible effect on their relative size, since brokers assume no risk.

There is no reason to believe that the determinants of transaction costs associated with commodity spot markets would be significantly different than those from the floor of a stock exchange. There are four key types of actors: 1) energy producers who seek to sell the commodity; 2) brokers that match buyers and sellers but do not take a position; 3) marketers that take positions, acting as market makers, as well as matching buyers and sellers; and 4) wholesale consumers that seek to purchase the commodity. Given that

[46] For an extended discussion of the problem of interpreting the demand for immediacy in the context of defining liquidity, the reader is referred to Grossman and Miller (1988).

both brokers and marketers operate in the energy market, observed buy and sell prices will reflect both brokerage fees and the demand for immediacy. Obviously, commodities with high storage costs or rapid decay, such as electricity, will have a higher demand for immediate transactions, independent from consumer patience.

Studies of commodity markets similar to those completed for securities markets have not been conducted, because, as explained by Smith and Whaley (1994): "Unlike securities markets, however, futures markets do not record the sequence of all transaction prices and bid/ask quotes during the trading day." Futures exchanges work on an "open outcry," system. In this structure pit traders act as agents to match buyers and sellers; a process that does not reveal an explicit bid-ask spread.[47] On the other hand, because of the sheer number and diversity of equity listings, stock exchanges frequently rely on specialists as market makers to provide liquidity.[48]

Even if bid-ask data were available from a futures market, it might not reveal differences in the transaction costs based on commodity characteristics. Financial instruments commonly traded on NYMEX and other exchanges are physically homogenous. Unlike the commodities themselves, they are stored, handled, transported, and protected in similar or identical ways. It is also important to note that futures markets are classic derivative markets. That is, they exist to shift risk, not allocate the physical resource. Thus, it is the cost of the price mechanism in physical markets that is of primary interest.

Some approaches to the study of transaction costs in securities markets are helpful to this study. Robert Neal (1992), for example, studied the variation in the cost of options trading in two different market structures. His study is similar to this investigation, which is a study of different commodities in the same market structure.

Although transaction cost data are not available from futures exchanges, Reuters conducts daily price surveys on buy and sell prices for crude oil, gasoline, heating oil, and natural gas in the physical market.[49] For the North American market, there are 49 distinct products and/or trading hubs. These data provide sufficient diversity to conduct cross-sectional regression analysis to explain the relationship between transaction costs (as measured by the spread between buy and sell prices), commodity characteristics and other independent variables.

7.4 Accounting for Liquidity

There is one major hurdle in any study of transaction costs in spot markets; data on the volume of trading are not available. As it turns out, however, trading volume does not

---

[47] Recognizing that bid-ask spread data are not available from futures exchanges, a number of studies [Smith and Whaley (1994) Locke and Venkatesh (1997), Wang, Yau, and Baptiste (1997), and Ding (1999)] have attempted to estimate the parameter based on observed price changes through the day of trading. In my view, however, these methodologies are too speculative to apply to an estimation of the impact of commodity characteristics on transaction costs.

[48] Some of the burgeoning electronic exchanges, notably EOL, used the market maker system. At any one time Enron was simultaneously offering to buy and sell a specific commodity.

[49] Most other price survey publications collect a single price or daily range, rather than buy and sell.

appear to be a critical component in estimating the impact of commodity characteristics on transaction costs in energy markets except in specific instances that can be proxied by dummy variables. As the data reveal, North American natural gas, crude oil, gasoline and heating oil markets are continuous across the continent. There are exceptions, which are explained below, but they are few and can be accounted for qualitatively.

The North American natural gas pipeline system is highly interconnected, with only one major bottleneck in this time period: capacity into California. The bottleneck was caused by a surge in the demand for gas-fired power generation associated with California's electricity crisis and by an explosion on the El Paso pipeline which stretches from Texas to California, ultimately leading to an increase in the variability of basis,[50] which Tussing and Tippee (1995 p. 271) define as "a geographic price differential between a particular market and the delivery point specified in an exchange-traded commodity contract such as a futures contract." This meant that California gas prices frequently diverged from those at Henry Hub, the marker for NYMEX futures trading.

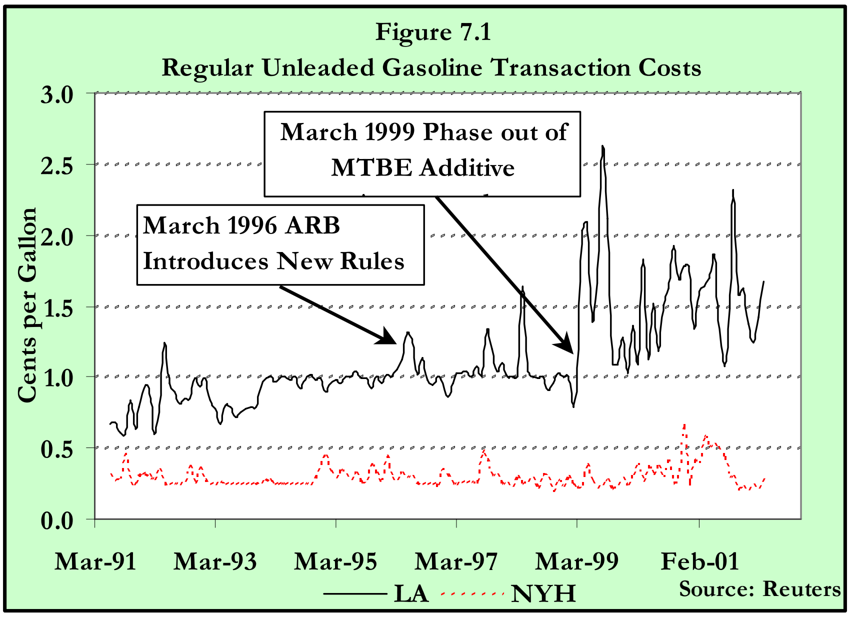
While Henry Hub is arguably the single most liquid and highest-volume market for natural gas trading, the marker for crude oil trading is less well defined. Originally, NYMEX chose WTI at Cushing, Oklahoma as the marker. Over the years, however, the volume declined and changes in the pipeline infrastructure caused WTI prices at Cushing to sometimes diverge from comparable crude oils in the region. Consequently, NYMEX shifted the marker to a more general categorization – "light sweet crude." While this resolved the problem of out-of-sync pricing, it also diffused liquidity among multiple crude oils and locations.

The crude oil market can be subdivided into two principal regions, the West Coast and the Gulf Coast. There are no major connecting crude oil or petroleum product pipelines across the Rockies and the only alternative is to ship oil by tanker to the Gulf Coast or export it to the Far East. Until recently, the U.S. banned the export of Alaska crude oil. The combination of surplus oil from Alaska, a ban on exports, and costly transportation to alternative markets resulted in a West Coast oil "glut," which frequently depressed prices relative to world standards (Van Vactor 1995 p. 18). However, Alaska North Slope crude oil production has declined by over 50% and the glut has disappeared. The West Coast is now a net importer of crude oil and the cost of moving Persian Gulf oil to California is about the same as moving it to Houston. As a consequence, West Coast and Gulf Coast crude oil prices are now closely linked and liquidity is similar in the two markets.

The volume of futures trading in crude oil and natural gas is much higher than gasoline and heating oil and there are greater regional differences that impact the use of petroleum products. The heating oil market is dominated by New England, which until recently was not widely served by natural gas. Heating oil also has a strong seasonal market. Concern about high seasonal prices and product availability led the U.S. government to establish a heating oil storage facility in New England. However, for all practical purposes, heating oil and diesel fuel (for transport) are effective substitutes, so

[50] Also see Hull (1993), pp. 34-36.

the problem of supplying New England may be exaggerated. Reuters also tracks heating oil prices in the Pacific Northwest (Oregon and Washington). This is the most isolated and smallest petroleum product market in the U.S. Four major refineries plus imports and surplus products from California serve the region. However, there are no major crude oil or petroleum product pipelines connecting the central markets of the Pacific Northwest to other regions.



Recently, U.S. environmental regulations have exacerbated segmentation of the gasoline market. According to API (2001): "Refiners and distributors must make and transport a growing number of products because of new government requirements, both federal and state. For example, many states have imposed or are considering imposing specification for gasoline sold there but in few other places." Such market segmentation increases transaction costs. Figure 7.1 illustrates the trend in transaction costs of regular gasoline sold in California, as compared to sales in New York Harbor. California's unique changes in product specification have not only increased transaction costs, but these costs are also far more volatile. In addition to the impact of environmental regulations, the gasoline market is also segmented by geography. Again, the Pacific Northwest is the exception – it is the most isolated market.

As the above discussion implies, there are known differences in the liquidity of oil and gas trading hubs in North America. Ideally, these differences should be quantified, but an objective measure is simply not available. As a consequence, relative liquidity is addressed through qualitative discussions of the hubs and regions and the addition of a dummy variable for petroleum products marketed in California and the Pacific Northwest.

7.5 Trading Hubs and Statistics

Before describing the statistical analysis of North American oil and gas hubs, it is useful to provide a brief description of their location. Figures 7.2 to 7.5 illustrate the geographic dispersion of natural gas, crude oil, heating oil, and gasoline hubs and trading locations. Also indicated in the illustration are the basis risk differentials (the variation of the difference between the regional product and the NYMEX marker for futures trading as a percentage of the price) for each product over the period from April 1, 2001 to March 31, 2002.

