ESSAYS IN FINANCIAL TRANSMISSION RIGHTS PRICING

A Thesis in
Energy, Environmental and Mineral Economics

by
Barry Posner

© 2006 Barry Posner

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2006
The thesis of Barry Posner was reviewed and approved* by the following:

Andrew N. Kleit  
Professor of Energy, Environmental and Mineral Economics  
Thesis Advisor  
Chair of Committee

Richard L. Gordon  
Professor Emeritus of Mineral Economics

Adam Z. Rose  
Professor of Geography

Anthony M. Kwasnica  
Associate Professor of Business Economics

Joseph Doucet  
Associate Professor of Regulatory Economics  
University of Alberta  
Special Member

Robert G. Crane  
Professor of Geography  
Interim Dean of the College of Earth and Mineral Sciences

*Signatures are on file in the Graduate School
ABSTRACT

This work examines issues in the pricing of financial transmission rights in the PJM market region. The US federal government is advocating the creation of large-scale, not-for-profit regional transmission organizations to increase the efficiency of the transmission of electricity. As a non-profit entity, PJM needs to allocate excess revenues collected as congestion rents, and the participants in the transmission markets need to be able to hedge their exposure to congestion rents. For these purposes, PJM has developed an instrument known as the financial transmission right (FTR). This research, utilizing a new data set assembled by the author, looks at two aspects of the FTR market.

The first chapter examines the problem of forecasting congestion in a transmission grid. In the PJM FTR system firms bid in a competitive auction for FTRs that cover a period of one month. The auctions take place in the middle of the previous month; therefore firms have to forecast congestion rents for the period two to six weeks after the auction. The common methods of forecasting congestion are either time-series models or full-information engineering studies. In this research, the author develops a forecasting system that is more economically grounded than a simple time-series model, but requires less information than an engineering model. This method is based upon the arbitrage-cost methodology, whereby congesting is calculated as the difference of two non-observable variables: the transmission price difference that would exist in the total absence of transmission capacity between two nodes, and the ability of the existing transmission to reduced that price difference. If the ability to reduce the price difference is greater than the price difference, then the cost of electricity at each node will be the same, and congestion rent will be zero. If transmission capacity limits are binding on the flow of power, then a price difference persists and congestion rents exist.

Three transmission paths in the Delmarva Peninsula were examined. The maximum-likelihood two-way Tobit model developed in Chapter One consistently predicts the expected responses to the independent variables that have employed, but the model as defined here does a poor job of predicting prices. This is likely due to the inability to
include system outages (i.e., short-term changes in the structure of the transmission grid) as variables in the estimation model.

The second chapter addresses the behavior of firms in the monthly auctions for FTRs. FTRs are a claim to congestion rent revenues along a certain path within the PJM grid, and are awarded in a uniform-price divisible-goods auction. Firms typically submit a schedule of bids for different amounts of FTR at different prices, akin to a demand curve. A firm bidding too high a price may cause the clearing price of the FTR to be higher than the realized value of the FTR, creating a loss from ownership of the FTR. A firm bidding too low means that it wins no FTRs, depriving itself of the ability to profit from ownership or to hedge against congestion. Several questions concerning firm behavior are addressed in this study. It is found that firms adjust their bids in response to new information that is obtained from past auctions: they raise or lower bids in accordance with changes in recent FTR prices and payoffs. Firms consistently bid below the value of the FTR (i.e., shade their bids.) This adds empirical evidence to the theoretically-posed notion that uniform-price auctions are not truth-telling, unlike the second-price auction for a non-divisible good. Firms employ greater bid-shading in response to increases in the volatility of both FTR clearing prices and realized FTR values. This validates the notion that firms are risk-averse. It is discovered that better-informed “insider” firms employ structurally different bidding strategies, but these differences do not lead to greater profits. However, profits do increase as firms gain more experience in these markets, lending credence to the notion that firms learn over time and that markets discipline poorly performing firms by either educating them or driving them out of the market. It is also found that firms that employ complicated bidding strategies enjoy greater profitability than firms which employ simple bidding strategies. A surprising corollary finding is that firm strategies do not converge to a common form, but that different firms continue to employ different strategies, and often move away from the seemingly dominant strategy. Firms can enter this market as either long-buyers or short-sellers, and it is discovered that long and short players display structurally divergent bidding strategies. This is perhaps unsurprising, given that long players can be either hedgers or speculators, but short players are overwhelmingly speculators.
# Table of Contents

List of figures vi
List of tables viii
Acknowledgements x

1 Modeling transmission congestion in the Delmarva Peninsula 1
   1.1 Introduction 1
   1.2 PJM’s locational marginal pricing model 3
   1.3 Congestion in transmission networks 9
   1.4 Congestion in the Delmarva Peninsula 14
   1.5 Literature review 24
   1.6 Economic model definition 27
   1.7 Econometric model definition 47
   1.8 Data 64
   1.9 Experimental procedure 73
   1.10 Results and diagnosis 76
   1.11 Conclusion 97
   1.12 References 98
   Appendix A: Network calculations 101

2 Bidder behavior in PJM FTR auctions 106
   2.1 Introduction 106
   2.2 Congestion in electricity transmission grids 108
   2.3 The PJM market for financial transmission rights 113
   2.4 Firm behavior in FTR auctions 117
   2.5 Literature review 134
   2.6 Economic model specification 153
   2.7 Econometric model specification 171
   2.8 Data 181
   2.9 Results and discussion 190
   2.10 Conclusion 200
   2.11 References 203
   Appendix B: Firm-level profitability rankings 212
# List of figures

1.1 PJM East utilities .......................... 4  
1.2 Weekly percentage of hours with single system-wide price .......................... 10  
1.3 Example network – low-load period ................................................................. 11  
1.4 Delmarva counties .................................................. 15  
1.5 Major eastern PJM generation ............................................................. 19  
1.6 Delmarva generation facilities ............................................................ 20  
1.7 Delmarva northern transmission facilities ............................................ 22  
1.8 Delmarva southern transmission facilities ............................................. 23  
1.9 Equilibria at two nodes .................................................. 27  
1.10 Equilibria with limited transmission capacity ........................................ 29  
1.11 Observed price difference with unobserved price components ............ 30  
1.12 Maximum, average and minimum hourly loads, PJM East ...................... 49  
1.13 Energy cost versus load, PJM East .................................................. 50  
1.14 Modeled supply and demand curves without transmission ................... 52  
1.15 Modeled supply and demand curves with transmission ........................ 52  
1.16 Supply curves at N and S nodes .................................................. 60  
1.17 Path 1 price (Wilmington – Dover) .................................................. 65  
1.18 Path 2 price (Dover – Salisbury) .................................................. 65  
1.19 Path 3 price (Salisbury – Eastville) .................................................. 66  
1.20 Four-hour mean temperature, New Castle Co. airport ......................... 69  
1.21 Maximum congestion price versus temperature ................................ 69  
1.22 Natural gas prices .................................................. 70  
1.23 Maximum congestion price versus natural gas price .......................... 71  
1.24 Estimated system components, Path 1 .................................................. 87  
1.25 Estimated system components, Path 2 .................................................. 88  
1.26 Estimated system components, Path 3 .................................................. 88  
A.1 Sample three-node network ........................................................................ 101  

2.1 PJM control area, May 2006 .................................................. 110  
2.2 Number of firms participating in monthly on-peak FTR auctions ......... 118  
2.3 Size of monthly on-peak FTR auction .................................................. 120  
2.4 Size of long monthly on-peak FTR market ............................................. 120  
2.5 Size of short monthly on-peak FTR market ............................................. 121  
2.6 Size of long market with and without outliers ........................................ 122  
2.7 Size of big player and residual short markets .......................................... 124  
2.8 Duration of firm participation in FTR market ........................................ 125  
2.9 Bid complexity, selected firms, part 1 .................................................. 126  
2.10 Bid complexity, selected firms, part 2 .................................................. 126  
2.11 Average monthly FTR profits .................................................. 129  
2.12 Firm-specific profit, long FTR market .................................................. 130  
2.13 Firm-specific profit, short FTR market .................................................. 131  
2.14 Herfindahl-Hirschman Index, long FTR market ..................................... 132  
2.15 Herfindahl-Hirschman Index, short FTR market ..................................... 133  
2.16 FTR market structure .................................................. 156
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.17</td>
<td>Discrete and continuously-modeled bids</td>
<td>158</td>
</tr>
<tr>
<td>2.18</td>
<td>Demand with different numbers of bidders</td>
<td>161</td>
</tr>
<tr>
<td>2.19</td>
<td>Mean correlation function of clearing price versus lagged FTR value</td>
<td>188</td>
</tr>
</tbody>
</table>
List of tables

1.1 Prices and welfare gains with and without congestion 13
1.2 Delmarva geographic and economic data 16
1.3 Congestion statistics 16
1.4 Delmarva generators 21
1.5 Installed generation capacity 49
1.6 Estimator parameters 63
1.7 Price data statistics 67
1.8 Degrees of congestion 67
1.9 Congestion correlation 67
1.10 Temperature data statistics 68
1.11 Natural gas price data statistics 70
1.12 Correlations for $\Delta P_N$ estimator variables 72
1.13 Path 1 (Wilmington - Dover) parameter estimates 76
1.14 Path 2 (Dover - Salisbury) parameter estimates 77
1.15 Path 3 (Salisbury - Eastville) parameter estimates 78
1.16 Reductive capacity constant term 79
1.17 SQRT parameter estimates 80
1.18 PP_DIFF parameter estimates 80
1.19 NG parameter estimates 81
1.20 ET parameter estimates 81
1.21 HTD and CTD parameter estimates 82
1.22 Autocorrelation parameter estimates 82
1.23 Positive and negative price streaks 83
1.24 Likelihood-ratio tests for autocorrelation 84
1.25 Path 1 impact analysis 86
1.26 Observed versus predicted prices, path 1 91
1.27 Observed versus predicted prices, path 2 91
1.28 Observed versus predicted prices, path 3 91
1.29 Skill scores 92
1.30 Forecast bias 92
1.31 Comparison of path 1 skill scores 96
A.1 Flows and prices, low-load period 103
A.2 Flows and prices, high-load period without line flow constraints 104
A.3 Flows and prices, high-load period with line flow constraints 105

2.1 Estimator parameters, bid price regression 174
2.2 Estimator parameters, bid size regression 177
2.3 Estimator parameters, clearing price regression 179
2.4 Summary statistics 185
2.5 Variable correlations, bid price regressions 186
2.6 Variable correlations, bid quantity regression 186
2.7 Variable correlations, clearing price regression 186
2.8 Parameter estimates, single-component bid price regressions 190
2.9 Parameter estimates, upper limit of bid price regressions 191
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.10</td>
<td>Parameter estimates, lower limit of bid price regressions</td>
<td>192</td>
</tr>
<tr>
<td>2.11</td>
<td>Parameter estimates, bid quantity regressions</td>
<td>195</td>
</tr>
<tr>
<td>2.12</td>
<td>Parameter estimates, clearing price regressions</td>
<td>196</td>
</tr>
<tr>
<td>2.13</td>
<td>Parameter estimates, value of FTR regression</td>
<td>198</td>
</tr>
<tr>
<td>2.14</td>
<td>Parameter estimates, winning percentage regressions</td>
<td>199</td>
</tr>
<tr>
<td>B.1</td>
<td>Firm-wise profitability rank index</td>
<td>213</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to take this opportunity to acknowledge the assistance of my thesis advisor, Dr. Andrew Kleit and the motivation provided by my boss, Dr. Hugh Murphy. I am also grateful for the patience of my employers in the completion of this document, the invaluable administrative assistance of Karen Royer from the School of Earth and Mineral Sciences, and the guidance and flexibility of Pauletta Leathers from the PSU Thesis Office.
1 Modeling transmission congestion in the Delmarva Peninsula

1.1 Introduction

Transmission refers to the long distance transportation of electricity at high voltages. It is one of the three core building blocks of an electricity system, the others being generation and distribution, where electricity is delivered at low voltages to end-users. Transmission is the vital link between the other two, given that direct generation-to-distribution systems of any size are rare. In the Mid-Atlantic region of the United States the transmission network is operated in a non-profit manner by an organization called PJM Interconnect, in accordance with federal regulations. A flat tariff is charged to users of the transmission grid. However, as with any transport medium, the grid can become crowded, and congestion can exist. Unlike a road system, congestion does not result in the slower movement of some product: since electricity is governed by immutable physical laws and is not easily stored, the amount transported is the amount generated and consumed in real-time, and the amount transported is subject to strict upper limits. Thus, sometimes there is competition for access to the electricity grid. PJM rations this limited access by giving electricity different prices at different places, based upon localized supply and demand and the availability of transmission.

Congestion creates significant difficulties for both load-serving entities and grid operators. It leads to price uncertainties for load serving entities and is a source of major operational inefficiencies for system operators, who are tasked with operating the electricity grid in a socially optimal manner. Additionally, as not-for-profit entities, the system operators have to redistribute any congestion rents collected to system users. To provide both a redistribution mechanism and price-risk mitigation tool, PJM sells access to the collected congestion rents by a device called a Financial Transmission Right (FTR). FTRs are sold in annual and monthly uniform price auctions to any interested PJM members firms. Chapter 2 of this study examines the FTR market in greater detail. The amount of congestion rent in a given month defines the value of an FTR. Like all financial instruments, an accurate mechanism for predicting the value of the FTR can
provide a firm with a competitive advantage in the FTR market. The majority of price-prediction algorithms are time-series based, in that the value of some variable at a certain time is assumed to be some function of previous values of that variable. A weakness of time-series prediction models is that they are strongly dependent on the starting value and usually converge to some steady state as the time span of the forecast increases. Thus, they are not particularly useful for predicting values that vary greatly over short periods of time, such as congestion rents over a link in electricity network, which can spike from zero to $200/MWh and back in short periods of time. In this essay, an alternative model for predicting the causes and levels of transmission congestion is posited and tested. An econometric model derived from the theory of arbitrage costs is specified, examining the cost of congestion in a transmission line as a function of two variables: the expected price difference in the absence of any transmission between the two locations and the ability of the existing transmission to reduce that price difference. If the ability to reduce costs is greater than the price difference, the cost of congestion will be zero. If it is less than the expected price difference, we will see congestion. Both of these components are unobservable in reality, but are modeled using readily-observable public information: electricity prices, generation fuel prices and weather conditions. In addition to assisting in the pricing of FTRs, such a model enables an estimation of the shadow value of transmission, which lends a yardstick towards the evaluation of the benefits of expansion of transmission capacity.

This essay is divided into the following sections. Section 2 is an overview of the PJM locational marginal pricing model. Section 3 contains a discussion of congestion, with an example of how congestion affects prices. Section 4 examines historical congestion in the Delmarva Peninsula. Section 5 provides a review of the literature in this area of study. Section 6 contains a definition of the economic model used to define transmission costs, and section 7 details how this model is translated into an econometric specification. In section 8, the data used in this study are described. The procedure used to obtain the results is presented in section 9, and the results are presented, diagnosed, and analyzed in section 10. Conclusions are drawn in section 11.
1.2 PJM’s locational marginal pricing model

PJM Interconnect L.L.C. is a not-for-profit corporation tasked with operating electricity markets in the Mid-Atlantic region of the US. PJM began in 1927 as a reserve-sharing agreement amongst three Philadelphia-area utilities: Philadelphia Electric Company (PECO), Public Service Electric and Gas (PSEG), and Pennsylvania Power and Light (PPL). What was then called the PA-NJ Agreement has since expanded to include several other utilities in Pennsylvania, New Jersey, Maryland and Delaware. Figure 1.1 details the scope of PJM before its April 1, 2002 absorption of Allegheny Power.

The area covered in Figure 1.1 is currently referred to as “PJM East,” as PJM has recently expanded to include utilities serving all or parts of Pennsylvania, West Virginia, Virginia, Ohio, Indiana, Michigan and Illinois. The topic of this essay is the Delmarva Peninsula, so PJM East is the only area of concern. More information about the current status of PJM’s expansion plans can be found at its website, www.pjm.com. PJM East comprises 540 generation units, over 8,000 miles of transmission lines and serves almost 10% of the US population.

Throughout the 1990s, the Federal Energy Regulatory Commission (FERC) promoted market-based reforms to the North American electricity infrastructure. In 1996, FERC issued Order No. 888 (FERC, 1996), which governed the formation of Independent System Operators (ISOs). In the past, transmission lines were owned and operated by vertically integrated utilities largely to serve each utility’s native load. With the rise of long-distance energy trading (retail wheeling), it became more important to have broader scale control of transmission networks: the existence of many locally-operated transmission networks was an obstacle to long-distance trade, as each
transmission company could charge a tariff to the trading partners – a practice known as “pancaking.” FERC promoted the model of ISOs, which are non-profit companies formed for the purpose of operating transmission networks in a coordinated fashion over a region that typically encompasses several smaller transmission companies. FERC defined the following fundamental characteristics of an ISO:

*Independence: An ISO must be able to act independently of any individual market participant (or class of participant.) All participants within the electricity infrastructure must have representation on the ISO governing
committees, and no ISO employee may have a financial interest in any participant.

**Access:** an ISO has to provide open-access, self-scheduled transmission at non-pancaked rates within the confines of a general, non-discriminatory tariff to all eligible users.

**Control:** an ISO must have supervisory control of grid reliability, which refers to control of the grid and generation facilities.

**Efficiency:** an ISO must operate with the general goal of maximizing the utility to electricity consumers within its area of control. Concurrent with this is a commitment to provide information to all participants in a timely manner via an electronic information network.

Given its long-standing history, PJM was able to meet FERC’s requirements and make the transformation to ISO status before any other organization, and as such was designated as the nation’s first ISO on November 25, 1997. However, Order 888 was not FERC’s last word on this matter. To attempt to capture greater regional efficiencies from the transmission grid, FERC issued Order 2000 in December 1999 (FERC, 1999). This order detailed the formation of Regional Transmission Organizations (RTOs), which may be ISOs, transmission companies, grid companies, or any combination thereof. Rule 2000 maintains the fundamental principles of Order 888, but adds the proviso that “size matters.” RTOs are proposed to be of an even broader scale than ISOs, and FERC is currently exploring the notion of super-RTOs, covering even larger areas. The end goal is to have the entire US (and Canada) broken up into four or five super-RTOs, each with wide geographic coordination of transmission networks. PJM was the first entity to obtain RTO status, on December 12, 2002 (PJM, 2002). Details of PJM’s journey to ISO and RTO status can be found in Lambert (2001).
PJM operates markets in several products related to the delivery of electricity (e.g., capacity, reliability, spinning reserve, reactive power) but the focus of this essay is one of PJM’s two energy markets. PJM uses a two-tiered model for pricing electricity as purchased from generators and delivered to load-serving entities: a day-ahead (DA) market and a real-time (or balancing) market. The DA market provides hourly prices at each of several thousand intersections (nodes) within the PJM control area. These prices are known as locational marginal prices (LMPs). DA-LMPs are defined by the accumulation of generation offers, demand bids, increment offers, decrement bids and bilateral transaction schedules submitted into the DA market. Bidding into the DA market closes at 12 noon on the day before the delivery day. At 4:00 pm, PJM issues generation schedules and LMPs for the next day based on the first round of bidding. Between 4:00 and 6:00 pm a second round of bidding is open into which any generating bids not accepted in the first round can be rebid. After 6:00 pm PJM issues the final DA-LMPs and generation unit commitment schedules for the next day. The real-time balancing market works on five-minute increments utilizing actual loads and generation availability to calculate real-time prices, which are needed because the actual use of electricity at any time will be different to that estimated in the day-ahead market.

Thus, PJM uses a two-settlement system to charge and pay grid users for electricity. A load-serving entity is obliged to pay the DA-LMP for the amount of energy it bid for in the DA market. The difference between the actual amount of energy consumed and that bid for in the DA market is billed (or credited) at the real-time price. A numerical example: a utility successfully bids for 100 MWh at some node in the day-ahead market, at a DA-LMP of $50/MWh. Thus, the utility is obliged to pay $50/MWh * 100 MWh = $5,000. If the actual average load over that given hour is 110 MWh, and the average real-time price is $65/MWh, then the utility is obliged to pay $65/MWh * (110 – 100)MWh = $650 for the real-time (balancing) energy, for a total of $5,650. On the other hand, if the utility uses less than its DA bid, it is credited at the real-time price. If the actual load is 95 MWh, then the utility receives a credit of (100-95)MWh * $65/MWh = $325, and pays a total of $4,675. Real-time prices can be higher or lower than DA-LMPs depending on whether load is above or below that forecast by the accumulation of commitments in the
DA market, and whether all scheduled generation and transmission is available. Details of PJM energy market procedures can be found in the PJM Scheduling Operations Manual (PJM, 2004).

PJM uses a linear optimization program to calculate the LMPs. The theory of electricity spot pricing was developed at MIT and detailed most comprehensively in Schwepe, et al (1988). In its basic form, as in Hogan (2000), the price-establishment model can be written as follows:

\[
g = \text{vector of power generation at all buses;}
q = \text{vector of power loads at all buses;}
q' = \text{vector of net loads, demand minus generation at each bus, } q' = q - g;
x = \text{the vector of transmission variables (e.g., transformer settings, reactive power inputs and voltages);}
B(q) = \text{bid-based benefit function for loads (i.e., demand curve);}
C(g) = \text{bid-based cost function for generation (i.e., supply curve); and}
K(x, y) = \text{vector of transmission constraints.}
\]

The bid-based, security-constrained, economic dispatch problem is defined as:

\[
\text{Max } B(q) - C(g)
\]

Subject to the constraints:

\[
q - g = q', \text{ and } K(x, y) \leq 0
\]

Essentially, PJM uses a linear programming algorithm to maximize the benefits (i.e., minimize the costs) to load-serving entities by scheduling the cheapest possible generation given the constraints of the transmission grid. The exact details of the PJM LMP model are available only to PJM members.

The output of the LMP model has two parts:

(i) a matrix of prices, one for each node (currently about 2200 in PJM East) for each hour of the following calendar day; and
(ii) a schedule of day-ahead commitments to buy or sell power for every load-serving entity and generator that has participated in the DA market.

As mentioned above, differences between the scheduled DA buy and sell commitments and the actual amounts of power consumed are priced by the real-time market. The smaller the difference between the day-ahead schedules and actual power consumption, the smaller the difference between day-ahead and real-time prices will be. This essay is concerned only with day-ahead prices.
1.3 Congestion in transmission networks

As mentioned above, the day-ahead market model calculates a set of prices, one for each node. These prices may be all identical, or they may vary. In a perfect world, generators would provide all the electricity required within the PJM system in merit-order (i.e., lowest-cost first), and the resulting system-wide single price would be the marginal cost - the cost of the most expensive generator. This is often the case: between June 1, 2000 and April 30, 2004, approximately 20% of the hours featured a single system-wide DA price across PJM East. Shown in Figure 1.2 is the percentage of hours within a week that a single system-wide DA price was observed in PJM East. As can be seen, this number was typically higher towards the beginning of the period of study, showed a marked decline over the period June 2002 – June 2003, and appears to be rebounding somewhat since June 2003.

Sometimes transmission constraints sometimes require a generator to be dispatched out of merit order. Because the cheapest possible power cannot be delivered to every spot in the grid, some more expensive generators have to be dispatched. Thus, the customers in the region downstream of the constrained transmission link must pay a higher price than those upstream of the constraint. The power consumed downstream of the constraint will be billed at the marginal cost of power on the downstream side of the constraint, but not all of the power consumed at this location will come from downstream generators: some will be supplied by lower-cost generators along the constrained transmission lines. Thus, the ISO collects more money from consumers than it pays to generators. This excess revenue is called “congestion rent.” It is distributed by the non-profit ISO back to its constituent members by use of a mechanism called a “financial transmission right.” Financial transmission rights (FTRs) are the focus of study in Chapter 2 of this document.
The amount of power that can be sent down a line is limited by two things: the thermal limits in the line in question and the thermal limits in other lines, which directly affect all lines in a transmission grid due to Kirchhoff’s Laws. Kirchhoff’s Current Law states that the algebraic sum of current (or power) flows into a node shall be zero, \( i.e., \) energy is neither created nor destroyed at a node, and Kirchhoff’s Voltage Law states that the algebraic sum of voltage drops around a loop shall be zero, \( i.e., \) the voltage drop between two points shall be the same regardless of the path between those two points. Taken together, these two laws govern the pattern of flow rates and voltage drops in any electrical network. Changing the current flow at any point in the network has an effect on the flow rates in all other links in the network. In the context of this study, the existence of a flow constraint (the thermal limit) in one line impacts the permissible flows in all other lines in the network. This effect is referred to, in economic terms, as the loop flow externality. For a more thorough discussion of Kirchhoff’s Laws and network flow
modeling, please consult an elementary electrical engineering text, such as Smith (1984.) A numerical example of congestion follows. This example is adapted from that presented in Chao and Peck (1996).

We begin with a simple three-node network. Generators are located at nodes 1 and 2, and a load-serving entity exists at node 3. The transmission lines between the nodes have the following thermal power-flow limits:

- Line 1-2: 100 MW.
- Line 1-3: 300 MW.
- Line 2-3: 220 MW.

Let $P$ = price and $Q$ = load.

The following are the supply and demand functions for low and high load periods:

- Generator 1: $P_1$ (= marginal cost) = $10 + 0.05Q_1$ (both high and low-load periods.)
- Generator 2: $P_2 = 20 + 0.1Q_2$ (low load), $P_2 = 35 + 0.1Q_2$ (high load.)
- Consumer: $P_3 = 80 - 0.2Q_3$ (low load), $P_3 = 140 - 0.2Q_3$ (high load.)

The network, in a non-congested low-load equilibrium, is shown in Figure 1.3. For simplicity, and without loss of generality, it is assumed that each line has the same resistance to flow and same length, so flow will be symmetrical.

![Figure 1.3: Example network – low-load period](image)
In Figure 1.3, the flow limits are shown in parentheses inside each line, and the actual flows are shown on the outside of the line. By minimizing the price paid at the consumption node, we can calculate the total power consumed to be 286 MW, at a price of $22.86. Of this, 257 MW will be generated at node 1 and 29 MW at node 2. Clearly, power has the same price at all nodes, as none of the flow constraints are binding.

Now consider a high load period. Absent any constraints, the results are:
\[
\begin{align*}
P & = 35.70; \\
Q_1 & = 514.3 \text{ MW}; \\
Q_2 & = 7.1 \text{ MW}; \text{ and} \\
Q_3 & = 521.4 \text{ MW}.
\end{align*}
\]
However, this results in the following current flows:
\[
\begin{align*}
Q_{1,2} & = 169.0 \text{ MW}; \\
Q_{1,3} & = 345.2 \text{ MW}; \text{ and} \\
Q_{2,3} & = 176.2 \text{ MW}.
\end{align*}
\]
The thermal limits of Lines 1-2 and 1-3 are exceeded in this output scenario.

We now include the system constraints: \(Q_{1,2} \leq 100\), \(Q_{1,3} \leq 300\) and \(Q_{2,3} \leq 220\), in addition to the generation and consumption equations. By solving the optimization problem with these conditions applied, we get the following results:
\[
\begin{align*}
P_1 & = 29.50; \\
P_2 = P_3 & = 44; \\
Q_1 & = 390 \text{ MW}; \\
Q_2 & = 90 \text{ MW}; \text{ and} \\
Q_3 & = 480 \text{ MW}.
\end{align*}
\]
Line flows are now:
\[
\begin{align*}
Q_{1,2} & = 100 \text{ MW}; \\
Q_{1,3} & = 290 \text{ MW}; \text{ and} \\
Q_{2,3} & = 190 \text{ MW}.
\end{align*}
\]
Line 1-2 is constrained at its thermal limit, and the other lines are unconstrained. The result is that 480 MW of power is billed at $44/MWh, but only 90 MW is paid for at this
rate. The power generated at node 1 is paid at the marginal cost of that generator, $29.50/MWh. In terms of the LMP model as described above, the utility function, $B(q) - C(g)$, is maximized subject to the constraints of the transmission network. Any other combination of generator outputs will result in either a lower net utility or a violation of one or more of the network constraints. These results, along with the producer and consumer surpluses, are summarized in Table 1.1. As we can see, the generator at node 2 benefits from congestion, and the other participants lose. Even if all congestion rents are reimbursed to the losing parties, a net loss in aggregate welfare is observed. For this reason, it is not sufficient for PJM to simply have an equitable mechanism for redistributing congestion rent. If we wish to see welfare from electricity usage maximized, it is necessary to minimize congestion. The question then becomes one of comparing the costs of new transmission construction with the scale of the welfare losses.

Table 1.1: Prices and welfare gains with and without congestion

<table>
<thead>
<tr>
<th>Property</th>
<th>Value without constraints</th>
<th>Value with constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$ ($/MWh)$</td>
<td>$35.70$</td>
<td>$29.50$</td>
</tr>
<tr>
<td>$P_2$</td>
<td>$35.70$</td>
<td>$44.00$</td>
</tr>
<tr>
<td>$P_3$</td>
<td>$35.70$</td>
<td>$44.00$</td>
</tr>
<tr>
<td>$Q_1$ (MW)</td>
<td>514.3</td>
<td>390</td>
</tr>
<tr>
<td>$Q_2$</td>
<td>7.1</td>
<td>90</td>
</tr>
<tr>
<td>$Q_3$</td>
<td>521.4</td>
<td>480</td>
</tr>
<tr>
<td>Node 1 Surplus</td>
<td>$6,609$</td>
<td>$3,803$</td>
</tr>
<tr>
<td>Node 2 Surplus</td>
<td>$2$</td>
<td>$405$</td>
</tr>
<tr>
<td>Node 3 Surplus</td>
<td>$27,191$</td>
<td>$23,040$</td>
</tr>
<tr>
<td>Total Surplus</td>
<td>$33,802$</td>
<td>$27,248$</td>
</tr>
<tr>
<td>Congestion Rent</td>
<td>$0.00$</td>
<td>$5,655$</td>
</tr>
<tr>
<td>Net Welfare loss</td>
<td>---</td>
<td>$900$</td>
</tr>
</tbody>
</table>

A detailed explanation of the optimization problem for this network model is contained in Appendix A.
1.4 Congestion in the Delmarva Peninsula

Delmarva is the peninsula that projects into the Atlantic Ocean, south of the Philadelphia area. It consists of all of the state of Delaware, and parts of Maryland and Virginia. As part of the PJM grid, Delmarva is somewhat unusual in that it is essentially a “limb”, connected to the greater network only at one end. Delmarva has also been the site of some of the most frequent and costly congestion in PJM over the past five years. Public outcry was so great that FERC launched a fact-finding investigation (FERC, 2003) and PJM has issued two internally commissioned reports on the causes and extent of congestion.

The Delmarva Peninsula shall be construed, for the rest of this study, to be synonymous with the control area of Delmarva Power and Light (DPL), one of PJM’s constituent utilities. This control area covers 14 counties: three in Delaware, nine in Maryland and two in Virginia. These are shown in Figure 1.4.

Table 1.2, below, contains some geographic and economic statistics about the Delmarva counties. The main urban concentration is in the extreme north of the peninsula, the metropolitan Wilmington area. New Castle (DE) and Cecil (MD) counties are classified as part of the Philadelphia-Camden-Wilmington Metropolitan Statistical Area by the US Census Bureau.1 As can be seen in Table 1.2, New Castle County, which includes the greater Wilmington area, contains almost half of the population, and provides half the wealth of the peninsula. Other sizeable urban concentrations are Dover, (Kent Co., DE), metro population 35,000; Cambridge (Dorchester Co., MD) at 11,000, Easton (Talbot Co., MD) at 12,000 and Salisbury (Wicomico Co., MD) at 24,000.

---

1 See [http://www.census.gov/population/estimates/metro_general/List1.txt](http://www.census.gov/population/estimates/metro_general/List1.txt) for a US Census Bureau listing of metropolitan and micropolitan statistical areas and components.
Figure 1.4: Delmarva counties
Table 1.3 contains statistics concerning congestion at some selected points throughout Delmarva. Some explanation of these statistics is necessary. The four locations listed were chosen to provide snapshots at both highly populated areas and geographically diverse locations. The last location, Eastville, is in Northampton County, VA, near the extreme southern end of the peninsula. This area suffers some of the most extreme congestion in the DPL control area, due to its distance from large coal-fired (and, hence, low marginal cost) generators.

Table 1.3 Congestion Statistics

<table>
<thead>
<tr>
<th>Location</th>
<th>Wilmington</th>
<th>Dover</th>
<th>Salisbury</th>
<th>Eastville</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average electricity price</td>
<td>$35.12</td>
<td>$35.85</td>
<td>$36.54</td>
<td>$42.54</td>
</tr>
<tr>
<td>Average congestion cost</td>
<td>$2.75</td>
<td>$3.48</td>
<td>$4.17</td>
<td>$10.17</td>
</tr>
<tr>
<td>Average congestion severity</td>
<td>$5.83</td>
<td>$6.86</td>
<td>$8.48</td>
<td>$18.36</td>
</tr>
<tr>
<td>% of hours congested</td>
<td>47.2%</td>
<td>50.8%</td>
<td>49.2%</td>
<td>55.4%</td>
</tr>
<tr>
<td>% of hours severely congested</td>
<td>18.8%</td>
<td>20.5%</td>
<td>21.2%</td>
<td>29.8%</td>
</tr>
</tbody>
</table>

The first row in Table 1.3 is simply the average hourly DA-LMP, in $/MWh, over the period June 1, 2000 to March 31, 2004. The second row is the average cost of congestion, that is, the average part of the price that is paid out as congestion rent, and not as energy cost. The third row lists the severity of congestion. This is the average value of
congestion given that congestion exists, or put another way, the cost of congestion averaged over only those periods when congestion cost is larger than zero. The last two rows list the percentage of time when congestion is present. The bottom row lists time when congestion is considered “severe.” There is no broadly accepted definition of what constitutes “severe” congestion. In this study, it is taken as any time that congestion accounts for over 10% of the DA-LMP. As can be seen from Table 1.3, congestion exists all over the DPL control area about half the time, but the severity and frequency of severe congestion both increase dramatically as one moves south away from Wilmington.

**Generation**

Prices downstream of a constraint will only be higher if the marginal costs of generation downstream are higher. This is, in some respects, a tautological statement: the reason why a region would want to import power in the first place is because its local generation is more expensive than remote generation. Nonetheless, it is instructive to look at the various types and costs of generation within Delmarva in order to gain an understanding of the cost profiles across the region. Electricity is priced at marginal cost, which is the per-unit cost of the most expensive generator dispatched at any given time. Different types of generators have different marginal costs. The primary variable is fuel type. Of the main sources of electricity, marginal cost is typically ranked (from lowest to highest) in the following order: hydroelectric, nuclear, coal, fuel oil, natural gas. Hydro and nuclear have marginal costs near zero. By comparison, coal, fuel oil and natural gas have typical marginal costs on the order of $1.25, $5 and $7 per million BTU respectively (at least, over the period of study in this essay)\(^2\). One benefit of natural gas, which partially offsets its high marginal cost, is that small natural gas “peaking” plants can be built and operated (ex-fuel) very cheaply. Coal and fuel oil all require significant size before economies of scale work in their favor. For this reason, many new small natural gas plants have been built in recent years because their capital cost requirements per installed MW are much lower than coal or fuel oil plants. Furthermore, an additional cost of generation is the acquisition of sulfur-dioxide emission permits. Natural gas units emit essentially no SO\(_2\), so this cost burden does not exist. Natural gas plants also have a much

---

\(^2\) For fuel cost data, see [http://www.eia.doe.gov/cneaf/electricity/page/ferc423.html](http://www.eia.doe.gov/cneaf/electricity/page/ferc423.html)
easier time passing federal New Source Review pollution requirements. None of the generation within the DPL control area is hydro or nuclear, although there are considerable amounts of nuclear (and some hydro) near the northern boundaries of Delmarva. All major (>500 MW) plants in the eastern half of PJM are shown in Figure 1.5. The second variable is size: economies of scale typically exist in electricity generation, and *ceteris paribus*, large plants tend to have lower marginal costs than smaller plants. The third variable is age. Newer technology provides more efficient, and hence, lower cost plants. The new technology is typically in the form of more sophisticated electronic control systems. As can be seen from Figure 1.5, there is not much large-scale, low-cost generation within the DPL control area, and none in the south.

A more detailed list of generation facilities in DPL is given in Figure 1.6 and Table 1.4. The numbers in the first column of Table 1.4, “Map Ref,” refer to the numbers circled in Figure 1.6. As mentioned above, Delmarva has a dearth of large, modern, low marginal-cost generation, and the generators in the south are very small and employ high cost fuels.
Figure 1.5: Major eastern PJM generation
Figure 1.6: Delmarva generation facilities
### Table 1.4: Delmarva generators

<table>
<thead>
<tr>
<th>Map Ref.</th>
<th>Operator</th>
<th>Plant Name</th>
<th>Town</th>
<th>County</th>
<th>Capacity (MW)</th>
<th>Fuel</th>
<th>In-service Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Conectiv Delmarva</td>
<td>Delaware City 10</td>
<td>Delaware City</td>
<td>New Castle, DE</td>
<td>16</td>
<td>Fuel Oil</td>
<td>1968</td>
</tr>
<tr>
<td>2</td>
<td>Motiva</td>
<td>Delaware City Plant</td>
<td>Delaware City</td>
<td>New Castle, DE</td>
<td>382</td>
<td>Natural Gas</td>
<td>1956-2000</td>
</tr>
<tr>
<td>3</td>
<td>Conectiv Delmarva</td>
<td>Christians</td>
<td>New Castle</td>
<td>New Castle, DE</td>
<td>43</td>
<td>Fuel Oil</td>
<td>1973</td>
</tr>
<tr>
<td>4</td>
<td>Conectiv Delmarva</td>
<td>R Madison</td>
<td>New Castle</td>
<td>New Castle, DE</td>
<td>11</td>
<td>Fuel Oil</td>
<td>1962</td>
</tr>
<tr>
<td>5</td>
<td>Conectiv Delmarva</td>
<td>West Station</td>
<td>West Wilmington</td>
<td>New Castle, DE</td>
<td>15</td>
<td>Fuel Oil</td>
<td>1964</td>
</tr>
<tr>
<td>6</td>
<td>Conectiv Delmarva</td>
<td>Edge Moor</td>
<td>Wilmington</td>
<td>New Castle, DE</td>
<td>260</td>
<td>Coal</td>
<td>1954-66</td>
</tr>
<tr>
<td>7</td>
<td>Conectiv Delmarva</td>
<td>Edge Moor</td>
<td>Wilmington</td>
<td>New Castle, DE</td>
<td>457</td>
<td>Fuel Oil</td>
<td>1963-73</td>
</tr>
<tr>
<td>8</td>
<td>Conectiv Delmarva</td>
<td>Hay Road</td>
<td>Wilmington</td>
<td>New Castle, DE</td>
<td>1060</td>
<td>Natural Gas</td>
<td>1989-93</td>
</tr>
<tr>
<td>9</td>
<td>Delaware Municipal Electric</td>
<td>NA1</td>
<td>Smyrna</td>
<td>Kent, DE</td>
<td>48</td>
<td>Natural Gas</td>
<td>2002</td>
</tr>
<tr>
<td>10</td>
<td>City of Dover</td>
<td>McKee Run</td>
<td>Dover</td>
<td>Kent, DE</td>
<td>136</td>
<td>Fuel Oil</td>
<td>1962-73</td>
</tr>
<tr>
<td>11</td>
<td>City of Dover</td>
<td>Van Sant Station</td>
<td>Dover</td>
<td>Kent, DE</td>
<td>39</td>
<td>Fuel Oil</td>
<td>1991</td>
</tr>
<tr>
<td>12</td>
<td>City of Lewes</td>
<td>Lewes</td>
<td>Lewes</td>
<td>Sussex, DE</td>
<td>2</td>
<td>Fuel Oil</td>
<td>1993</td>
</tr>
<tr>
<td>13</td>
<td>City of Seaford</td>
<td>Seaford</td>
<td>Seaford</td>
<td>Sussex, DE</td>
<td>7</td>
<td>Fuel Oil</td>
<td>1947-1993</td>
</tr>
<tr>
<td>14</td>
<td>DuPont</td>
<td>Seaford Delaware Plant</td>
<td>Seaford</td>
<td>Sussex, DE</td>
<td>7</td>
<td>Coal</td>
<td>1939</td>
</tr>
<tr>
<td>15</td>
<td>Indian River Operations</td>
<td>Indian River</td>
<td>Millsboro</td>
<td>Sussex, DE</td>
<td>767</td>
<td>Coal</td>
<td>1957-80</td>
</tr>
<tr>
<td>16</td>
<td>Indian River Operations</td>
<td>Indian River</td>
<td>Millsboro</td>
<td>Sussex, DE</td>
<td>17</td>
<td>Fuel Oil</td>
<td>1968</td>
</tr>
<tr>
<td>17</td>
<td>Consolidated Edison</td>
<td>Rock Springs</td>
<td>Rock Springs</td>
<td>Cecil, MD</td>
<td>663</td>
<td>Natural Gas</td>
<td>2003</td>
</tr>
<tr>
<td>18</td>
<td>Easton Utilities Comm</td>
<td>Easton 1 &amp; 2</td>
<td>Easton</td>
<td>Talbot, MD</td>
<td>60</td>
<td>Fuel Oil</td>
<td>1934-95</td>
</tr>
<tr>
<td>19</td>
<td>VEE Operations</td>
<td>Vienna Operations</td>
<td>Vienna</td>
<td>Dorchester, MD</td>
<td>170</td>
<td>Fuel Oil</td>
<td>1968-72</td>
</tr>
<tr>
<td>21</td>
<td>Maryland Environmental Service</td>
<td>Eastern Correctional Inst</td>
<td>Westover</td>
<td>Somerset, MD</td>
<td>2</td>
<td>Fuel Oil</td>
<td>1988</td>
</tr>
<tr>
<td>22</td>
<td>Maryland Environmental Service</td>
<td>Eastern Correctional Inst</td>
<td>Westover</td>
<td>Somerset, MD</td>
<td>2</td>
<td>Pulp Liquor</td>
<td>1988</td>
</tr>
<tr>
<td>23</td>
<td>Conectiv Delmarva</td>
<td>Crisfield</td>
<td>Crisfield</td>
<td>Somerset, MD</td>
<td>10</td>
<td>Fuel Oil</td>
<td>1968</td>
</tr>
<tr>
<td>24</td>
<td>Commonwealth Chesapeake</td>
<td>Commonwealth Chesapeake</td>
<td>New Church</td>
<td>Accomack, VA</td>
<td>312</td>
<td>Fuel Oil</td>
<td>2000-01</td>
</tr>
<tr>
<td>25</td>
<td>Old Dominion Electric Corp</td>
<td>Diesel Group 1</td>
<td>Parksley</td>
<td>Accomack, VA</td>
<td>12</td>
<td>Fuel Oil</td>
<td>2002</td>
</tr>
<tr>
<td>26</td>
<td>Conectiv Delmarva</td>
<td>Tusley</td>
<td>Tusley</td>
<td>Accomack, VA</td>
<td>26</td>
<td>Fuel Oil</td>
<td>1972</td>
</tr>
<tr>
<td>27</td>
<td>Conectiv Delmarva</td>
<td>Bayview</td>
<td>Bayview</td>
<td>Northampton, VA</td>
<td>12</td>
<td>Fuel Oil</td>
<td>1963</td>
</tr>
</tbody>
</table>
Transmission

Figures 1.7 and 1.8 show most of the transmission lines in Delmarva. Not all 69 kV lines in the Wilmington area are shown due to their density in that area. 69 kV lines are shown on the rest of the two figures where they exist as part of the network between points of interest, but some 69 kV “spurs” that extend out from the grid to isolated locations not addressed in this study are omitted. All appropriate 69 kV lines are shown, and all lines of other voltages are shown in their entirety. The numbered circles in Figure 1.7 and 1.8 refer to the same generation facilities referenced in Table 1.4 and Figure 1.6.

![Figure 1.7: Delmarva northern transmission facilities](image-url)
As can be seen, three of the four locations referred to in Table 1.3 can be considered to be part of a network, where there are several possible transmission paths between any two points. However, Eastville is connected to the rest of the Delmarva grid by only a single 69 kV line. Should the line between Eastview and the Tasley power plant (No. 18 in Figure 1.8) become severed for any reason Eastville and all points south become a “transmission island”, served only by the Bayview generation facility (No. 19).
1.5 Literature review

Given the move towards deregulation and the introduction of market mechanisms in transmission concurrent with FERC’s recent rulings, there has been a spurt of papers addressing various aspects of deregulated transmission. Many of these come from two primary sources: the University of California Energy Institute, and the Electricity Policy Group at Harvard University. However, many of the papers from these groups focus upon market power issues, addressing the independence aspect of FERC policy, or examine market based incentives for expanding the transmission grid, which remains a centrally-planned function of the ISO under present FERC guidelines. Few papers empirically examine market microstructure issues such as the scale, scope and causes of congestion in specific regions.

One paper in which congestion is addressed, albeit indirectly, is Eynon, et al (2000). The authors, staffers at the Energy Information Administration, performed a multi-regional analysis examining the congestion effects of increased inter-regional bulk power trades. While markets as efficient as possible can be developed within the control area of a single ISO, there are significant “seam” issues, whereby maximizing the efficiency within an ISO may cause additional difficulties to those who wish to trade across ISO borders. This is one of the issues that FERC is striving to address in its push towards super-RTOs. However, the Eynon et al paper is strictly an engineering paper, using flow-modeling software to examine long-term changes in regional loads and the effects of such on inter-regional trade, with the existence of congestion being included in the model as a constraint on such trades. This paper does not address economic issues. Another broad-based federal government technical study is the National Transmission Grid Study, performed by the Department of Energy (DOE, 2002.) The primary goals of this study were to identify transmission “bottlenecks”, define why they must be eliminated, and provide description of some methods to improving grid efficiency. This study employed modeling software to aggregate actual transmission lines and networks into paths connecting 69 nodes across the country. The software provided estimates of the frequency and cost of congestion along the large paths of this pseudo-network. However,
the extreme aggregation employed in the DOE study renders the quantitative results useless for little more than general policy discussions.

As mentioned, PJM commissioned two studies of Delmarva congestion; one in-house and one external. The in-house study (Whitehead, 2002) estimated the cost of congestion by measuring the average price difference between the DPL and Peco zones and multiplying that by the DPL aggregate load. He also modeled the shadow price of each constraint and estimated total congestion cost by multiplying the load times the shadow cost at each constraint. Whitehead focused on transmission outages as the causes of congestion. Two troubling points arose from examination of this study: Whitehead claims that the majority of congestion is caused by a small number of constraints and then concentrates his examination upon a small number of constraints in northern Delmarva, which overlooks the most severe congestion in the south; secondly, Whitehead concentrates on outages as the root of much congestion but then reports that almost half of the congestion existed without any outages. Whitehead estimates the aggregate cost of congestion in the DPL area at $7 million in 1999, $20 million in 2000, $59 million in 2001 and $16 million in 2002. This is somewhat in accordance with Figure 1.2, which shows fewer congestion-free hours in 2001 than 2000 or 2002. As Whitehead is a PJM employee making a report to a Federal court (this study was an exhibit in FERC’s examination of Delmarva congestion) one is not surprised that PJM’s success in combating congestion is be emphasized. However, 2003 suffered more congestion than 2002. Whitehead makes it clear that PJM has focused upon addressing constraints in the highly populated northern part of Delmarva, where small reductions in the scale of congestion have larger aggregate welfare benefits. However, those paying the highest congestion charges, in the south, are largely overlooked in Whitehead’s study.

The other PJM study (Mitsche, 2002) was performed by PowerGEM consultants. Mitsche used real-time (not day-ahead) prices from all 288 nodes in DPL to calculate price differences along every constrained path at each hour. This report then focused on the most costly constraints. Once again, the majority of these were within the high-load northern part of the peninsula, even though these were not the site of the largest price
differences. Mitsche’s three largest-valued constraints all occur at interconnections with
the 500 KV transmission line running across New Castle and Cecil Counties, the primary
connector between DPL and the rest of PJM. Hence, these constraints affect the price
difference between DPL and PJM but have minimal effect on internal congestion. A
variety of constraints further south in the peninsula, mostly between Dover and Salisbury
are also examined, but the large southern price disparities are once again largely
overlooked.

One way of modeling congestion is to use a full-information engineering model. This
involves creating an analog of the full grid, with cost and demand functions at all
generation and consumption nodes, and then examining load patterns at various supply
and demand scenarios. This method was employed in the two PJM studies mentioned
above, and by some consulting firms that provide training services to participants in
deregulated energy markets that use the LMP methodology. Best known of these is the
firm Tabors Caramanis and Associates, founded by two of the authors of the seminal
work in spot pricing (Schweppe et al, 1988). Such modeling techniques are very data
intensive and much of the data required are proprietary or, at the very least, not generally
available in the public domain. One method of modeling prices between different markets
that is much less data-intensive is the arbitrage cost model, as employed by Spiller and

the shortfalls of time-series modeling, and use a simulation-based model, which
economists would call a full-information structural model. Hong and Hsiao (2001) and
Ma, et al (2004) describe neural network prediction algorithms: These can be thought of
as hybrids of full-information simulations with a learning component that is basically an
adaptive time-series element. Yang, Meliopoulos and Stefopolous (2005) examine the
effects on LMPs of remedial actions in networks. These are actions such as capacitor
shifting and transformer adjustment. Yang, et al demonstrate how such actions can
sometimes be used to abate congestion, instead of changing the dispatch order of
generators.
1.6 Economic model definition

The cost of transmission can be modeled as a combination of two unobserved factors. First, we consider the difference in prices between two locations if there were no transmission links between them. If each location is assumed to be part of an electricity network, then we are modeling the effects of linking these two otherwise autonomous networks. The prices at the two locations would be functions of the composition of each location’s respective network: the mix of generators, transmission facilities and loads.

If the locations in question are defined as node 1 and node 2 and we assume (without loss of generality) that the no-transmission price at node 2 is higher than that at node 1, then the transmission-free price difference is defined as:

\[ \Delta P_{12}^N = P_2^N - P_1^N \]  

where: \( P_i^N \) = price at node i in the absence of a transmission link with node j.

\( \Delta P_{12}^N \) is shown in Figure 1.9 as the difference between equilibrium prices \( P_2^N \) and \( P_1^N \):

![Figure 1.9: Equilibria at two nodes](image)

\[ \Delta P_{12}^N \]
We now consider the effects of creating a link, or set of links, between nodes 1 and 2; a set of links between the two previously autonomous networks. If there were no limits to the amount of transmission capacity, then users at node 2 would simply import all of their power from node 1. This results in the same composite supply and demand curves at both nodes, with a common price at nodes 1 and 2. In other words, the price difference is reduced to zero. Hence, transmission has the inherent capability to reduce or eliminate the “no-transmission” price difference between two nodes. This capability shall henceforth be referred to as the “Reductive Capacity”, denoted R.

We have addressed cases of no transmission and unlimited transmission. In reality, we usually have limited transmission. The amount of power that can be sent down a line is limited by two things: the thermal limits in the line in question, and the thermal limits in other lines, which directly affect all other lines in a transmission network due to the loop flow externality. If a transmission line between nodes 1 and 2 is in constrained operation, \( i.e., \) it is flowing the maximum possible amount, given its own thermal constraints and loop flow considerations, we say that the line is congested. Instead of being able to satisfy all demand at node 2, it is only able to partially supply it. The observed price difference (denoted \( y_{12} \)) will differ from the no-transmission price by how much the two price equilibria shift. If some power is supplied to the high-cost side of the link by the transmission line, then the demand curve on the low-cost side shifts outwards and the demands curve on the high-cost side shifts inwards by an equal amount. This is illustrated in Figure 1.10, below. The no-transmission demand curves are shown in grey. Demand in the low-cost zone (node 1) shifts outwards by an amount that is exactly equal to the transmission capacity between the two nodes. Demand in the high-cost zone shifts inwards by an equal amount. Assuming positively sloped supply curves, this results in a higher price at node 1 and a lower price at node 2.
The observed price difference can be modeled as the no-transmission price difference minus the reductive capacity. We have assumed that the price is higher at zone 2, thus yielding necessarily positive prices. If a negative price is observed on path 1-2, then node 2 is the low cost node, and power is flowing from the zone containing node 2 to the zone containing node 1.

Maintaining our assumption that $\Delta P_{12}^N > 0$ (and hence, flow is from node 1 to node 2) and letting $y_{12}$ denote an observed price difference between nodes 1 and 2, then:

$$y_{12} = \max(\Delta P_{12}^N - R_{12}, 0)$$

So $y = f(\Delta P^N, R)$. Now assume that we have linear estimators for the two components of price, $\Delta P^N$ and $R$. At first, assume $\Delta P^N$ is defined using some (as yet unspecified) linear estimator of exogenous variables $X$, parameters $\beta$ and disturbance $\epsilon_P$:

$$\Delta P^N = X \beta + \epsilon_P, \quad \epsilon_P \sim N(0, \sigma_P^2)$$

Similarly, assume the reductive capacity $R$ is defined by a linear estimator of exogenous variables $Z$, parameters $\alpha$ and disturbance $\epsilon_R$:

$$R = Z \alpha + \epsilon_R, \quad \epsilon_R \sim N(0, \sigma_R^2), \quad \epsilon_R \text{ truncated below at } -Z \alpha$$

$R$ is lower constrained at zero: the capacity of some line to reduce prices cannot be less than zero: if the line is out of service (which means that it does not exist in the grid model), then it has zero effect on the reductive capacity, not a negative effect.
Figure 1.11 illustrates this behavior in the $X\beta$-$y$ plane:

![Diagram of observed price difference with unobserved price components]

**Figure 1.11: Observed price difference with unobserved price components**

The blue line in Figure 1.11 indicates the no-transmission price difference. Since it is modeled as a linear function of $X\beta$, it will necessarily be a line of constant slope and intercept zero. The red lines refer to the reductive capacities. Note that there are two reductive capacities shown in Figure 1.11: a negative ($R^-$) and a positive ($R^+$) one. This convention is adopted to allow modeling of negative prices without having to concern ourselves with transforming actual negative price data into some positive form. The purple line in Figure 1.11 is the observed price: the no-transmission price minus the reductive capacity, with the constraint of being equal to zero when the absolute value of the appropriate $R$ is greater than the absolute value of $\Delta P^N$.

Thus, we have three states of observed prices: positive, negative or zero. As can be seen from Figure 1.11, the observed price is not a continuous function. Thus, a single linear estimator for the entire range of prices would be extremely inefficient. Instead, each of the three separate regions needs to be estimated separately, and those three estimators combined into a composite maximum-likelihood estimator. The three parts are described as follows.
**Price = 0 estimator**

Firstly, recall the estimators for the model parameters:

\[
\Delta P^N = X\beta + \varepsilon_p, \quad \varepsilon_p \sim N(0, \sigma^2_p) \quad (3)
\]

\[
R^+ = Z\alpha^+ + \varepsilon_{R+}, \quad \varepsilon_{R+} \sim N(0, \sigma^2_{R+}), \ varepsilon_{R+} \text{ truncated below at } -Z\alpha^+ \quad (5)
\]

\[
R^- = Z\alpha^- + \varepsilon_{R-}, \quad \varepsilon_{R-} \sim N(0, \sigma^2_{R-}), \ varepsilon_{R-} \text{ truncated above at } -Z\alpha^- \quad (6)
\]

A zero price is observed when the transmission free price difference is less than the positive reductive capacity and greater than the negative reductive capacity. Thus:

\[
Pr(y=0) = Pr(R^+ > \Delta P^N > R^-)
\]

This can be decomposed into two halves: those parts where \(\Delta P^N > 0\), and those where \(\Delta P^N < 0\).

Since \(R \leq 0\), if \(\Delta P^N > 0\) then \(\Delta P^N > R^-\)

Likewise, since \(R^+ \geq 0\), if \(\Delta P^N < 0\) then \(\Delta P^N < R^+\)

These conditions ensure that cases such as \(0 < R^- \Delta P^N\) do not occur.

**P = 0: Case where \(\Delta P^N > 0\):**

\[
Pr(y=0 \cap \Delta P^N > 0) = Pr(R^+ > \Delta P^N \cap \Delta P^N > 0) = Pr(Z\alpha^+ + \varepsilon_{R+} > X\beta + \varepsilon_p \cap X\beta + \varepsilon_p > 0) \quad (7)
\]

This probability is calculated by integrating over the joint bivariate normal probability surface defined by:

\[
\frac{1}{\sigma_p \sigma_{R+}} f \left( \frac{\varepsilon_p}{\sigma_p}, \frac{\varepsilon_{R+}}{\sigma_{R+}} \right)
\]

Since the surface is truncated in the \(R\) dimension as per the specification in Equation (5), we have to divide by the area left after truncation, which is \(F \left( \frac{Z\alpha^+}{\sigma_{R+}} \right) \). It is then necessary
to integrate over the appropriate space. The limits of the appropriate space can be derived from Equation (7):

- The first half of this statement defines the bound on $\varepsilon_{R+}$ as $\varepsilon_{R+} > X\beta - Z\alpha^* + \varepsilon_p$
- The second half defines the bound on $\varepsilon_p$ as $\varepsilon_p > -X\beta$

In other words, the lower bound for integration over the $\varepsilon_{R+}$ variable is $X\beta - Z\alpha^* + \varepsilon_p$, and the lower bound for integration over the $\varepsilon_p$ variable is $-X\beta$. In both cases, the upper bound is positive infinity.

Remembering to include the correction for the truncation as mentioned above, the area can be integrated as follows:

$$\Pr\left( y = 0 \cap \Delta P^N > 0 \right) = \frac{1}{\sigma_p \sigma_{R+}} \int_{A} \int_{\mathcal{D}} f\left( \frac{\varepsilon_p}{\sigma_p}, \frac{\varepsilon_{R+}}{\sigma_{R+}} \right) d\varepsilon_{R+} d\varepsilon_p$$

(8)

Where $A = X\beta - Z\alpha^* + \varepsilon_p$ and $B = -X\beta$

We are assuming that the error terms are independent. If two variables are independent, then the joint probability function $f(a,b)$ becomes $f(a)f(b)$. Thus, Equation (8) can be written as:

$$\Pr\left( y = 0 \cap \Delta P^N > 0 \right) = \frac{1}{\sigma_p \sigma_{R+}} \int_{A} \int_{\mathcal{D}} f\left( \frac{\varepsilon_p}{\sigma_p} \right) f\left( \frac{\varepsilon_{R+}}{\sigma_{R+}} \right) d\varepsilon_{R+} d\varepsilon_p$$

(9)

Stepwise solution of a double integral involves first solving for the “inner” integral. In Equation (9), the inner integral is the expression:

$$\int_{X\beta - Z\alpha^* + \varepsilon_p}^{\infty} f\left( \frac{\varepsilon_p}{\sigma_p} \right) f\left( \frac{\varepsilon_{R+}}{\sigma_{R+}} \right) d\varepsilon_{R+}$$

(10)
Since we are integrating with respect to $\varepsilon_{R^+}$, the other integrating variable ($\varepsilon_p$) is held constant, thus Equation (10) becomes:

$$
\int_{-\infty}^{\infty} R_{SA}^+ \varepsilon_p \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f\left(\frac{\varepsilon_p}{\sigma_p} \right) f\left(\frac{\varepsilon_{R^+}}{\sigma_{R^+}} \right) d\varepsilon_{R^+} d\varepsilon_p
$$

(11)

Consider that the cumulative density function for some distribution function is defined by:

$$
F\left(\frac{X}{\sigma}\right) = \int_{-\infty}^{x} \frac{1}{\sigma} \int_{-\infty}^{\frac{x}{\sigma}} f\left(\frac{x}{\sigma}\right) dx
$$

Hence, we are simply integrating the normal density function between $(X\beta - Z\alpha^+ + \varepsilon_p)/\sigma_{R^+}$ and infinity, which is the same as the cumulative standard normal from negative infinity and $(Z\alpha^+ - X\beta - \varepsilon_p)/\sigma_{R^+}$, commonly written as $F((Z\alpha^+ - X\beta - \varepsilon_p)/\sigma_{R^+})$

Thus,

$$
\Pr\left( Y = 0 \cap \Delta P_N > 0 \right) = \frac{1}{\sigma_p \sigma_{R^+}} \int_{-\infty}^{\infty} f\left(\frac{\varepsilon_p}{\sigma_p} \right) \int_{-\infty}^{\infty} f\left(\frac{\varepsilon_{R^+}}{\sigma_{R^+}} \right) d\varepsilon_{R^+} d\varepsilon_p
$$

is rewritten as

$$
\Pr\left( Y = 0 \cap \Delta P_N > 0 \right) = \frac{1}{\sigma_p \sigma_{R^+}} \int_{-\infty}^{\infty} f\left(\frac{\varepsilon_p}{\sigma_p} \right) \left[ \int_{-\infty}^{\infty} f\left(\frac{\varepsilon_{R^+}}{\sigma_{R^+}} \right) d\varepsilon_{R^+} \right] d\varepsilon_p,
$$

and the expression within the square brackets replaced with $F\left(\frac{Z\alpha^+ - X\beta - \varepsilon_p}{\sigma_{R^+}}\right)$, yielding the simplified form of the integral equation:
\[
\Pr(y = 0 \cap \Delta P^N > 0) = \frac{1}{\sigma_p F\left(\frac{Z\alpha^+}{\sigma_{R^+}}\right)} \int_{-\infty}^{\infty} f\left(\frac{\epsilon_p}{\sigma_p}\right) f\left(\frac{Z\alpha^+ - X\beta - \epsilon_p}{\sigma_{R^+}}\right) d\epsilon_p. \tag{12}
\]

**P=0: Case where \(\Delta P^N < 0\):**

By a method similar to that above, we get:

\[
\Pr(y=0\cap\Delta P^N<0) = \Pr(R^- < \Delta P^N \cap \Delta P^N < 0) = \Pr(Z\alpha^- + \epsilon_{R^-} < X\beta + \epsilon_p \cap X\beta + \epsilon_p < 0) \tag{13}
\]

In this case \(\epsilon_{R^-}\) is truncated above, which means the truncation factor is \(F\left(\frac{-Zz^-}{\sigma_{R^-}}\right)\).

From Equation (13), the first half of the statement defines the bound on \(\epsilon_{R^-}\) as:

\[\epsilon_{R^-} < X\beta - Z\alpha^- + \epsilon_p\]

The second half defines the bounds on \(\epsilon_p\) as \(\epsilon_p < -X\beta\)

In other words, the upper bound for integration over the \(\epsilon_{R^-}\) variable is \(X\beta - Z\alpha^- + \epsilon_p\), and the upper bound for integration over the \(\epsilon_p\) variable is \(-X\beta\). In each case, the lower bound is negative infinity.

Thus, the expression for the probability can be written as follows:

\[
\Pr(y = 0 \cap \Delta P^N < 0) = \frac{1}{\sigma_p \sigma_{R^-} F\left(\frac{-Z\alpha^-}{\sigma_{R^-}}\right)} \int_{-\infty}^{\infty} \int_{-a}^{a} d\epsilon_p d\epsilon_{R^-} f\left(\frac{\epsilon_{R^-}}{\sigma_{R^-}}\right) f\left(\frac{\epsilon_p}{\sigma_p}\right) \tag{14}
\]

Where \(A = X\beta - Z\alpha^- + \epsilon_p\)

and \(B = -X\beta\)

Performing the “inner” integration as above, Equation (11) can be reduced to:
\[
\Pr(y = 0 \cap \Delta P^N < 0) = \frac{1}{\sigma_p} F\left(\frac{-Z\alpha^-}{\sigma_{\alpha^-}}\right) - \int_{-\infty}^{XH} f\left(\frac{\varepsilon_p}{\sigma_p}\right) F\left(\frac{X\beta - Z\alpha^- + \varepsilon_p}{\sigma_{\alpha^-}}\right) d\varepsilon_p
\]

(15)

Because \([Pr(\Delta P^N > 0), Pr(\Delta P^N < 0)]\) is a mutually exclusive and completely exhaustive set, then \(Pr(y = 0) = Pr(y = 0 \cap \Delta P^N > 0) + Pr(y = 0 \cap \Delta P^N < 0)\). Thus, the entire expression for the probability of observing a zero price is:

\[
\Pr(y = 0) = \frac{1}{\sigma_p} F\left(\frac{Z\alpha^+}{\sigma_{\alpha^+}}\right) - \int_{-\infty}^{XH} f\left(\frac{\varepsilon_p}{\sigma_p}\right) F\left(\frac{Z\beta - X\alpha^- + \varepsilon_p}{\sigma_{\alpha^-}}\right) d\varepsilon_p + \frac{1}{\sigma_p} F\left(\frac{-Z\alpha^-}{\sigma_{\alpha^-}}\right) - \int_{-\infty}^{XH} f\left(\frac{\varepsilon_p}{\sigma_p}\right) F\left(\frac{X\beta - Z\alpha^- + \varepsilon_p}{\sigma_{\alpha^-}}\right) d\varepsilon_p
\]

(16)

**Price > 0 estimator**

The likelihood of observing a given positive price is the likelihood that \(y = \Delta P^N - R^+\).

Substituting the estimators from Equations (3) and (5) into the likelihood function gives:

\[
LH[y = X\beta + \varepsilon_p - (Z\alpha^+ + \varepsilon_{R^+})]
\]

Rearranging, this gives:

\[
LH[\varepsilon_p = y + Z\alpha^+ - X\beta + \varepsilon_{R^+}]
\]

(17)

The likelihood of observing any point \((\xi, \psi)\) in the joint bivariate distribution space, assuming independence of \(\xi\) and \(\psi\), is:

\[
LH = \frac{1}{\sigma_\xi \sigma_\psi} f\left(\frac{\xi}{\sigma_\xi}\right) f\left(\frac{\psi}{\sigma_\psi}\right)
\]

Including the truncation factor as shown in Equation (5), and substituting our variables, for any \(\varepsilon_p\), this likelihood is given by:
\[ LH = \frac{1}{\sigma_p \sigma_{R+} F\left( \frac{Z\alpha^+}{\sigma_{R+}} \right)} \int_{-Z\alpha^+}^{\infty} f\left( \frac{y + Z\alpha^+ - X\beta + \varepsilon_{R+}}{\sigma_p} \right) f\left( \frac{\varepsilon_{R+}}{\sigma_{R+}} \right) d\varepsilon_p \]  

(18)

To gain the complete likelihood of price \( y_{12} = y | y > 0 \) requires integration of the likelihood function across the relevant space of \( \varepsilon_p \).

Given the truncation of \( R^+ \), we know that \( \varepsilon_{R+} > - Z\alpha^+ \).

Thus, the lower bound on \( \varepsilon_{R+} \) is \( - Z\alpha^+ \). The upper bound is positive infinity.