Figure 7.2: North American Heating Oil Hubs and Basis Risk vs. NYH

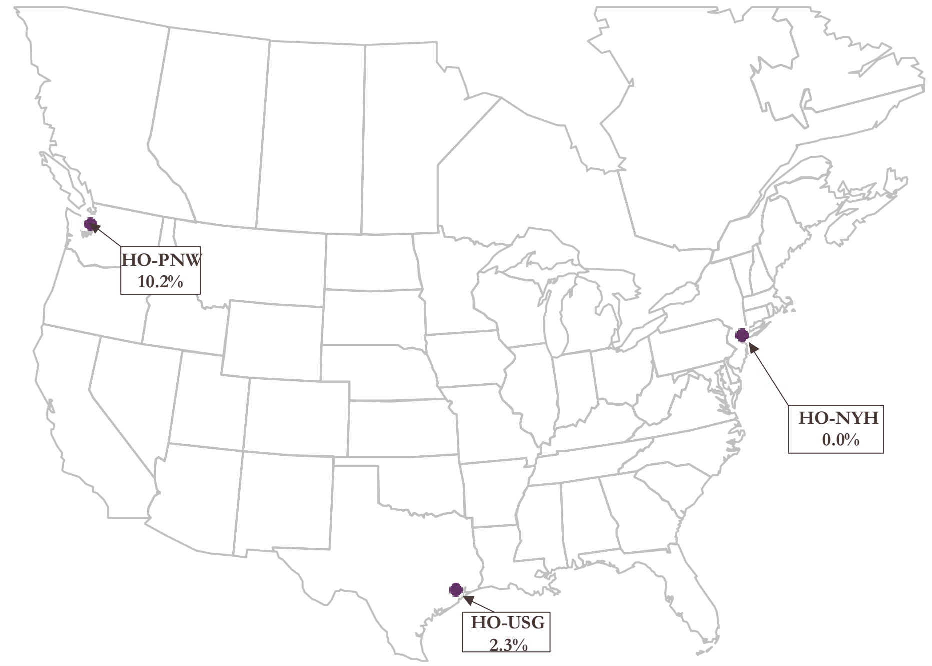


Figure 7.3: North American Gasoline Hubs and Basis Risk vs. NYH

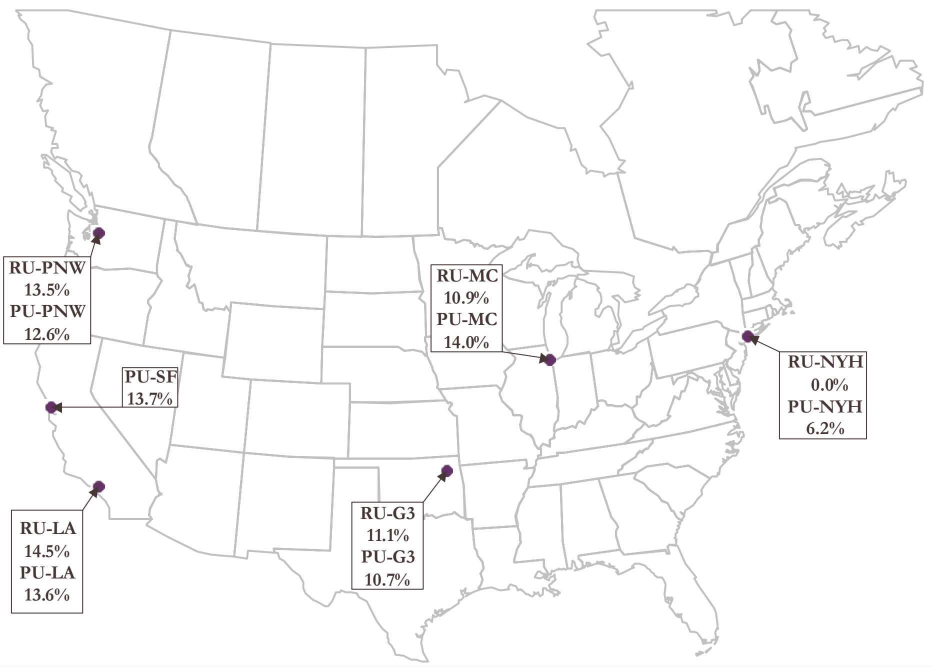


Figure 7.4: North American Crude Oil Hubs and Basis Risk vs. NYMEX Marker

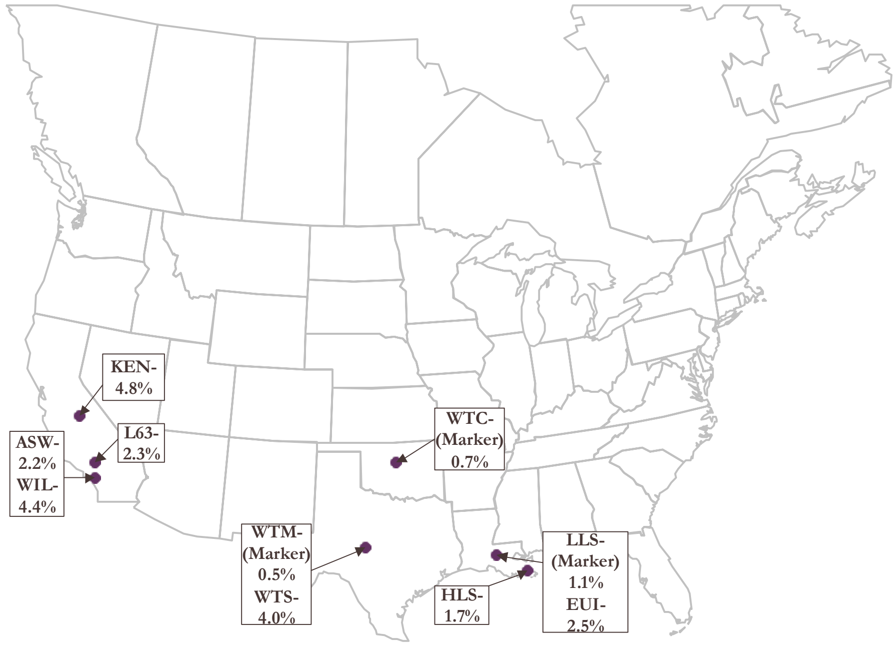


Figure 7.5: North American Natural Gas Hubs and Basis Risk vs. Henry Hub

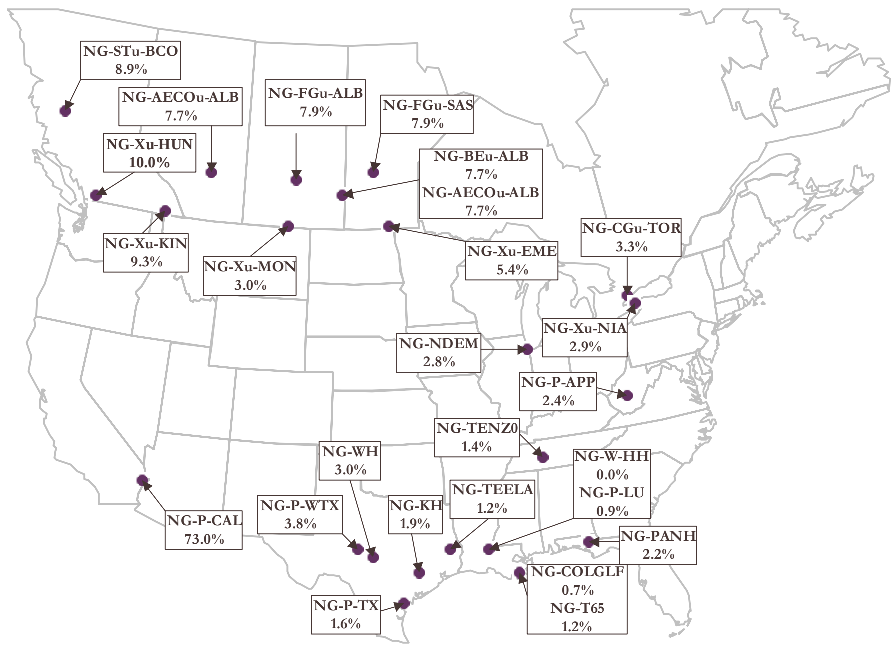


Table 7.1 provides summary statistics on oil and gas trading hubs in North America for the period April 1, 2001 to March 31, 2002. The commodity with the lowest transaction cost (TC%) was West Texas Intermediate at Midland, Texas (WTM-) at 0.17% of the buy price. Natural gas at the California border (NG-P-CAL) was the commodity with the highest transaction cost, at 3.44%, twenty times higher than the lowest. Except for petroleum products traded on the West Coast, the transaction cost of trading natural gas was consistently higher than crude oil or petroleum products. As noted earlier this is likely due to the constraints imposed by the fixed pipeline distribution system and the high cost of storage, which arise from commodity characteristics.