Including these bounds, the complete likelihood function is given by:

\[ LH(y_{12} = y | y > 0) = \frac{1}{\sigma_p \sigma_{R+} F\left( \frac{Z\alpha^+}{\sigma_{R+}} \right)} \int_{-Z\alpha^+}^{\infty} f\left( \frac{y + Z\alpha^+ - X\beta + \varepsilon_{R+}}{\sigma_p} \right) f\left( \frac{\varepsilon_{R+}}{\sigma_{R+}} \right) d\varepsilon_p \]  

(19)

**Price < 0 estimator**

Using a method similar to that for the positive price estimator, we get:

\[ LH(y_{12} = y | y < 0) = \frac{1}{\sigma_p \sigma_{R-} F\left( \frac{-Z\alpha^-}{\sigma_{R-}} \right)} \int_{-Z\alpha^-}^{\infty} f\left( \frac{y + Z\alpha^- - X\beta + \varepsilon_{R-}}{\sigma_p} \right) f\left( \frac{\varepsilon_{R-}}{\sigma_{R-}} \right) d\varepsilon_p \]  

(20)

Now assign three indicator variables corresponding to the three possible states of \( y_{12} \):

- Let \( j1_t = 1 \) if \( y_{12,t} < 0, 0 \) otherwise.
- Let \( j2_t = 1 \) if \( y_{12,t} = 0, 0 \) otherwise.
- Let \( j3_t = 1 \) if \( y_{12,t} > 0, 0 \) otherwise.
- Let \( t \) be the time index.

The composite likelihood function will be

\[ LH_t = \prod_{t=1}^{T} \left[ L^+(y_{12,t})^{j1_t} \right] \left[ Pr(y_{12,t} = 0)^{j2_t} \right] \left[ L^-(y_{12,t})^{j3_t} \right] \]  

(21)

Taking logs and summing over all \( t \) yields the maximum likelihood estimator:
\[ LLF = \sum_{i=1}^{T} (j_1 t \cdot \ln(L^i(y_{1z_j})) + j_2 t \cdot \ln(\Pr(y_{1z_j} = 0)) + j_3 t \cdot \ln(L^i(y_{1z_j})) \]  

(22)

**Autocorrelation**

If we wish to include autocorrelation in the specification of the estimators, then the first three equations in this document become:

\[ \Delta P^N_t = X_t \beta + \rho_P \Delta P^N_{t-1} + \varepsilon_{P,t}, \quad \varepsilon_P \sim N(0, \sigma_P^2) \]  

(23)

\[ R^+_t = Z_t \alpha^+ + \rho_{R^+} R^+_{t-1} + \varepsilon_{R^+,t}, \quad \varepsilon_{R^+,t} \sim N(0, \sigma_{R^+}^2 | \varepsilon_{R^+,t} > -Z\alpha^+) \]  

(24)

\[ R^-_t = Z_t \alpha^- + \rho_{R^-} R^-_{t-1} + \varepsilon_{R^-,t}, \quad \varepsilon_{R^-,t} \sim N(0, \sigma_{R^-}^2 | \varepsilon_{R^-,t} < -Z\alpha^-) \]  

(25)

where \( \rho_P, \rho_{R^+}, \) and \( \rho_{R^-} \) are the correlation coefficients for the three lagged variables.

In each case, the endogenous variable is unobserved. Therefore, it is necessary to calculate the expected value of the lagged variable. This is equal to the value derived from the estimators (i.e., \( X\beta, Z\alpha^+ \) and \( Z\alpha^- \)) plus the expected value of the error term:

\[ E(\Delta P^N_{t-1}) = X_{t-1} \beta + E(\varepsilon_{P,t-1}) \]  

(26)

\[ E(R^+_{t-1}) = Z_{t-1} \alpha^+ + E(\varepsilon_{R^+,t-1}) \]  

(27)

\[ E(R^-_{t-1}) = Z_{t-1} \alpha^- + E(\varepsilon_{R^-,t-1}) \]  

(28)

For any continuous variable \( \xi \) distributed according to probability density function \( g(\xi) \), the expected value of \( \xi \) is defined by:

\[ E(\xi) = \int_{-\infty}^{\infty} \xi \cdot g(\xi)d\xi \]  

(29)
This specification assumes that the area of the distribution sums to 1. If it does not, we have to normalize this specification by dividing by the area of the distribution function:

\[
E(\xi) = \frac{\int_{-\infty}^{\xi} \xi \cdot g(\xi) d\xi}{\int_{-\infty}^{\infty} g(\xi) d\xi}
\]

(30)

Furthermore, if the distribution is discontinuous, such that separate parts need to be integrated separately, then the areas of the separate parts of the distribution must be taken into consideration. Assuming a discontinuity at \( \xi = A \) and distributions \( g(\xi) \) and \( h(\xi) \) on either side of the discontinuity, the two parts of the expected error are:

\[
E(\xi | \xi < A) = \frac{\int_{-\infty}^{A} \xi \cdot g(\xi) d\xi}{\int_{-\infty}^{\infty} g(\xi) d\xi}
\]

(31)

and:

\[
E(\xi | \xi > A) = \frac{\int_{A}^{\infty} \xi \cdot h(\xi) d\xi}{\int_{A}^{\infty} h(\xi) d\xi}
\]

(32)

We need to sum Equations (31) and (32), each weighted by their respective proportion of the total event space. The total space is defined by the summed probability function:

\[
\int_{-\infty}^{\infty} f(\xi) d\xi = \int_{-\infty}^{A} g(\xi) d\xi + \int_{A}^{\infty} h(\xi) d\xi
\]

(33)

Thus, adding the two parts of \( E(\xi) \) with the appropriate weighting becomes:

\[
E(\xi) = \frac{\left[ \int_{-\infty}^{A} \xi \cdot g(\xi) d\xi \right] \int_{-\infty}^{A} g(\xi) d\xi}{\int_{-\infty}^{\infty} g(\xi) d\xi} + \frac{\left[ \int_{A}^{\infty} \xi \cdot h(\xi) d\xi \right] \int_{A}^{\infty} h(\xi) d\xi}{\int_{A}^{\infty} h(\xi) d\xi}
\]

Which, when multiplied out and factored, yields:
Looking at the variables in question in this study, the distribution of the error term depends on the state of the system. That is, the value of any $E(\varepsilon)$ depends upon whether $y_t < 0$, $y_t = 0$ or $y_t > 0$.

Since there are three error terms and three possible states of $y_t$, then there are nine possible combinations of error term and state. However, some combinations do not occur in the model: if $y_t > 0$, then the likelihood function does not contain the error term $\varepsilon_R$, likewise if $y_t < 0$ then the likelihood function does not contain the error term $\varepsilon_{R+}$.

This is summarized as follows:

<table>
<thead>
<tr>
<th>State of $y_t$</th>
<th>Expected error terms to be calculated</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 0</td>
<td>$E(\varepsilon_P)$, $E(\varepsilon_{R+})$</td>
</tr>
<tr>
<td>= 0</td>
<td>$E(\varepsilon_P)$, $E(\varepsilon_{R+})$, $E(\varepsilon_R)$</td>
</tr>
<tr>
<td>&lt; 0</td>
<td>$E(\varepsilon_P)$, $E(\varepsilon_R)$</td>
</tr>
</tbody>
</table>

**Price = 0 estimator with autocorrelation**

First we will look at the case when $y_t = 0$.

As shown above, the probability of observing a zero price is broken up into two regions, one of which is a function of $\varepsilon_P$ and $\varepsilon_{R+}$, and the other a function of $\varepsilon_P$ and $\varepsilon_R$. Hence, $E(\varepsilon_P)$ must be estimated over two discontinuous regions, whereas the “$R$” error terms each exists in only one region. The generic form of this error term is as shown in Equation (34), above.
**Expected \( \varepsilon_p \) when \( \Delta P^N > 0 \) and \( y = 0 \)**

Equation (12) provides the distribution function for \( \varepsilon_p \):

\[
\Pr(y = 0 \cap \Delta P^N > 0) = \frac{1}{\sigma_p F\left(Z\alpha^+/\sigma_{R^+}\right)} \int_{-\infty}^{\infty} \varepsilon_p \cdot f\left(\frac{\varepsilon_p}{\sigma_p}\right) \frac{Z\alpha^+ - X\beta - \varepsilon_p}{\sigma_{R^+}} d\varepsilon_p.
\] (12)

Hence, omitting for simplicity the time subscripts, the expected value of \( \varepsilon_p \) for the region \( \Delta P^N > 0 \) is found by:

\[
E(\varepsilon_p | \Delta P^N > 0, y = 0) = \frac{1}{\sigma_p F\left(Z\alpha^+/\sigma_{R^+}\right)} \int_{-\infty}^{\infty} \varepsilon_p \cdot f\left(\frac{\varepsilon_p}{\sigma_p}\right) \frac{Z\alpha^+ - X\beta - \varepsilon_p}{\sigma_{R^+}} d\varepsilon_p
\] \[=\frac{\Pr(y = 0 \cap \Delta P^N > 0)}{\Pr(y = 0 \cap \Delta P^N > 0)}
\] (35)

**Expected \( \varepsilon_p \) when \( \Delta P^N < 0 \) and \( y = 0 \)**

By the same methods as described above, the expected value of \( \varepsilon_p \) when \( \Delta P^N < 0 \) is given by:

\[
E(\varepsilon_p | \Delta P^N < 0, y = 0) = \frac{1}{\sigma_p F\left(-Z\alpha^-/\sigma_{R^-}\right)} \int_{-\infty}^{\infty} \varepsilon_p \cdot f\left(\frac{\varepsilon_p}{\sigma_p}\right) \frac{X\beta - Z\alpha^- + \varepsilon_p}{\sigma_{R^-}} d\varepsilon_p
\] \[=\frac{\Pr(y = 0 \cap \Delta P^N < 0)}{\Pr(y = 0 \cap \Delta P^N < 0)}
\] (36)

Weighting and adding Equations (35) and (36) gives:

\[
E(\varepsilon_p | y = 0) = \frac{E(\varepsilon_p | \Delta P^N < 0, y = 0) \cdot \Pr(y = 0 \cap \Delta P^N < 0) + E(\varepsilon_p | \Delta P^N > 0, y = 0) \cdot \Pr(y = 0 \cap \Delta P^N > 0)}{\Pr(y = 0 \cap \Delta P < 0) + \Pr(y = 0 \cap \Delta P < 0)}
\]

\[
E(\varepsilon_p | y = 0) = \frac{g() + h()}{\Pr(y = 0)}
\] (37)
Where 
\[ g(\cdot) = \frac{1}{\sigma_p F\left(-Z\alpha^-/\sigma_{\epsilon^-}\right)} \int_{-\infty}^{\chi p} \epsilon_p \cdot f\left(\frac{\epsilon_p}{\sigma_p}\right) F\left(\frac{X\beta - Z\alpha^- + \epsilon_p}{\sigma_{\epsilon^-}}\right) d\epsilon_p \]

and 
\[ h(\cdot) = \frac{1}{\sigma_p F\left(Z\alpha^+/\sigma_{\epsilon^+}\right)} \int_{-\infty}^{\chi p} \epsilon_p \cdot f\left(\frac{\epsilon_p}{\sigma_p}\right) F\left(\frac{Z\alpha^+ - X\beta - \epsilon_p}{\sigma_{\epsilon^+}}\right) d\epsilon \]

**Expected \( \epsilon_{R+} \) when \( y = 0 \)**

The probability density function for this error term is given by Equation (8):

\[ \Pr\left(y = 0 \cap \Delta P^N > 0\right) = \frac{1}{\sigma_p \sigma_{\epsilon^+} F\left(Z\alpha^+/\sigma_{\epsilon^+}\right)} \int_{B}^{\infty} \int_{A}^{\infty} f\left(\frac{\epsilon_{R+}}{\sigma_{\epsilon^+}}\right) f\left(\frac{\epsilon_{R+}}{\sigma_{\epsilon^+}}\right) d\epsilon_{R+} d\epsilon_{R+} \]

(8)

Where \( A = X\beta - Z\alpha^+ + \epsilon_p \) and \( B = -X\beta \)

Hence, the expected value of \( \epsilon_{R+} \) over this space will be:

\[ E(\epsilon_{R+} \mid y = 0) = \frac{1}{\sigma_p \sigma_{\epsilon^+} F\left(Z\alpha^+/\sigma_{\epsilon^+}\right)} \int_{B}^{\infty} \int_{A}^{\infty} f\left(\frac{\epsilon_{R+}}{\sigma_{\epsilon^+}}\right) f\left(\frac{\epsilon_{R+}}{\sigma_{\epsilon^+}}\right) d\epsilon_{R+} d\epsilon_{R+} \]

(38)

The denominator has been calculated above as \( \Pr(y=0 \cap \Delta P^N>0) \). Since \( \epsilon_{R+} \) only enters into the function when \( \Delta P^N>0 \), we need not worry about the other “half” of the \( y=0 \) space until we seek to calculate \( E(\epsilon_{R+}) \).

We would like to simplify the numerator into a single integral form. However, in the form shown in Equation (6), the inner integral is
\[
\int_{\mathbb{D}}^\infty \mathcal{E}_{R_+} \cdot f\left( \frac{\mathcal{E}_p}{\sigma_p} \right) f\left( \frac{\mathcal{E}_{R_+}}{\sigma_{R_+}} \right) d\mathcal{E}_{R_+} 
\]  

(39)

Holding \( \mathcal{E}_p \) constant yields:

\[
\int_{\mathbb{D}}^\infty \mathcal{E}_{R_+} \cdot f\left( \frac{\mathcal{E}_p}{\sigma_p} \right) f\left( \frac{\mathcal{E}_{R_+}}{\sigma_{R_+}} \right) d\mathcal{E}_{R_+} 
\]  

(40)

For which there is no analytic solution. Of course, there is no analytic solution to the integral of the standard normal, but approximations of the cumulative normal are readily available in many software packages, which is the reason for wanting to simplify the double integral to a single integral involving the cumulative density function, \( F(\cdot) \).

This problem can be addressed by reversing the order of the integrands in the numerator of Equation (38). However, since we are working on a bounded, non-rectangular region of the probability space, we also need to modify the integrating intervals in the numerator of Equation (38).

That is, we need the numerator of Equation (38) in the form:

\[
\frac{1}{\sigma_p \sigma_{R_+}} F\left( \frac{Z\alpha^+}{\sigma_{R_+}} \right) \int_{A}^{B} \int_{C}^{D} \mathcal{E}_{R_+} \cdot f\left( \frac{\mathcal{E}_p}{\sigma_p} \right) f\left( \frac{\mathcal{E}_{R_+}}{\sigma_{R_+}} \right) d\mathcal{E}_p d\mathcal{E}_{R_+} 
\]  

(41)

What are the modified integrating intervals? Firstly, looking at A and B, the bounds on the space of \( \mathcal{E}_p \): from Equation 4 we have

\[
\mathcal{E}_{R_+} > X\beta - Z\alpha^+ + \mathcal{E}_p 
\]  

(42)

which rearranges to \( \mathcal{E}_p < \mathcal{E}_{R_+} + Z\alpha^* - X\beta \).
Equation (3) also tells us that $\varepsilon_p > -X\beta$.

Thus, in Equation (41) we have $A = -X\beta$ and $B = \varepsilon_{R+} + Z\alpha^+ - X\beta$.

Now we look at $C$ and $D$, the bounds on $\varepsilon_{R+}$. Equation (42) can once again be applied. We also know from above that the lower bound of $\varepsilon_p = -X\beta$.

Substituting for $\varepsilon_p$ in Equation (42) yields:

$$\varepsilon_{R+} > X\beta - Z\alpha^+ - X\beta$$

Reducing, we get:

$$\varepsilon_{R+} > -Z\alpha^+$$

There is no specified upper bound on $\varepsilon_{R+}$.

Thus, in Equation (41) we have $C = -Z\alpha^+$ and $D = \infty$.

Now we can perform the integration of Equation (42). The inner integral is:

$$\int_{-X\beta}^{dR+Z\alpha-X\beta} \varepsilon_{R+} \cdot f \left( \frac{\varepsilon_p}{\sigma_p} \right) f \left( \frac{\varepsilon_{R+}}{\sigma_{R+}} \right) d\varepsilon_p \quad (43)$$

Holding $\varepsilon_{R+}$ constant gives:

$$\varepsilon_{R+} \int_{-X\beta}^{dR+Z\alpha-X\beta} f \left( \frac{\varepsilon_p}{\sigma_p} \right) f \left( \frac{\varepsilon_{R+}}{\sigma_{R+}} \right) d\varepsilon_p \quad (44)$$

The integral can now be evaluated as the difference of the normal cumulative density function at the two bounds:
\[ F((\varepsilon_{R^+} + Z\alpha^+ - X\beta)/\sigma_p) - F(-X\beta/\sigma_p). \]

Thus, Equation (44) can be rewritten as:

\[ \varepsilon_{R^+} f\left(\frac{\varepsilon_{R^+}}{\sigma_{R^+}}\right) \left[ F\left(\frac{\varepsilon_{R^+} + Z\alpha^+ - X\beta}{\sigma_p}\right) - F\left(-\frac{X\beta}{\sigma_p}\right)\right] \]

And Equation (41) as:

\[ \frac{1}{\sigma_{R^+} F\left(\frac{Z\alpha^+}{\sigma_{R^+}}\right)} \int_{-Z\alpha}^{\infty} \varepsilon_{R^+} f\left(\frac{\varepsilon_{R^+}}{\sigma_{R^+}}\right) \left[ F\left(\frac{\varepsilon_{R^+} + Z\alpha^+ - X\beta}{\sigma_p}\right) - F\left(-\frac{X\beta}{\sigma_p}\right)\right] d\varepsilon_{R^+} \quad (45) \]

So, the expected value of \( \varepsilon_{R^+} \) when \( y = 0 \) is:

\[ E(\varepsilon_{R^+} \mid y = 0) = \frac{\int_{-Z\alpha}^{\infty} \varepsilon_{R^+} f\left(\frac{\varepsilon_{R^+}}{\sigma_{R^+}}\right) \left[ F\left(\frac{\varepsilon_{R^+} + Z\alpha^+ - X\beta}{\sigma_p}\right) - F\left(-\frac{X\beta}{\sigma_p}\right)\right] d\varepsilon_{R^+}}{\sigma_{R^+} F\left(\frac{Z\alpha^+}{\sigma_{R^+}}\right) \cdot \text{Pr}\{y = 0 \cap \Delta P^N > 0\}} \quad (46) \]
Expected \( \varepsilon_R \) when \( y = 0 \)

By the same method as in the immediately preceding section, the expected value of \( \varepsilon_R \) when \( y = 0 \) is:

\[
E(\varepsilon_{R-} \mid y = 0) = \frac{-\frac{\varepsilon_{R-}}{\sigma_{R-}} \cdot \int_{-\infty}^{-\frac{X\beta}{\sigma_p}} F\left(\frac{-X\beta}{\sigma_p}\right) - F\left(\frac{-\varepsilon_{R-} + Z\alpha^- - X\beta}{\sigma_p}\right) \, d\varepsilon_{R-}}{\sigma_{R-} \cdot \int_{-\infty}^{-\frac{Z\alpha^-}{\sigma_{R-}}} - F\left(\frac{-Z\alpha^-}{\sigma_{R-}}\right) \cdot \Pr\left(y = 0 \cap \Delta P^y < 0\right)}
\]

(47)

Expected \( \varepsilon_p \) when \( y > 0 \)

The expected value of \( \varepsilon_p \) when \( y > 0 \) takes the form of Equation (30), modified for a non-infinite lower limit:

\[
E(\xi) = \frac{\int_{\xi}^{y} \xi \cdot g(\xi) \, d\xi}{\int_{\xi}^{\infty} g(\xi) \, d\xi}
\]

(48)

The objective function \( \int_{\xi}^{\infty} g(\xi) \, d\xi \) is given by the likelihood function for \( \varepsilon_p \mid y > 0 \), Equation (19):

\[
LH(y_{12} = y \mid y > 0) = \frac{1}{\sigma_p \sigma_{R_+} F\left(\frac{Z\alpha'}{\sigma_{R_+}}\right)} \int_{-\infty}^{\infty} \frac{\varepsilon_p}{\sigma_p} \cdot f\left(\frac{X\beta - Z\alpha' - y + \varepsilon_p}{\sigma_{R_+}}\right) \, d\varepsilon_p
\]

(19)

Substituting (19) into (48) yields the expected value of \( \varepsilon_p \) when \( y > 0 \):

\[
E(\varepsilon_p \mid y > 0) = \frac{1}{\sigma_p \sigma_{R_+} \cdot F\left(\frac{Z\alpha'}{\sigma_{R_+}}\right)} \int_{-\infty}^{\infty} \frac{\varepsilon_p}{\sigma_p} \cdot f\left(\frac{X\beta - Z\alpha' - y + \varepsilon_p}{\sigma_{R_+}}\right) \, d\varepsilon_p
\]

\[
LH(y_{12} = y \mid y > 0)
\]

(49)
**Expected \( \varepsilon_R \), when \( y > 0 \)**

Using the same method as employed above, the expected value of \( \varepsilon_R^+ \) when \( y > 0 \) is given by:

\[
E(\varepsilon_R^+, y > 0) = \frac{1}{\sigma_p \sigma_{R^+} F\left(Z\alpha^+ / \sigma_{R^+}\right)} \int_{-\infty}^{\infty} \varepsilon_p \cdot f\left(\frac{\varepsilon_p}{\sigma_p}\right) f\left(\frac{X\beta - Z\alpha^+ - y + \varepsilon_p}{\sigma_{R^+}}\right) d\varepsilon_p
\]

\[
LH(y_{12} = y|y > 0)
\]

(Note: the objective function could be rewritten and integrated in \( \varepsilon_R^+ \), and this would be the necessary form if we were integrating analytically, but since the integration is being performed numerically, rewriting and integrating in \( \varepsilon_R^+ \) is an unnecessary step.)

**Expected \( \varepsilon_R \), when \( y < 0 \)**

Using the same method as above, and instead using the likelihood function for \( y_{12} = y|y < 0 \), Equation (20), then \( \varepsilon_R \) when \( y < 0 \) is given by:

\[
E(\varepsilon_R^-, y < 0) = \frac{1}{\sigma_p \sigma_{R^-} F\left(-Z\alpha^- / \sigma_{R^-}\right)} \int_{-\infty}^{\infty} \varepsilon_p \cdot f\left(\frac{\varepsilon_p}{\sigma_p}\right) f\left(\frac{X\beta - Z\alpha^- - y + \varepsilon_p}{\sigma_{R^-}}\right) d\varepsilon_p
\]

\[
LH(y_{12} = y|y < 0)
\]

**Expected \( \varepsilon_R \), when \( y < 0 \)**

Combining the previously employed methods, the expected value of \( \varepsilon_R \) when \( y < 0 \) is given by:

\[
E(\varepsilon_R^-, y < 0) = \frac{1}{\sigma_p \sigma_{R^-} F\left(-Z\alpha^- / \sigma_{R^-}\right)} \int_{-\infty}^{\infty} \varepsilon_p \cdot f\left(\frac{\varepsilon_p}{\sigma_p}\right) f\left(\frac{X\beta - Z\alpha^- - y + \varepsilon_p}{\sigma_{R^-}}\right) d\varepsilon_p
\]

\[
LH(y_{12} = y|y < 0)
\]
1.7 Econometric model definition

It is now necessary to define the components of the three linear estimators contained in Equations (3), (5) and (6): those for the transmission-free price difference and for the positive and negative reductive capacities. It is good modeling practice to only include as variables those factors that can be shown to be economically justifiable. As such, it is wise to begin with a little simple economic modeling of the two estimators. Initially, it will be assumed that power flows from node 1 to node 2 (that is, the price is higher at node 2). To model $\Delta P^N$, we assume that both nodes in question are connected to two completely separate networks, unlinked in any way. As described in the previous section, the components of the model are functions of the supply and demand curves at the two nodes, which are the unobserved supply and demand functions for the two separate networks.

First, consider the structure of the supply curve. Electricity supply curves are generally modeled as having a step-like form. This is due to some of the technical aspects of electricity generation. Any form of supply curve will necessarily have the lowest cost-of-supply goods closest to the origin. These are the cheapest goods to produce, and thus they are the ones that will be consumed first as demand increases from zero upwards. The level of the supply curve at any point is the marginal cost of producing that good – that is, the cost of producing an additional Watt of electricity. Hence, the supply curve will slope upwards to the right as we move from the lowest-marginal-cost source of electricity to the highest. The marginal cost of producing hydroelectric power is essentially zero. While consideration of the ability to produce power in the near future given falling reservoir levels in the absence of expected rainfall or snow pack must be made in the western US electricity market (this was a factor in the California energy crisis of 1999 and 2000), the reality of the situation for PJM is that there are two hydroelectric facilities on the lower Susquehanna river, and river elevation varies minimally from season to season. The cost of generating an extra Watt from hydro is the cost of opening a sluice valve by some minute increment, plus some incremental increase in expected maintenance costs. For all meaningful purposes, this amount is zero. The same is largely
true of nuclear power – to generate more power, more control rods are raised out of the reactor, and the water circulation rate through the reactor must be stepped up slightly. The marginal cost is the extra power used to pump more water, and the increased rate of usage of the radioactive medium, the fuel. Exact cost data are notoriously difficult to obtain in the public domain, but nuclear is largely held to have a marginal cost that is close to zero. The remaining sources of electricity are all fossil-fuel based, overlooking the as-yet minuscule market share of wind, solar and other fringe technologies. When considering fossil-fuel operations, the marginal cost is essentially the cost of additional fuel. Between each form of fuel there is a sharp increase in marginal cost: as all power from a particular source is used, the next unit must come from a higher cost fuel source. Thus, over the span of a network, the price of power may change very little over a wide range of demand, but then increase or decrease sharply as one of the discontinuities in the supply curve is encountered.

This is the theory. How about practice? Table 1.5 lists the installed summer capacity of all generation facilities within PJM East as per the Energy Information Administration 2003 Annual Electric Generator Report. The summer generating capacity is reported because summer is the peak-demand season, and summer capacity is slightly below winter capacity for most steam-turbine based generators due to reduced cooling capacity in the summer – as it gets hotter outside, cooling facilities are able to condense less steam, thus placing a lower cap on the amount of power that can be generated. Some modern facilities use condensing turbines, which allow a change of phase of the steam to water, but most steam turbines will be severely damaged by the presence of liquid water. The classes of generation are listed in the assumed increasing order of marginal cost.
Table 1.5: Installed generating capacity

<table>
<thead>
<tr>
<th>Fuel Type</th>
<th>Summer Capacity, MWh</th>
<th>Cumulative Capacity, MWh</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydro</td>
<td>3,130</td>
<td>3,130</td>
<td>4.5%</td>
</tr>
<tr>
<td>Nuclear</td>
<td>13,134</td>
<td>16,264</td>
<td>19.1%</td>
</tr>
<tr>
<td>Waste/Renewables</td>
<td>1,389</td>
<td>17,653</td>
<td>2.0%</td>
</tr>
<tr>
<td>Coal</td>
<td>20,926</td>
<td>38,579</td>
<td>30.4%</td>
</tr>
<tr>
<td>Fuel Oil</td>
<td>12,770</td>
<td>51,349</td>
<td>18.5%</td>
</tr>
<tr>
<td>Natural Gas</td>
<td>17,559</td>
<td>68,908</td>
<td>25.5%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The data in Table 1.5 can be compared to the loads typically seen in PJM East. These are illustrated in Figure 1.12.

![Figure 1.12: Maximum, average and minimum hourly loads, PJM East](image)

Although there is not a strict one-to-one correspondence between load and generation within PJM East – some power will be transmitted across the borders of the region, in both directions, we can look at Table 1.5 and Figure 1.12 to get an idea of how load is served in PJM East. The minimum hourly load seen in most months is in the area of
20,000 MW. From Table 1.5, we see that the combined total output of the very-low marginal cost plants - hydro, nuclear and waste/renewable – is not capable of meeting even the minimum loads in the region. For this reason, they are basically irrelevant to consideration of the marginal cost of power. Thus, the determining factors in the price of energy in PJM East are the costs of fossil fuels – coal, fuel oil, and natural gas. This is the part of the supply curve we will “see” – not the part below 20,000 MW, which is provided by the near-zero marginal cost fuels. So, what does the supply curve look like in the region in question? Figure 1.13 shows the minimum zonal price versus average hourly load in PJM over a one-week period in July, 2003. A period of this length was chosen because using more data points tends to crowd the plot without providing meaningful additional information. The minimum zonal price was obtained from PJM day-ahead price data. PJM publishes the average price in each of its constituent utilities’ service areas. The minimum of these values was chosen to attempt to remove the costs of congestion, leaving only energy costs. All data were retrieved from the PJM website.

![Figure 1.13: Energy cost versus load, PJM East](image-url)
Figure 1.13 is consistent with the notion that lower-cost energy is supplied first. The lowest cost energy, around $10/MWh is priced at about the lowest cost of coal: in July 2003, the average cost of coal in PJM was $1.26 per million BTU. Assuming a plant thermodynamic efficiency of 40%, this translates to an energy cost of about $11/MWh. 40% is the efficiency of the newest, largest coal-fired plants. Older and smaller plants have lower efficiencies, and this is observed in the increasing price as the load increases. When we reach a load of about 36,000 MWh, some sharply higher prices are observed (although not always). This coincides with the point in Table 1.5 where we start to run out of coal-fired generation and must then bring fuel oil and natural gas plants on-line. In July 2003, fuel oil prices were approximately 4 times higher than coal, on a per BTU basis, and natural gas was about 6 times higher. These explain the marginal costs of between $40 and $70 seen during the higher load periods. As can be seen from Figure 1.13, the step-function behavior is absent. This is because we do not have a single marginal cost for all coal plants, one for all oil plants, and one for all gas plants. Depending on age, size and technology, different plants have different efficiencies. There is a large number of different plants in PJM – well over 500 separate generation machines, very few with an individual capacity of over 1000 MW. This segregation of generation leads to a supply curve that has many small steps, enough that the curve can be considered to have a near-continuous form, as opposed to the step-like series of zero-slope regions as theorized. Thus, for the rest of this discussion, a continuous supply curve can be considered to be a reasonable approximation to reality in the region of the supply curve (20,000 to 50,000 MWh) that we are examining.

Recalling Figure 1.9, we have a market defined by two sets of supply and demand curves. This model can be expanded upon to extract the desired causative variables. The model shown in Figure 1.9 can be simplified somewhat. In the absence of demand-side load response, which can be assumed to be the case for most load in Delmarva, the demand for power at either node can be modeled as a vertical line: in the short-run, people do not moderate their demand for power in response to wholesale price changes, as people do not typically observe those short-term price changes. As mentioned above, since we are examining systems which consist of many generators, it is reasonable to abstract the
stepped supply function to a sloped one. Such a system is displayed in Figure 1.14, below. $\Delta P^N$ is simply the difference between the node 1 and node 2 equilibrium prices.

![Figure 1.14: Modeled supply and demand curves without transmission](image)

The observed price difference, $y_{12}$, is shown in Figure 1.15. In this case, it is assumed that a transmission link exists between nodes 1 and 2, but this link is not capable of fully satisfying the load at node 2. The demand curve at node 2 moves to the left, from $Q_2$ to $Q_2^*$, by the amount that can be served by transmission, shown as $\Delta Q$ in Figure 1.15.

![Figure 1.15: Modeled supply and demand curves with transmission](image)

The demand curve at node 1 moves to the right, from $Q_1$ to $Q_1^*$ by the same amount: $\Delta Q$. 

52
Some load has been transferred from the high-price network at node 2 to the low price network at node 1. Thus, the price at node 1 increases and the price at node 2 decreases. Because the load to node 2 has not been fully transferred to the low price network there is congestion, and there will still be a price difference between the two nodes. If we assume the supply curve at node 1 to have slope $m_1$ in the vicinity of $Q_1$, and the supply curve at node 2 to have slope $m_2$ in the vicinity of $Q_1$ then the difference between $y_{12}$ and $\Delta P^N$ will be the sum of the price changes at nodes 1 and 2, or $\Delta Q(m_1 + m_2)$. Since $R = \Delta P^N - y_{12}$, then this is the definition of the reductive capacity:

$$R = \Delta Q(m_1 + m_2).$$

where $\Delta Q =$ security constrained transmission limit between nodes 1 and 2
$m_1 =$ slope of node 1 network supply curve.
$m_2 =$ slope of node 2 network supply curve.

We can define $\Delta P^N$ in a similar fashion. $\Delta P^N = P_2 - P_1$. $P_1$ is the node 1 system load times the implied slope of the system supply curve. The implied slope is simply $P_1/Q_1$, and is labeled $M_1$. Note that $M_1$, the average slope between the origin and equilibrium, may not be the same as the slope in the region of the equilibrium, $m_1$. The same holds true for $P_2$, $Q_2$, and $M_2$.

$$\Delta P^N = M_2 Q_2 - M_1 Q_1$$

where $Q_2 =$ load of node 2 system.
$M_2 =$ implied slope of node 2 system supply curve
$Q_1 =$ load of node 1 system.
$M_1 =$ implied slope of node 1 system supply curve

Ideally, we would like to be able to structurally model, separately, the individual components of $R$ and $\Delta P^N$, the slopes, prices and loads. However, the system of equations as specified in Section 1.6 will only enable us to model the two “combined” components and not their individual parts. Furthermore, Equations (53) and (54) are non-linear, which leads to a complication of error terms. For example, were we able to linearly model $\Delta Q$, $m_1$ and $m_2$ independently, we would get the model $R = (\Delta \hat{Q} + e_{\Delta Q})(\hat{m}_1 + e_{m_1} + \hat{m}_2 + e_{m_2})$, making it impossible to isolate the error terms, and thus making maximum likelihood
estimation impossible. Thus, we have to define the unobserved dependent variables, \( \Delta P \), \( R^+ \) and \( R^- \), as linear estimators and not the products of estimators of their component parts.