Table 7.1
Key Statistics Concerning North American Oil and Gas Hubs
April 1, 2001 to March 31, 2002

Product	TC%	Average Price	Average Basis	Basis Risk	Price Volatility
Heating Oil (Per Gallon)					
HO-NYH (Marker)	0.36%	$0.67	$0.00	0.0%	41.1%
HO-PNW	2.35%	$0.71	$0.04	10.2%	53.2%
HO-USG	0.36%	$0.64	($0.02)	2.3%	45.6%
Gasoline (Per Gallon)					
RU-NYH (Marker)	0.50%	$0.70	$0.00	0.0%	49.1%
PU-G3	0.98%	$0.82	$0.12	10.7%	57.7%
PU-LA	1.89%	$0.87	$0.17	13.6%	63.3%
PU-MC	1.28%	$0.83	$0.13	14.0%	61.2%
PU-NYH	0.60%	$0.77	$0.07	6.2%	51.1%
PU-PNW	1.85%	$0.86	$0.16	12.6%	42.4%
PU-SF	1.91%	$0.87	$0.17	13.7%	66.0%
RU-G3	0.59%	$0.77	$0.08	11.1%	61.1%
RU-LA	2.05%	$0.81	$0.11	14.5%	67.4%
RU-MC	0.84%	$0.77	$0.07	10.9%	57.6%
RU-PNW	2.11%	$0.80	$0.10	13.5%	42.6%
Crude Oil (Per Barrel)					
LLS- (Marker)	0.24%	$24.57	$0.15	1.1%	43.9%
WTC- (Marker)	0.28%	$24.45	$0.05	0.7%	44.1%
WTM- (Marker)	0.17%	$24.22	($0.20)	0.5%	44.0%
ASW-	0.76%	$22.24	($2.16)	2.2%	48.7%
EUI-	0.74%	$23.24	($1.20)	2.5%	54.3%
HLS-	0.42%	$24.28	($0.14)	1.7%	46.9%
KEN-	0.60%	$17.72	($6.69)	4.8%	39.8%
L63-	0.81%	$21.74	($2.67)	2.3%	50.2%
WIL-	0.55%	$18.95	($5.45)	4.4%	38.9%
WTS-	0.49%	$22.21	($2.20)	4.0%	52.9%
Natural Gas (Per mmBtu)					
NG-W-HH (Marker)	0.70%	$3.20	$0.00	0.0%	76.0%
NG-AECOu-ALB	1.41%	$2.84	($0.36)	7.7%	109.1%
NG-BEu-ALB	1.36%	$2.86	($0.34)	7.7%	107.2%
NG-CGu-TOR	1.66%	$3.35	$0.14	3.3%	72.3%
NG-COLGLF	1.77%	$3.15	($0.05)	0.7%	81.3%
NG-FGu-ALB	1.42%	$2.77	($0.43)	7.9%	111.5%
NG-FGu-SAS	1.43%	$2.77	($0.43)	7.9%	111.8%
NG-KH	1.77%	$3.18	($0.02)	1.9%	89.7%
NG-NDEM	1.79%	$3.14	($0.07)	2.8%	88.0%
NG-PANH	1.83%	$3.08	($0.12)	2.2%	90.1%
NG-P-APP	1.67%	$3.37	$0.17	2.4%	83.8%
NG-P-CAL	3.44%	$5.48	$2.28	73.0%	187.9%
NG-P-LU	1.79%	$3.16	($0.04)	0.9%	83.3%
NG-P-TX	1.82%	$3.10	($0.11)	1.6%	88.5%
NG-P-WTX	1.84%	$3.06	($0.14)	3.8%	105.8%
NG-STu-BCO	1.39%	$2.84	($0.36)	8.9%	100.9%
NG-T65	1.74%	$3.22	$0.02	1.2%	83.0%
NG-TEELA	1.79%	$3.15	($0.06)	1.2%	84.8%
NG-TENZ0	1.81%	$3.10	($0.11)	1.4%	88.6%
NG-WH	1.81%	$3.12	($0.09)	3.0%	97.5%
NG-Xu-EME	1.83%	$3.08	($0.12)	5.4%	84.4%
NG-Xu-HUN	1.98%	$2.92	($0.29)	10.0%	102.3%
NG-Xu-KIN	1.98%	$2.91	($0.29)	9.3%	101.2%
NG-Xu-MON	2.07%	$3.02	($0.18)	3.0%	88.6%
NG-Xu-NIA	1.77%	$3.39	$0.18	2.9%	78.6%

Source: Reuters

In each category of Table 7.1, the first product listed is the marker commodity for futures trading. In the case of heating oil, NYMEX's marker is based on trading in New York harbor. During this period the average wholesale price of New York heating oil was $0.67 per gallon and the average basis difference was, of course, zero. Heating oil in the Pacific Northwest had an average of $.71 per gallon during the period, for an average basis difference of $0.04 per gallon. Basis risk, the fifth column in the table, is computed as the standard deviation of the daily basis difference as a percentage of the average price. Generally, crude oils have the lowest basis risk, followed by heating oil, gasoline, and natural gas. Basis risk in natural gas is the most variable, reflecting the risk of pipeline bottlenecks (as was the case with California deliveries in 2000 and 2001).

The final column of Table 7.1 is price volatility. This series is calculated using the standard formula for volatility based on historical prices (Hull 1993 pp. 214-17). Begin with a set of historical prices for a given commodity:

$$\{P_0, P_1, \dots, P_T\}$$

First, relative log prices are computed, representing the percentages that prices change between each observation expressed as continuously compounded rates for t from 1 to T (in this instance T=20 business days):

$$R_t = \ln(P_t/P_{t-1})$$

Then, compute the mean of the R_t as follows

$$\bar{R} = \frac{\Sigma R_t}{T}$$

From this compute the variance, v^2, or the sum of the squared difference between relative log prices and their mean over the sample.

$$v^2 = \frac{\Sigma(R_t - \bar{R})^2}{(T-1)}$$

The final volatility calculation is then calculated as follows:

$$\sigma = v\sqrt{N}$$

Where σ = volatility, v = standard deviation, and N = number of price observations in one year (252). "v" is the square root of the variance, v^2.

The above formula is widely used by the industry to manage risk. Crude oil and heating oil have the lowest level of price volatility, followed by gasoline. Natural gas has the highest level of price volatility.

7.6 Modeling Transaction Costs

As noted, studies of securities markets have found that price volatility and trading volume are the key determinants of the spread between bid and ask prices. Commodity markets, however, have an additional constraint: they must account for regional and quality diversity. Traders try to avoid what they describe as being "squeezed." That is, they do not want to be caught with the obligation to buy or sell in a thin market with few trading opportunities characterized by high basis risk. In general, the more idiosyncratic the product and its delivery point, the greater the risk of a bargaining disadvantage. If the basis is stable, then traders can use the futures market to hedge risk and prevent being squeezed. If, on the other hand, the basis is unpredictable, then the risk of trading at the hub in question is much higher. The level of basis risk has other important implications. One of the key functions of a futures market is price discovery. During the open hours of the exchange, trading is continuous and prices are widely disseminated. If these prices are a reliable indicator of value at a local hub, then information and search costs will be far lower and the negotiation range between buyers and sellers will be significantly narrowed. All things equal, there should be a strong link between transaction costs and basis risk – the greater the risk, the higher the expected cost.

The first step of analyzing the differences between the transaction costs of trading oil and gas was to conduct a simple t-test of 24 observations from both categories. This test resulted in a t value of 4.91 and standard deviation of 0.00581, indicating a probability of 0.001 of falsely rejecting the null hypothesis; that is, that the mean of the two samples is the same.

The second step was to include other determinants of transaction costs in the analysis. The data in Table 7.1 can also be used to estimate the relationship between transaction costs and relevant independent variables across a cross-section of commodities in crude oil, petroleum products, and natural gas markets. This has the added advantage of allowing other important variables to be considered, so that a more predictive model can be formed.

The key variables are summarized as follows:

T = transaction cost as percentage of buy price;
B = basis risk as standard deviation of basis difference, as a percent of average price;
D1 =dummy variable for oil (1) as differentiated from gas (0);
D2 =dummy variable for West Coast petroleum products (1) as compared to all other products (0).

Where

$$T = f(B, D1, D2)$$

It is expected that the sign of the coefficient of the independent variables will be as follows: $B>0$, $D1<0$, $D2>0$.

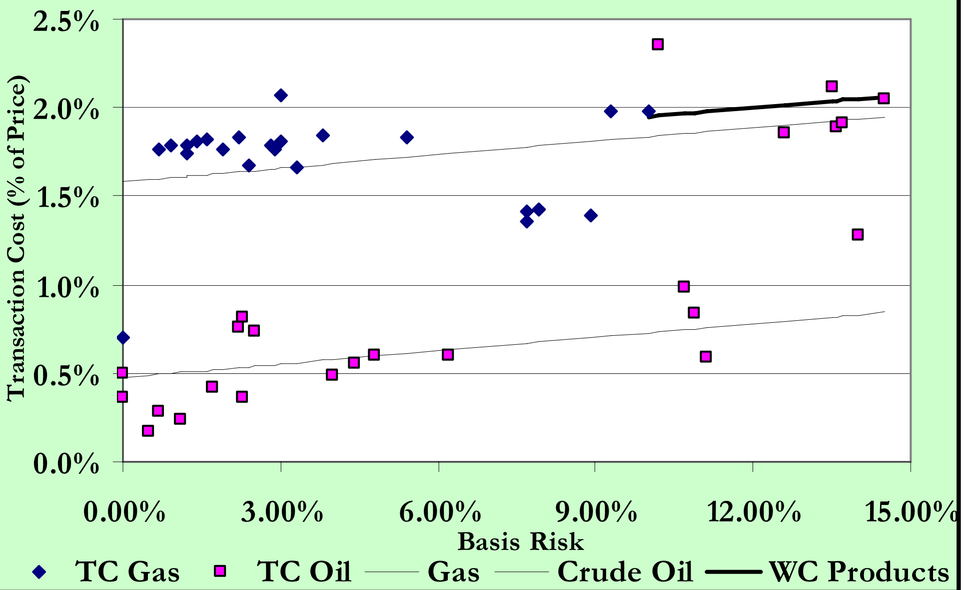
Ordinary least-square estimates of the above coefficients were calculated with the results reported below:

Table 7.2 Summary Statistics of Transaction Cost Estimating Model			
Dependent Variables	**Coefficient**	**Standard Error**	**T-Ratio**
Intercept	.0158	.0006	27.5
Basis Risk	.0254	.0036	7.0
Oil-Gas Dummy	-.0110	.0008	-13.7
WC Products Dummy	.0122	.0013	9.7

The overall R^2 value of the estimation is .872, and T-Ratio statistics for all of the variables were significant. The null hypothesis for interpretation of the results can be stated as follows: transaction costs are unrelated to commodity characteristics and basis risk. These results indicate, as with the t-test, a very low probability of falsely rejecting the null hypothesis. Other statistical indicators of the regression are included in the appendix, but there did not appear to be any substantive biases. Since this is a very simple and straightforward estimation, it is also represented graphically in Figure 7.6.