In the region of consideration, the slope of the supply curves is considered constant. Thus, reductive capacity varies only with transmission limits, and so to model \( R^+ \) and \( R^- \), we need to consider which variables affect transmission capacity.

**Independent variables for estimation of reductive capacities**

The reductive capacity is related to the amount of power that can flow along a line. Thus, to model reductive capacity, we need to model those factors that can reduce flow along a line. As power flow increases, resistive losses increase (with the square of current flow), and resistive losses cause the line to heat. This heat causes thermal expansion, which allows lines to sag, and possibly contact foreign objects, such as trees. Thus, the ambient temperature is a factor in how much a line can flow. A lower ambient temperature means that heat is dissipated into the environment quicker, thus reducing the approach of the line to its thermal capacity. Increased winds will also increase the cooling rate of the ambient air on the line, but localized wind data are not obtainable. The square root of temperature is used in this estimator. The simplified explanation for using a square root transformation is as follows. The rate of heat flow away from a wire by simple conduction into the air is given by:

\[
\dot{\mathcal{H}} = \mathcal{S}(T_{\text{wire}} - T_{\text{air}})
\]

Where:
- \( \dot{\mathcal{H}} \) = heat flow from wire
- \( \mathcal{S} \) = specific heat capacity of air
- \( T_{\text{wire}} \) = maximum allowed wire temperature
- \( T_{\text{air}} \) = ambient air temperature

The rate of heat generated in a wire at maximum operating temperature is given by:

\[
\dot{\mathcal{H}} = Q^2 \Omega
\]

Where:
- \( \dot{\mathcal{H}} \) = rate of heat generation in wire
- \( Q \) = power flow along wire
\[ \Omega = \text{resistance of wire.} \]

At a steady temperature, the rate of heat generation and the rate of heat flow out of the wire must be equal, so we can equate (55) and (56), and rearrange for Q, giving us:

\[ Q = \frac{\amp}{\Omega} \sqrt{(T_{\text{wire}} - T_{\text{air}})} \]  \hspace{1cm} (57)

Given that \( \amp, \Omega \) and \( T_{\text{wire}} \) are all assumed to be constants, then the maximum allowable power flow declines by the square root of the difference between the maximum allowable wire temperature and the air temperature. Within the range of observed ambient air temperatures, this can be approximated by \( Q = k_1 - k_2 \sqrt{T_{\text{air}}} \), where \( k_1 \) and \( k_2 \) are constants. Hence, fundamental physics requires the use of the square-root transformation of ambient temperature in the linear estimator. Other factors, such as forced and natural convection and cooling from wind currents, affect line temperature but are very difficult to model concisely. Thus, a variable SQRT, the square root of the ambient temperature, was defined.

The other limit on how much a line can flow comes from the loop flow externality. Transmission lines are part of a network, and if one part of a network is congested, the maximum allowable flow in other lines close to the congested line is limited. That is, congestion in other parts of a network can limit the capacity of a given line to satisfy load demands, even if those demands are below the line’s thermal limits. A proxy was necessary to model “other congestion,” as including all other lines in the model would present an unwieldy specification. However, using the prices in other links in the network presents an endogeneity problem: for an estimator to have maximum efficiency, we wish for all independent (right-hand side) variables to be independent of the dependent (left-hand side) variable. However, given the loop flow externality, flows in all lines are dependent upon the flows in all other lines in a network, and thus the prices in each line are related to the prices in all other lines. It was necessary to define an estimator for congestion in other parts of the network that is independent of the price in some line in the Delmarva region. Thus, a variable henceforth labeled as PP_DIFF was defined this to estimate the effect of the loop-flow externality that avoids the endogeneity problem that is inherent with using other path prices as causative variables. This variable was
calculated by subtracting the PJM system-wide average hourly price from the hourly PECO zonal price. This variable measures how congested PECO is relative to the rest of PJM. Since almost all power that flows from the rest of PJM into the Delmarva region has to come through the PECO service territory, it is a measure of how hard it is to get power into Delmarva. The endogeneity problem is removed because while Delmarva prices are figured into the PJM system price, they are such a small component that they do not effect any significant change on the load-average system price. This statement can be justified as follows:

- The generation capacity in Delmarva is approx 3,300 MW, compared to 69,800 MW for all of PJM. Thus, DPL provides 4.7% of the generation.

- According to the PJM Annual Load Forecasts, the load in DPL is typically about 6% of the total load of the PJM RTO.

Thus, by either of these measures, DPL contributes a small proportion to the PJM system price. In this case, there is no concrete definition of “small”, and there will be some amount of two-way causation, but statistically speaking, this variable is less affected by endogeneity factors than any other considered at this time.

The transmission capacity of the line in question will have a necessary minimum value of zero (when the line is down) and a real positive value at all other times.

The line transmission capacity is a model for the reductive capacity of the line. Therefore, the functional form of Equations (5) and (6) will be:

\[
R^+ = \alpha_0^+ + \alpha_1^+ (\text{SQRT}) + \alpha_2^+ (\text{PP\_DIFF}) + \epsilon^+_R \\
R^- = \alpha_0^- + \alpha_1^- (\text{SQRT}) + \alpha_2^- (\text{PP\_DIFF}) + \epsilon^-_R.
\] (58) (59)
Expectation on parameter estimates for reductive capacity

The sign convention adopted in this essay defines flow from low-cost node 1 to high-cost node 2 as having positive values. In reality, when examining any pair of nodes, sometimes the flow will be from 1 to 2, and sometimes it will be from 2 to 1: there are very few (if any) pairs of nodes in PJM that exhibit constant one-way flow. One of these directions must be defined as positive, and the other as negative. The definition of positive-flow direction in each pair of nodes examined was based on historic price patterns: if one price was more frequently higher than the other, then the nodes were labeled as low-cost (1) and high-cost (2) accordingly. When flow is in the positive direction, \( i.e., \) when \( P_2 > P_1 \), we need to estimate the positive reductive capacity, \( R^+ \). If \( P_1 > P_2 \), then we need to model the negative reductive capacity, \( R^- \). Thus, \( R^+ \) will be necessarily defined as a positive number, and \( R^- \) as a negative number.

Thus, in the absence of other variables, we expect the constant term \( \alpha_0^+ \) to be positive and the constant term \( \alpha_0^- \) to be negative.

As mentioned above, an increase in ambient temperature leads to a reduction in capacity of the line. This means that we expect the parameter \( \alpha_i^+ \) to be negative. Given the inversion of the sign convention for the \( R^\) estimator we expect \( \alpha_i^- \) to be positive, \( i.e., \) it will decrease the absolute value of a negative number.

When PP_DIFF is positive, we have a higher price in PECO than the rest of PJM. Thus, PECO will want to “pull” power in from all regions, and will constrain all flows into DPL. Thus, if PP_DIFF is strongly positive it will reduce point-to-point congestion within Delmarva (all lines will be underutilized), and will entice power to flow in the opposite direction to the usual north-to-south direction. When PP_DIFF is negative, power is relatively cheap in PECO, and thus there will be a greater demand for it in DPL, with an expected increase in congestion. Thus, the parameter \( \alpha_2^+ \) is expected to be negative and \( \alpha_2^- \) to be positive.
Independent variables for estimating transmission-free price difference

As can be seen in Figure 1.9, the transmission-free price difference is a function of the average slopes of the supply curves and the loads at each node. If we assume similar average slopes of supply at either node, then the price difference is reduced to being a factor of the respective loads at the two nodes. Thus, we need variables that will model loads. Load demand shifts are largely predictable, on a daily and hourly basis – usage patterns are well established, and change based primarily upon two factors: additions of new loads to the grid (e.g., new residences, additions of new plant and equipment in industrial facilities), which are typically known and anticipated, and changes in response to temperature. Hot weather increases loads from cooling, which is almost completely electrically powered. Cold weather increases heating loads. Heating loads are not as electricity-intensive – only a small percentage of heaters are direct-electric heaters, most others use natural gas, propane, fuel oil, or only use electricity to move air or some other heat transfer medium. Shorter daylight hours in winter lead to greater lighting loads. Daily-load patterns are largely contingent upon domestic use: peaking during breakfast and supper hours. Thus, to model load, it is necessary to include the following variables:

ET: Elapsed time from the beginning of the study was included to model for expected population-driven load growth.

Temperature: we need separately to model warm temperatures to capture cooling loads and cold temperatures to capture heating loads. Hourly temperature data for the period in question from the weather station at New Castle County Airport were purchased from the US National Climate Data Center. The heating temperature departure ($HTD$) for each hour was calculated, using the formula:

$$HTD = \max(65 - T, 0), \text{ where } T \text{ is measured in degrees Fahrenheit.} \tag{60}$$

The cooling temperature departure ($CTD$) for each hour was calculated, using the formula:

$$CTD = \max(T - 65, 0), \text{ where } T \text{ is again measured in } ^\circ \text{F.} \tag{61}$$

$HTD$ and $CTD$ values were normalized and squared, yielding temperature proxies $HTD^2$ and $CTD^2$, as in the work of Kleit (2001).
Time-of-day variables were initially employed but found to be almost never statistically
significant and were thus omitted to improve the speed of the modeling process. The time
of year was strongly correlated with temperature data and thus was not employed.

Vertical displacement of the supply curves is caused by changes in fuel prices. Although
we have generation from coal, natural gas and fuel oil, the prices of all three fuels are
correlated to some degree. For this reason, it was decided that the cost of natural gas, in
the form of the variable NG, would be used as a proxy for all fuel prices.

Expectation of parameters in transmission-free price difference estimator
The expected value of the constant term, \( \beta_0 \), is positive, given that the pathways are
defined with the downstream node as the generally higher priced one. If the opposite
were true, we would expect to see price drop as we travel south along Delmarva, which is
generally not the case.

The expectation on the parameter for elapsed time, \( \beta_1 \), is not intuitively obvious before
the fact. As time passes, loads will tend to increase with population growth and industrial
activity, but we are interested in looking at changes in load differentials in combination
with growth in generation. That is, are loads and generation growing at different rates in
different locations? A cursory examination of US census data showed that population
growth was highest in Kent and Sussex counties in Delaware (about 10% growth from
2000 – 2004), medium in New Castle, DE and most of the MD counties (4-5%) and low
in the very south (1-2%). Most of the generation growth was in the north.

An increase in CTD or HTD signifies a movement to higher cooling or heating loads, and
thus a greater load differential. The expectation on the temperature variables is
ambiguous: they depend on the slopes of the respective supply curves in the area of the
temperature-based load shifts. If the upstream supply curve is steeper than the
downstream one, then they will be positive, and vice-versa.
The expectation on the natural gas parameter is also somewhat indeterminate before the fact. To model what might happen, one might consider the following thought experiment. Think of a model with two nodes, labeled N and S. Node N (for north) has a load that is many times larger than that at S. All of the generation at node S is fuel-oil fired.

Node N has a diversified generation slate with natural gas as the peaking fuel. We can assume, without loss of generality, that node N is served by coal as a baseline fuel and natural gas as a peaking fuel.

Now assume that the price if natural gas > price of fuel oil > price of coal.

So the supply curves at the two nodes will be similar to those shown in Figure 1.16:

![Figure 1.16: Supply curves at N and S nodes](Image)

Now, let us consider the case when $P(S) > P(N)$. In Delmarva, this corresponds to having southbound congestion, which is the “normal” case (i.e., we see southbound congestion more frequently than northbound).

If $P(S) > P(N)$, we have one of three scenarios:
1. South node is operating on fuel oil below its capacity constraint, north node is operating on coal. In this case, the price of natural gas is irrelevant to the model.

2. South node is at its capacity constraint, north node operating on coal. Once again, the price of natural gas is irrelevant.

3. South node at capacity constraint, north node operating on natural gas. This results in a very high price at node S, depending on slope of demand curve. If demand curve is vertical, then \( P(N) = \infty \), but in reality it will be finite but large – certainly larger than \( P(NG) \).

Thus, scenario (3) is the only one where the price of NG is relevant to the model. Assuming \( P(S) > P(NG) \), then as the price of natural gas increases, the transmission-free price difference will decrease. Thus, we expect the coefficient on natural gas to be negative when \( P(S) > P(N) \).

Now consider the case when \( P(N) > P(S) \). There are two possible scenarios here:

4. North node is at capacity constraint, south node operating on fuel oil. Price of natural gas is irrelevant to the model in this case.

5. North node operating on natural gas, south node on fuel oil.

Thus, scenario (5) is the only one where NG is a relevant variable. In this case, as the price of natural gas increases relative to the price of fuel oil, we would expect to see an increase in the transmission-free price difference. Since we are not including the price of fuel oil in this regression, the model will tell us that as \( P(NG) \) increases, \( \Delta P^N \) will increase. In other words, we have the expectation of a positive sign on the NG coefficient in the regression.
So, the two relevant cases involving the price of natural gas give us different expectations on the sign of the NG coefficient. When we are at moderately high load, and both sides are operating on their peaking fuel, we expect $\Delta P^N$ to be negative and the coefficient on NG to be positive. However, as soon as the south node reaches its capacity constraint, $\Delta P^N$ switches to a positive value with an expected negative coefficient on the price of natural gas. (Since the north node is connected to the national grid, we can assume that the north node never reaches a capacity constraint.)

So now we must consider the relative effects of these two scenarios. When the south node is in capacity constraint, then $P(S) >> P(N)$. Thus, a change in the price of natural gas will have, percentage-wise, a smaller effect on the price difference than in the case when $P(N) > P(S)$. However, this is a linear model, so we are concerned with absolute, and not percentage effects. The absolute effect of a change in the price of natural gas is equal in magnitude and opposite in sign for the two scenarios.

The estimated sign will be affected by the relative frequency of the two scenarios. If scenario (3) occurs more frequently than scenario (5) we will get a negative sign, and vice-versa. If scenarios (3) and (5) are both rare compared to the full set of scenarios, then we can expect the coefficient on NG to not be statistically significant.

The variables included in the estimators for $R$ and $\Delta P^N$ and their expected signs are listed in Table 1.6, below.
Table 1.6: Estimator parameters

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Expectation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_0^+)</td>
<td>Constant</td>
<td>Positive</td>
</tr>
<tr>
<td>(\alpha_1^+)</td>
<td>SQRT, Square root of temperature</td>
<td>Negative</td>
</tr>
<tr>
<td>(\alpha_2^+)</td>
<td>PP_DIFF, difference between PECO and PJM zonal prices</td>
<td>Negative</td>
</tr>
<tr>
<td>(\alpha_0^-)</td>
<td>Constant</td>
<td>Negative</td>
</tr>
<tr>
<td>(\alpha_1^-)</td>
<td>SQRT, Square root of temperature</td>
<td>Positive</td>
</tr>
<tr>
<td>(\alpha_2^-)</td>
<td>PP_DIFF, difference between PECO and PJM zonal prices</td>
<td>Positive</td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>Constant</td>
<td>Positive</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>ET, Elapsed time</td>
<td>Indeterminate</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>NG, Natural gas price</td>
<td>Indeterminate</td>
</tr>
<tr>
<td>(\beta_3)</td>
<td>HTD2, Heating temp departure squared</td>
<td>Indeterminate</td>
</tr>
<tr>
<td>(\beta_4)</td>
<td>CTD2, Cooling temp departure squared</td>
<td>Indeterminate</td>
</tr>
</tbody>
</table>

Estimated paths

In order to capture the effects of congestion on consumers at various points in Delmarva, it is necessary to model congestion over a variety of different links. In this study, the price was modeled across three links connecting the four locations defined in Table 1.3: Wilmington, Dover, Salisbury, and Eastville. Additionally, a “rest of PJM” node was created using the average price in the PECO region, which is the bridge between most of the DPL network and the rest of the country. Using these locations provides coverage of all the major metropolitan areas in Delmarva, as well as giving complete north-south geographic coverage. The paths are numbered as follows:

Path 1: Wilmington, DE to Dover, DE
Path 2: Dover, DE to Salisbury, MD
Path 3: Salisbury, MD to Eastville, VA.

For locations of these cities, please refer to Figures 1.7 and 1.8.
1.8 Data

Day-ahead locational marginal price data

All hourly on-peak day-ahead LMPs for the period June 1, 2000 (the beginning of the DA market) until March 31, 2004 for all nodes within PJM East were downloaded from the PJM website. When the decision was made to restrict this study to Delmarva congestion and to four locations in Delmarva, the following nodes were selected for study: PJM and PECO zonal prices (for calculation of the PP_DIFF variable); City of New Castle (for Wilmington area prices); North Street, Dover; North Salisbury; and Eastville. All prices quoted are in $/MWh. Time-series plots of the price along each path are shown in Figures 1.17 though 1.19. As we are interested in modeling price differences, the price at the upstream side of each link (node 1) was subtracted from the downstream price (node 2) to yield an observed price difference. These hourly data were then averaged over four-hour periods to decrease the computational burden – reducing 15,648 observations to 3,912. These data are employed in estimators that use logarithms of normal distribution functions. Evaluating the normal probability density function at extreme values yields values of zero, which makes taking logarithms impossible. This problem is typically solved by using a log transformation on large primitives to reduce them in scale, but since we have both positive and negative prices, and a large number of zero-valued observations, this was not an option. The prices as supplied by PJM are in $/MWh, and some of the price difference values observed in this study were over $100 – clearly not something that can be used in a normal distribution function. The extreme values were removed by censoring all low-price data to the value of the 0.1\textsuperscript{th} percentile value and all high-price data to the 99.9\textsuperscript{th} percentile. That is, the most extreme 0.1\% of prices on either end of the distribution were censored. The resulting data were normalized to a range of [0, 1].

64
Figure 1.17: Path 1 price (Wilmington - Dover)
Note: price spike to $196/MWh in August 2002 omitted for clarity

Figure 1.18: Path 2 price (Dover - Salisbury)
The paths were labeled as described above. Table 1.7 contains some statistical information about the transformed price difference data. As can be seen from Table 1.7, congestion existed in between 50 and 85% of the time – as we get further south, congestion is less frequent (but as seen before, more severe). Table 1.8 displays the concentrations of combined congestion in the network: the number of hours when congestion existed in a given number of the three paths. In every case, prices tended to increase as we move south: the +/- bias of greater than one in Table 1.7 means that positive transmission costs are more common in the southward direction. In addition, we can see that congestion over the whole peninsula is more common than having it isolated in certain parts, but when it is isolated, it is in the north. That is, congestion in the south requires congestion in the north, and localized causes of congestion in the south were rare. Correlations for the existence of congestion between each path was calculated, whereby a value of one was assigned if congestion existed, zero otherwise for each of the three paths. The results are contained in Table 1.9.
Temperature data

Hourly temperature data from the weather station at New Castle County Airport (located about 5 miles south of central Wilmington) for the period in question were purchased from the National Climate Data Center. The data were averaged into four-hour blocks, and then three transformations were performed:

- The square root of temperature (in °F) was calculated, and these values were transformed to a range of [0, 1] by subtracting the minimum value and dividing
by the range of the data (maximum minus minimum). Since the minimum observed temperature was 4.5°F, taking roots of negative values was not a concern.

- The heating temperature departure (HTD) was calculated by subtracting the observed temperature from 65°F. Any negative values of HTD (i.e., any temperatures above 65°F) were replaced by zero. These values were squared, and then transformed to a range of [0, 1].

- The cooling temperature departure (CTD) was calculated by subtracting 65°F from the observed temperature. Any negative values of CTD (i.e., any temperatures below 65°F) were replaced by zero. These values were squared, and then transformed to a range of [0, 1].

The four-hour average temperatures are shown in Figure 1.20. Table 1.10 contains some statistical data concerning the temperature data. The variables denoted with an asterisk in Table 1.10 are the transformed values. Figure 1.21 shows the maximum observed absolute congestion price versus untransformed temperatures. As can be seen, there is not an obvious correlation between high temperatures and high congestion.

### Table 1.10: Temperature data statistics

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Temp, °F</th>
<th>Temp^{0.5}</th>
<th>Temp^{0.5}*</th>
<th>HTD</th>
<th>HTD*</th>
<th>CTD</th>
<th>CTD*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>54.3</td>
<td>7.25</td>
<td>0.66</td>
<td>13.97</td>
<td>0.23</td>
<td>3.23</td>
<td>0.10</td>
</tr>
<tr>
<td>Minimum</td>
<td>4.5</td>
<td>2.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>97.5</td>
<td>9.9</td>
<td>1</td>
<td>60.5</td>
<td>1</td>
<td>32.75</td>
<td>1</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>18.12</td>
<td>1.29</td>
<td>0.17</td>
<td>14.29</td>
<td>0.24</td>
<td>5.83</td>
<td>0.18</td>
</tr>
<tr>
<td>Non-zero occurrences</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>5,572</td>
<td>5,572</td>
<td>2,827</td>
<td>2,827</td>
</tr>
</tbody>
</table>
Figure 1.20: Four-hour mean temperature, New Castle Co. airport

Figure 1.21: Maximum congestion price versus temperature

Note: single $350/MWh price point omitted for clarity
Natural Gas Prices
FERC requires all power plants of capacity greater than 50 MW to file Form 423, entitled "Monthly Report of Cost and Quality of Fuels for Electric Plants." The compiled data from these forms are reported on the Energy Information Administration’s website. This was the data source for monthly average prices of natural gas delivered to generators in PJM. Prices are reported in dollars per million BTUs. Fuel prices are plotted in Figure 1.22. These data were subsequently normalized to a range of [0, 1].

![Figure 1.22: Natural gas prices](image)

Table 1.11 contains statistical data concerning the price of fuel.

<table>
<thead>
<tr>
<th>Fuel</th>
<th>Raw prices</th>
<th>Normalized prices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean price, $/mmBTU</td>
<td>6.39</td>
<td>0.44</td>
</tr>
<tr>
<td>Minimum Price</td>
<td>4.51</td>
<td>0</td>
</tr>
<tr>
<td>Maximum Price</td>
<td>8.78</td>
<td>1</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.00</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Figure 1.23 shows the maximum observed congestion cost versus natural gas cost. As can be seen, there is no strong observable relationship between peaking-fuel price and congestion costs.
Figure 1.23: Maximum congestion price versus natural gas price
Note: single $350/MWh price point omitted for clarity

Time data
Elapsed time was normalized from the beginning to the end of the study to a range of [0, 1].

Independent variable correlations
The efficiency of any linear estimator is improved if the causal variables are independent of each other. This independence is tested by performing a correlation analysis of each variable against each other. Tables of the Pearson correlation coefficients for the variables employed in the $P_N$ regression are shown in Tables 1.12. To obtain the $R^2$ value, simply square the correlation coefficients. As these coefficients are symmetrical, the spaces below the diagonals are left blank. For the two variables used in the $R^+$ and $R^-$ regressions, PP_DIFF and SQRT, the correlation coefficient is 0.143.
Table 1.12: Correlations for $\Delta P^N$ estimator variables

<table>
<thead>
<tr>
<th></th>
<th>ET</th>
<th>HTD2</th>
<th>CTD2</th>
</tr>
</thead>
<tbody>
<tr>
<td>NG</td>
<td>0.088</td>
<td>-0.091</td>
<td>0.059</td>
</tr>
<tr>
<td>ET</td>
<td>0.171</td>
<td>-0.076</td>
<td></td>
</tr>
<tr>
<td>HTD2</td>
<td></td>
<td>-0.330</td>
<td></td>
</tr>
</tbody>
</table>

**Autocorrelation**

The algorithm employed in this essay is specifically not a time-series process. Indeed, arbitrage-cost models are often presented as an alternative to time-series models. However, there may be some meaningful time series effects present. As such, an extra set of parameter estimates was performed; including each of the aforementioned variables as well as a single-period lagged dependent variable. That is, estimated values of $R$ and $\Delta P^N$ from the previous period were included as causal variables.
1.9 Experimental procedure

The core of the analysis is finding the three vectors of parameters, $\alpha^+$, $\alpha^-$ and $\beta$, and the standard deviations $\sigma_{R+}$, $\sigma_{R-}$ and $\sigma_p$ which maximize the value of Equation 22:

\[
LLF = \sum_{i=1}^{T} \left( j_1 \cdot \ln(L_i(y_{12,i})) + j_2 \cdot \ln(\Pr(y_{12,i} = 0)) + j_3 \cdot \ln(L^*(y_{12,i})) \right)
\]  

(22)

Where:

\[
L^* = \frac{1}{\sigma_p \sigma_{R+} F \left( -Z \frac{\alpha^-}{\sigma_{R-}} \right) \left( y_{12} \right)^{-X^\beta}} \left[ \int_{-\infty}^{X^\beta} \left( \frac{\epsilon_p}{\sigma_p} \right) f \left( X^\beta - Z \frac{\alpha^- - y + \epsilon_p}{\sigma_{R-}} \right) d\epsilon_p \right] + 
\]

(19)

\[
Pr(y = 0) = \frac{1}{\sigma_p F \left( Z \frac{\alpha^+}{\sigma_{R+}} \right) \left( y_{12} \right)^{-X^\beta}} \left[ \int_{-\infty}^{X^\beta} \left( \frac{\epsilon_p}{\sigma_p} \right) f \left( Z \frac{\alpha^+ - X^\beta - \epsilon_p}{\sigma_{R+}} \right) d\epsilon_p \right] + 
\]

(16) and

\[
L^* = \frac{1}{\sigma_p \sigma_{R} F \left( Z \frac{\alpha^+}{\sigma_{R+}} \right) \left( y_{12} \right)^{-X^\beta}} \left[ \int_{-\infty}^{X^\beta} \left( \frac{\epsilon_p}{\sigma_p} \right) f \left( X^\beta - Z \frac{\alpha^+ + \epsilon_p}{\sigma_{R+}} \right) d\epsilon_p \right] + 
\]

(20)

Equation (22) is maximized when its first derivative with respect to any parameter is zero. The Newton-Raphson root-finding technique was used in this study. This involves finding the first and second derivatives of the likelihood function (22).

The first and second derivatives of the likelihood function are either extremely difficult or impossible to obtain analytically and instead must be estimated numerically. A central-differencing method was used to estimate the derivatives, as described below. The method of central differencing assumes that any derivative over a small interval can be approximated by the slope of a line segment which is defined by evaluations of the basic
function at two points: one just above and one just below the point in question. That is, given some variables $X$ and $Y$, and an estimation interval $h$:

$$\frac{\partial f(X,Y)}{\partial X} \approx \frac{f(X+h,Y)-f(X-h,Y)}{2h}$$  \hspace{1cm} (62)$$

The second derivative is found by simply taking the derivative of Equation (62):

$$\frac{\partial^2 f(X,Y)}{\partial X \partial Y} \approx \left[ \frac{f(X+h,Y+h)-f(X-h,Y+h)}{2h} - \frac{f(X+h,Y-h)-f(X-h,Y-h)}{2h} \right] \frac{1}{2h}$$

which reduces to:

$$\frac{\partial^2 f(X,Y)}{\partial X \partial Y} \approx \frac{f(X+h,Y+h)-f(X-h,Y+h)-f(X+h,Y-h)+f(X-h,Y-h)}{4h^2}$$  \hspace{1cm} (63)$$

Thus, for each first derivative, the function is evaluated twice, and for each second derivative, four times. This method is described in more detail in Gerald and Wheatley (1990).

Given the peculiar nature of a maximum-likelihood function that contains normal distribution functions, a good initial estimate of the parameters is required to obtain convergence of the model. With most functions the Newton-Raphson technique converges rapidly, with a correction factor shrinking to the order of $1 \times 10^{-8}$ within four or five iterations. This is true in this case if the initial estimate is good. If not, the model will “blow up”. The more variables in a model, the easier it is to obtain a divergent correction factor. Given $n$ variables and assuming a symmetrical Hessian, we must evaluate $(n^2+n)/2$ second derivatives, and if only one of these second derivatives points in a direction away from the maximum point, the model will diverge.

To arrive at a parameter estimate that would not blow up in this model, a grid search technique was employed. This routine is only reliable if there are no local maxima in this likelihood function: it must be strictly monotone declining as one retreats from the global
maximum in any direction. This was ascertained non-analytically for this function by using a loop routine to compute the value of the function versus various values of several of the parameters and examining a plot of the results. In each case the function had only a global maximum, thus, using a grid search will not steer us towards a local maximum that is lower than the global maximum. The grid search routine in question evaluated the likelihood function at some value of a parameter \(B\), and at some value \(B+h\). If \(L(B+h)>L(B)\), then the function was evaluated at \((B + 2h)\) and so on until a maximum was reached. This routine was executed for each parameter. Such a routine finds an approximate maximum by varying only one parameter at a time, so the whole routine was run several times over, until such a point when no changes in the parameter values were exhibited. The interval size, \(h\), was then decreased by an order of magnitude, and the process started over again. The initial value of \(h\) was 0.1 for all parameters, and values of 0.01 and 0.001 were also applied. After such a routine was completed, the error between the observed likelihood function and the maximum value was typically between 0.5 and 1.0. It was necessary to get this close to the maximum for the Newton-Raphson routine to converge. When this point was reached, a Microsoft Excel spreadsheet was employed, using the Excel Solver add-in to fine-tune the parameters to within about 0.01 of the maximum value.

Even if a gridsearch routine or a spreadsheet application provided the necessary accuracy for parameter estimates, it was still required to evaluate the Newton-Raphson routine to obtain the Hessian matrix. This is because the Hessian is the basis of the information matrix required to estimate the parameter variances. The information matrix is the inverse of the negative of the Hessian, and the parameter variances are the diagonal terms of the information matrix. For more details on this procedure, see Greene (2000). It is possible to estimate the information matrix without taking the Hessian using the BHHH routine (Berndt, et al, 1974), which essentially is the outer product of the gradient. The autocorrelation estimations were sufficiently onerous that the BHHH algorithm was applied to estimate the information matrix and standard errors.
1.10 Results and diagnosis

Tables 1.13 through 1.15 contain the parameter estimates and t-statistics for each of the three paths of study, both with and without autocorrelation. Entries in bold-face type are statistically significant to a 95% confidence level (i.e., they have a t-statistic of absolute value greater than 1.96.) The first five parameters in each table refer to the $R^+$ regression, the next five to the $R^-$ regression, and the last seven to the $\Delta P^{N}$ regression. The value of the composite likelihood function is shown at the bottom of each table.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>With autocorrelation</th>
<th>Without autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^+$ regression</td>
<td>Value</td>
<td>t-stat</td>
</tr>
<tr>
<td>Constant</td>
<td>0.049</td>
<td>4.63</td>
</tr>
<tr>
<td>Lag of $R^+$</td>
<td>0.177</td>
<td>2.78</td>
</tr>
<tr>
<td>PP_DIFF</td>
<td>-0.048</td>
<td>-4.53</td>
</tr>
<tr>
<td>SQRT</td>
<td>-0.014</td>
<td>-36.45</td>
</tr>
<tr>
<td>$\sigma_{R^+}$</td>
<td>0.001</td>
<td>-659.24</td>
</tr>
<tr>
<td>$R^-$ regression</td>
<td>Value</td>
<td>t-stat</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.002</td>
<td>-2.75</td>
</tr>
<tr>
<td>Lag of $R^-$</td>
<td>0.987</td>
<td>325.74</td>
</tr>
<tr>
<td>PP_DIFF</td>
<td>0.023</td>
<td>4.81</td>
</tr>
<tr>
<td>SQRT</td>
<td>0.001</td>
<td>0.97</td>
</tr>
<tr>
<td>$\sigma_{R^-}$</td>
<td>0.002</td>
<td>-2.60</td>
</tr>
<tr>
<td>$\Delta P^{N}$ regression</td>
<td>Value</td>
<td>t-stat</td>
</tr>
<tr>
<td>Constant</td>
<td>0.174</td>
<td>4.87</td>
</tr>
<tr>
<td>Lag of $\Delta P^{N}$</td>
<td>0.074</td>
<td>0.41</td>
</tr>
<tr>
<td>NG</td>
<td>-0.203</td>
<td>-4.87</td>
</tr>
<tr>
<td>ET</td>
<td>-0.034</td>
<td>-1.45</td>
</tr>
<tr>
<td>HTD$^2$</td>
<td>-0.144</td>
<td>-3.25</td>
</tr>
<tr>
<td>CTD$^2$</td>
<td>-0.211</td>
<td>-5.37</td>
</tr>
<tr>
<td>$\sigma_p$</td>
<td>0.266</td>
<td>-204.01</td>
</tr>
<tr>
<td>LLF</td>
<td>-1373.5</td>
<td>-1409.7</td>
</tr>
</tbody>
</table>
Table 1.14: Path 2 (Dover - Salisbury) parameter estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>With autocorrelation</th>
<th>Without autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value</td>
<td>t-stat</td>
</tr>
<tr>
<td>R⁺ regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.250</td>
<td>3.26</td>
</tr>
<tr>
<td>Lag of R⁺</td>
<td>0.034</td>
<td>0.18</td>
</tr>
<tr>
<td>PP_DIFF</td>
<td>-0.452</td>
<td>-3.90</td>
</tr>
<tr>
<td>SQRT</td>
<td>-0.101</td>
<td>-1.63</td>
</tr>
<tr>
<td>σ⁺</td>
<td>0.091</td>
<td>-9.04</td>
</tr>
<tr>
<td>R⁻ regression</td>
<td>Value</td>
<td>t-stat</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.003</td>
<td>-1.18</td>
</tr>
<tr>
<td>Lag of R⁻</td>
<td>1.000</td>
<td>185.33</td>
</tr>
<tr>
<td>PP_DIFF</td>
<td>0.101</td>
<td>7.16</td>
</tr>
<tr>
<td>SQRT</td>
<td>0.011</td>
<td>2.30</td>
</tr>
<tr>
<td>σ⁻</td>
<td>0.026</td>
<td>-14.29</td>
</tr>
<tr>
<td>ΔPᴺ regression</td>
<td>Value</td>
<td>t-stat</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.052</td>
<td>-2.04</td>
</tr>
<tr>
<td>Lag of ΔPᴺ</td>
<td>0.142</td>
<td>0.79</td>
</tr>
<tr>
<td>NG</td>
<td>0.087</td>
<td>3.61</td>
</tr>
<tr>
<td>ET</td>
<td>0.124</td>
<td>3.75</td>
</tr>
<tr>
<td>HTD²</td>
<td>-0.002</td>
<td>-0.04</td>
</tr>
<tr>
<td>CTD²</td>
<td>0.251</td>
<td>5.16</td>
</tr>
<tr>
<td>σ⁻</td>
<td>0.278</td>
<td>-92.90</td>
</tr>
<tr>
<td>LLF</td>
<td>-2143.7</td>
<td>-2158.6</td>
</tr>
</tbody>
</table>
Table 1.15: Path 3 (Salisbury - Eastville) parameter estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>With autocorrelation</th>
<th>Without autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value</td>
<td>t-stat</td>
</tr>
<tr>
<td>$R^+$ regression</td>
<td>1.381</td>
<td>3.54</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag of $R^+$</td>
<td>-0.254</td>
<td>-0.86</td>
</tr>
<tr>
<td>PP_DIFF</td>
<td>-1.185</td>
<td>-3.62</td>
</tr>
<tr>
<td>SQRT</td>
<td>-1.500</td>
<td>-3.82</td>
</tr>
<tr>
<td>$\sigma_{R^+}$</td>
<td>0.503</td>
<td>-5.35</td>
</tr>
<tr>
<td>$R^-$ regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.008</td>
<td>0.00</td>
</tr>
<tr>
<td>Lag of $R^-$</td>
<td>-0.009</td>
<td>0.00</td>
</tr>
<tr>
<td>PP_DIFF</td>
<td>0.000</td>
<td>0.00</td>
</tr>
<tr>
<td>SQRT</td>
<td>0.020</td>
<td>0.00</td>
</tr>
<tr>
<td>$\sigma_{R^-}$</td>
<td>0.003</td>
<td>-0.05</td>
</tr>
<tr>
<td>$\Delta P^N$ regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.318</td>
<td>6.26</td>
</tr>
<tr>
<td>Lag of $\Delta P^N$</td>
<td>0.312</td>
<td>3.01</td>
</tr>
<tr>
<td>NG</td>
<td>0.315</td>
<td>6.52</td>
</tr>
<tr>
<td>ET</td>
<td>-0.434</td>
<td>-6.48</td>
</tr>
<tr>
<td>HTD$^2$</td>
<td>0.401</td>
<td>3.57</td>
</tr>
<tr>
<td>CTD$^2$</td>
<td>-0.222</td>
<td>-4.53</td>
</tr>
<tr>
<td>$\sigma_{\Delta P}$</td>
<td>0.225</td>
<td>-51.74</td>
</tr>
</tbody>
</table>

Analysis of parameter estimates

It is instructive to examine each of the parameter estimates across all paths and comment on the commonality of the parameters, and the expected changes they make on the dependent variable. Remember that we are estimating two unobservable phenomena: the transmission-free price difference, and the capacity of the transmission line to reduce the transmission-free price. A further reduction will tell us that we are estimating the hourly loads (demand curves), the shapes of the supply curves and the transmission capacities. While expectations can be formed about what variables should have which effects on these estimates, the same cannot be said about expected prices. An increase in loads may result in an increase or decrease in the transmission-free price difference.
Variables affecting reductive capacity

These variables affect the capacity to reduce prices. There are two sets of variables: one for the assumed positive direction of power flow (from North to South) and one for the assumed negative direction.

Constant term

The expectation is that this will be positive: when temperature is low and there is no congestion in the rest of the network (*i.e.*, when other variables are valued at zero) then we expect the line to be able to reduce congestion rent by a positive amount. In the $R^-$ direction we expect the constant to be negative. The resulting parameter estimates are shown in Table 1.16.

**Table 1.16: Reductive capacity constant term**

<table>
<thead>
<tr>
<th>Path</th>
<th>R+ With autocorrelation</th>
<th>R- With autocorrelation</th>
<th>R+ Without autocorrelation</th>
<th>R- Without autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.049 (4.63)</td>
<td>-0.002 (-2.75)</td>
<td>0.066 (3.71)</td>
<td>-0.248 (-7.50)</td>
</tr>
<tr>
<td>2</td>
<td>0.250 (3.26)</td>
<td>-0.003 (-1.18)</td>
<td>0.425 (5.89)</td>
<td>-0.180 (-2.73)</td>
</tr>
<tr>
<td>3</td>
<td>1.381 (5.54)</td>
<td>0.008 (0.00)</td>
<td>1.023 (6.54)</td>
<td>0.033 (0.00)</td>
</tr>
</tbody>
</table>

As can be seen, we have the expected sign for paths 1 and 2. For path 3, we get the expected behavior in the positive direction, but insignificant behavior in the negative.

Square root of temperature

Expectations about this variable are obvious: an increase in ambient temperature should decrease the security-constrained transmission capacity of a line. Thus, the parameter should be negative in the $R^+$ direction, positive in the $R^-$ direction. The estimates are shown in Table 1.17.
Table 1.17: SQRT parameter estimates

<table>
<thead>
<tr>
<th>Path</th>
<th>With autocorrelation</th>
<th>Without autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R+</td>
<td>R-</td>
</tr>
<tr>
<td>1</td>
<td>-0.014 (-36.45)</td>
<td>0.001 (0.97)</td>
</tr>
<tr>
<td>2</td>
<td>-0.101 (-1.63)</td>
<td>0.011 (2.30)</td>
</tr>
<tr>
<td>3</td>
<td>-1.500 (-3.28)</td>
<td>0.020 (0.00)</td>
</tr>
</tbody>
</table>

Observation of Table 1.17 tells us that the temperature variable behaves as predicted in 11 of the 12 cases. In the only case where it does not meet expectations, it is not statistically significant.

**Congestion in PECO region:** As described in Table 1.6, we expect this parameter to follow the same sign behavior as the SQRT variable: negative in the positive direction, and vice versa. The estimates are shown in Table 1.18.

Table 1.18: PP_DIFF parameter estimates

<table>
<thead>
<tr>
<th>Path</th>
<th>With autocorrelation</th>
<th>Without autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R+</td>
<td>R-</td>
</tr>
<tr>
<td>1</td>
<td>-0.048 (-4.53)</td>
<td>0.023 (4.81)</td>
</tr>
<tr>
<td>2</td>
<td>-0.452 (-3.90)</td>
<td>0.101 (7.16)</td>
</tr>
<tr>
<td>3</td>
<td>-1.185 (-3.62)</td>
<td>0.000 (0.00)</td>
</tr>
</tbody>
</table>

As can be seen from Table 1.18, this parameter follows expectations in every case, and is significant in 10 of 12 cases.
Variables affecting transmission-free price difference

Natural gas prices

Parameters for the natural gas parameter, NG, are contained in Table 1.19, below. As previously mentioned, there is no unambiguous expectation on the sign of the NG parameter.

Table 1.19: NG parameter estimates

<table>
<thead>
<tr>
<th>Path</th>
<th>Parameter</th>
<th>t-stat</th>
<th>Parameter</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.203</td>
<td>-4.78</td>
<td>-0.191</td>
<td>-10.07</td>
</tr>
<tr>
<td>2</td>
<td>0.087</td>
<td>3.61</td>
<td>0.065</td>
<td>3.20</td>
</tr>
<tr>
<td>3</td>
<td>0.315</td>
<td>6.52</td>
<td>0.314</td>
<td>13.18</td>
</tr>
</tbody>
</table>

The values of the natural gas parameter tell us something about the supply curves at the two ends of the path. If the parameter is negative, then the upstream price is increasing relative to the downstream one. If it is positive, vice versa.

Elapsed time

Similar to the NG parameter, there is no unambiguous expectation on the sign of the ET parameter. The calculated values are shown in Table 1.20.

Table 1.20: ET parameter estimates

<table>
<thead>
<tr>
<th>Path</th>
<th>Parameter</th>
<th>t-stat</th>
<th>Parameter</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.034</td>
<td>-1.45</td>
<td>-0.056</td>
<td>-2.28</td>
</tr>
<tr>
<td>2</td>
<td>0.124</td>
<td>3.75</td>
<td>0.091</td>
<td>5.36</td>
</tr>
<tr>
<td>3</td>
<td>-0.434</td>
<td>-6.48</td>
<td>-0.434</td>
<td>-18.47</td>
</tr>
</tbody>
</table>

According to these data, as time passed the transmission-free price differences increased on path 2 and declined on paths 1 and 3. This is consistent with reality: much work was done to reduce chronic (but not so severe) congestion in the crowded northern area, and on the infrequent, but severe, congestion in the south. In the central region, we saw the largest population growth in the region.
Temperature-related parameters

Table 1.21 shows the heating- and cooling-temperature departure parameter estimates in the price-difference regression.

**Table 1.21: HTD and CTD parameter estimates**

<table>
<thead>
<tr>
<th>Path</th>
<th>With autocorrelation</th>
<th>Without autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HTD2</td>
<td>CTD2</td>
</tr>
<tr>
<td>1</td>
<td>-0.144 (-3.25)</td>
<td>-0.211 (-5.37)</td>
</tr>
<tr>
<td>2</td>
<td>-0.002 (-0.04)</td>
<td>0.251 (5.16)</td>
</tr>
<tr>
<td>3</td>
<td>0.401 (3.57)</td>
<td>-0.222 (-4.53)</td>
</tr>
</tbody>
</table>

Once again, instead of having expectations about the signs of these parameters based upon the theory, the signs inform us about the actual supply curves seen in the market. A negative number means that price difference decreases with load, which means that the upstream supply curve is steeper than the downstream supply curve. A positive number tells us the opposite. If the HTD and CTD numbers have the same sign, then it means that the supply curves have similar relationships in the summer and winter, when we are seeing different loads. A change in the sign means that the supply curves have different relative slopes at different parts of the curves.

**Autocorrelation**

Table 1.22, below, contains the autocorrelation parameters for the three regressions. As can be seen, in no case are all the autocorrelation parameters significant.

**Table 1.22: Autocorrelation parameter estimates**

<table>
<thead>
<tr>
<th>R^2 regression</th>
<th>R^2 regression</th>
<th>ΔP^N regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter t-stat</td>
<td>Parameter t-stat</td>
<td>Parameter t-stat</td>
</tr>
<tr>
<td>Path 1 0.177 2.78</td>
<td>0.987 325.74</td>
<td>0.074 0.41</td>
</tr>
<tr>
<td>Path 2 0.034 0.18</td>
<td>1.000 185.33</td>
<td>0.142 0.79</td>
</tr>
<tr>
<td>Path 3 -0.254 -0.86</td>
<td>-0.009 -0.00</td>
<td>0.312 3.01</td>
</tr>
</tbody>
</table>
Four of the nine parameters are statistically significant. Two parameters are negative, but not statistically significant. There is no obvious underlying economic justification for some components along some path exhibiting autocorrelation – we would expect the occurrence to either exist or not in some consistent fashion. One possible cause for this problem is as follows: the non-continuity of the data. Since we are looking only at on-peak data, there is a gap of eight hours every night where data were not analyzed, since 11 pm to 7 am are off-peak hours. Also, weekends and NERC holidays are off peak. Thus, when looking at autocorrelation, some of the data points are contiguous, and others are separated by between 8 and 80 hours (or 2 to 20 four-hour intervals). Another problem arises from the non-continuity of the calculation: when price is positive, the negative reductive capacity does not enter into the log-likelihood function, and vice-versa. If we go for long periods observing a positive or negative price, we have long periods where the opposing reductive capacity is not only non-observable, but also non-existent. At these points, we can use the estimated parameters to generate a value for the opposing reductive capacity, but because those points do not enter into the calculation, we are essentially extrapolating a value on a point that exists outside our data sample set.

Table 1.23 contains some information about positive and negative price streaks. As can be seen, long positive-price streaks are far more common than negative ones on paths 1 and 3, but about as frequent on path 2. Furthermore, the longest observed positive-price streaks are much longer than the negative-price ones. This lends some evidence for the differences between the positive and negative reductive capacity autocorrelation estimates, as well as the different behavior on path 2. In either case, the non-continuity of the on-peak data leaves renders the use of autocorrelation questionable, due to the differential length between so-called “consecutive” observations.

Table 1.23: Positive and negative price streaks

<table>
<thead>
<tr>
<th>Path:</th>
<th>1 - Wilmington-Dover</th>
<th>2 - Dover-Salisbury</th>
<th>3 - Salisbury-Eastville</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price:</td>
<td>Positive</td>
<td>Negative</td>
<td>Positive</td>
</tr>
<tr>
<td>10-period streaks</td>
<td>61</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>50-periods streaks</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Longest streak</td>
<td>79</td>
<td>31</td>
<td>55</td>
</tr>
</tbody>
</table>
In addition to the aforementioned points, the validity of excluding autocorrelation in the model can be empirically tested by using a likelihood-ratio test. Not employing autocorrelation can be viewed as the same thing as employing autocorrelation but restricting the autocorrelation parameters to a value of zero. We can test whether applying this restriction is statistically significant or not by comparing the calculated values of the log-likelihood function. For the case of autocorrelation, the likelihood ratio test statistic is defined as:

$$LR = 2[LLF(\alpha^+, \alpha^-, \beta, \rho) - LLF(\alpha^+, \alpha^-, \beta, 0)]$$  

(64)

The likelihood-ratio test statistic is distributed according to the $\chi^2$ distribution. The results of the LR test are shown in Table 1.24. The p-value was calculated using the Microsoft Excel CHIDIST function assuming 14 degrees of freedom, this being the number of non-restricted parameters.

<table>
<thead>
<tr>
<th>Path:</th>
<th>1 - Wilmington-Dover</th>
<th>2 - Dover-Salisbury</th>
<th>3 - Salisbury-Eastville</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrestricted LLF</td>
<td>-1373.5</td>
<td>-2143.7</td>
<td>-2061.1</td>
</tr>
<tr>
<td>Restricted LLF</td>
<td>-1409.7</td>
<td>-2158.6</td>
<td>-2062.2</td>
</tr>
<tr>
<td>LR statistic</td>
<td>72.4</td>
<td>29.8</td>
<td>1.1</td>
</tr>
<tr>
<td>p-value</td>
<td>0.00</td>
<td>0.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 1.24: Likelihood-ratio tests for autocorrelation

Thus, at the 95% confidence level, we fail to reject the null hypothesis that $\rho = 0$ for path 3, but not paths 1 and 2. Given the high p-values for one of the three paths and the non-continuity of the data, it is perhaps wise to prefer the model without the inclusion of autocorrelation.

**Impact analysis**

When using an ordinary least-squares regression, a parameter estimate can be easily transformed into an elasticity, describing the percent change in the dependent variable given a percent change in some independent variable. This is helpful in defining the relative impact of certain variables. However, such an analysis is not nearly as straightforward in a maximum-likelihood formulation. The dependent variables in this analysis are the log-likelihood values. To calculate the impact of changes in some
dependent variable, we need to examine changes in the value of the log-likelihood function, which is not linearly related to the parameter estimates. Instead, the function has to be evaluated at different values of the independent variable to calculate the impact. In this case, the likelihood function has three parts corresponding to three different price regimes, so a change in the composite likelihood function is a result of changes in the three constituent parts. Thus, it is more instructive to observe the impact on each price regime separately; that is, how does a change in some $X$ affect the probability that $y = \theta$, the likelihood that $y = y_{i[y>\theta]}$ and the likelihood that $y = y_{i[y<\theta]}$.

It is not necessary to test the impact of changing SQRT, HTD2 and CTD2 separately, since all are derived from the same variable: temperature. The standard deviation of temperature over the course of the study was $18.3^\circ F$, so the impact of both increasing and decreasing every observation of temperature by this amount was calculated. Both positive and negative changes were examined in order to address asymmetries in the response. The impact of a change in the price of natural gas was also tested. The standard deviation of the price of natural gas over the course of the study was $\$1.00/mmBTU$, and the responses to both positive and negative changes by this amount were calculated. The normal practice in an impact analysis is to set all the independent variables to their average value and then modify them one at a time by adding (subtracting) the appropriate standard deviation to (from) the average value. This is not an appropriate method in this case, because if we take the average values, we have only a single observation at a point that falls into one of the three price regimes. As mentioned above, we are interested in examining the effect of a change in a variable on the three separate parts of the log-likelihood function. Thus, to measure the impact of a change instead we take each of the observed values and add or subtract the appropriate standard deviation. Instead of observing the impact on an average observation, we have the average of the impacts on each observation. The results of an impact analysis on Path 1 are shown in Table 1.25, below. The first column contains the (default) maximum likelihood values for the probability and likelihoods. In the subsequent columns, the percent change on each of the default values resulting from the indicated one-standard deviation change in the independent variable is shown.
Table 1.25: Path 1 impact analysis

<table>
<thead>
<tr>
<th>Perturbation:</th>
<th>Default</th>
<th>NG + σ</th>
<th>NG − σ</th>
<th>T + σ</th>
<th>T − σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pr(y = 0)</td>
<td>0.704</td>
<td>-0.7%</td>
<td>+1.3%</td>
<td>+4.5%</td>
<td>-1.8%</td>
</tr>
<tr>
<td>L(y &gt; 0)</td>
<td>0.746</td>
<td>+31.7%</td>
<td>-30.2%</td>
<td>+53.8%</td>
<td>+25.3%</td>
</tr>
<tr>
<td>L(y &lt; 0)</td>
<td>0.661</td>
<td>-13.6%</td>
<td>+14.1%</td>
<td>-20.7%</td>
<td>-1.0%</td>
</tr>
</tbody>
</table>

From Table 1.25 we can see that the results are not symmetrical: for example, the positive likelihood values are affected far more by all perturbations than the negative likelihood, and this much more than the zero probability in most cases. Notably, $L(y>0)$ is improved and $L(y<0)$ damaged by either increasing or decreasing temperature.

Some conclusions can be drawn from examination of the parameters:

1. In the reductive capacity regressions, the sign of the constant term is consistent expectations in every case in which it is significant.
2. Temperature increases cause the reductive capacity to shrink, consistent with theory.
3. Congestion on other paths leads to a reduction in reductive capacity. This is as anticipated.
4. Natural gas prices are significant, but vary in sign. The parameter is negative in the north but positive in the south.
5. Transmission-free price differences increase with time in the central region, but decrease with time elsewhere. This is consistent with population growth patterns.
6. The effects of heating and cooling loads are also different in different parts of the region of study.
7. Autocorrelation does not appear to be an appropriate estimation tool, given discontinuities of the data. The test statistic in a likelihood ratio test is significant for paths 1 and 2, but not for paths 3.

Increased loads do not appear to create higher transmission-free price differences, but factors that cause high loads (time of day, temperature) tend to decrease the reductive
capacity of a network. The second of these two statements corroborates the claim of more congestion at high-load times, but for congestion to exist, we need transmission-free price differences. Thus, the two above statements seem to be at odds with each other. For congestion to exist, we need elevated $\Delta P^N$ and reduced $R^+$ and $R^-$. The model as explained above only clearly identifies one half of this pair of requirements. Clearly, something is missing in the analysis. Further discussion of this follows.

The model as predictor of prices

Figures 1.24 through 1.26 display the 12-period moving average of modeled values of the three unobserved parameters for each of the three paths. The moving average is presented to smooth the data and make the figure more legible. Whenever the $Xb(\Delta P^N)$ line is outside the bounds of the two $Za$ ($R$) lines, we observe a non-zero price.

![Figure 1.24: Estimated system components, Path 1](image)
Figure 1.25: Estimated system components, Path 2

Figure 1.26: Estimated system components, Path 3
Examination of the above figures yields some interesting observations not obvious from the parameter estimates. The negative reductive capacity component (the blue line in the graphs) behaves in a similar fashion along paths 1 and 2, exhibiting strong seasonal dependencies. The positive reductive capacities also exhibit clear seasonal oscillation. The transmission-free price difference exhibits a weak time trend with some observable seasonality along path 2 but not on paths 1 and 3. However, the behavior of the reductive capacities along path three deviates greatly from those on paths 1 and 2. As can be seen, the negative capacity is basically constant at a value of zero, and the positive capacity shows a much stronger seasonal oscillation. The reasons for this different behavior are not obvious. During the modeling process, paths 1 and 2 converged to the optimal parameter values with relative ease. However, there was great difficulty in obtaining convergence along path 3. Various initial estimates would converge to different optima, indicating the presence of local maxima in the log-likelihood function. This, obviously, leads to the possibility of obtaining a false optimum point. It is impossible to perform a quick test for the existence of multiple critical points in a system with 16 variables. Each of the variables, independently, showed a single optimal value. Since the problem was optimized using numerical estimation techniques, these derivatives were not found in analytical form, so it was not possible to prove uniqueness of a maximum. However, the function was evaluated at several values over a wide range for each parameter, and in each case a single maximum was observed. I am unaware as to whether the existence of a unique maximum with respect to each variable is a sufficient condition for the existence of a single multivariate optimum.

Examination of Figures 1.17 – 1.19, the observed prices along the paths, shows that path 3 prices appear structurally different to those on paths 1 and 2, especially for the period from February to May 2001, where we have the 198-period positive price streak. This is the only obvious clue from the data that we might observe drastically different behavior along the different paths. The ill-behaved outcome for path 3 may be the result of an aberrant local maximum, but it is also possible that a remnant of the numerical modeling methodology has yielded a corner solution. Examining the parameter estimates in $R^2$ regression in Table 1.14, the extremely low value of the variance stands out. In both of
the other paths, the $\sigma_R$ value is a couple of orders of magnitude smaller than the $\sigma_{R+}$ and $\sigma_P$ values, but for path 3 it is four orders of magnitudes smaller. This is a clue to the existence of a corner solution. A further clue is the fact that the derivative of the log-likelihood function with respect to the variance is very close to non-continuous at this point. Since we are evaluating a normal distribution function, at extreme values the value of the function goes to zero. Since we are modeling numerically with discrete intervals, at extreme values the function moves, step-wise, from some finite value to zero. Aside from the problem of estimating a derivative of a non-continuous function that is non-continuous at the critical value, this also leads to either a “divide by zero” or “log of zero” error. This problem was addressed by the standard numerical modeler’s trick of adding a small constant value to the denominator of a fraction and to any primitive of a logarithmic function. Thus, in the places where the function wanders into extreme territory, constraints are artificially imposed by the modeler, possibly resulting in a perverse or erroneous result. Many months were spent seeking convergence on the path 3 estimator, and the extreme value and non-continuity problems were frequently encountered. It is possible that there is, in fact, no “true” optimum value for the log-likelihood function along path 3, and the ill-behaved result that was obtained is in fact an artifact of the modeler’s attempt to torture the data into yielding an answer where none existed. As has been shown, information obtained from torture is usually wrong, and that possibility certainly exists in this case.

Maximum-likelihood modeling is primarily used for identifying significant causal variables. One reason for identifying these is to create a model that can be used to forecast future behavior. In the meteorological world, the practice of examining the accuracy of past predictions is known as forecast verification. Many methods of binary and discrete forecast verification have been devised and are explained in Wilks (1995). Employing some of these tools to examine the predictive capacity of the model as devised in this essay can be used. When assessing the value of a model as predictor of future behavior, it is necessary to examine how well it forecasts the past; that is, do the predicted values match the observed values?
For each of the paths studied in this essay, a three by three matrix was produced. The columns of the matrix correspond to the signs of the actual, observed prices. The rows correspond to the sign of the predicted values of the prices. For examples, if the actual price at some time was positive, and the predicted value, using the modeled values of $\Delta P^N$, $R^+$ and $R^-$, was also positive at the same time, then we have an entry into the cell in the matrix corresponding to the “positive, positive” pair. Tables 1.26 through Table 1.28 contain the observation – prediction matrices for the three paths, using the estimated parameters without autocorrelation. Employing autocorrelation had negligible effects on the results.

**Table 1.26: Observed versus predicted prices, path 1**

<table>
<thead>
<tr>
<th></th>
<th>&gt; 0</th>
<th>= 0</th>
<th>&lt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted price</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 0</td>
<td>1259</td>
<td>233</td>
<td>504</td>
</tr>
<tr>
<td>= 0</td>
<td>980</td>
<td>296</td>
<td>471</td>
</tr>
<tr>
<td>&lt; 0</td>
<td>73</td>
<td>6</td>
<td>90</td>
</tr>
</tbody>
</table>

**Table 1.27: Observed versus predicted prices, path 2**

<table>
<thead>
<tr>
<th></th>
<th>&gt; 0</th>
<th>= 0</th>
<th>&lt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted price</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 0</td>
<td>278</td>
<td>27</td>
<td>145</td>
</tr>
<tr>
<td>= 0</td>
<td>1160</td>
<td>1325</td>
<td>968</td>
</tr>
<tr>
<td>&lt; 0</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 1.28: Observed versus predicted prices, path 3**

<table>
<thead>
<tr>
<th></th>
<th>&gt; 0</th>
<th>= 0</th>
<th>&lt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted price</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 0</td>
<td>990</td>
<td>872</td>
<td>482</td>
</tr>
<tr>
<td>= 0</td>
<td>291</td>
<td>915</td>
<td>355</td>
</tr>
<tr>
<td>&lt; 0</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>
The data in the above tables can be converted to a simple skill score, which is the sum of the diagonal terms divided by the sum of all entries. The diagonal entries are “correct” predictions, so the simple skill score is the frequency of correctly predicting the sign of the price. The partial skill scores can then be calculated: for each category of observed values, we look at the frequency of correct predictions. Any phenomenon that has a large percentage of occurrences in one state will often have a large simple skill score; the partial skill scores highlight the accuracy in each category. Table 1.29 lists the simple and partial skill scores for each of the three paths.

**Table 1.29: Skill scores**

<table>
<thead>
<tr>
<th></th>
<th>Simple</th>
<th>Positive partial</th>
<th>Zero partial</th>
<th>Negative partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path 1</td>
<td>41.6%</td>
<td>53.7%</td>
<td>55.3%</td>
<td>8.5%</td>
</tr>
<tr>
<td>Path 2</td>
<td>41.0%</td>
<td>19.2%</td>
<td>98.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Path 3</td>
<td>48.8%</td>
<td>77.2%</td>
<td>51.2%</td>
<td>0.7%</td>
</tr>
</tbody>
</table>

From Table 1.29 we can see that that the simple skill scores are consistent for each path: the model predicts the correct sign 40-50% of the time. However, the partial scores are more revealing: the model can be excellent at predicting when the price will be zero or positive, but the prediction of negative prices is particularly poor – they almost never predicted. The bias of the prediction is shown in Table 1.30. The bias is defined as the number of predicted occurrences of a certain price regime divided by the number of observed occurrences. If it is greater than one the model is over-forecasting a certain regime; if less than one, it is under-forecasting.

**Table 1.30: Forecast bias**

<table>
<thead>
<tr>
<th></th>
<th>Positive price bias</th>
<th>Zero price bias</th>
<th>Negative price bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path 1</td>
<td>0.87</td>
<td>3.25</td>
<td>0.16</td>
</tr>
<tr>
<td>Path 2</td>
<td>0.29</td>
<td>2.55</td>
<td>0.00</td>
</tr>
<tr>
<td>Path 3</td>
<td>2.19</td>
<td>0.61</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Examination of Table 1.30 tells us that prices of zero are greatly over-forecasted along paths 1 and 2, and positive prices are over-forecasted on path 3. In every case, negative prices are greatly under-forecasted.

Since it is the quality of forecast that we are interested in generating from this entire exercise, instead of maximizing a likelihood function, a possible alternative model is a “maximum skill score” model, where the objective function is the simple skill score, and this is maximized with respect to all the estimated parameters. However, developing such a function presents several difficulties. Most obviously, the estimated price is not a continuous function: if $|\Delta P^N| > |R|$, then $y > 0$, else $y = 0$. Secondly, the entries into the observed-predicted values matrix are not continuous variables, but discrete entities. We are faced with a function with two multivariate “if” statements. Furthermore, the skill score is not a continuous function; moving an observation from one cell of the matrix to another causes a step-wise change in the skill score. Hence, the derivative of the skill score with respect to a parameter is not continuous. This is not an insurmountable problem: since we are using a finite-differencing method for estimating derivatives, with a discontinuous estimating interval, we should be able to find optimal points. However, extra care would have to be taken within such a model to ensure that we do not get false predictions of zero-valued derivatives. It remains to be seen whether a maximum-skill-score model will yield higher simple and partial skill scores than a maximum-likelihood model.

As mentioned above, the most obvious feature of Tables 1.26-1.28 is the almost complete absence of negative price predictions. Over all three paths, a total of 3,021 negative prices were observed. However, only 185 negative predictions were made – 7.4% of all cases – and 169 of these were made over path 1. We saw 1,214 cases of a positive-price prediction: the wrong sign showed up over five times as often as the right one. Why is this? In order to predict negative signs, we need to have a negative value of $\Delta P^N$, as well as a low $R$. The model as developed here does a poor job of predicting negative prices. One question that can be asked: is the lack of negative prices because of non-negative values of $\Delta P^N$, or values of $R$ that are too large? A quick analysis of the negative prices
along path 2 was undertaken. This path was chosen because of the extremely low number of negative predictions. There were 1,112 observed negative prices along this path. In those 1,112 cases a negative value of $\Delta P^N$ was calculated 462 times. That is, the model correctly predicted the sign of $\Delta P^N$ 41.5% of the time. Examining these 462 cases, the value of $\Delta P^N$ was compared with that of $R$. It was discovered that in only nine of 462 cases was $R$ smaller than $\Delta P^N$ (in absolute-value terms). This yields nine predicted negative prices in 1,112 instances. The most obvious statement that can be made here is that either the reductive-capacity values calculated are consistently too large, or that we do not exhibit enough short-term variability in the reductive capacity to provide us with sufficient sign-consistent estimates. Given that reductive capacity is unobservable, this is a difficult statement to empirically test, beyond what has already been done in this essay. However, it is clear that something is missing from the analysis.

What is the missing variable? Outages. The model, as developed above, assumes that the network is the same at all times, or changes only gradually over time. However, this is not the case. Much of the time, some part of the transmission network is out of service. For example, in his Delmarva congestion study for PJM, Whitehead noted that “2/3 of constrained hours had concurrent transmission outages.” Further analysis of Whitehead’s paper shows that 34% of the congestion was caused by construction-related outages, 10% by maintenance-related outages, and 13% by forced outages (i.e., unforeseen system failures.). The other 43% of the congested hours see no outages – that is, all congestion is caused by demand for transmission that exceeds installed capacity. In this model, all of the prices caused by system outages must reside in the error term. This is the source of much of the error here. When PJM generates the schedule of day-ahead LMPs, it can be assumed that this calculation is being performed over a network model that includes all scheduled outages. A naive interpretation of Whitehead’s numbers tells us that only 23% (13/57) of the outages are truly stochastic and unforeseen. Clearly, the model as developed here forces a large amount of known information into the error term. By “known information”, I am referring to that which is known to the system operator and the market participants when bidding for power, but not to the author of this essay.
Another weakness of this model lies in the modeling of the supply curves. As shown in Figure 1.15, changes in transmission-free price difference are a function of the two loads, which can be reasonably assumed to be vertical lines, and the height and slopes of the supply curves at the given loads. In this study, only a single parameter is being used to model this complicated phenomenon. As we move along the supply curve, the slope changes, and if load gets high enough, we face capacity constraints. At this point, the supply curve “goes vertical”, and faced with the intersection of two near-parallel lines, we have extreme price hikes. In addition to the concept of physical capacity constraints, we also have to consider strategic behavior by generators. Bidding into the day-ahead market is a repeated game that is played once daily. Over time, players develop a familiarity with certain aspects of the game. If a generator observes that one of his facilities is frequently the marginal power supplier, i.e., the price setter, then he can modify his bids to maximize his profits. None of this information is captured by using the price of natural gas as the sole estimator for the supply curve. Furthermore, the price of natural gas in this study is a monthly value as reported by the generators to FERC. Thus, absolutely none of the short-term changes in the supply curve can be captured.

**Ex-ante information**

All estimators used in this study are known before the fact except for the PECO-PJM price difference. This variable is included to model the effects of the loop flow externality, and due to the real-time nature of electricity flow, ex-ante data is meaningless. The loop-flow externality is an important component of the behavior of electricity networks, and the fact that the PP_DIFF variable is always statistically significant in the estimates in this essay is testament to that fact. However, if one wishes to have a true forecasting model, it is necessary to model using only *ex-ante-facto* information. What would happen if the loop-flow externality were not considered? What if the reductive capacities were modeled as constants that varied only with temperature? Such a system was modeled for Path 1. The value of the log-likelihood function for this restricted function was -1431.3, compared to a value of -1409.7 for the unrestricted estimate. Employing a likelihood-ratio test we get a p-value of 0, implying that the PP_DIFF parameters are statistically important. Another metric that has been employed
above is the skill score. The skill scores for the restricted and unrestricted estimators are shown in Table 1.31. We can see that excluding PP_DIFF damages the ability of the model to predict positive prices, but does not hurt that already very low ability to predict negative prices. The overall skill score declines by about 7%.