Figure 7.6

Regression of Oil and Gas Transaction Costs



Oil products are represented in Figure 7.6. The relationship between the various variables is strictly linear. Double log and partial log versions of the model were also tried; however, linear specifications provided the best fit. The oil-gas dummy variable estimates that the regression line for natural gas should be shifted down by 1.1 percentage points for oil products, maintaining the same slope. That is, for any given level of basis risk, transaction costs of natural gas would be expected to be 1.1 percentage points higher. In turn, the regression line of the six West Coast oil products is shifted up by 1.2 percentage points above other oil products. Figure 7.6 does not illustrate transaction costs and basis risk for California natural gas, which were 3.4% and 76%, respectively. Surprisingly, it makes no significant difference to the estimate of the slope of the regression, or the overall R^2, whether or not California gas is included.

Within a given category, the lowest transaction costs are usually associated with the marker for futures trading. This represents both low basis risk and high volumes of trading at that hub. Five natural gas observations from Western Canada have a relatively low transaction cost, given their level of basis risk. These are trading and export points surrounding Alberta's AECO hub. Alberta's NGX offers a basis contract between AECO and Henry Hub that can be used to mitigate adverse price movements. This, along with new pipeline development, appears to have enhanced the liquidity of the Alberta market and may explain the relatively lower transaction costs.

In addition to the least squares model chosen here, a number of variations were considered and rejected. For example, independent variables for volatility, a dummy variable differentiating oil and gas, and a dummy variable for futures marker hubs provided reasonable R^2 and t-values, but had some statistical problems. Likewise a model based on basis risk, the volatility of the marker price for each type of fuel, and associated dummy variables produced similar results. Instead of calculating transaction costs as a percentage, Neal (1992) and others have specified the variable in original units and included the associated price of the commodity as an independent variable. This approach was also tried, but appeared to exhibit problems of heteroskedasticity and the R^2 and t-values were not as high. In virtually all cases, however, t-values of the oil-gas dummy appeared to be significant. Seasonal demand volatility estimates for each fuel at each hub were also substituted for the oil-gas dummy variable as an alternative explanation, but without statistically significant results.[51]

7.7 Transaction Costs of Electricity Trading

Exepting one series in the U.K. reported in Chapter 6, there are no readily available data on buy and sell prices for electricity that are comparable to natural gas. ICE, however, now makes trading data available on both natural gas and electricity. ICE had active trading at 27 natural gas hubs and 7 electricity hubs for the period April 1, 2001 to March 31, 2002. Details on recent trading are contained in Table 7.5 at the end of this chapter. Volume weighted averages are summarized in Table 7.3.

These data summarize what is well known among energy traders: electricity prices and costs are far more volatile than any other commodity. The various measures of volatility are consistently high for the individual trading hubs reported in Table 7.5. Basis risk calculations for electricity are problematic, since dominant trading hubs have yet to emerge. In this instance, basis risk was calculated from the difference between Midwest hubs and "Cinergy," which is in Ohio. Cinergy is the most actively traded hub, and is

[51] Overall natural gas has stronger seasonal demand than oil. A figure based on the monthly standard deviation of consumption can be substituted for the oil-gas dummy variable with reasonable results. However, the relationship breaks down when seasonality at individual hubs is considered. Natural gas seasonal demand in the southern states matches that of oil. Heating oil has strong seasonal demand in New England, if it is considered different from diesel fuel. Gasoline demand is quite seasonal in the Pacific Northwest, but not in California.

central to the region. Basis risk for Nepool (New England Power Pool) is based on prices in PJM, the most active hub in the Mid-Atlantic.

Table 7.3
Summary Statistics from the Intercontinental Exchange (ICE)

Commodity	Price Volatility	Daily Price Range	Basis Risk
Natural Gas	97%	4.3%	6.6%
Electricity	221%	12.6%	20.9%

The data in Table 7.3 imply high transaction costs for electricity trading. If the model's coefficients derived for natural gas are applied to the ICE basis risk calculation for electricity, the model would predict transaction costs at 2.1%, plus any costs added due to the special characteristics of electricity. Of course, electricity is likely to have generally higher transaction costs than natural gas, due to immediate decay and the necessity to balance the grid. This is particularly true for "real-time" trading (Van Vactor and Blumsack 2002).

The evidence presented in this study suggests that a commodity's physical characteristics do have an impact on transaction costs and that the more complex those characteristics, the higher will be the cost of trading. No evidence was uncovered to accept the null hypothesis, that transaction costs are unrelated to commodity characteristics. Further, the regression analysis suggests that the immobility and inflexibility of the natural gas delivery system has a larger impact on transaction costs than does variation in crude oil quality. That is, higher negotiating costs for natural gas from lower liquidity at individual hubs substantially outweigh measurement and other costs arising from crude oil diversity.

In addition to a commodity's characteristics, transaction costs appear to be determined by basis risk. This is in contrast to studies of transaction costs in securities markets where the key determinants are price volatility and the volume of trading. The difference is explained by the fact that stock and bond prices do not vary with respect to where the certificates are stored or transferred and that the trading volume of each type of security can be measured. In contrast, trading volume of commodities at the various hubs is unknown and basis risk substitutes as a measure of market integration, an indicator of liquidity.

Although this study has not investigated the relationship between transaction costs and the competitiveness of markets, there is a strong inference that the two are linked. Clearly in the semi-regulated U.S. natural gas and electricity wholesale markets, concern about the exercise of market power is most closely associated with those markets that exhibit the highest transaction cost, i.e., California. Again this comes down to market segmentation. During California's electricity and gas crises, there were serious bottlenecks between the State and other regions. Because pipelines and power transmission facilities had inadequate capacity, outside competitors could not economically enter the market. This created high prices, high basis risk, low liquidity and high transaction costs.

Table 7.4

Ordinary Least Squares Estimation

****

Dependent variable is TC
49 observations used for estimation from 1 to 49

****

Regressor	Coefficient	Standard Error	T-Ratio[Prob]
INPT	.015825	.5749E-3	27.5270[.000]
BASRSK (Basis Risk)	.025377	.0036261	6.9985[.000]
D1 (oil-gas)	-.011077	.8073E-3	-13.7209[.000]
D2 (WC pet. Prod.)	.012206	.0012633	9.6621[.000]

****

R-Squared	.87180	R-Bar-Squared	.86325
S.E. of Regression	.0025965	F-stat. F(3, 45)	102.0016[.000]
Mean of Dependent Variable	.013591	S.D. of Dependent Variable	.0070215
Residual Sum of Squares	.3034E-3	Equation Log-likelihood	224.2836
Akaike Info. Criterion	220.2836	Schwarz Bayesian Criterion	216.4999
DW-statistic	2.2195		

****

Diagnostic Tests

****

Test Statistics	LM Version	F Version
A:Serial Correlation	CHSQ(1)= .96690[.325]	F(1, 44)= .88572[.352]
B:Functional Form	CHSQ(1)= .87689[.349]	F(1, 44)= .80176[.375]
C:Normality	CHSQ(2)= 11.9016[.003]	Not applicable
D:Heteroscedasticity	CHSQ(1)= .0078169[.930]	F(1, 47)= .0074991[.931]

****

A:Lagrange multiplier test of residual serial correlation
B:Ramsey's RESET test using the square of the fitted values
C:Based on a test of skewness and kurtosis of residuals
D:Based on the regression of squared residuals on squared fitted values
[Based on a null hypothesis of homoscedasticity]

Table 7.5
Comparison of Key Data From the Intercontinental Exchange

Natual Gas Hubs	Average Price	Volume (mmBtu) or MWh	Volatility*	Standard Deviation of Price	Daily Price Range	Standard Deviation of Range	Basis Risk
HUN-SMS	$2.35	30,897,200	125%	23%	4.1%	2.9%	12.0%
TGC-ELP	$2.93	33,032,500	93%	31%	3.7%	3.4%	1.9%
NGPCA-NPCG	$3.04	36,855,100	89%	31%	2.7%	2.9%	2.1%
TCGPC-Z6	$3.35	37,386,400	127%	27%	6.1%	8.3%	7.8%
EP-SJB	$2.60	46,503,300	150%	33%	6.3%	6.4%	12.7%
MC-CG	$3.13	50,216,000	87%	32%	3.3%	3.5%	2.8%
TXE-WLA	$2.95	52,655,100	92%	31%	3.6%	3.7%	1.5%
TXE-M3Z	$3.36	53,320,900	95%	29%	4.0%	4.4%	3.5%
NGPCA-MCP	$2.88	54,388,200	94%	32%	3.5%	3.6%	2.0%
NNG-DP	$2.95	54,747,300	93%	32%	3.2%	3.7%	2.8%
PGE-CG	$3.93	62,002,900	194%	73%	6.2%	7.9%	51.1%
WH-WTX	$2.86	65,547,700	101%	32%	4.0%	7.8%	3.2%
OP-TG	$2.46	65,872,900	154%	33%	7.3%	5.5%	14.9%
SCAL-BOR	$4.69	66,442,100	171%	85%	5.6%	7.3%	65.2%
TCGTC-ZSLFT	$3.00	66,747,500	80%	31%	3.7%	3.1%	0.8%
TCGPC-S65	$3.04	73,648,500	90%	31%	3.7%	3.7%	1.7%
CGTC-OP	$3.00	73,908,600	82%	32%	4.0%	4.1%	0.9%
ANRPC-SETP	$2.96	77,779,600	89%	32%	3.8%	3.8%	1.1%
TGPC-ZL800	$2.92	80,557,800	88%	32%	4.1%	4.1%	1.6%
D-SP	$3.22	80,791,500	89%	31%	3.9%	3.7%	2.8%
TXE-ELA	$2.98	82,929,200	90%	31%	3.9%	3.1%	1.3%
TGPC-ZL500	$2.96	90,548,100	87%	32%	4.1%	3.2%	1.5%
NGPCA-LP	$2.98	92,457,700	86%	32%	4.6%	7.5%	1.0%
EP-KP	$2.88	106,096,000	107%	34%	4.8%	3.4%	3.8%
CGC-TCOP	$3.18	156,378,500	84%	31%	3.7%	3.1%	2.4%
NGPCA-NCG	$3.04	292,699,600	88%	32%	4.1%	3.4%	2.1%
HH-TGLA	$3.02	361,923,600	80%	31%	4.6%	3.4%	0.0%
Wtd. Avg.	**$3.07**	**148,442,469**	**97%**	**34%**	**4.3%**	**4.2%**	**6.6%**
Correlation With Volume			**-26%**	**-5%**			**-15.6%**
Power Hubs							
SP-15	$68.39	1,788,000	231%	131%	10.5%	22.4%	NA
Comed	$29.34	2,368,800	228%	58%	10.6%	17.1%	9.0%
TVA	$29.74	2,616,800	233%	49%	9.1%	14.5%	10.8%
Nepool	$40.33	2,689,600	192%	36%	3.7%	5.6%	18.6%
Entergy	$30.68	6,735,200	152%	45%	13.3%	12.6%	29.9%
PJM-West	$35.94	13,451,200	218%	79%	8.1%	26.5%	0.0%
Cinergy	$30.04	20,512,800	246%	56%	17.4%	40.9%	0.0%
Wtd. Avg.	**$33.58**	**13,355,875**	**221%**	**62%**	**12.6%**	**28.2%**	**20.9%**
Correlation With Volume			**24%**	**-10%**			**-67.5%**