Table 1.31: Comparison of path 1 skill scores

<table>
<thead>
<tr>
<th></th>
<th>Simple</th>
<th>Positive partial</th>
<th>Zero partial</th>
<th>Negative partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrestricted</td>
<td>41.6%</td>
<td>53.7%</td>
<td>55.3%</td>
<td>8.5%</td>
</tr>
<tr>
<td>Restricted</td>
<td>38.5%</td>
<td>49.1%</td>
<td>52.9%</td>
<td>8.4%</td>
</tr>
</tbody>
</table>

This illustrates the non-linearity between the log-likelihood function and the skill-score estimator: in the above case, the log-likelihood function decreased only 1.5%, but the skill score declined 7%.
1.11 Conclusions

This paper studies the use of an arbitrage-cost methodology employing three non-observable components for estimating the causative factors of transmission congestion in the Delmarva Peninsula. The results of this study were mixed: three transmission paths were examined, and results that appeared to be prima facie plausible were observed for two of these paths. The results for a third path appear to show a corner solution that does not properly describe the real-life observations.

The results are consistent with the physics of electricity networks: an increase in the ambient temperature causes a reduction in the ability of transmission lines to mitigate price differences by reducing the maximum capacity for power flow. Furthermore, the loop-flow externality is real and observable: congestion in some other part of the network has measurable and significant effects on the ability of the transmission lines in question to reduce congestion.

Price differences in the absence of transmission are function of local supply and demand curves, or given the assumption of vertical demand curves, functions of local loads. For the model to be able to work well it is necessary to be able to model loads and generation cost schedules at each end of a path. This is difficult to do without a large number of variables. Load was modeled as a function of either heating or cooling loads, with long-term changes modeled by an elapsed time variable. The sole variable for estimating the cost function was the price of natural gas. Perhaps the greatest weakness of the model as specified is that supply-cost modeling is so crude: since natural gas prices are assumed to change only monthly, there is no short-term variation in the supply curves. This is not consistent with reality: as different generators are brought on- and off-line the cost of supplied power changes drastically. Furthermore, there is no capacity for modeling capacity constraints or system gaming by the generation owners, which are important factors in explaining short-term price spikes.
1.12 References


Appendix A: Network calculations

We have a triangular three-node network with two generation nodes, labeled Node 1 and Node 2, and a consumption node labeled Node 3. There are security-constrained limits on the flows in the three lines connected the nodes: 100 MW on Line 1-2, 220 MW on Line 2-3 and 300 MW on Line 1-3.

Flows along each line are defined by the symbol Q subscripted with the beginning and end nodes of the line. If a power flow is injected into or removed from the network the other end of the path in question will be the “outside world”, defined as node zero. Thus, the injection of power into the network by the generator at Node 1 is labeled as $Q_{01}$, that injected at Node 2 as $Q_{02}$, and that removed at Node 3 as $Q_{30}$. Note that in this convention $Q_{30} = -Q_{03}$, and so on. This network, with the three binding flow constraints listed in parentheses beside each line, is shown in Figure A.1.

![Sample three-node network](image)

Figure A.1: Sample three-node network
Case A1: low-load period.

We wish to solve for the power flows and nodal prices during a low-load period. Injected and removed power flows are governed by the supply and demand functions. During the low-load period, we have the following supply and demand functions:

Node 1 supply function: \( P_1 = 10 + 0.05Q_{01} \) \( (A1) \)

Node 2 supply function: \( P_2 = 20 + 0.1Q_{02} \) \( (A2) \)

Node 3 demand function: \( P_3 = 80 - 0.2Q_{30} \) \( (A3) \)

The problem is solved by minimizing the price paid by consumers. The amount consumed is equal to the sum of the amount generated:

\[ Q_{01} + Q_{02} + Q_{03} = 0 \]

Which, when written with our sign convention evolves to:

\[ Q_{30} = Q_{01} + Q_{02} \] \( (A4) \)

By substitution of Equation (A4) into Equation (A3), the independent system operator’s objective function becomes:

\[ \min_{Q_{01}, Q_{02}} \left( 80 - 0.2(Q_{01} + Q_{02}) \right) \] \( (A5) \)

The first constraint comes from the fundamental concept of marginal pricing: the consumer shall pay the amount that is equal to the highest dispatched generator:

\[ P_3 = \max(P_1, P_2) \]

Written in terms of load, by substituting the supply and demand functions, we get:

\[ 80 - 0.2(Q_{01} + Q_{02}) = \max(10 + 0.05Q_{01}, 20 + 0.1Q_{02}) \] \( (A6) \)

We have the further constraints that generators do not consume any power:

\[ Q_{01} \geq 0, \text{ and:} \]
\[ Q_{02} \geq 0 \] \( (A7) \)

\( (A8) \)

Summarizing, the optimization problem involves solving the objective function:

\[ \min_{Q_{01}, Q_{02}} \left( 80 - 0.2(Q_{01} + Q_{02}) \right) \] \( (A5) \)

Subject to the constraints:

\[ 80 - 0.2(Q_{01} + Q_{02}) = \max(10 + 0.05Q_{01}, 20 + 0.1Q_{02}) \] \( (A6) \)

\[ Q_{01} \geq 0, \text{ and:} \]
\[ Q_{02} \geq 0 \] \( (A7) \)

\( (A8) \)
We are also interested in calculating the flow in each of the transmission lines. This is done by applying Kirchhoff’s Voltage law, which states that the sum of voltage drops around a loop shall be zero. Assuming, for simplicity, symmetry in the length and impedance of the three lines, we obtain the three following equations:

\[
Q_{12} = \frac{Q_{01}}{3} - \frac{Q_{02}}{3} \quad (A9)
\]

\[
Q_{13} = \frac{2Q_{01}}{3} + \frac{Q_{02}}{3} \quad (A10)
\]

\[
Q_{23} = \frac{Q_{01}}{3} + \frac{2Q_{02}}{3} \quad (A11)
\]

The optimization problem described by Equations (A8) – (A11) was solved using the Microsoft Excel Solver add-in. The calculation results are shown in Table A.1.

<table>
<thead>
<tr>
<th>Node</th>
<th>Flow, MW</th>
<th>Price, $/MW</th>
<th>Revenue, $ (Payment)</th>
<th>Path</th>
<th>Flow, MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>257.14</td>
<td>22.86</td>
<td>(5,877.48)</td>
<td>1-2</td>
<td>76.19</td>
</tr>
<tr>
<td>2</td>
<td>28.57</td>
<td>22.86</td>
<td>(653.12)</td>
<td>1-3</td>
<td>180.95</td>
</tr>
<tr>
<td>3</td>
<td>285.71</td>
<td>22.86</td>
<td>6,530.60</td>
<td>2-3</td>
<td>104.76</td>
</tr>
</tbody>
</table>

As can be seen, all prices are the same, which means there is no congestion rent. This is corroborated by the fact that the sum of the revenues is zero ($6,530.60 - $5,877.48 - $653.12 = $0.00). That is, all monies paid by consumers go to the generators, and none to the transmission operator.

**Case A2: high-load period, no constraints on path flow**

We wish to solve for the power flows and nodal prices during a high-load period, but without considering the effects of line flow constraints. During the high-load period, we have the following supply and demand functions:

Node 1 supply function: \( P_1 = 10 + 0.05Q_{01} \) (same as low-load period) \( (A1) \)

Node 2 supply function: \( P_2 = 35 + 0.1Q_{02} \) \( (A12) \)

Node 3 demand function: \( P_3 = 140 - 0.2Q_{30} \) \( (A13) \)
As in Case A1, the problem is solved by minimizing the consumer price. Summarizing, this optimization problem involves solving the objective function:

$$\min_{\theta_0, \theta_0} \left( 140 - 0.2(Q_{01} + Q_{02}) \right)$$

Subject to the constraints:

$$140 - 0.2(Q_{01} + Q_{02}) = \max\left(10 + 0.05Q_{01}, 35 + 0.1Q_{02}\right)$$

$$Q_{01} \geq 0, \text{ and:}$$

$$Q_{02} \geq 0$$

This optimization problem was solved using the Microsoft Excel Solver add-in. The calculation results are shown in Table A.2.

**Table A.2: Flows and prices, high-load period without line flow constraints**

<table>
<thead>
<tr>
<th>Node</th>
<th>Flow, MW</th>
<th>Price, $/MW</th>
<th>Revenue, $ (Payment)</th>
<th>Path</th>
<th>Flow, MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>514.29</td>
<td>35.71</td>
<td>(18,367.41)</td>
<td>1-2</td>
<td>169.05</td>
</tr>
<tr>
<td>2</td>
<td>7.14</td>
<td>35.71</td>
<td>(255.05)</td>
<td>1-3</td>
<td>345.24</td>
</tr>
<tr>
<td>3</td>
<td>521.43</td>
<td>35.71</td>
<td>18,622.47</td>
<td>2-3</td>
<td>176.19</td>
</tr>
</tbody>
</table>

As can be seen, all prices are the same, which means there is no congestion rent. In a network where no flow constraints are being violated, we always expect to see a uniform price and no congestion rent, since we always have the optimal merit-order power dispatch. However, we may now wish to compare the calculated line flows to the constraints displayed in Figure A.1. As we can see, the flows along Line 1-2 and Line 1-3 exceed the capacity constraints on those lines (169.05 > 100 and 345.24 > 300). Thus, we need to solve the case with binding line capacity constraints.

**Case A3: high-load period, binding constraints on path flow**

In this example, the optimization problem is identical to that in Case A2 except we have the three line-flow constraints applied to the objective function. For Case A3 the optimization problem is fully specified as:

$$\min_{\theta_0, \theta_0} \left( 140 - 0.2(Q_{01} + Q_{02}) \right)$$

(A14)
Subject to the constraints:

\[ P_3 = \max(P_1, P_2) \]  \hspace{1cm} (A15)
\[ Q_{01} \geq 0 \]  \hspace{1cm} (A7)
\[ Q_{02} \geq 0 \]  \hspace{1cm} (A8)
\[ Q_{12} \leq 100 \]  \hspace{1cm} (A15)
\[ Q_{13} \leq 300 \]  \hspace{1cm} (A16)
\[ Q_{23} \leq 220 \]  \hspace{1cm} (A17)

This optimization problem was solved using the Microsoft Excel Solver add-in. The calculation results are shown in Table A.3.

**Table A.3: Flows and prices, high-load period with line flow constraints**

<table>
<thead>
<tr>
<th>Node</th>
<th>Flow, MW</th>
<th>Price, $/MW</th>
<th>Revenue, $ (Payment)</th>
<th>Path</th>
<th>Flow, MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>390.00</td>
<td>29.50</td>
<td>(11,505.00)</td>
<td>1-2</td>
<td>100.00</td>
</tr>
<tr>
<td>2</td>
<td>90.00</td>
<td>44.00</td>
<td>(3,960.00)</td>
<td>1-3</td>
<td>290.00</td>
</tr>
<tr>
<td>3</td>
<td>480.00</td>
<td>44.00</td>
<td>21,120.00</td>
<td>2-3</td>
<td>190.00</td>
</tr>
</tbody>
</table>

In this case, the flow constraint on Line 1-2 is binding. The amount of congestion rent collected is the sum of revenues: $21,120 – $3,960 – $11,505 = $5,655.
2 Bidder behavior in PJM FTR auctions

2.1 Introduction

Congestion sometimes exists in electricity transmission grids. The organization responsible for operating the transmission network in the Mid-Atlantic region of the United States, PJM Interconnect L.L.C. (PJM) uses a node-based pricing system to value access to its grid, and during periods of congestion, different nodes in the grid will exhibit different prices. This is in line with Federal Energy Regulatory Commission (FERC) initiatives promoting market mechanisms in all aspects of the electricity business in the United States (FERC, 1996). While establishing market-based prices for network transmission, this system presents two difficulties. Firstly, the system operator is a not-for-profit organization, and as such, must redistribute revenue collected to the system customers. Secondly, congestion can lead to great geographical variation in prices. Load-serving entities (i.e., companies that deliver energy to end-users) are typically regulated by state-level utilities commissions and are barred from passing these extra costs through to their customers. Thus, it is desirable for load-serving entities to have a mechanism for hedging against congestion-based price volatility in the wholesale energy market.

An instrument has been developed to address both of these issues: the financial transmission right (FTR). FTRs were first proposed and since developed by Hogan (1991, 2002.) A simple but thorough overview is presented by Alsac, et al. (2004). FTRs are sold in competitive annual and monthly auctions conducted by PJM. Member firms submit bid schedules for FTRs and are awarded them based upon whether they bid above or below some common market clearing price. Possession of an FTR is a claim against congestion revenues collected by PJM.

In this chapter the actions of participants in monthly FTR auctions are examined. Several assertions about firm behavior in uniform-price divisible-goods auctions contained in the auction-theory literature are empirically examined in the context of PJM FTR markets for the first time. Among the questions that are examined are:
• How do firms adjust to information from past auctions?
• Do firms consistently bid below the value of the auctioned good?
• Do firms with better information outperform others in the auction? (Corollary: are there firms with better information than others?)
• Do hedgers and speculators behave differently?
• Do firms perform better with increased experience in the FTR market?
• Do firms that submit complicated bids perform better than those that submit simple bids?

This chapter is organized as follows. In Section 2, congestion in transmission grids is discussed in more depth. Section 3 contains details of the PJM FTR market mechanism. In Section 4, activity in recent PJM FTR auctions is examined on a firm-level basis. A literature review is contained in Section 5. The economic formulation for studying firm bidding behavior is developed in Section 6, and Section 7 contains details of the econometric model specifications. Section 8 contains descriptions of the data used in this analysis. The results are presented and discussed in Section 9, and conclusions are drawn in Section 10.
2.2 Congestion in electricity transmission grids

As described in Chapter 1, electricity networks suffer from congestion. Every component in a transmission grid has an upper limit on the rate at which electrical energy is allowed to flow across that component. These limits are established by physical and economic considerations. The flow rate of electrical energy is called the power flow and is measured in megawatts (MW). As the power flow along a line (or across a transformer or switch) increases, the resistance to flow increases in a quadratic fashion. The energy used to overcome this resistance generates heat, resulting in higher operating temperatures, and, as temperatures climb, the probability of component failure increases. The energy that drives this temperature rise is the electricity flowing across the line, so in addition to stressing equipment, pushing more power down a line will lead to greater energy losses in the form of heat transfer to the environment. For these reasons, upper limits on power flows are imposed upon all components of an electricity grid. These are known as the security-constrained limits.

Congestion occurs when network users would like to send an amount of power along a line that exceeds the security-constrained limit. When this happens, transmission becomes a scarce and competitive resource, and in modern power markets access to scarce transmission is allocated by willingness to pay. Thus, when congestion exists and transmission is scarce, its value is established by the prices bid by those seeking to use the grid. Prices at all points in the network are adjusted until the power flows demanded are equal to (or less than) the security-constrained limit in every component of the grid.

A numerical example of congestion in a simple electricity grid is shown in Chapter 1 and Appendix A of this study.

Under conditions of congestion, electricity will be priced at different levels in different parts of the grid. In the region being examined in this study - that covered by PJM Interconnect - the Independent System Operator (ISO) is tasked with collecting revenues from load-serving entities and power-wheeling customers and paying generators for
supplied electricity. The grid is modeled as a collection of flow paths and nodes – locations where two or more flow paths are joined. Locations where energy is injected into the grid are known as generation nodes, and locations where energy is removed from the grid and transformed to lower voltages for distribution to end-use customers are referred to as consumption nodes. There are also nodes where the PJM grid connects to neighboring transmission grids, which are called interconnection nodes. Since prices will necessarily be higher at consumption nodes (which are on the downstream end of congested paths) than at generation nodes, the ISO will collect more revenue than it will pay out to power providers. Under economic orthodoxy, these extra revenues, termed “transmission congestion charges”, or congestion rents, are the profit gained by the transmission owner from providing access to a scarce resource. However, under the current FERC framework, the ISO does not own the transmission grid. Instead, it is taskled with operating it in a socially optimal and not-for-profit manner. Transmission owners agree to cede control of their facilities to the ISO and are paid a per MW tariff for use of those facilities that is not related to congestion rents. Since it is operating on a not-for-profit basis, the ISO also claims some portion of the tariff sufficient to cover previously agreed-upon operating costs and is not entitled to keep the congestion rents. Instead, they are redistributed to the users of the transmission grid. The mechanism that has been developed for the redistribution of congestion rents is called the financial transmission right (FTR).

The PJM control area as of May 2006 is shown in Figure 2.1. The time period addressed in this study (2000 – 2003) was before expansion into the areas covered by ComEd, AEP, Dayton Power & Light, Duquesne Light and Dominion. Allegheny Power’s service territory came under PJM control in April, 2002.
An FTR is a financial instrument – it does not involve the exchange of electricity. However, the structural framework of the FTR system is similar to that of an actual electricity network: the two defining characteristics of an FTR are the path over which power flows (i.e., a pair of nodes, referred to as the source and sink nodes), and the amount of power that flows across that path. This is because the FTR is a claim to revenues from the flow of power along specific paths in specific amounts.

The FTR is a claim to congestion rents that are created by congestion in the day-ahead energy market. The day-ahead market is that created by advance purchase agreements for electricity between generators and load-serving entities and by retail-energy-wheeling customers between two interconnection nodes. “Retail wheeling” refers to energy that is neither produced nor consumed within the PJM service territory – it is merely using the PJM grid as a highway for trans-shipment from a non-PJM generator to a non-PJM load-
serving entity. The day-ahead market is just that: the aggregation of trades agreed upon on the day before the power actually flows. After all requested trades have been made, they are entered into a simultaneous feasibility model that checks the requested flows against physical system capacity. This models congestion in the system and provides, as its output, a matrix of prices – one for each node in the system during each hour of the following day. These are referred to as the day-ahead locational marginal prices (DA-LMP). If the day-ahead scheduled flows do not exceed the physical limit in any component of the grid, then all nodes have the same DA-LMP, and there will be no transmission congestion charges, and no congestion rents to be redistributed. However, if any capacity constraint becomes binding, then rationing of access by price occurs, and different DA-LMPs are observed at different locations. These give rise to congestion rent, which must be redistributed by the FTR system. In addition to the day-ahead market, there is a real-time balancing market, which prices the differences between the day-ahead scheduled flows and the actual amount of power that flows. The FTR system does not allocate congestion rents collected in the real-time balancing market, only those collected in the day-ahead market.

An FTR has the following components:

- Source Node: the beginning of the flow path.
- Sink Node: the termination point of the flow path.
- MW Reservation: the size of the FTR.

The unit of measurement of a DA-LMP is $/MW. Thus, the amount paid out to an FTR owner is the difference in the DA-LMPs between the source and sink nodes times the MW reservation of the FTR. For example, if the DA-LMP at the source node is $55/MW and that at the sink node is $64/MW, and an FTR has a MW reservation of 4 MW, then the amount paid to the owner of the FTR is ($64/MW - $55/MW) x 4MW = $9/MW x 4MW = $36. The aggregate value of an FTR is the sum of all hourly payments over the duration period of the FTR:
\[ V(P_A, P_B, MWR) = MWR \sum_{h=1}^{H} (P_{B,h} - P_{A,h}) \]

Where

\( V \) = value of FTR, $

\( MWR \) = Megawatt reservation, the size of the FTR

\( P_{A,h} \) = DA-LMP at source node during hour \( h \)

\( P_{B,h} \) = DA-LMP at sink node during hour \( h \)

\( H \) = total number of hours in duration of FTR

The total number of hours in the duration of an FTR is defined by two factors: the duration of the FTR and the peak-status of the FTR. FTRs are either month-long or year-long. Some are only valid during peak hours, which are 7 am – 11 pm, Monday to Friday\(^3\). There are also FTRs valid only for off-peak periods and those valid 24 hours per day.

Note that in Equation (65) there is no qualification that the price at the sink node be higher than the price at the source node. While it may seem to be intuitive that this must be true, there is no limit on the definition of the source and sink nodes: it is possible to purchase an FTR where the source node is a consumption node, and the sink is a generation node – this is commonly referred to as a “contra-flow path”. It is also allowable for both nodes to be consumption (or generation) nodes. Additionally, an FTR defined by two interconnection nodes may have a higher DA-LMP at the source than at the sink. Thus, it is possible that during some or all hours within the duration of the FTR, the value may be negative. When this is the case, the owner of the FTR is obligated to make a payment to the ISO. This gives rise to the definition of this specific financial instrument as an obligation. In addition to the existence of FTR obligations, there are FTR options. The valuation of an option is similar to that in Equation (65) with the exception that the value can never be less than zero:

\[ V(P_A, P_B, MWR) = MWR \sum_{h=1}^{H} \max [(P_{B,h} - P_{A,h}), 0] \]

\(^3\) Except on National Electricity Reliability Council (NERC) holidays: New Year’s Day; Memorial Day; Independence Day; Labor Day; Thanksgiving and Christmas Day. When New Year’s Day, Independence Day or Christmas Day fall on a Sunday then the following Monday is observed as the holiday.
2.3 The PJM market for financial transmission rights

There are three ways to obtain FTRs:

- FTRs with year-long durations are auctioned in an annual four-stage auction
- FTRs with monthly durations are auctioned in monthly auctions
- FTRs obtained in either of the above auctions can be resold in an electronic secondary FTR market.

The annual auction offers for sale the entire long-term transmission capacity of the PJM grid. In the annual auction, allowable nodes are: generation nodes; interconnection nodes; hubs; zones; and aggregates. Hubs, zones and aggregates are virtual nodes defined by the load-weighted prices of the individual nodes that comprise them. Individual consumption nodes are not valid in the annual auction, but are in the monthly auctions. This greatly reduces the number of available nodes in the annual auction: in the most recent auctions, there were 1,265 valid nodes in the annual auction and 7,665 in the monthly auction. Thus, much of the capacity auctioned in the monthly auctions is between aggregated virtual nodes and physical consumption nodes, as well as any generator-to-aggregate capacity that was not allocated in the annual auction.

PJM also operates an on-line bulletin-board based secondary market for FTRs, allowing them to be traded at any time.

A bid in an FTR auction consists of a source node, a sink node, a MW reservation, and a bid price. A bidder may enter multiple bids on the same FTR path, at different prices and for different MW reservations. MW reservations are defined in minimum increments of 0.1 MW. All bids in an auction are collected and are subjected to a simultaneous feasibility test (SFT). That is, they are aggregated into a virtual direct-current (DC) network, and the flow rates in this network are compared to the security-constrained limits in the real PJM network. If congestion exists, low-priced bids are sequentially discarded until all virtual flow rates satisfy the security constraints. The SFT contains an optimization routine that maximizes the “quote-based value” of the FTRs awarded in each auction. In other words, the model maximizes revenue obtained in the FTR auction.
The outcome of the SFT is a vector of virtual locational marginal prices. To decide whether a bid is successful, one subtracts the virtual source LMP from that at the sink: this is the clearing price of the FTR path. If the bid price is greater than the clearing price, the bid is successful, and the bidder pays the clearing price and receives an FTR entitlement equal to the requested MW reservation. If the bid price is identical to the clearing price, it is possible that a partial MW reservation FTR will be awarded. This phenomenon occurred in some of the early monthly auctions but has become rarer as the number of bids and bidders has increased over time. It is rare for a clearing price to be exactly equal to the lowest winning bid along a path.

**Contra-flow FTRs**

As mentioned in Section 2, the price of an FTR may be negative. In such a case, a firm does not make a payment to PJM for winning the FTR in the auction, but instead receives a credit equal to the price difference times the MW reservation. Over the duration of the FTR, the hourly values of the FTR as calculated in Equation (65) are added to the purchaser’s balance. For any hour that the sink price is lower than the source price, the owner of the FTR is obliged to make a payment equivalent to the congestion amount, and money is deducted from the credit received at the time of purchase. If the source price is higher than the sink, as expected with a conventional (positively priced) FTR, then the owner of the FTR receives an additional credit. Use of the term “contra-flow” implies that the market believes that the predominant direction of current flow will be in the opposite direction to that specified in the FTR definition. Thus, contra-flow is defined by market expectations, and not any engineering definitions. It is possible that some paths will have different-signed values from month to month.

The purchase of a normal (positively priced) FTR is consistent with the expectation of congestion along the path: indeed, the buyer is gambling that the total of the hourly congestion rents collected along the path in question will be greater than the amount paid for the FTR. If it were less, the buyer would be paying more for the FTR than the congestion-rent revenue he is entitled to from owning the FTR. Conversely, a buyer of a contra-flow (negative-priced) FTR is gambling that there will be less congestion than
expected by the market consensus. If the actual congestion rent is greater than the value of the FTR, then the buyer will have to pay more to PJM in contra-flow rents than he will receive for “buying” the FTR. In this context, the purchaser of a contra-flow FTR is akin to a short seller in any other financial market: he is selling now in the anticipation of buying at a lower price in the future. Thus, in this essay, any FTR purchased at a positive price shall be referred to as a “long” FTR, and any one purchased at a negative price shall be called a “short” FTR.

The above is true for obligations. For options, the market-clearing price is based upon the load-weighted shadow price of each binding constraint in the network defined by the optimal outcome of the SFT. The shadow price is defined as the reduction in the amount of congestion rent collected on a path given an incremental increase in the security constrained limit – or symbolically:

\[
SP = \left. \frac{d(TCC)}{dQ} \right|_{Q=Q^*}
\] (67)

Where: \(SP\) = shadow price of constraint

\(TCC\) = transmission congestion charges over congested link

\(Q\) = power flow along congested line

\(Q^*\) = security-constrained line limit

Option prices are not directly observable from the vector of virtual LMPs in the way that obligation prices are. Clearly, it is not possible to purchase a negatively-priced, contra-flow FTR option – all options will be positively priced, whereas the price of an obligation may take any value, positive or negative. Furthermore, the number of valid paths for options is restricted in order to allow convergence of the SFT. For any given path, the clearing price of an option will not be less than the clearing price of an obligation.

The first monthly FTR auction took place for May 1999, and the first annual auction took place in 2003. Options and 24-hour FTRs were first offered in May 2003. The specific rules and practices of the PJM FTR market are contained in the PJM Financial Transmission Rights Manual (PJM, 2005). Monthly FTR auctions take place on the
closest business day to the 15th of the month prior to the active duration of the FTR. For example, for monthly FTRs for the month of March, the auction takes place on February 15th. This means that the forecasting interval for future day-ahead prices for FTR bidders is the period between two and six weeks in the future.

**Auction revenue rights**

Financial transmission rights are a tool for distributing congestion rents collected by the non-profit ISO. However, the process of selling FTRs generates revenue which also has to be redistributed to market participants by the ISO. The mechanism for disbursing FTR auction revenue is the Auction Revenue Right (ARR). Auction revenue rights are not sold but are allocated to PJM network transmission customers based upon usage patterns over the previous year. Thus, auction revenues accrue to the actual users of the PJM grid, unlike FTR revenues, which can be purchased by any qualified bidder (a qualified bidder is a PJM member firm that has established a sufficient credit level to purchase the FTRs it wishes.) Like FTRs, ARRs are defined along specific paths and have specific MW reservations. Ownership of ARRs can have a distinct effect on a firm’s bidding behavior: if it is bidding on an FTR along a path for which it owns an ARR, then there is no incentive for the firm to enter a low bid, as the revenue from the sale of the FTR will go to the firm in question. Any amount bid comes back to the company, so bidding high to ensure a winning bid is a risk-free strategy. In fact, PJM has developed a mechanism whereby an ARR can be automatically converted to an FTR along the same path without any cash flows occurring. This is akin to the bidder purchasing the FTR at the clearing price and receiving the clearing price revenue simultaneously. Inclusion of ARRs into an economic study of FTRs would add significant complexity to the analysis. However, this study focuses only on the monthly FTR auctions, and ARRs are only applicable to the annual FTR auctions. This greatly simplifies the analysis. The details of the ARR allocation and disbursement mechanisms are also contained in the PJM FTR manual (PJM, 2005).
2.4 Firm behavior in FTR auctions

The PJM FTR market began operation May 1999 and continues to this day. This study examines firm behavior from October 2000 to April 2003, a period of 31 months. The reasons for the selection of this time period are as follows:

- PJM commenced the day-ahead market in June 2000. Before this date FTR values were established by a non-published, non-market-based set of nodal prices. That is, the data required to calculate the return to owning an FTR were not available for before June 2000.

- This model employed in this study requires past day-ahead prices and past auction results as inputs. The period from June – September 2000 is used to provide this initial data to the model. Hence, the analysis of firm decisions begins with decisions made for the October 2000 FTR auction.

- PJM provides access to all bid data in the FTR market but employs a six-month delay in publishing this information to protect the confidentiality of member firms. In these delayed bid data files, each bidder was identified by a unique two-digit code, which remained the same for each firm over time. By comparing the published results of the FTR auctions, in which the names of the winning companies are shown, with the bid data, it was possible to divine which two-digit code referred to which member firm, except in the rare case when a bidder was never successful in winning an FTR. To study bidder behavior, it is necessary to compare bidding actions and auction results. However, beginning in May 2003 PJM stopped including the two-digit company code in the delayed FTR bidding data files. Thus, it became impossible to do two things:
  - o discover whether a series of bids on the same FTR path were part of a combination bid from one firm, or separate bids from different firms; and
  - o to combine the bid data with auction results data. For this reasons, the period of analysis stops with the auction for April 2003.

However, using April 2003 as a cut-off point is meaningful in some other ways: in May 2003 PJM began to allow firms to bid on FTR options. Including FTR options would complicate the analysis considerably.
During the 35-month period for which data were collected, 52 firms submitted bids into monthly on-peak FTR auctions. 48 of these firms were successful in winning at least one FTR. The number of bidders in any given month varied from 6 to 31, as seen in Figure 2.2.

![Figure 2.2: Number of firms participating in monthly on-peak FTR auctions](image)

**Figure 2.2: Number of firms participating in monthly on-peak FTR auctions**

**Market size**
A true measure of the size of the FTR market is difficult to define. One can look at the number of bids made, but the size of a bid must be considered. The size of a bid is measured in MW, but this measure does not take into consideration how much of the grid capacity a certain bid-size equates to. A true measure of market-size would examine how much capacity was available for purchase along every line in the PJM grid, and how much of this was bid upon and won. However, this would require a detailed analysis of the network, and integration with annual FTR markets to evaluate the remaining capacity available after annual FTRs have been awarded. The data required to perform such an analysis are not available in the public domain. Put another way, using the MW sum of bids as a measure of market size does not take into account the length of the FTR path or the effect of a bid on the entire market. For example, a 1-MW FTR that covers a 5 foot-long interconnect between two transformers in a switching yard has the same weighting as a 1-MW FTR between two interconnection nodes hundreds of miles apart. Clearly,
these two bids will have different effects on the outcome of the SFT. However, if one is interested in month-to-month comparisons, using total MW bid may be applicable if it can be assumed that the distribution of path lengths does not change much from month to month. A cursory analysis was performed looking at the number of awarded FTRs sorted by the zonal locations of their source and sink nodes, and it was found that FTRs that had their source and sink in the same zone were by far the most common type, and that FTRs within certain zones (PECO, PSEG, DPL, Allegheny Power, and PennElec) were always the most frequently observed ones. This is not surprising, for the following reason: FTRs for the “backbone” of the grid are awarded in the annual FTR auctions, and consumption nodes are not eligible nodes in the annual auction. Instead, bidders must use aggregate or zonal nodes or hubs as substitutes for consumption nodes in the annual auction. An aggregate is a virtual node, a load-weighted average of prices at individual nodes. A zonal node is merely an aggregate that covers all consumption nodes in a given utility’s operating region, and a hub is a larger aggregate, covering all the nodes in a multi-zone region of PJM. In the monthly auction, paths from the virtual aggregate node to the actual consumption nodes are available. Thus, much of the monthly traffic will be from a hub, zone or aggregate to consumption nodes within the same zone. Of the twenty most commonly observed zone-to-zone paths, ten are within the same zone, seven are between a hub and a zone, and only three are between different zones. Those three inter-zonal paths are PECO - PSEG, PSEG - JCPL and MetEd – PennElec: all neighboring zones. The point to be taken from the above discussion is this: even though the MW sum of bids does not incorporate any measure of distance, its use as a measure of market size can be seen as legitimate because we are only interested in month-to-month comparisons in the same market, not comparisons between different markets, and the month-to-month traffic flows show consistent patterns.

The monthly MW-based measures of market size are shown in Figure 2.3. This figure contains data for all tendered bids and all successful bids. It includes both long and short FTRs. Note that the unit of measurement is Gigawatts, or thousands of Megawatts.
The amount of bids placed seems to vary greatly in Figure 2.3. Decomposing the market into short and long bids yields some explanation of this behavior. For purposes of this study, a short bid is any bid with an average value equal to or less than zero. The long and short markets are shown in Figures 2.4 and 2.5.
The actions of single firms are responsible for some of the extreme swings in Figures 2.4 and 2.5. First, examining Figure 2.4, the long FTR market, we see three distinct peaks: December 2000-January 2001, July-September 2001 and January 2003. All of these can be attributed to single firms. The first two are caused by El Paso Merchant Energy. El Paso rose to infamy in the wake of the California Energy Crisis of 2000 and 2001 based upon accusations of manipulation of the transmission market, including the creation of phantom congestion (and congestion rents) by submitting massive “wash” trades onto small transmission lines. At the same time, El Paso became very active in the PJM market, which appears to have been more robust to manipulation attempts than the poorly designed California market. Often mentioned in the same breath as Enron, El Paso was a small regional natural-gas producer and distributor that grew into a diversified multinational energy company through aggressive acquisitions from 1996-2001. It became actively involved in the nascent energy-trading business in the late 1990s. El Paso tentatively entered the PJM long market in September 2000, submitting about 1% of the MW bids, but over the following few months it was routinely submitting 50-75% of all bids, on both an MW and number of bid basis. Its participation waned to zero in June 2001 but came back up to the 50-70% range for seven of the next eight months. In the
wake of the California energy crisis and the collapse of Enron, El Paso came under increasing scrutiny by politicians, regulators, and investors. The company began to lose money, and a new management team was brought in. These managers decided to divest El Paso of all but its core businesses, and today it is involved only in natural gas production and pipelining (El Paso, 2004). The last spike in Figure 2.4 is caused by the activities of Citadel Energy Products, L.L.C. This firm is a subsidiary of Citadel Investments Group, a Chicago-based private equity management firm. Citadel actively trades in all major equity and commodity markets and started Citadel Energy as an active power marketing company. Citadel entered the PJM long market in November 2002, bidding about 5% of the market. It has stayed around this level except for two months, January and March 2003 when it bid 40% and 27% of the market respectively.

If one isolates the activities of these two firms, the long market appears as shown in Figure 2.6.

![Figure 2.6: Size of long market with and without outliers](image-url)
With the two aforementioned firms isolated, we can begin to see some stability in the market. There are two basic regimes: a relatively constant market size of 10-20 GW from June 2000 to June 2002, a short period of growth, and a new steady state at about 30 GW after July 2002. The shift from one regime to another is not related to the entry of any one firm but is the result of step-like increases in the activity of three long-time large players in the market: Allegheny Energy (the trading arm of Allegheny Power), Williams Energy Marketing and Trading, and Morgan Stanley Capital Group. What notable event took place in July 2002? This was the first month after the first annual FTR auction, the beginning of the annual FTR duration. Before this time, annual FTRs were allocated to load-serving entities by PJM based upon historic usage patterns. It would appear that the amount of FTRs allocated in the annual auction was not as great as those allocated under the previous regime, and thus, a greater number of bids are made in the market. Perhaps coincidentally, PJM expanded its control area to include the Allegheny Power service territory in April 2002, but it seems odd that there was a three-month lag before this addition was reflected in the size of the FTR market. This addition made several hundred more nodes and transmission paths eligible for the monthly FTR market, but there was no noticeable increase in the MW sum of bids for April, May, or June 2002. A closer examination of the behavior of the three firms mentioned above did not reveal any consistent patterns: Allegheny added some new paths, but kept the size of all other bids the same. Morgan Stanley increased some bid sizes and added some new paths. Williams had largely different slates of bids over the two months. What is clear is that the aggregate bidding behavior of the long-standing market players changed in July 2002.

Looking at Figure 2.5, the size of the short market, we see two periods of pronounced increases in activity: one from February-June 2001, and one from August 2002 until June 2003. The February 2001 spike is the result of Niagara-Mohawk Energy Marketing making a brief entry into the short market. In this month, they were responsible for 25,000 of the 29,000 MW that was bid in the short market. After this month, it mysteriously retreated from the market, making only one small appearance again in April 2001. The spike that culminates in June 2001 was from El Paso activity. From this point, the market was basically quiet for a year. The August – November 2002 bump was from
the activity of two companies entering the market: FPL Energy Power Marketing, a branch of Florida Power and Light, and Outback Power Marketing, a trading company on which information was not available anywhere in the public domain. After having large trading volumes for a couple of months, these companies went down to bidding under 1,000 MW per month. For the last five months of this study, a number of trading firms entered the short FTR market in considerable volume. These include the aforementioned Citadel Energy Products, as well as DC Energy, a similar type of firm based in Vienna, VA, and for the last two months the busiest trader of all, Coral Power, the North American energy-marketing subsidiary of oil giant Royal Dutch Shell. Figure 2.7 displays the size of the short market segregated into these big players and the residual market.

![Figure 2.7: Size of big player and residual short markets](image)

A distinction can be made between the big players and the residual market: most of the former are non-utilities from outside the PJM control area who are clearly speculating on congestion being less severe than expected. The residual consists mostly of PJM member utilities who are likely hedging their exposure to contra-flow congestion.
The FTR market is notable for the volatility of participation: many firms enter and leave the market. As can be seen in Figure 2.8, almost half of all participants stay in the market less than 6 months. Only three firms have participated in every FTR auction in this period: Exelon (previously known as PECO), PP & L Energy Plus and Morgan Stanley. PSEG participated in 34 of the 35 months.

![Figure 2.8: Duration of firm participation in FTR market](image)

It is perhaps surprising that a majority of the PJM native utilities do not make regular use of FTRs as hedging devices for the power they consume. Many of the short-time companies are non-native firms, often power traders who own no generation or distribution, who are speculators in the market. It would appear that many of these firms leave the market after failing to attain the profits they project, although several that entered towards the end of this study, including DC Energy and Citadel, appear to be having some success, and are still active three years after the end of this study period.

**Bidding strategies**

As the allowable amount of FTRs along any given path can be broken up into increments as small as 0.1 MW, it is possible to submit a schedule of bids with different prices for different quantities of the FTR in question. Different firms choose to undertake different
strategies in this respect. The average monthly numbers of bid parts for two selected groups of firms are shown in Figures 2.9 and 2.10.

**Figure 2.9: Bid complexity, selected firms, part 1**
Note: Exelon’s October 2001 observation (value = 91) omitted for clarity

**Figure 2.10: Bid complexity, selected firms, part 2**
These two sets of firms were chosen to illustrate that even similar firms have different ideas about what the optimal bidding strategy is, and that the choices concerning optimal strategies do not exhibit convergent evolution. Figure 2.9 contains data for three firms that have been involved in every long market in the period of study: Exelon, Morgan Stanley and Pennsylvania Power & Light (PPL). Exelon and PPL are both native PJM load-serving utilities that also own generation, whereas Morgan Stanley is a power reseller that neither owns generation nor directly serves loads. We might expect a marketer to act differently to a vertically integrated utility. PPL consistently employs a strategy of employing single or two-part bids, rarely more. Exelon, on the other hand, has extremely complex bids. Exelon’s strategy increased in number of parts to an extreme of 91 in October 2001. It has since trended down but still remains over 20 per bid. Morgan Stanley exhibits an unchanging strategy of using bids broken into about five parts. Figure 2.10 displays the strategies of three firms that entered the markets at a later date. Allegheny Energy and Constellation PowerSource (the rebranded image of Baltimore Gas and Electric) are also both PJM utilities, but not only do they have different strategies, their choices are moving on opposite directions. Allegheny’s bid complexity started at over 20 parts/bid and has consistently trended down to below 10, whereas Constellation started below five and has now reached 15. Williams Energy, a power marketer, is following a similar path to Constellation: getting more complex over time, although the last observation shows a significant decline. As is addressed in more detail later in this chapter, auction theorists have examined the issue of bid complexity, and have stated quite unambiguously that complex bids strictly dominate simple ones in profit-maximizing terms, at least in the models developed by the researchers (Scott and Wolf, 1979.) The divergent array of strategies exhibited in the PJM FTR market does, at a first glance, raise some skepticism about these findings: as the old economist’s adage goes, people seldom leave money sitting on the sidewalk. If complexity is strictly better, we would expect to see more of it and movements towards it, not away from it.

**Profitability**

At first glance, the definition of profit in the FTR market is not straightforward. It is complicated by the fact that for some players an FTR is only part of a portfolio that
includes auction revenue rights, physical power deliveries and payment of congestion rent, whereas others purchase it as a stand-alone financial product with the expectation of positive revenue flows. Fortunately, ARRs are only applicable to the annual FTR auction, and this study addresses only monthly FTR auctions.

Consider a load-serving entity. This is a firm that must deliver retail power to end users, regardless of the locational marginal prices it has to pay. Therefore, regardless of whether the firm does or does not own an FTR, it has to purchase power from the ISO at the given DA-LMP and must pay the congestion rents associated with the LMPs at the consumption node in question. These utilities do not have the option of buying power or not based upon the price: they must purchase whatever amount is necessary to satisfy their customers’ real time demand load. Thus, we can reasonably assume that congestion rent paid is an exogenously determined charge that the utility has no control over: its existence and size cannot be affected by any actions the utility can make in the short run. This, too, simplifies the analysis of the profitability from owning an FTR. The model now reduces to a question of two cash flows: the amount paid in the FTR auction and the amount received as reimbursed congestion rent. This is the same scenario as for a power trader with no load-serving obligations, or, for that matter, a pure speculator.

PJM employs a single-settlement system for paying companies for FTRs: when the FTR is purchased, the firm’s account with PJM is debited the amount of the payment. As the period of the FTR passes, congestion rent accumulates in the firm’s account. At the end of the FTR period, PJM makes a payment to the firm based upon the net of the FTR payment and the collected rent. If the collected rent is less than the purchase price, the firm is obligated to make a payment to PJM; otherwise, PJM pays the firm. For the sake of comparing the value of different investments with a common measure, the conventional method of expressing profit from owning a financial instrument is the annualized yield; that is, how much the owner makes as a function of the cost of the instrument and the time between payment and return. Since we have only a single cash flow, it is impossible to calculate a yield. Any profit from owning an FTR must be
expressed in absolute currency terms. The only financial constraint in participating in the FTR market is the establishment of a sufficient line of credit with PJM.

Thus, for the firm, the profit to owning an FTR is simply amount of congestion rent minus the clearing price of the FTR. What sort of profitability is typically exhibited in PJM? One approach is to look at aggregate monthly profit.

![Figure 2.11: Average monthly FTR profits](image)

Figure 2.11 contains the aggregated monthly profit in the long and short markets. The total profit was divided by the total number of MW awarded and the number of hours in the month to yield a comparative average value. As can be seen, there is a lot of volatility in profit over time. Over the lifetime of the study, the mean monthly profit from owning a long FTR was $0.77/MWh with a standard deviation of $1.35/MWh. For short FTRs the mean was $0.35/MWh with a standard deviation of $1.83/MWh. For perspective, the average on-peak cost of electricity in PJM over this period was approximately $44/MWh with a standard deviation of $27/MWh. If, for simplicity, one assumes that profits are normally distributed then we have a 28% probability of losing money on a long FTR and
a 42% probability of losing money on a short one. Over the period in question, PJM made total net payments of $42.91 million to holders of long FTRs and $5.05 million to holders of short FTRs.

**Firm-specific profits**

Figures 2.12 and 2.13 display the average profits from owning FTRs in the long and short markets. The profits are measured in terms of $/MWh. To get some understanding of the gross profits, the amount of FTRs awarded over the length of the period of study is also included. The amounts awarded are the columns in the figures, and are measured on the left scale. The average profits are shown as points along a curve and are measured on the right scale. The firms are sorted in ascending order of average profit.

![Figure 2.12: Firm-specific profit, long FTR market](image-url)
As can be seen from both figures, firms that are most active in both markets tend to have average profits in the range of $0 - $1/MWh. This is not surprising: given the law of large numbers, the more a firm participates, the more we would expect its profits to tend towards the mean. This observation also implies that firms are not able to maintain greater than average profits over time, a prediction consistent with the idea of an efficient market.

**Market concentration**

It is informative to examine how concentrated the FTR markets are. The most broadly accepted measure of market concentration is the Herfindahl-Hirschman Index (HHI), which is calculated by summing up the squares of the percent market-share of each competitor. The HHI has a maximum value of 10,000 if there is only one player in the market, and a lower bound of zero in the case of atomistic competition. The US Federal Trade Commission (FTC) considers a market with an HHI of greater than 1,800 to be
“concentrated” (FTC, 1997.) Figures 2.14 and 2.15 contain the HHI for both the bids into the FTR markets and for the number of FTRs awarded, in the long and short FTR markets.

Figure 2.14: Herfindahl-Hirschman Index, long FTR market
Figure 2.15: Herfindahl-Hirschman Index, short FTR market

Note: no short FTRs awarded in June 2001, thus no HHI available

In both markets, we see both HHIs trending down as more firms enter the market. We can also see that the HHI for the bids tendered is generally higher than that for FTRs awarded, although divergence in these two measures shrinks as time passes: as we move into 2002, the distribution of bids and awards are very close to each other. This implies that firms that perform poorly in the FTR market either learn to improve their performance over time, or are replaced by better-performing entrants. The short market is more concentrated than the long one, given that there are fewer firms in this market. In the last year of this study the long market moved below the FTC’s “concentrated” threshold, into the “moderately concentrated” category. Towards the end of this study, it is skirting the 1,000 mark, which is the point at which the FTC considers a market to be “not concentrated.”
2.5 Literature review

The auction-theory literature contains many specialized terms with meanings that are not intuitively obvious to the non-specialist. Therefore, before describing the literature, definitions of some of the technical terms are presented.

**Reserve price**
The minimum acceptable sale price for the good. It will be specified in advance by the seller. It can be either published or kept private. Without a reserve price, a good will definitely sell. With a reserve price, it may not.

**Indivisible good**
This is a good that is sold as a single, whole entity. Parts of it are not sold separately.

**Divisible good**
This is a good that can be broken down into parts. In a divisible-goods auction, bidders may bid on part or all of the good, and those with the highest bids that correspond to the sum of the parts of the good win a share. FTRs are divisible goods.

**Multi-good auction**
This is similar to a divisible-good auction in that a bidder is competing for more than one indivisible unit of a good. A treasury-securities auction can be seen as either a multi-unit auction, whereby the bidders are competing for a multiple number of T-Bills, or it can be seen as a divisible-goods auction, whereby the bidders are competing for a share of the total amount.

**English auction**
Sometimes referred to as an open-outcry ascending-price auction, this is the familiar model whereby bidders keep submitting bids until the reservation price of the bidder who places the highest value on the good is reached. The final price paid will be somewhere between the values placed on the good by the two highest bidders.
Dutch auction
In contrast to the English auction, the Dutch auction is one where the auctioneer starts at a high price, and announces prices in descending order. The auction ends when a bidder signals the auctioneer to stop, with the winning bidder paying the price at which the auctioneer stops. Thus, it is an open-outcry descending-price auction. The name comes from the fact that this auction format was first used to auctions lots in Dutch flower markets. Oddly enough, in Holland it is referred to as a “Chinese auction.”

First-price auction
This is an auction for an indivisible good in which the highest bid wins the good, and the winner pays the amount that he bids.

Second-price auction
This is an auction for an indivisible good in which the highest bid wins, but the winner pays an amount equal to the second-highest bid.

Discriminatory-price auction
This is a divisible-good auction where each winning bidder pays his bid price. If we have a descending slate of bids, each winner will pay a different price for his portion of the good, hence the use if the term “discriminatory.” It is analogous to the first-price indivisible good auction.

Uniform-price auction
This is a divisible-good auction where each winning bidder pays the same price. The uniform price may be defined as the lowest-priced successful bid or the highest-priced unsuccessful bid. Theorists have shown that these two cases yield different equilibrium solutions, but in practical terms they are usually so close together as to be indistinguishable. The PJM FTR auction is a uniform-price auction with the clearing price equal to the lowest successful bid.
Private value auction

This is an auction for a good in which each bidder places a personal, private value on the good in question. It can also be called an endogenous value, as the valuation comes from within the bidder, and not some external consideration, such as yield or resale value. For example, a piece of art that is valued for its esthetic appeal to the bidder.

Common-value auction

This is an auction for a good where the value to the winner is determined by some physical or market process. It is the opposite extreme to the private value auction. For example, the value of an oilfield lease is the net profit gained by a firm from producing and selling the oil and is held to be generally the same for all bidders. Other examples include the yield from a US Treasury security or, in the case at point, the stream of revenues from ownership of a financial transmission right. The value is generally thought to be exogenously determined. A weaker case is the “affiliated values” model, whereby bidders have different private values but those values are partly contingent upon other bidder’s values. This is sometimes described as having “envious” bidders: the value that one person places on a good is partly determined by the value other people place on it. The use of the term “common value” can be confusing, especially if one thinks of “values” as variables in one’s bidding function. That is, the “value” is the bidder’s ex-ante expectation of how much the good is worth, whereas the “common value” in the definition refers to the actual ex-post return from owning the good. The term is not clear in this distinction. It may add lucidity to the debate to instead talk of “common-payoff” auctions, since the return to the winner is the common factor, not the value he places on the good before the auction. Another point to make is that the payoffs are not necessarily common: the payoff is often partly endogenous, as it can be affected by the decisions of the winning bidder. The value of a Treasury note changes as the prevailing interest rates change, and the winner’s timing decisions about selling or holding the note will affect his payoff. Likewise, the payoff to owning a mineral lease will change depending upon which exploitation technologies the winner uses, as well as his timing choices in production. However, for an FTR, the payoff can be assumed to be entirely exogenous: there is no decision or action on the part of the winner that will affect the stream of cash
flows that accrue to the owner of the FTR. Violation of this assumption means that market participants can exercise market power in the setting of day-ahead locational marginal prices. This assumption is addressed at more length later in this section.

**Bidding strategy**
This is some function that takes the value that the bidder places upon the good as its argument and has the actual bid as the result. For an indivisible good, the function will typically be some constant multiplier of the bidder’s value, and the bid will be a number. In a divisible-good auction, it will be a function of value and quantity, and the output will be either a continuous function over the range of the good or in the more common discrete formulation, a series of bids for different portions of the good at different prices.

**Optimal strategy**
This refers to a Bayesian Nash equilibrium whereby a bidder’s profit is maximized by employing a strategy that is identical to the strategies employed by all other bidders. Some auction formats have a single dominant optimal strategy, but others have either none or an infinite number.

**Proprietary information**
In a common-value auction, the common value of the good is not normally known in advance. Instead, it will be the realization of some random distribution of possible values. This distribution will often not be known, although in repeated auction scenarios the bidders form priors based upon observation of previous auctions. It is often assumed that the bidders have some knowledge of the distribution of the values and thus form posterior distributions. This knowledge, referred to as “proprietary information”, is the knowledge set that the bidder has about the distribution of the actual payoff. It is a common assumption that the proprietary information for each bidder is a random sample drawn from the same distribution of information as all other bidders. In such a case, the bidders are said to be independent.
**Symmetric bidder**

Considering the proprietary information discussed above, one can assume that each bidder has proprietary information that is independent and identically distributed (\(i.i.d\)), that is, each bidder’s proprietary knowledge can be modeled as a random sample from some distribution of all knowledge. If one assumes this, then the bidders are said to be symmetric. This does not mean that all bidders have the same information, but that over repeated auctions, they all randomly receive information that is from the same distribution. Within a given auction, this will lead to each bidder placing a different value on his expected payoff, but, over repeated auctions, the law of large numbers takes over and all bidders will have the same mean assessment of the value of the good over time. If the bidder’s proprietary information is not \(i.i.d\). then we expect some bidders to form different mean valuations over time. This could arise if one bidder is unequally privy to inside information, or if one bidder has a more accurate prediction algorithm than his competitors. In such a case, we describe bidders as being asymmetric. Asymmetry leads to great complications in the mathematical analysis of auctions. One assumption that is tested in this chapter is whether certain firms (specifically, PJM native utilities) have better information about the expected payoffs to owning FTRs and the expected clearing price in the auction.

**Bid shading**

This is the practice of submitting a bid that is below one’s expected value of the good. In a first-price auction, it is the profit-maximizing equilibrium strategy, assuming rational bidders. In other auctions, it can be seen as either a non-collusive equilibrium or a result of bidder collusion – often it is mathematically difficult to show which the case is.

Much of the literature concerning divisible-goods auctions is focused upon the primary markets for government securities, which are the largest and most frequent auctions of divisible goods. Friedman (1960) makes the first recorded call for the adoption of a uniform-price auction for US Treasury securities. He felt that the discriminatory auction limited participation in the primary market to a few specialists and well-financed investment houses, because of the extremely high costs of making an overly optimistic
bid. This resulted in a lack of competition, and even if outright collusion was not provable, monopsonistic market power existed and resulted, in Friedman’s opinion, in systematic underpricing of the issued securities. Friedman’s findings were first questioned by Brimmer (1962), who presents a model in which the Treasury receives greater revenues from discriminatory pricing than uniform pricing. This model is criticized by Friedman (1963), as he feels it overlooks the explicit point that he was making: that discriminatory pricing causes revenue losses not because of the some idealized market structure, but because of monopsony and bidder collusion. Goldstein (1962) qualitatively examines Friedman’s assumptions, and finds them unsubstantiated. He claims that it is difficult to show any benefit to switching formats. Nonetheless, Friedman’s view was generally held, in an informal manner, by most observers of the treasury auctions for most of the next two decades. The question of which method maximizes revenues to the Treasury was first addressed in a mathematically rigorous manner by Smith (1966). Smith modeled the primary and secondary buyers’ bids as simple functions of price (i.e., single bids) and claimed that in either case, unambiguously, a discriminatory auction leads to lower bids (and prices, and revenues) than a uniform-price one. This bolsters Friedman’s case. However, in the Treasury auction, a bidder is not limited to a single bid but can make a schedule of bids at different prices for different quantities. That is, the bid is a function of price and quantity. In this model, Smith was not able to reach a concrete conclusion. He posits a specific example where a discriminatory auction yields a lower price but does not claim that this result is tractable as a general solution. Referencing Smith’s work, Scott and Wolf (1979) show that multi-unit bids are more profitable for the bidder than single-price bids and that actual participants in the primary treasury auctions employ multi-unit bid schedules, although the pricing algorithms employed by these firms are simplified versions of the optimal algorithm, and thus the bids are sub-optimal. The authors claim that this is due to time constraints in the auction process. Scott and Wolf do not attempt to compare bidding strategies in discriminatory versus uniform-price auctions. They do, however, verify that the actual market is characterized by the indeterminate case in Smith’s paper. Bikhchandani and Huang (1989) create a model in which the primary and secondary treasury markets are informationally linked. They reach the conclusion that a uniform-
price auction will generate more revenue for the Treasury, concurring with Friedman’s beliefs. Cammack (1991) performed one of the first empirical studies of bidding strategies in the discriminatory treasury auction. She finds that bidding behavior in these markets is consistent with the theory, as developed by Wilson (1977, 1979) and Reece (1978) that a first-price auction without an infinite number of bidders will yield less than the value of the good to the seller. She tests this by comparing auction results with secondary market values, and finds a consistent downwards-biasing of the auction prices. Despite repeated demonstration to the contrary, Friedman (1991) reiterated his support for the uniform-price auction, as did Chari and Weber (1992), who mistakenly transferred the properties of the non-divisible good auction to the divisible good. An excellent general summary of the literature of treasury auctions is contained in Das and Sundaram (1997).

Vickrey (1961) wrote what is broadly considered the seminal paper in modern auction theory, often said to be two decades ahead of its time. The author focuses on Pareto-optimal allocations under different auction formats. There are several “firsts” in this paper, such as the derivation of the dominant strategies in first- and second-price auctions, and the proof of revenue-equivalence of the first-price and open-outcry auction (under the assumption of independent and symmetric bidders). He demonstrates that, given certain assumptions about rationality, independence and symmetry, bid-shading is the equilibrium strategy in the first-price auction and that a person’s bid should be \((N-1)/N\) times his true value, where \(N\) = the number of bidders. However, bidding one’s true value is the dominant strategy in a second-price auction (hence, second-price auctions are described as being “truth-telling.”) He also addresses bidder asymmetry, with the encouraging statement that “if the assumption of homogeneity among the bidders is abandoned, the mathematics of a complete treatment become intractable.” He demonstrates that under certain auction rules, bidder asymmetry can result in an inefficient (non-Pareto-optimal) allocation, although the effect on seller revenue is not addressed. This position is reinforced by Greismer, Levitan, and Shubik (1967.) He also addresses the notion of bid-shading in a divisible-good auction and reinforces Friedman’s contention that a uniform price has the same “true-value” revelation properties as the
second-price auction. He does mention that with a shortage of bidders, the uniform-price format is more susceptible to underpricing because of bidder collusion.

Several papers examine the revenue considerations of different types of auctions. Ramsey (1980) demonstrates that a second-price auction can yield either lower or higher revenues than the first-price auction for offshore leases depending upon the structure of the auction. This bolsters the idea that there is not a deterministic closed-form general case for revenue maximization. Harris and Raviv (1981) examine a multi-unit auction with single bids, and claim that with risk-neutral bidders (i.e., utility linear with profit) the revenue to a seller will be identical in both discriminatory and uniform-price auctions, but they do not model multiple bids. They claim that with risk-averse bidders, the discriminatory auction raises more revenue. Milgrom and Weber (1982) develop a model for an indivisible good showing that when bidders are risk-neutral, the second-price auction generates a higher price than either the first-price or Dutch auction. However, Milgrom and Weber clearly state that these results are not directly transferable to divisible-goods auctions. McAfee, McMillan, and Reny (1989) devise an auction structure similar to the Vickrey (second-price) auction where they show that with the correct design, this auction will generate the highest price for the seller and, in a realistic situation with a finite number of bidders and discrete value distributions, will generate “nearly” maximum prices in every case. Ausubel and Cramton (1998) take a slightly different approach to most papers, examining ascending-price and modified Vickrey auctions and comparing them to discriminatory and uniform-price sealed bid auctions. They are able to devise a novel auction structure that has a single equilibrium that yields greater revenue than the “multiplicity of equilibria” in uniform-price and discriminatory auctions. The Vickrey auction is a special case of the sealed-bid auction where each winner pays the second-highest price for each different unit of good he buys. It is akin to a discriminatory auction where the discriminatory prices are the highest losing bid, not the winning bid. Ausubel and Cramton (2002) further extend this idea with the inclusion of reserve prices into the Vickrey-multi-unit auction, and claim that it is superior to the uniform-price auction on both revenue and efficiency grounds. The authors claim that truth-telling (bidding your value) is a dominant strategy in this format of auction.
Wilson (1967) examines the question of bidder symmetry. He examines a two-bidder sealed-bid first-price auction where one bidder knows the *ex-post* value of the good with certainty. The context is that of an offshore oil-lease auction. This work builds upon that of Woods (1965). Wilson lays out a solution methodology which provides the first-order condition for an equilibrium as a function of the distribution of the uninformed bidder’s estimate of the value of the good. Depending upon the distribution, the solution may or may not be analytically determinate. Wilson lays out a simple example using a uniform distribution and shows that there is extreme asymmetry in the expected payoffs of the two bidders: the informed party’s expected profit is 8.5 times that of the uninformed bidder, with an expected profit of 62%. This is an early demonstration of the value of perfect information\(^4\). This work is an extension of Wilson (1968), which was done earlier but published later. In his 1968 work, Wilson lays out a formulation where two bidders have access to a distribution of proprietary information, and their bids are informed by this distribution. Wilson once again lays out the framework of the general solution and provides a specific example assuming that both bidders have access to the same information set. The general solution does not require that each bidder have access to the same information, but Wilson does not solve such an example.

Capen, Clapp, and Campbell (1971) make the first explicit mention of the phrase “the winner’s curse” in the literature. The authors were all engineers working for Atlantic-Richfield (ARCO) in the 1960s and were involved in that company’s offshore exploration efforts on the Outer Continental Shelf (OCS) region of the Gulf of Mexico. The authors noted that ARCO, as well as all other companies exploring the OCS, suffered horrendously low returns on investment in the OCS and were able to pin down the idea that they were over-paying for the exploration rights, which were sold in first-price auctions.

\(^4\) Weverbergh (1979) addresses Wilson’s 1967 paper, pointing out that Wilson’s “easy example” solution is in error. He claims that Wilson’s answer is not a Nash equilibrium because Wilson mistakenly modeled a first-mover game, and not a simultaneous-move game.
In a remarkable paper, Wilson (1979) launches the first mathematically-rigorous challenge to the long-held assumptions about the transferability of qualities between second-price indivisible-goods auctions and uniform-price divisible-goods auctions. The framework is an extension of his 1960s papers, in which he studied indivisible goods. Once again, he derives the first-order conditions for bidder profit-maximization, which are functions of a pair of distributions: the distribution of the \textit{ex-post} value of the good and the distribution of a set of proprietary information about the good. As in his previous papers, he follows up with some examples using simple distributions. He shows that the price in a uniform-price divisible-good auction will be considerably lower than if the good was sold as an indivisible unit – as low as half the price, depending upon one’s assumptions. He claims that switching to a discriminatory auction will not affect seller revenue: the price of the last good sold will be lower, but the distribution of goods and the seller’s revenue will not change. This is the first recorded claim that the format of a divisible-goods auction – discriminatory versus uniform-price – does not have any effect on the seller’s revenue. However, Wilson is clear to note that this conclusion is based upon specific examples, and he is not able to provide a mathematical proof of the generality of the conclusion. Wilson does not address bidder asymmetry in this paper – he once again assumes that the bidders randomly draw from some common distribution of proprietary information. Extending on this work, Pesendorfer and Swinkels (1997) claim that the price probabilistically converges to the value only in the case where there are both an infinite number of goods to be sold and an infinite number of bidders. Any finite amount of goods or bidders will yield a lower price in a uniform-price auction. Wilson (2002) maintains his preference for uniform-price auctions. Engelbrecht-Wiggans and Kahn (1998) and Noussair (1995) have attempted to examine the structural reasons for underbidding in a multi-unit uniform-price auction and have described declining marginal utility as being responsible for “demand reduction” and that bids on the first, highest-priced good may be true values, but subsequent bids will be underpriced.

In a survey piece, McAfee and McMillan (1987) sketch out a “benchmark” auction based upon four assumptions: risk-neutrality; independence of private values; symmetry of bidders and payment being a function of bids alone. They then examine results from the
relaxation of these assumptions. They address bidder asymmetry, and show that in such a case, the optimal (revenue-maximizing) auction involves bid discrimination, in much the same way that first-degree price discrimination maximizes profit when segments of the demand curve can be segregated. The authors also claim that such an auction will result in a non-Pareto-optimal allocation, but this can be solved by the presence of a secondary market.

Back and Zender (1993) compare a sealed-bid uniform-price auction with a sealed-bid discriminatory auction (DA), with the assumption of perfect divisibility of a good. This was a response to the “experiment” of using the uniform-price format to sell 2-year and 5-year notes by the US Treasury beginning in September 1992. This experiment was in response to the Salomon Brothers scandal of 1991, when that brokerage was able to purchase 92% of the 2-year notes, in violation of the rule that no more than 35% of any issue go to one purchaser. It has been shown that these notes were significantly overpriced in the secondary market for several months afterwards (Jegadeesh, 1993; Jordan and Jordan, 1996.) There has been much discussion of the merits of uniform-price versus discriminatory formats, and many writers have generalized the theoretical results from the first- and second-price indivisible-goods auction to the divisible-good discriminatory and uniform-price formats, which are seen as analogous. Back and Zender show that the indivisible-goods-auction results are not analogous. While not presenting a specific Bayesian-Nash equilibrium solution, they show that in at least one specific example there can be collusive outcomes which lead to prices much lower than realized payoffs. Back and Zender (2001) claim that many of the problems of the uniform-price divisible good auction can be corrected by having an endogenous supply, allowing the seller to pull some amount of a divisible good off the market after bids have been made. McAdams (2002) extends upon this idea. He speaks of “collusive-seeming” equilibria, which are the oft-exhibited clearing prices in uniform-price auctions that are far below realized item values. He claims that adding a small payment discount to the price will remove this equilibrium, and argues that post-bid perturbations of the supply curve, as claimed by Back and Zender, also work to maximize revenue. Damianov (2005) further extends Back and Zender’s work claiming that if the goods are rewarded to all winning
bidders on an equal-share basis, and not a pro-rata basis, in the presence of endogenous supply then all goods sold in the uniform-price auction will be sold at the competitive price, and all collusive-seeming low-price equilibria will vanish.

Maskin and Riley (2000) directly address the effects of asymmetry in the expectation of bidders in a common-values auction. They note that given Vickrey’s findings about the revenue-equivalence of different auction types, there is not a diversity of use of different auction types within specific fields. They hypothesize that this may be due to asymmetric information leading to violation of the revenue-equivalence theorem. Maskin and Riley do not state that certain markets may have organically settled into certain auction types as an unconscious response to this effect, although this can be seen to be a natural outcome of the hypothesis they are testing. They examine only a non-divisible good, and compare outcomes between first- and second-price auctions given different asymmetries. These include two bidders having completely different expected value distributions (one uniformly distributed over the interval [0, 1] and one over the interval [2, 3]), two bidders having similar distributions but where one has been “stretched” to have a higher upper bound, and bidders having different point-wise distributions. The first model leads to a price that is half of the low-value bidder’s value in the second-price auction and equal to the upper bound of the lower bidder’s value in the first-price auction. Revenue is below the high-bidder’s expected value, either in a first or second price auction. Bid-shading is an equilibrium outcome. In the second case, they claim that a first-price auction will yield a higher price than a second-price auction, with the gap increasing as the “stretch” of the distribution increases. In the third example, the second-price auction is shown to yield higher revenue. One important case that is not modeled is the case where both bidders have the same expected mean value, but different variances in their distribution. That is, one bidder is more confident than the other about the value of the good. Hausch (1987) examines bidder asymmetry in a common-values unit auction. He models the asymmetry as having one bidder consistently receiving “better” information from the distribution of proprietary information that is so frequently used in the literature. First, he models both bidders as having the symmetric signals and shows that the well-known revenue-ordering (second-price over first-price) holds, but when relaxing the assumption of symmetry, the
ordering is not necessarily maintained. That is, the second-price auction sometimes generates lower bids in the presence of bidder asymmetry. Hausch also shows that the seller may not want the bidders to be symmetrically informed, which is in contrast to most of the earlier work. The intuitive answer is that there are cases where a poorly-informed bidder is likely to bid more than the value of the good. Kagel and Levin (1986) also examined this idea, showing that in a market with an obvious winner’s curse, full public information reduces average revenue to the seller. Their work was later clarified by Cox, Dinkin, and Smith (1999) and Campbell, Kagel, and Levin (1999).