*Volatility, based on 20 day moving average, calculated for most recent 11 months only.

Table 7.5 Continued
Key to Gas Hubs

HubName	CodeName
American Natural Resources Pipeline Co. - SW Pool	ANRPC-SWP
American Natural Resources Pipleline Co. - SE Transmission Pool	ANRPC-SETP
Carthage Hub - Tailgate	CH-T
Columbia Gas Co. - TCO Pool (Appalachia)	CGC-TCOP
Columbia Gulf Transmission Co. - Mainline Pool	CGTC-MP
Columbia Gulf Transmission Co. - Onshore Pool	CGTC-OP
Consumers Energy Citygate	CE-CG
Dominion - South Point	D-SP
El Paso - Keystone Pool	EP-KP
El Paso - San Juan Basin, Blanco Pool (non-Bondad)	EP-SJB
Henry Hub - Tailgate, Louisiana	HH-TGLA
Katy - Exxon Plant Tailgate	KT-EPT
Michigan Consolidated Citygate	MC-CG
Natural Gas Pipeline Co. of America - Louisiana Pool	NGPCA-LP
Natural Gas Pipeline Co. of America - Mid-Continent Pool	NGPCA-MCP
Natural Gas Pipeline Co. of America - Nicor Citygate	NGPCA-NCG
Natural Gas Pipeline Co. of America - Nipsco Citygate	NGPCA-NPCG
Natural Gas Pipeline Co. of America - TXOK East Pool (Gulf Coast)	NGPCA-TXOK
Northern Natural Gas - Demarcation Pool	NNG-DP
Northwest Pipeline Corp. - Stanfield Pool	NPC-SP
Huntingdon/Sumas	HUN/SMS
Opal Plant Tailgate	OP-TG
PG&E - Citygate	PGE-CG
Pacific Gas Transmission - Malin	PGT-M
Panhandle Eastern Pipe Line Co. - Pool Gas	PEPLC-PG
Peoples Gas Light and Coke Citygate	PGLC-CG
SoCal Border	SCAL-BOR
Tennessee Gas Pipeline Co. - Zone 0	TGPC-Z0
Tennessee Gas Pipeline Co. - Zone L, 500 Leg Pool	TGPC-ZL500
Tennessee Gas Pipeline Co. - Zone L, 800 Leg Pool	TGPC-ZL800
Texas Eastern - East LA	TXE-ELA
Texas Eastern - M3 Zone	TXE-M3Z
Texas Eastern - West LA	TXE-WLA
Texas Gas Transmission Corp. - Zone SL FT Pool	TCGTC-ZSLFT
Transcontinental Gas Pipeline Corp. - Station 65	TCGPC-S65
Transcontinental Gas Pipeline Corp. - Zone 6 (NY)	TCGPC-Z6
Trunkline Gas Company - East Louisiana Pool	TGC-ELP
Union Gas - Dawn	UG-D
Waha Hub - West Texas	WH-WTX
Westcoast Energy Inc - Station 2, B. C.	WEI-S2BC

8. *Conclusion*

8.1 *Overview of Results*

8.1.1 *Statement of the Thesis*

Transaction costs (the cost of using the price mechanism) is thought be an important determinant of firm size and structure. In the last two decades wholesale markets for oil, natural gas, and electricity have undergone a radical transformation in which the cost of trading has dropped dramatically. Such an extraordinary change in the cost of procuring raw materials ought to have had a major impact on the organizational structure of energy firms; in the oil industry there should have been a shift from vertical integration to greater specialization. The oil industry has changed, but the change has not been as complete as might be expected. All of the Major oil companies remain vertically integrated to some degree, although there is now greater variation in their activities. The energy industries investigated – oil, gas, and electricity – are, however, heavily constrained by government regulations, which has an important impact on firm size and structure, independent of economic incentives. Moreover, the newly fashioned energy markets have substantial price volatility, which provides a motive for firms to diversify, as conglomerates or by integrating upstream and downstream.

The impact of changing transaction costs on the structure of energy firms may have been ambiguous, but its influences on the institutions of the marketplace are not. When OPEC broke apart the oil industry's integrated structure, it replaced it with a high-cost and untrustworthy system of contracting. As a consequence, a traditional commodity market styled after agricultural and financial markets developed. The institutions of the new market – futures trading, price transparency through trade-press reporting, standardized contracts, the entry of marketers, etc. – have systematically reduced the cost of buying and selling wholesale energy products and created an efficient system of trade.

The oil market, North American and U.K. natural gas markets, and a number of electricity markets throughout the world have adopted a commodity market structure. There is, however, great variation in the specific institutions that facilitate these markets and great variation in the cost of trading. The differences in transaction costs are explained, in part, by differences in the characteristics of each commodity. Energy products are sold in radically different forms – oil is a liquid, at room temperature natural gas is a gas, and electricity flows in the form of electrons through wires. Commodities can be classified as to their degree of complexity and the more complex a commodity, the higher the transaction cost of bringing it to market. Electricity has the highest trading costs, because it cannot be stored and the grid must constantly be balanced. The cost of trading energy commodities forms a hierarchy; oil has the lowest transaction costs, followed by gas and electricity.

8.1.2 *An Important Caveat*

Scientific investigations typically begin with a hypothesis and proceed by analytical methods to demonstrate its consistency or inconsistency with observed behavior.

Testing a hypothesis is a challenge in economics, because most economic behavior is a function of multiple changes in a wide array of variables. Nothing is ever quite the same, so that the ability of an experimental model to give appropriate weights to the appropriate set of observable values is always suspect. This enquiry is no different and in many respects exceptionally difficult since the analysis covers three decades of the energy industry, involving three very different commodity groups on two continents. Nonetheless, it is worth stating where the enquiry began and where it ended. It began with the hypothesis that transaction-cost economics could explain the industrial structure of the energy industries, because they are rife with "specific" assets that leave buyers and sellers open to opportunistic behavior, thus providing an incentive for vertical integration. Closer examination, however, suggested that there were fewer of such asset types than sometimes thought and modest regulatory changes can all but eliminate the problem. Most importantly, high volume trading and a liquid market exist for crude oil, petroleum products, North American and U.K. natural gas, and Northern Europe electricity; such an observation is antithetical to the proposition that high transaction costs create the crucial incentive for vertical integration. Thus, the analysis became broader, an examination of why energy markets have been transformed.

8.1.3 Industrial Structure

Although this investigation began with the idea that transaction costs were a determinant of the size and structure of energy firms, it concludes that while such costs are an aspect of organizational structure, they are probably not the crucial one, particularly when there is time to adjust. Organizational structures are often ad hoc: in little more than two decades, the oil industry has undergone two major waves of mergers and acquisitions; the regulatory environment of the North American gas industry has completely changed and its power industry is poised for restructuring; the gas industry in the U.K. has been privatized and its electricity industry reorganized. These changes have accompanied, and arguably have been provoked, by major shifts in the cost of trading resources. The provocations have had a striking impact on the nature of wholesale energy markets, but with the exception of the U.K. electricity industry, relatively little impact on the typical structure of firms observed in the energy industries.

Fan (1999 pp. 2-3) succinctly stated the traditional view of transaction-cost economics on industrial structure and specifically vertical integration:

> The transaction cost theory of Williamson (1971, 1975, 1979) and Klein, Crawford and Alchian (1978) maintains that vertical integration is a response to asset specificity caused by specialized investment that has lower value outside a given transaction. If a contract is drawn to govern the transaction, the specialized investment creates an ex post bilateral bargaining situation in which opportunistic rent seeking, or holdup, by the transactors may occur. Vertical integration is proposed as a solution to the holdup problem, because the possibility of holdup is suppressed under the common ownership.

The foregoing explanatory approach to industrial structure has only limited relevance in the oil, gas, and electricity industries. There are economies of scale in oil and gas pipelines and high-voltage transmission grids. As a consequence, in some circumstances the control of key transportation infrastructure (or fear of someone else's control) may be a motive for integration. This impact cannot, however, be substantial. Most oil refineries are now located at tidewater, where they can be served by the highly flexible system of marine tankers. Further, most oil pipelines are now regulated, guaranteeing access. Likewise, the North American gas pipeline system is extensively interconnected and regulatory authorities guarantee access to third parties. In most cases movements to liberalize electricity and gas markets have begun with the concept of ensuring open access to transportation facilities, thus freeing up competition in the supply sector.

Asset specificity does not appear to play a substantial role in the structure of the energy industries, but a parallel concept – component products – is significant. Because both producers and consumers must make specialized investments in inflexible assets, energy demand and supply schedules are highly price-inelastic in the short term. As a consequence, energy products have high price volatility. Price fluctuations are comparable to agricultural products where unpredictable price swings are due to exogenous variation in weather, crop diseases, etc.

The drive to moderate price volatility has been a powerful motive for oil companies to integrate into upstream and downstream sectors. Simultaneous investments in crude oil producing and refining assets are thought to be a natural hedge in profitability and returns in the two sectors are often inversely correlated. When OPEC asserted control over prices, however, it broke up the integrated structure; refiners were suddenly vulnerable to radical price movements. Vertical integration is not the only way to solve the problem of price volatility. Commodity markets combine the physical trading of the commodity with opportunity for price hedging by means of forward transactions and financial swaps. Thus, OPEC's intervention into the upstream crude-oil market created necessary and sufficient incentives for the industry to devise an alternative market structure.

It could be argued that the heterogeneity of crude oil and variation in refining capabilities creates a specific-asset problem. However, refinery upgrades aimed at increasing the proportion of gasoline and middle distillates have simultaneously increased flexibility in the number and types of crude oils that can be refined. And, as pointed out in Chapter 4, while Major integrated oil companies own both refining and producing assets, there is frequently no physical connection between the operations; they trade for the oil they refine.

The structure of the North American natural gas industry has been largely determined by intensive regulation since the 1930s. Since then an extensive gas interstate transmission system has been constructed that interconnects hundreds of gas producers, pipelines, and consumers. Pipelines are classic specific assets in that they have little alternative use, exhibit strong economies of scale (limiting competition,) and are location specific. However, regulatory changes that grant third-party access to the pipeline infrastructure have effectively eliminated the problem of specific assets – as long as gas producers have guaranteed access to transmission pipelines at regulated rates they have little or no motive to integrate, unless there are other economic advantages.

Since gas producers and pipeline companies were effectively barred from integrating, the North American gas industry traded gas though long-term contracts, with extensive regulatory oversight. This structure could not, however, endure the dramatic change in energy prices of the 1970s. Consequently, as in the oil market, a classic commodity market structure arose. In the new structure the bulk of the gas trading shifted to month-long and daily spot pricing. Price volatility was managed separately through futures contracts and over-the-counter swaps. The rapid growth of the futures market contributed to declining costs of gas trading, which further reduced the incentive to integrate.

Relative to the natural gas industry, the North American electric industry is by and large vertically integrated. In contrast, when the U.K. restructured its nationalized power industry, policy makers chose to separate it into three principal sectors: distribution, grid management, and generation. In retrospect, it is clear that there were two errors in the initial organization. First, there were too few suppliers (an effective duopoly). Second, in contrast to the oil and gas commodity markets, the structure was designed "inside-out." That is, a transparent auction was used for the daily trading of the physical commodity, while opaque over-the-counter financial transactions were expected to moderate price risk. Because daily prices were volatile and the "derivative" market in contracts for differences was illiquid, the structure created an incentive for distribution companies and generators to merge rather than trade; by 2002 eighty-three percent of the distribution companies were owned by generators. In contrast, transparent prices and increasing liquidity obviate the need for vertical integration. Nord Pool integrated forward, futures, and daily trading into its electricity exchange. Combining financial and physical trading into a single organization helps reduce transaction costs in electricity trading, because settlement accounting is especially complex and costly. In addition, the Scandinavian power industry is not highly integrated. As a consequence, Nord Pool has been successful in creating a liquid and transparent market, while the U.K. and many other experimental markets have so far faltered.

8.1.4 Market Transformation

When OPEC wrested control over the world's most prolific oil reserves, it need not have provoked the development of a commodity market. Indeed, the companies' preference, and initial response, was to replace vertical integration with long-term contracts within an evolving system of administered prices. The Iranian Revolution, however, created a worldwide oil-supply shortage and enticed most OPEC members into opportunistic behavior, destroying the confidence of crude-oil buyers in long-term transactions, and significantly increasing transaction costs. The high cost of buyers and sellers searching out each other and striking a deal stimulated the development of a wide array of new market institutions: e.g., trade-press price reporting, standardized contracts, and two new futures contracts on exchanges in New York and London. In retrospect the speed at which the industry restructured itself is remarkable; all of the fundamental institutions of the new market changes emerged between 1979 and 1983.

Buyers and sellers in the North American natural gas market faced a comparable situation to that faced by oil traders. Local and seasonal shortage episodes had been becoming more widespread and longer in duration from the early 1970s, and peaked in the winter of 1976-77 as rising oil prices created massive distortions in the regulated natural gas market. Unlike the oil industry, U.S. regulations effectively deprived gas producers, pipelines, and distribution companies of any expectation of price or supply benefits from integrating. Thus, the breakdown of inflexible contracts created the demand for an alternative market structure. In the crude-oil market the main barrier to trading had been the high cost of information. In the natural gas market, the problem was enforcement. Many of the high-priced contracts transmission pipelines had entered into as resellers in the late 1970s and early 1980s, had been premised on growing demand that did not materialize, and consequently, they were threatened with bankruptcy.

Ultimately, over about a decade, federal regulators allowed contracts to be modified and prices to be deregulated in a new market framework that offered buyers and sellers third-party access to the pipeline infrastructure. Pipelines became transportation companies instead of merchants. This transformation took longer than that observed in the oil industry, but the change has been every bit as radical. The North American natural gas market is now considered one of the most efficient commodity markets in the world.

Transformation of electricity markets is, at best, incomplete. The U.K. initiated the trend in 1990 when it restructured its market. The fundamental revisions of the market structure undertaken from 1998 through 2000 were aimed at making it more like a traditional commodity market. So far, however, the transaction costs of trading are exceptionally high and futures contracts have failed. Restructuring in North America, if anything, has been less successful. The breakdown of the California power market is well known. Less well known is that before restructuring California was already part of a functioning integrated wholesale electricity market covering western U.S. States and Canadian provinces. The California crisis inhibited development of what might have been a viable prototype for electricity trading. Elsewhere, power pools in the eastern U.S. are meeting with some success, but the collapse of the eastern grid on August 14, 2003 raises a host of unanswered questions. Moreover, electricity futures markets established throughout the country have yet to succeed. So far, Nord Pool has emerged as the most viable type of market structure, but in many important respects it differs from traditional commodity market models.

8.1.5 Differences in Energy Commodities

Energy commodities have very different characteristics with respect to their natural form, cost of storage, transmission systems, handling costs, etc. Myopic regulation and inappropriate institutional structures may explain why the transaction cost of trading electricity has remained high and the market has not developed as intended. But an equally compelling explanation is the characteristics of the commodity itself. Electricity is far and away the most complex of the energy commodities. The uncertain nature of grid congestion combined with the lack of suitable energy-storage technologies or facilities requires system operators to match demand and supply continuously, create extraordinary problems of grid management and seem to require involvement of system operators in purchase and sale decisions.

Data on energy markets are inadequate to assess the impact of electricity's characteristics on trading costs, but they are adequate to analyze the impact on natural gas and oil. In North America both of these markets are at about the same level of maturity and the volume of trading is comparable. Nonetheless, the transaction costs of trading natural gas are substantially higher than oil. Natural gas is homogenous as compared to crude oil, so this should lower the transaction costs of trading. On the other hand, oil is significantly more mobile than natural gas. It can be moved in a variety of containers and transportation systems. Natural gas is most economically delivered (except in bulk shipments for inter-continental distances) in a stationary pipeline network. The inability to shift gas cheaply by alternative means from one node to another reduces liquidity (and raises transaction costs) at trading points in the network.

A large number of studies in the financial literature have determined that transaction costs are positively correlated to price volatility and negatively correlated to the volume of trade in securities exchanges. Chapter 7 extends this analysis to commodity markets. In gas and oil markets, data on the difference between buy and sell prices are available for various trading points in the physical market, but not for futures exchanges. The volume of sales for each trading point is not available, but can be approximated using dummy variables. The statistical analysis conducted for Chapter 7 concludes that transaction costs are a function of the commodity (gas or oil), basis risk (variation in the difference between local prices and the futures exchange marker price) and the approximation of volume.

8.1.6 Summary Conclusions

In summary the conclusions of this thesis are:

- The proposition from transaction-cost economics that the avoidance of opportunistic behavior is an important motive for vertical integration is not supported for important segments of the energy industry.

- High transaction costs were, however, a primary motivating factor in the change in market institutions that govern the wholesale trading of energy products, particularly for oil and natural gas.

- The physical characteristics of a commodity matter with respect to the transaction costs of trading and the optimum market structure.

8.2 Contributions to the Field of Energy Economics

The following lists and describes the contributions of this thesis to the field of energy economics:

- Chapters 4, 5, and 6 of the thesis contain a narrative description of the development of oil, gas, and electricity markets since 1973. Other authors have completed descriptions of these industries but this is the first attempt to compare and contrast the development of the three major energy markets in a systematic fashion. Chapter 4 begins with a description of the oil market at the point that OPEC asserted its control over the pricing and disposition of crude oil. It explains that the market structure that replaced vertical integration was inefficient and laden with high transaction costs. This provoked major changes and the oil industry adopted a commodity market structure similar to the agricultural and financial industries. Chapter 5 describes how the North American natural gas industry was faced with similar convulsive changes and adopted a market structure similar to that of the oil industry. Traders and pundits anticipated a third wave of reform, which would make similar changes in the electricity markets, but Chapter 6 describes how the characteristic of the commodity may cause a different type of structure to emerge.

- Chapter 3 of the thesis examines the definition of transaction cost and its development in academic literature. The chapter identifies and analyzes various commodity characteristics that impact transaction costs. The analysis builds on analytical work from institutional economics and the analysis of commodity markets.

- Chapter 3 of the thesis also extends the analysis of asset specificity, which is frequently used to explain firms' vertical integration, to a parallel concept: "component products." All energy products are component products – where both sellers and buyers must make specialized investments to produce or use a specific type of energy commodity. The resulting price inelasticity of both demand and supply creates special problems for the energy industry that impact transaction costs and market structures.

- The thesis identifies and quantifies the transaction costs of trading energy commodities, using a definition developed for the securities industry. This definition reveals that transaction costs of trading energy have changed over time and that they vary from one commodity to another.

- Chapter 7 analyzes the impact of differences in oil and gas commodity characteristics on the transaction costs of trading in North America. It extends the analysis completed by other analysts on securities markets. It concludes that the type of commodity, level of trading activity, and the extent of basis risk are the key determinants of the transaction costs of trading commodities.

Bibliography

Adelman, M.A. (1972): *The World Petroleum Market*, John Hopkins University Press.

Adelman, M.A. (1995): *The Genie out of the Bottle, World Oil since 1970*, The MIT Press.

Akerlof, G. (1970): "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, pp. 488-500.

Antill, N., (1999) "Vertical Integration and Gigantism," *Oxford Energy Forum*, Issue 36, pp. 3-4.

Barrera-Rey, F. (1995): "The Effects of Vertical Integration on Oil Company Performance," Oxford Institute for Energy Studies

Black, D. G., (1986): "Success and Failure of Futures Contracts: Theory and Empirical Evidence," Solomon Brothers Center for the Study of Financial Institutions, NYU, Monograph Series in Finance and Economics: Monograph 1986-1.

Blackmon, G. (1985): "A Futures Market for Electricity Benefits and Feasbility," Harvard Energy & Environmental Policy Center, Discussion Paper, E-85-07.

Blair, J. M. (1978): *The Control of Oil*, Vintage Books.

Bower, J. (2002): "Why Did Electricity Prices Fall in the England and Wales?" Oxford Institute for Energy Studies

Bradley, R. (1989): *The Mirage of Oil Protection*, Cato Institute, University Press of America.

British Petroleum (2003): *BP Statistical Review of World Energy 2003*, BP Website.

Boiteux, M. (1949): "La tarification des demand en point: application de la theorie de la vent au cout marginal," Reveue Géneral de Electricité, 321-40, translated as "Peak-Load Pricing," *Journal of Business*, v. 12, 298, 33, 157-7.

Buchanan, J. (1986): *Liberty, Market and State Political Economy in the 1980s*, New York University Press.

Bushnell, J. and Blumstein, C. (1994): "A Reader's Guide to the Blue Book: Issues in California's Electric Industry Restructuring and Regulatory Reform," University of California Energy Institute, Berkeley, California.

Campbell, C.J. (1997): *The Coming Oil Crisis,* Multi-Science Publishing Company and Petroconsultants S.A.

Carlton, D. (1984): "Futures Markets: Their Purpose, Their History, Their Growth, Their Successes and Failures," The Journal of Futures Markets, vol 1, no. 3, pp 237-71.

Chalker, J. (2003): "The European Union's Security of Energy Supply and Russian Gas," International Association for Energy Economics (IAEE) 26th International Conference Proceedings.

Cheshire, J. (1996): "U.K. Electricity Supply and Public Ownership," *The British Electricity Experiment Privatization: the Record, the Issues, the Lessons*, John Surrey, Earthscan Publications, London, pp-14-39.

Cheung, S. (1998): "The Transaction Cost Paradigm," 1998 Presidential Address, Western Economic Association, *Economic Inquiry*, vol. 36, issue 4, pp. 514-21.

Cicchetti, C. J., Gillen, W. J., Smolensky, P. (1977): *The Marginal Cost and Pricing of Electricity, An Applied Approach,* Report to the National Science Foundation, Ballinger Publishing Company, Cambridge Massachusetts.

Coase, R. H. (1937): "The nature of the firm," *Economica*, 4: 386-404.

Coase, R. H. (1960): "The problem of social cost," *Journal of Law and Economics*, 3: 1-44.

Coase, R. H. (2000): "The acquisition of Fisher Body by General Motors," *Journal of Law and Economics*, vol. 43, 15-31.

Collins, R. (2003): "New PJM Electricity Futures Contracts," *Energy in the News,* NYMEX Volume 2, pp 2-6.

Cooter, R. and Ulen, T. (1996): *Law and Economics*, Second Edition, Addison-Wesley.

Demsetz H. (1968): "The Cost of Transacting," *Quarterly Journal of Economics* vol. 82, 33-53.

Dietrich, M. (1994): *Transaction Cost Economics and Beyond*, Routledge.

Energy Information Administration (EIA 2001): *U.S. Natural Gas Markets: Midterm Prospects for Natural Gas Supply*, EIA Website.

Enron Corporation (1997): *Annual Report*, Enron

Evans, J. and Green, R. (2003): "Why did British Electricity Prices fall after 1998?" Draft Paper, University of Hull.

Fan, J. (1999): "Price Uncertainty and Vertical Integration: An examination of Petrochemical Firms," Hong Kong University of Science and Technology, School of Business and Management, unpublished paper.

Fesharaki, F. and Rzavi, H. (1986): *Spot Oil, Netback and Petroleum Futures*, Economist Intelligence Unit, London.

Frankel, P. H. (1969): *Essentials of Petroleum*, Frank Cass & Co. Ltd.

General Accounting Office (GAO 1980): *The United States Exerts Limited Influence on the International Crude Oil Spot Market*, Report to the Congress by the Comptroller General.

Granitz, E. and Klein, B. (1996): "Monopolization by 'Raising Rivals' Costs': The Standard Oil Case," *Journal of Law Economics,* vol. XXXIX (April 1996).

Grossman, S. J. and Miller, M. H. (1988): "Liquidity and Market Structures," *The Journal of Finance*, vol. 43, no. 3, 617-633.

Gordon, R.L. (1974): "Optimization of Input Supply Patterns in the Case of Fuels for Electric Power Generation," *The Journal of Industrial Economics,* Vol. 23, No 1, 19-37

Gordon, R.L (1979): "The Powerplant and Industrial Fuel Use Act of 1978—An Economic Analysis," *Natural Resources Journal*, vol 19, 871-884.

Gordon, R.L. (1992): "The Public Utility Holding Company Act," *Cato Review of Business and Government*, Winter, 58-65.

Gordon, R.L. (2001): "Don't Restructure Electricity; Deregulate," Cato Journal, Winter, 327-59.

Henney, A. (1994): *A Study of the Privatisation of the Electricity Supply Industry in England & Wales*, EEE Limited, London.

Ho, T. and Stoll, H. R. (1980): "On Dealer Markets Under Competition," *The Journal of Finance*, vol. 35, no. 2, 259-267.

Houthakker, H. S., 1959, "The Scope and Limits of Futures Trading," in Abramovitz, Moses, et al., The Allocation of Economic Resources, Stanford: Stanford University Press, 134-59.

Hull, J.C., (1993): *Options, Futures, and other Derivative Securities*, Prentice-Hall, Englewood, N.J.

International Energy Agency, (IEA 2001): *Regulatory Institutions in Liberalized Energy Markets,* Organization for Economic Cooperation and Development, Paris.

Johnson, W.A., Messick, R.E., Van Vactor, S.A., Wyant, F.R (1976): *Competition in the Oil Indusry*, George Washington University.

Joskow, P. (1997): "Restructuring, Competition and Regulatory Reform in the U.S. Electricity Sector," *Journal of Economic Perspectives*, 11:3, pp. 119-138.

Klein, B.; Crawford, R.A.; Alchian, A.A. (1978): "Vertical integration, appropriable rents, and the competitive contracting process," *Journal of Law and Economics*, vol. 21: 297-326.

Klein, B. (2000): "Fisher-General Motors and the nature of the firm," *Journal of Law and Economics*, vol. 43, 105-141.

Lee, R. (1998): *What is an Exchange?* Oxford University Press.

Lehr, D. and Van Vactor, S. A.: (1997) "Evolution of Wholesale Power Price in the Western Power Market," *The U.S. Power Market*, Risk Publications, pp. 233-250.

Locke, P. and Venkatesh, P. C. (1997): "Futures Market Transaction Costs," *The Journal of Futures Markets*, vol. 17, no. 2, 229-245.

London Economics (1997): "Overview of Nord Pool," available on: http://faculty-gsb.stanford.edu/wilson/archive/E542/classfiles/nordpool.pdf

Lovins, A. (1976): "Soft Energy Paths," *Foreign Affairs.*

Mabro, R., (1992): "OPEC and the Price of Oil," *The Energy Journal*, Volume 13, Number 2.

Mabro, R., (1975-76): "OPEC after the Oil Revolution," *Millennium*, London School of Economics Journal of International Studies, 191-199.

Mango, B., Woodley, J. (1994): "The Inevitable Commoditisation of Electric Power Markets," *Public Utilities Fortnightly*, February 1.

Millman, G.J. (1993): "Futures and Options Markets," *The Fortune Encyclopedia of Economics*, Time, Inc., D.R. Henderson, Editor.

McGee, J. S. (1958): "Predatory Price Cutting: The Standard Oil (N.J.) Case." *Journal of Law and Economics* 1: 137-69.

Mitchell, J., (1999): "Turns Integration 90°," *Oxford Energy Forum*, Issue 36, pp. 7-9.

Mulcare, A. (2002): "The Impact of the Belgium-U.K. Interconnector," www.interconnector.com.

Mulherin, J. Harold (1986): "Specialized Assets, Governmental Regulation, and Organizational Structure in the Natural Gas Industry," *Journal of Institutional and Theoretical Economics* (142/3).

Neal R. (1992): "A Comparison of Transaction Costs between Competitive Market Structure and Specialist Market Structures," *Journal of Business*, vol. 65, no. 3, 317-334.

Newbery, D. (2003): PowerPoint Presentation at the IAEE Meetings, Prague, June.

Newbery, D. and McDaniel, T. "Auctions and trading in energy markets – an economic analysis," CRI Regulatory Review 2002/2003, pp. 1-25.

Nord Pool (1997): "The Spot Market," www.nordpool.no

Nord Pool (2003): "The Nordic Power Market – Transition toward a Unified Nordic Power Market" and "History of Nord Pool's Financial Markets," www.nordpool.no

North, D. C., (1997): "Transaction cost through time," *Transaction Cost Economics: Recent Developments*, Menard, C. and Elgar, E.

OECD (1976): *Energy Balances of OECD Countries,* OECD, Paris.

O'Hara, M. (1995): *Market Microstructure Theory*, Blackwell Publishers, Inc. Cambridge, Mass.

Oil and Gas Journal, (2001): December 24.

OPEC Secretariat (1986): *OPEC Annual Report 1985*.

OPEC Secretariat (1980-83): Press releases as published in the *OPEC Bulletin*.

Outhred, H. (2000): "The Competitive Market for Electricity in Australia: Why it Works so Well," *Proceedings of the 33rd Hawaii International Conference on system Science*.

Parker, M. (1996): "Effects on Demands for Fossil Fuels," *The British Electricity Experiment Privatization: the Record, the Issues, the Lessons*, John Surrey, Earthscan Publications, London, pp-120-137.

Safer, A. (1979): *International Oil Policy*, Lexington Books, Lexington, Mass.

Sampson, A. (1975): *The Seven Sisters,* Bantam Books – Viking Press.

Schmidt, Paul F. (1951): *Fuel Oil Manual,* The Industrial Press, New York, New York.

Schweppe, F.C., Tabors, R. D., Kirtley, J. L., Outhred, H. R., Cox, A. J. and Pickel F.H. (1980): "Homeostatic Utility Control," *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS 99, No. 3

Schweppe, F. C., Caramanis, M. C., Tabors, R. D., and Bohn, R. E., (1988): *Spot Pricing of Electricity*, Kluwer Academic Publishers.

Silber, W. L., (1981): "Innovation, Competition, and New Contract Design in Futures Markets," *The Journal of Futures Markets*, 2: 123-167.

Smith, A. (1776): *The Wealth of Nations,* Random House Modern Library Edition, 1937.

Smith, T. and Whaley, R. E. (1994): "Estimating the Effective Bid/Ask Spread from Time and Sales Data," *The Journal of Futures Markets*, vol. 14, no. 4, 437-455.

Stern, J. P. (1998): *Competition and Liberalization in European Gas Markets,* The Royal Institute of International Affairs.

Stigler, G. (1951): "The Division of Labor is Limited by the Extent of the Market," *The Journal of Political Economy*, Vol. 59, No. 3, pp. 185-193.

Stoft, S. (2002): *Power System Economics, Designing Markets for Electricity,* John Wiley and Songs.

Stoll, R. H., and Whaley, R. E., (1990): "Stock Market Structure and Volatility," *The Review of Financial Studies*, Volume 3, Issue 1, 37-71.

Teece, D. J. (1976): "Vertical Integration in the U.S. Industry," in *Vertical Integration in the Oil Industry*, Edited by Edward J. Mitchell, pp. 105-189.

Thomas, S. (1996): "The Privatization of the Electricity Supply Industry," *The British Electricity Experiment Privatization: the Record, the Issues, the Lessons*, John Surrey, Earthscan Publications, London, pp-40-66.

Treat, J. E. (2000): *Energy Futures Trading Opportunities*, Penwell Books.

Turvey, R. and Anderson D., (1977): *Electricity Economics, Essays and Case Studies,* World Bank, The John Hopkins Press

Tussing, A. R. and Tippee, B. (1995): *The Natural Gas Industry*, second edition, Penwell Books.

Tussing, A. R. and Hatcher, D. (1994): "Prospects for an Electricity Futures Market: Lessons from Petroleum and Natural Gas," *Resources Policy*, 20 (2) p. 135.

Tussing, A. R. (2004): Direct correspondence with the author.

van Oostvoorn, F. and Boots, M.G. (1999): Impacts of Market Liberalization on the EU Gas Industry, Conference paper.

Van Vactor, S.A. and Tussing, A. (1987): "Retrospective on Oil Prices," *Contemporary Policy Issues*, Volume V, Number 3, pp. 1-19.

Van Vactor, S. A., (1997): "Time to End the Alaskan Oil Export Ban," CATO Institute, Policy Analysis, No. 227.

Van Vactor, S. A., (1998): "Electricity restructuring in the United States," *Financial Times Energy Economist*, November.

Van Vactor, S. A., (2000) "East vs. West: Comparing Electric Markets in California and PJM," *Public Utilities Fortnightly*, July 15, 2000.

Van Vactor, S. A., and Blumsack, S. B., (2002): "How to Make Power Markets Competitive," *Electricity Pricing in Transition,* editors A. Faruqui and B.K. Eakin, Khuwer Academic Publishers.

Verleger, Philip K., (1987): "The Evolution of Oil as a Commodity," *Energy Markets and Regulation*, edited by R.L. Gordon, H.D. Jacoby, and M.B. Zimmerman, MIT Press.

Williamson, O. E. (1985): *The Economic Institutions of Capitalism*, The Free Press.

Wolkoff, Neal, Executive Vice President, NYMEX: Testimony before the New York Senate, February 1, 2000.

World Bank (2003): *Infrastructure Regulations Promises, Perils and Principles*, Draft Report.

Yergin, D. (1991): *The Prize, The Epic Quest for Oil, Money, and Power*, Simon & Schuster.

Acronyms

AC	Alternating Current
AECO	Alberta Energy Company
ANGTS	Alaska Natural Gas Transportation System
API	American Petroleum Institute
APIG	API Gravity
APX	Automated Power Exchange
Aramco	Arabian American Company
BG	British Gas
BPA	Bonneville Power Administration
BT	British Telephone
Btu	British thermal units
CBOT	Chicago Board of Trade
CC	Combined Cycle
CEC	California Energy Commission
CEGB	Central Electricity Generating Board
CEO	Chief Executive Officer
CFD	Contracts for Differences
CFTC	Commodity Futures Trading Commission
CHP	Combined Heat and Power
CPUC	California Public Utility Commission
CT	Combustion Turbine
DC	Direct Current
DOJ	Dept. of Justice
DTE	Developing and Transition Economy
EC	European Community
EEI	Edison Electric Institute
EEX	European Energy Exchange
EIA	Energy Information Administration
EIS	Environmental Impact Statement
EPRI	Electric Power Research Institute
ERA	Energy Regulatory Administration
EU	European Union
FERC	Federal Energy Regulatory Commission
FPC	Federal Power Commission
FSU	Former Soviet Union
FTC	Federal Trade Commission
IEA	International Energy Agency
IPE	International Petroleum Exchange
IPP	Independent Power Producer
ISDA	International Swaps and Derivatives Association
ISO	Independent System Operator
LADWP	Los Angeles Dept. of Water and Power
LMP	Locational Marginal Prices
LNG	Liquid Natural Gas
LPG	Liquefied Petroleum Gas
LSE	Load-Serving Entity

MIT	Massachusetts Institute of Technology
MOIP	Mandatory Oil Import Program
MW	Megawatt
MWh	Megawatt-hour
NEA	National Energy Act
NEPOOL	New England Power Pool
NERC	North American Electric Reliability Council
NETA	New Electricity Trading Arrangements
NGA	Natural Gas Act
NGC	National Grid Company
NGPA	Natural Gas Policy Act
NGX	Natural Gas Exchange
NSHEB	North of Scotland Hydro-Electric Board
NYCE	New York Commodity Exchange
NYMEX	New York Mercantile Exchange
NYPP	New York Power Pool
OECD	Organization for Economic Cooperation and Development
OM	Scandinavian Electronic Exchange Co.
OPEC	Organization of Petroleum Exporting Countries
OTC	Over-the-Counter
PG&E	Pacific Gas and Electric Co.
PGT	Pacific Gas Transmission
PJM	Pennsylvania-New Jersey-Maryland Hub
PUC	Public Utility Commission
PUHCA	Public Utility Holding Company Act
PURPA	Public Utility Regulatory Policy Act
PX	Power Exchange
QF	Qualifying Facility
REC	Regional Electric Company
RTO	Regional Transmission Organization
SCE	Southern California Edison
SDG&E	San Diego Gas & Electric Co.
SEC	Securities and Exchange Commission
SG	Specific Gravity
SMP	System Marginal Price
SMUD	Sacramento Municipal Utility District
SSEB	South of Scotland Electricity Board
TEOR	Thermally Enhanced Oil Recovery
TOCOM	Tokyo Commodities Exchange
TPA	Third Party Access
TSO	Transmission System Operator
WSPP	Western Systems Power Pool
WTI	West Texas Intermediate