Examining asymmetry in a first-price indivisible-good independent-private-value auction, Lebrun (1999) proves the existence of an equilibrium. However, as others have shown, like so many other properties of the indivisible-good auction, this equilibrium is not transferable to a divisible-good auction.

Compte and Jehiel (2002) examine the welfare effect of adding extra bidders in an asymmetric-information setting. They examine a non-divisible good in English open-outcry and sealed-bid second-price auctions (note that Vickrey claimed revenue equivalence for these two types of auction.) The value of the good is a private value with asymmetric knowledge of some attributes of the good. They model the addition of a third bidder to a two-bidder model and claim that if that third bidder is symmetric to the original two bidders, there is no net effect on welfare. The third bidder is not bringing any new information to the model. However, when the third bidder has better information but a lower private value, the authors claim that welfare is decreased by the addition of the third bidder. In a second-price setting, the intuition behind this claim is that sometimes the bidder who does not have the highest private value sometimes wins the good. Thus, Pareto-efficiency is violated, and aggregate welfare is not maximized. The result is contingent upon this being a private-value auction – it is not true of a common-value auction.

Tenorio (1993, 1995) takes a closer look at the functional form of bidding functions. He models a foreign-exchange market in which no single bidder is large enough to change
the equilibrium price. He claims that in a discriminatory auction, the bidder will submit a bid that is equal to the marginal product of the quantity he is bidding for, whereas in a uniform-price auction he will submit a bid that is the average product of the quantity bid upon. Like many other papers in this field, Tenorio presents an example using greatly simplified distributions of values and information and shows that use of a discriminatory auction will result in lower-priced bids than an equivalent uniform-price auction. He states that due to the “lumpy” distribution of bids, it is possible that a non-Pareto-efficient allocation may be made but makes no comment on the relationship between bid price and ultimate value of the good. Like many other papers in the field, he makes clear that the results can not be assumed to have generality. Wang and Zender (2002) examine bidder asymmetry in divisible-goods auctions. They make the strong claim that given symmetric information and risk-neutrality there always exist equilibria that yield lower prices in a uniform-price auction than a discriminatory one. However, including asymmetry in the model leads to the absence of any sort of general equilibrium on revenue ordering. When including risk-aversion, the revenue order can be reversed – uniform-price auctions dominate discriminatory ones. In an earlier paper examining bidder asymmetry, Myerson (1981) claimed that in a market with asymmetric bidders, revenue maximization for the sellers is achieved by setting a different reserve price for each bidder based upon the quality of his information. Zheng (2002) claims that Myerson’s findings cannot be achieved without a resale market, likely because without this market knowledge of the quality of information is unobtainable.

In the area of experimental economics, Goswami, Noe, and Rebello (1996) examine bidder behavior in a divisible-good auction. They find that when bidders are able to communicate before the auction, collusion causes prices in a uniform-price auction to fall to lower than those in a discriminatory auction.

One of the earliest papers describing auctions in electricity markets is Duann (1991), examining auctions for merchant generation. He finds that a second-price, sealed-bid auction with special cost-sharing provisions appears to be the optimal method. Wolfram (1998) empirically examines bidding into generation markets in the UK. These are daily
multi-unit auctions for the supply of electricity, and the bids take the form of supply curves, not demand curves as in the FTR auction. These auctions are analogous to uniform-price auctions in that all generation is paid at the marginal (clearing) price. Wolfram finds evidence that strategic bidding raises marginal prices, given that firms can estimate their competitors’ cost schedules with the information from many previous auctions. Kamat and Oren (2002) look at various auction designs in another supply-bidding electricity-related market; that for spinning reserve. They model uniform-price and discriminatory auctions using social-cost and private-cost minimization criteria. They discover that using a social-cost, uniform-price framework maximizes allocative efficiency. Hortascu and Puller (2004) also examine supply-bidding, this time in the ERCOT (Texas) market. Comparing generators’ bidding data with marginal operating costs, the authors examine the optimality of the bidding structures. They discover that large generators’ bids were close to optimal, although smaller bidders deviated from optimality in a number of ways. They also claim the existence of learning over time.

There have been a small number of papers empirically examining uniform-price divisible-goods auctions in recent years. Berg et al. (2000) examine the functional form of the aggregate bidding schedule (i.e., for all bidders added together) in various treasuries markets (Israel, Norway, and Switzerland.) They discover that bids can be very closely fit to a logistic curve. This finding is robust over several different markets using several different auction formats. Vargas (2002) replicates their work in the Argentinean market, and finds that the logistic-curve format also exists there.

Nyborg, Rydqvist, and Sundaresan (2002) examine three bidder choice variables in 400 Swedish treasuries auctions. They regress bid-shading, bid dispersion and bid size on market volatility. They show that all three aspects of a bid are changed in respond to increased market volatility. This paper provides the foundation for the work performed in the essay at hand.

Février, Préget, and Visser (2002) attempted to model econometrically the unknown distributions employed in Wilson’s 1979 paper. The authors employed bid data from
French treasury auctions to develop non-parametric (i.e., “bin-based” or kernel) estimates of the distribution of the quantities bid upon (that is, the demand curve.) Using these data, the authors estimate parametric distributions for the user values and the proprietary knowledge. They discover that a modified Gamma distribution best fits the data. They then go on to show that the uniform-price auction provides the seller with about 5% less revenue in the markets they examine. This is the only known attempt at developing a structural econometric model for a multi-bidder auction.

Some general conclusions can be drawn from the auction theory literature:

• In an indivisible-good auction, a second-price format yields higher revenue than a first-price format. This relationship is not transferable to a divisible-good auction.

• Bidding your value is a dominant strategy in second-price unit-good auction, but not in a uniform-price divisible-goods auction.

• In the presence of asymmetric information, there is no deterministic way of ranking the auction formats in terms of revenue or Pareto-efficiency.

• In an auction with asymmetric information, revenue is maximized when there is reserve-price discrimination.

• Using a Vickrey pricing method in a divisible goods auction can maximize revenue.

• In a divisible-goods auction, bidding a schedule or demand curve strongly dominates presenting a single bid.

• Demand functions can be consistently modeled by a three-parameter logistic function in treasury auctions.
One may notice that none of the aforementioned literature addresses FTR markets. Several papers have been written about FTR markets, although none specifically address the PJM FTR auction, and only one author specifically addresses any FTR auctions at all. Most of the electricity transmission literature addresses issues of market power or how firms can be incentivized to build new transmission.

The basic framework of how an FTR system works, and a demonstration of how it can maximize aggregate welfare is drawn up by Chao and Peck (1996). Much of what was included in this paper was adopted into the PJM FTR system, although one interesting aspect was not: the idea that consumers could rank their reliability requirements by making differential payments for transmission insurance. This topic is further addressed by Woo, Horowitz, and Martin (1998) and again by Chao and Peck (1998). Bushnell and Stoft (1997) further explain how using a nodal pricing system can provide incentives for private investment in new transmission, questioning the received wisdom that transmission must be tightly regulated. The physical constraints placed on flow in electricity grids by Kirchhoff’s laws are the focus of a paper by Cardell, Hitt, and Hogan (1997), who show that it may be possible for generators to exploit transmission constraints by increasing output, as opposed to the normal monopolistic behavior of restricting output.

Oren (1997) claims that while an FTR system may work well in defining market prices at consumption nodes, over time generators will be able to capture all congestion rents by strategically bidding up the price of generation until we see a single price across the grid. That is, the lowest-cost bidders in a repeated game will discover their competitors’ cost schedules and increase their bids into the day-ahead market. Each firm will increase their bids to the level of the firm above them, until the aggregate supply curve becomes a horizontal line at the highest marginal cost normally observed anywhere in the grid. This proposition is similar to that presented by Wolfram. Were this true in the PJM market, we could expect to see transmission revenues shrinking over time, and the average system-wide price approaching the typical maximum value. This hypothesis has not yet
undergone empirical analysis in the PJM market, and may be an interesting basis for a future paper.

Borenstein, Bushnell, and Stoft (1998) examine the impact of transmission capacity on competition amongst generators. The authors find that there is not necessarily a strict relationship between the amount of power flowing along a transmission line and the competitive effects that line has on generation. The follow-on is that small additions to transmission in regions where it is currently constrained can have large aggregate welfare benefits. Joskow and Tirole (2000) examine how the addition of a transmission rights regime to a market affects the decisions of generators. They examine both physical and financial transmission rights. They claim that the addition of transmission rights can affect market power, but the exact effects differ according to the microstructure of the rights. They claim that physical rights can have much worse effects than financial ones.

Léautier (2001) simulates a three-node triangular transmission network and reaches the rather unremarkable conclusion that increased transmission availability reduces the average price of power by increasing competition amongst generators and thus generators have no incentive to build additional transmission.

Willems (2002) examines different methods of allocating constrained transmission in the framework of a Cournot duopoly in generation. He claims that a locational marginal pricing model is optimal in a perfect competition setting but not in an unregulated duopoly.

The only work uncovered that addresses bidding behavior in FTR markets is that performed by Zang (2005). She examines bidder behavior in the Texas transmission grid. Zang focuses on the relationship between owning FTRs and bidding into the day-ahead power market, studying the idea that owning FTRs can affect behavior in the DA market, and the effects of the DA bidding can affect how firms bid for FTRs. The Texas market is different from PJM in that it uses a small number of zonal prices and thus has a small
number of FTR pathways. It also has a small number of competitors and no retail wheeling, since it is almost completely isolated from the rest of the national grid.
2.6 Economic model specification

When addressing questions about bidder behavior in uniform-price divisible-goods auction, we need to look at the foundation of the bidders’ strategies. The basic assumption is that firms strive to maximize their profits from participating in the FTR market.

The simplest formulation for profit is that it equals the difference between revenue and cost, or in the context of the FTR market, the difference between the value of the FTR and the clearing price of the FTR:

\[
\Pi_{A,B,M} = V_{A,B,M} - CP_{A,B,M}
\]

(68)

Where:

- \( \Pi_{A,B,M} \) = profit from owning FTR with source \( A \) and sink \( B \) during month \( M \)
- \( V_{A,B,M} \) = revenue of FTR with source \( A \) and sink \( B \) during month \( M \)
- \( CP_{A,B,M} \) = clearing price of FTR with source \( A \) and sink \( B \) during month \( M \)

The above formulation assumes normalization of FTR size and duration, that is, the profit is measured in units of $/MWh. With perfect information, the expected value of the FTR will be common knowledge. However, given that the FTR auction takes place two weeks before the beginning of the month-long duration of the FTR, this information is not knowable with certainty. Nodal FTR price differences display significant variability over time in a way that is not obviously predictable. The expected value of the FTR for each bidder will be a product of the prediction algorithm employed by the bidder.

We may wish to assume that the realized value of an FTR (i.e., the future flow of rents) is exogenously determined. That is, an FTR holder cannot maximize the value of the congestion rent collected in the following month by modifying his own behavior. If the realized value is assumed to be exogenously determined, then we have a common-value auction. This is similar to the treasury auction, where the realized value of the auction good is the resale value of the treasury note, in contrast to a private-value auction for
something like a piece of art, where the value to the bidder is neither exogenously determined nor readily monetized.

Thus, given the above exogeneity assumption, the expected value derived by each bidder will be his forecast of the average value of a pair of nodal prices in the following month. His expectation of the value v is some realization of a random variable V which has the cumulative density function \( G(v) = Pr(V \leq v) \). In the auction-theory literature, differentiation of the expectation of rival bidders is modeled by the assumption of proprietary knowledge: each bidder knows something about the expected value that is not known (or may not be known) by other bidders. This proprietary knowledge is typically modeled as knowledge of a sample of \( V \) and is summarized as an estimate \( s_i \) that is the realization of a random variable \( S_i \), which has the conditional cumulative density function \( F(s, v) = Pr(S_i \leq s_i|V = v) \). In this framework, the ability of a bidder to accurately model expected values is the proprietary knowledge contained in the estimation process: some bidders may have better access to past information about flows and congestion and, given a certain level of information, may have more sophisticated algorithms for converting that ex-post information into accurate predictions of the future.

**Bidding strategy**

In a non-divisible item auction, this is the amount by which a bidder shades his bids. In the divisible-good uniform-price auction, this takes the form of a schedule of bids defined by \( q = q(p, s_i) \), that is, the bidder requests amount \( q \) at price \( p \) based upon the proprietary information he has received. A couple of conditions on these bids apply: \( \sum_p q(p, s_i) \leq 1 \): that is, the sum of the shares for a single bidder over his entire bid schedule should not be greater than the total amount of divisible good available. It is possible to violate this condition, but the only result is that a bidder’s lowest priced bids (equal to the lowest \( \Sigma q – 1 \) share of his bids) will be pushed off his bid schedule by his higher priced bids. To analogize to an indivisible good: a bidder may be able to enter multiple different priced bids on the same good in a sealed-bid auction, but, by doing so all but the highest priced bid are invalidated. From the above condition, it follows that each bid shall be on the interval \( (0 < q(p, s_i) \leq 1) \).
If we assume that bidder $i$ is using bidding strategy function $y = y(p, s_j)$ and every other bidder is using strategy $q$, then the clearing price, $p^*$, will be that for which

$$y(p^*, s_i) + \sum_{j \neq i} q(p^*, s_j) = 1$$  \hspace{1cm} (69)

One important observation to draw from this is that the clearing price, $p^*$, depends upon the information samples of every bidder, and not just $s_i$. Thus, there is uncertainty for bidder $i$ as to what the clearing price shall be, since he is not privy to other bidders’ proprietary information, $s_{j, j \neq i}$.

Wilson (1979) defines the clearing price as the realization of a conditional cumulative density function, as follows:

$$H(p; v, y) = \Pr\{p^* \leq p \mid V = v, y(p, s_i) = y\}$$  \hspace{1cm} (70)

Assuming that utility is linear with profit, the first-order condition for profit maximization can be shown to be:

$$0 = E\left[(V - p)\frac{\partial H}{\partial p}_{y=y(p, s_i)} + y(p, s_i)\frac{\partial H}{\partial y}_{y=y(p, s_i)}\right]$$  \hspace{1cm} (71)

An optimal strategy is one where $y(\cdot) = q(\cdot)$.

Results from the above formulation rest upon the functional forms and the values of the parameters of the density functions $F(\cdot)$ – the distribution of proprietary information, and $G(\cdot)$ – the distribution of the values.

The firm maximizes its profits by placing bids into the FTR market. The profit function is non-continuous: if the bid is lower than the clearing price of the FTR, then the profit is
zero, assuming that bidding is a costless exercise. If the bid is greater than or equal to the clearing price, then the profit will be equal to the ex-post realized value of the FTR minus the clearing price.

**Modeling the FTR market**

The market for FTRs can be modeled as a residual and fringe market, where the fringe is the bidder we are studying, and the residual market is the aggregate of all other bidders. If we abstract the bidder’s demand functions as continuous, the FTR market is as described in Figure 2.16. There are two demand curves: one for bidder X, defined by $P = c - dQ_X$, and one for the rest of the market, defined by $P = a - bQ_R$. Added together, these two curves represent the bidding market for FTRs. It can be generally assumed that the quantity that bidder X bids is less than the total amount available and that he will not typically bid a complete curve all the way down to $P = 0$, but instead will have a truncated demand, as shown in Figure 2.16. Thus, bidder X’s bid ranges from a maximum bid of $P = P_{X^*}$ at $Q = 0$ to a minimum bid of $P_X$ at $Q = Q_X^*$.

![Figure 2.16: FTR market structure](image)
For bidder X there are three possible outcomes, dependent upon the positioning of his bid and the quantity of FTRs available. In Figure 2.16, three possible supply curves are shown. The supply curve is vertical, as the capacity of power flow between two points is fixed. Later in this essay, it is shown that this assumption can relaxed, but for the present time assuming a vertical supply curve causes no loss of generality in the solution.

Firstly, assume that we have supply curve $Q = Q_{F1}$. This intersects the combined demand curve at price $P = CP_1$. Thus, the clearing price is higher than bidder X’s highest bid, $P_X^H$, and in this case, bidder X will not win any FTRs.

The second case is illustrated by supply curve $Q = Q_{F2}$. This intersects the composite demand at $P = CP_2$, which is in between bidder X’s highest and lowest bids. This means that part of his bid will be successful: he will be awarded some of the FTRs he bid for, the fraction being calculated by $(CP_2 - P_X^L)/(P_X^H - P_X^L)$.

The third case is illustrated by supply curve $Q = Q_{F3}$. This curve intersects the composite demand at $P = CP_3$, which is below the lowest amount of bidder X’s bid. This means that all of his FTR bids were at a price higher than the market clearing price, and as such all will be successful. Bidder X wins FTRs in the amount of $Q^*_X$.

The profit from winning the FTRs is obtained by multiplying the net profit, the value of the FTR minus the clearing price, by the quantity of FTRs awarded.

**Estimators for bids**

In the above explanation bidding demand functions were abstracted as continuous functions. This is not the case in reality. The residual-demand curve is composed of parts of all of the bids for all FTRs, and given that there are often several thousand bids in a month, the curve may reasonably be assumed to be continuous. However, an individual’s bidding function will be a series of discrete bids, each one offering to purchase a certain amount of FTRs at a certain price. Such a set of bids is difficult to functionalize, and must be transformed to a “continuous-seeing” proxy. If we model a continuous form,
the bid can be fully specified by three parameters: the maximum price, the minimum price, and the sum of the quantity of all bids. This formulation loses some detail in that it implies a linear bidding structure. Our discrete bidding format is described as linear if the bidder submits an equal sized quantity for each bid component, with equal sized bid increments between each bid component. In reality, it is not necessary for a bid structure to be linear; indeed, in the work of Berg, et al., (2000) and Vargas (2002) both describe how the aggregate-demand curve for treasury securities can be accurately be modeled as a logistic curve. However, it will be shown in the following section that the vast majority of composite bids in the FTR market are, in fact, linear; and thus representation as a single sloped line does not result in a significant loss of information. Single component bids are even simpler to represent, as the maximum and minimum prices are identical, and the slope of the demand function is zero. This idea is explored in Figure 2.17.

![Figure 2.17: Discrete and continuously-modeled bids](image)

In Figure 2.17 we have two series of bids. The six red lines define a linear series of bids: each component of the bid is shown by a solid red horizontal line, and these lines are all the same length and are equally spaced vertically (i.e., price-wise) between 1 and 6. They can be modeled as a continuous function, shown as the dashed red line. This line accurately reflects the collection of the individual bids. The black lines in Figure 2.17 display a non-linear bid, one that closer matches a logistic bidding strategy: the middle
component of the bid is larger than the two outer components, and the price-spacing is not equal. Replacing this discrete bidding function with the continuous function (the dashed black line) results in a loss of information about the structure of the bidding function. As mentioned above, it will be shown that the vast majority of the actual bids in this market are linear in nature.

**Evaluation of bidding components**

As per the previous section, all bids shall be abstracted to a continuous function with three components: the minimum price, the maximum price and the sum of the quantities of the individual bids. Thus, a bid is described by:

\[ B_{X,m,n,t} = f(P_{X^-}, P_{X^+}, Q_X) \]  

(72)

Where:

- \( B_{X,m,n,t} \) = bid of firm \( X \) on path \( m-n \) in month \( t \)
- \( P_{X^-} \) = price of lowest-priced bid component
- \( P_{X^+} \) = price of highest-priced bid component
- \( Q_X \) = sum of quantities of bid components

Let the three components be modeled by linear estimators:

\[ P_{X^-} = X_1 \beta^- + \varepsilon_{P^-}, \quad \varepsilon_{P^-} \sim N(0, \sigma_{P^-}) \]  

(73)

\[ P_{X^+} = X_2 \beta^+ + \varepsilon_{P^+}, \quad \varepsilon_{P^+} \sim N(0, \sigma_{P^+}) \]  

(74)

\[ Q_X = Y \alpha + \varepsilon_Q, \quad \varepsilon_Q \sim N(0, \sigma_Q) \]  

(75)

\( X_1, X_2 \) and \( Y \) are vectors of exogenous regressor variables. These three components are all mutually independent; save the necessary condition that \( P_{X^+} \geq P_{X^-} \). The size of the bids is in no way constrained by either of the price bounds.

What are the causative factors to be considered when making bids? First, we shall consider the prices. The maximum price is constrained by the value of the FTR. It makes no sense to place a bid that is higher than one’s expected value of the FTR, as it is possible that the closing price will be higher than the value of the FTR, and the bidder has then purchased something at a price higher than what he expects the return from owning
this good to be. In an indivisible-good second-price auction, this results in the person bidding what he thinks the good to be worth: if he wins, he is guaranteed to pay no more than what he values the good at. Since this is a divisible good, we are not constrained by a single expected value. A composite bid is essentially a submitted demand curve, and as such it necessarily slopes downwards. Demand curves for non-divisible goods slope down because of the concept of declining marginal utility: as one consumes more of a good, the benefit from consuming an additional unit declines. Thus, we are less willing to pay as much for the next unit of the good, because we do not get as much out of it. Clearly, this does not apply in the FTR market: the marginal value of every unit of FTR is the same. So why does the demand curve slope downwards? The answer lies in the variable expectation of the utility from owning the FTR. The value of the FTR is a realization of some randomly distributed variable. According to Wilson’s work, profit is maximized by submitting a bid that is a function of this distributed random variable.

Wilson (1979) showed that bidding the expected value (or expected distribution of the value) is not the profit-maximizing strategy in a uniform-price divisible-good auction. This is contrary to the early-held belief of people like Friedman that the truth-telling property of the second-price indivisible-good auction could be directly transferred to the uniform-price auction. This can be shown why, intuitively, with a simple example. Assume that a person ascribes to some divisible good a future value that is uniformly distributed over the interval [0, 1]. If we see the good as being perfectly divisible into any fraction between 0 and 1, then the truth-telling bidder would assume the optimal bidding function to be a straight line from the points (0, 1) to (1, 0). The quantity available is 1. With a single bidder, the clearing price would be zero. With two bidders the clearing price would be 0.5, which is the expected value of the good, and each bidder would win half of the good. However, if there were three bidders all submitting the same bid schedule the clearing price would be 2/3, and each bidder would win 1/3. Generalized, with \( N \) bidders, the clearing price can be shown to be \((N-1)/N\), with each winner winning a fraction of the good equal to \( 1/N \). Everybody wins and everybody pays more than the expected value of the good: the winner’s curse becomes a winners’ plague. If we wish the clearing price to be the expected value of 0.5, then each bidder should multiply his bid
function by a correction factor equal to \( N/(2(N-1)) \). As \( N \) approaches infinity, this correction factor approaches 0.5. This is concordant with Wilson’s slightly more complicated example, where he assumes a Weibull distribution for the value of the good. He still yields an optimal bidding strategy of 0.5. This behavior is shown in Figure 2.18, below.

![Figure 2.18: Demand with different numbers of bidders](image)

In Figure 2.18 there are several composite demand curves, corresponding to 1, 2, 3, 5 or 10 bidders, each with the same uniform distribution of values over the range \([0, 1]\). The clearing price at each number of bidders is the intersection of the vertical supply curve with the demand curve in question. As we can see, as the number of bidders increases, the clearing price increases to above the expected value of 0.5.

The previous example is generalizable for any symmetrical distribution with upper and lower limits, although the form of the “correction factor” depends on the shape of the distribution. The uniform distribution is the ultimate “fat-tailed” distribution; as a distribution becomes more centrally distributed (i.e., increasingly leptokurtic) then the “correction factor” should decrease. However, this simple example does not consider bidder asymmetry: different bidders may have different knowledge and beliefs about the expected value of the FTR, and the functional form of its distribution. A perfectly-
informed bidder would have a point-wise distribution; a less informed one a fatter-tailed distribution. Furthermore, the distributions may be asymmetric, reflecting risk-aversion. The point being, as raised by several authors in the auction-theory literature, that asymmetry in the bidder population greatly increases the mathematical complexity of modeling expected auction outcomes and renders a closed-form solution essentially unobtainable. To model an equilibrium strategy, we need *ex-ante* knowledge of each bidder’s demand curves – information usually not obtainable.

**Upper limit of bid**

Thus, the upper limit of the bid should be the expected upper limit of the expected value of the FTR. We are assuming here that the value of the FTR is exogenously determined by the sum of the congestion in the following month. By exogenously determined, we mean that actions of the FTR holder cannot affect the value of the FTR. On its face, this is not an unreasonable assumption. The holder of a long FTR is betting on there being more congestion that expected by the consensus FTR auction outcome. But who can generate more congestion? A load-serving entity is essentially slave to its customers’ demand curve. Power marketers are in the same boat, typically having long-term contracts to supply power at pre-determined prices. There are two types of person who wish to see greater congestion: generators in high-priced regions and FTR speculators. If the generator owns an FTR with the sink in his high-priced local area and the source in some low-priced area, he benefits if less power is imported into the high-price region. However, it has been shown (Zang, 2005) that this generator’s profit-maximizing behavior will be the same with or without ownership of the FTRs. He benefits from the congestion in two ways: selling power at a higher price and collecting congestion-rent revenue: if he owns the FTRs, he will make more money than if he does not own them, but his profit-maximizing operating point does not change with ownership of the FTR. Thus, owning an FTR will not change a generator’s behavior. The FTR speculator does not participate in the day-ahead energy market, so he is unable to affect congestion. Thus, the assumption of congestion being exogenous to the ownership of an FTR is a defensible one.
Given that we are assuming that the value of the FTR is exogenous to any decisions by the owner, we must then consider how the bidder forecasts the value of the FTR. As shown in Chapter One of this study, congestion is not easy to forecast. It was shown that as load increases, the propensity for congestion to occur increased, but it was also shown that a great deal of congestion was caused by system outages that occur stochastically.

**Lower limit of bid**

For a non-divisible good, the lowest-priced rational bid would be at the expected value of the clearing price of the FTR. Bidding below the clearing price means that one is making a bid that one expects to be rejected. As with the value of the FTR, the clearing price is the realization of a randomly distributed variable and given the divisible nature of the FTR, the bidder has some freedom to structure his bid contingent upon the form of the clearing price distribution. The important consideration here is the interaction between the expected clearing price and the expected value of the FTR. If the clearing price is greater than the FTR payout, then anybody who wins the FTR will be a guaranteed money-loser. However, this limit is defined by the upper price, and we are now concerned about the lower price limit, which is necessarily less than or equal to the upper limit. There is also a lower limit to the expectations of the value of the FTR – bidding a lower limit than this makes no sense: a lower bid decreases the probability of winning without increasing the probability of a positive payoff, thus it unambiguously reduces the expected payoff. However, the low end of the bid schedule is much more likely to be affected by the clearing price than by the value of the FTR: by definition of the probability distribution of the value, the distribution function of the expected clearing price will dominate the payoff function, and thus the decision about where to cut off the bidding. The expected profit is the expected winning percentage of all bids times the difference between the expected value and the expected clearing price: therefore, the distribution of expected profit is contingent upon the distribution of the value and the clearing price. Our winning percentage is unambiguously increased by increasing the lower value of the bid, but the profit is not. As certainty increases, we can see that the lower and upper limits of the bid will converge; the greater the uncertainty about these distributions, the greater the spread of the possible profits, which means a larger area of
the curve at values of profit less than zero. Once again, the model is complicated by the lack of knowledge of the bidders’ assumptions made about probability distributions. What is clear is that we can expect the past behavior of clearing prices to have some bearing on the lower limit of the bid. Of course, we expect the clearing price to be affected by market players’ beliefs about what the value of the FTR will be.

The upper and lower price levels will be estimated using the ordinary least squares methodology.

**Bid quantity**
The obvious comment about bid quantity is that it does not make sense to bid a quantity that is greater than the total amount available. If one does so, all one is doing is disqualifying one’s lower-priced bids. On the other hand, bidding for less than the full amount of FTRs is leaving money on the table if one assumes that the FTR will be profitable. An increase in uncertainty about the clearing price and value might cause a speculator to reduce his bid quantity, but for a hedger it makes no sense to bid anything but the amount of FTRs available. However, it is not clear how many FTRs are available on a certain path. This is due to the loop-flow externality and the interaction of all flows in a network. An FTR is specified as a point-to-point entity, but in reality, power that is injected at a source and removed at a sink does not flow in a point-to-point fashion. Instead, part of that power flows through every line in the network. Due to impedance and line-loss considerations, the further away a line is from either source or sink node, the less the effect on the flow pattern.

Considering an FTR as a series of flows through each line in a network, we begin to understand how different bids compete with each other. It is frequently observed in PJM that only one company bids on the FTRs on a specific path, and thus a layperson might assume that there is no competition for that FTR. But basically the FTR is a bundled collection of mini-FTRs, one for each line in the network. If parts of the network are uncongested, then the value of the FTR along those lines will be zero, but in congested regions there is a price to be paid for using every link in a network. For example, it is
common to see many different nodal prices within a small geographic region. Thus, a single FTR is actually a weighted sum of the flow though each line in the network times the difference in the nodal LMPs on either side of that line. Since every MW of power injected into the grid splits itself up and flows through every line in the grid, and since the FTR simultaneous feasibility test is modeled on a DC flow model, then every FTR bid competes with every other FTR bid. What sometimes looks like a thin market on a point-to-point basis is in fact almost perfectly competitive, from the context that every bidder (indeed, every single bid) is competing for shares of every link in the network.

Therefore, if somebody wants to win more FTRs along a certain path, then that person has to outbid all other bidders for ownership of the rights along many shared paths. This tends to invalidate the assumption of a vertical supply curve. Instead, the supply curve is upward sloping: higher bids are analogous to an outward shift of the demand curve, and with an upwards sloping supply curve this means that more FTRs are awarded. Note that more FTR capacity along one path means less along another path: if a bidder for a different FTR wants more on his path, he must bid higher. The result here is that the clearing price of an FTR and the available capacity on that path are causally connected: *ceteris paribus*, a higher clearing price means more FTR capacity.

Another consideration is the presence of short-sellers. This is somebody who sells now and buys later. If a person makes a negative-priced bid on a contra-flow FTR, then he is essentially supplying additional FTR capacity in the opposite direction. Therefore, the presence of a lot of people making short bids has the tendency to shift the supply curve down, lowering the clearing price. A short-seller affects the market if his bid is accepted. If it is not, such as when a short-seller tenders a very large negative bid, then he does not because he is only affecting the supply curve far to the right of the equilibrium point.

Thus, when attempting to model the quantity bid upon we assume that a rational bidder will attempt to bid for the full quantity of FTRs available along the path in question, and that the available FTR capacity is affected by the clearing price (higher = more FTRs)
and the presence of short sellers. The only constraint upon a bidder is his credit limit with PJM. Bid quantity is estimated using the OLS methodology.

**Modeling market outcomes**

The first stage of the analysis undertaken in this essay involves modeling the three choice variables any bidder has to decide upon: the upper and lower prices and the quantity bid upon. That is, we are examining individual firms’ decisions. The second part of the analysis involves examining the market outcomes that arise from the decisions that these firms make. The two outcome variables are the firm’s winning percentage (this baseball terminology is employed for familiarity, even though, like in baseball, what is modeled is actually a fraction and not a percentage) and the profits that accrue to the firm from winning the FTR.

The FTR process being modeled here has three distinct stages:

1. Firms formulate and enter bids into the FTR auction
2. Aggregate of all bids in the SFT defines the allocation of FTRs
3. Congestion rent revenues distributed to FTR holder

There is a strict temporal ordering to these events: outcomes of the auction cannot affect a firm’s bids, and the amount of congestion in the market does not affect the allocation of FTRs. As such, it is applicable to employ a staged estimation process, examining each phase of process as a function of the inputs from the previous stage.

The bid formulation, defined by the size and maximum and minimum prices, is described above.

In the second stage of the process, the first-stage estimates of the bid parameters shall be used as exogenous regressors in the maximum-likelihood estimators for winning percentage in the FTR auction. That is, our estimated bids define our estimate of the auction outcome.
The third stage of the process looks at the profitability of owning FTRs after they have been awarded. The profits are modeled over the set of all FTRs awarded.

This is a modified three-stage least-squares (3SLS) methodology, expect the second and third stages are not least-squares, but maximum-likelihood. Thus, this model can be called a three-stage least-squares/maximum-likelihood combination model, abbreviated as 3SLS/ML. For more information on multiple stage least-squares models, please consult any econometrics textbook, such as Greene (2000) or Intriligator, Bodkin, and Hsiao (1996). The use of OLS for modeling of the first-stage variables, as well as the practice of modeling them individually, and not jointly, is consistent with the other empirical work performed in this area (i.e., Nyborg, et al., and Cammack.)

**Winning percentage estimator**

The winning percentage of a composite bid takes one of three discrete forms:

\[
WP = \begin{cases} 
0 & \text{if } CP > P_{X^+} \\
1 & \text{if } CP < P_{X^-} \\
\frac{P_{X^+} - CP}{P_{X^+} - P_{X^-}} & \text{if } P_{X^-} > CP > P_{X^+} 
\end{cases}
\] (76)

Where:  
\( WP = \) winning percentage, fraction of bid quantity that is awarded  
\( CP = \) clearing price  
\( P_{X^-} = \) lower price limit of bid  
\( P_{X^+} = \) upper price limit of bid

The variable to be modeled in this equation is the clearing price:

\[
CP = Z^\gamma + \varepsilon_{CP}, \varepsilon_{CP} \sim N(0, \sigma_{CP})
\] (77)

\( Z \) is a vector of exogenous variables. Looking at the three regimes in Equation (76), and substituting Equation (77), we get the following: the probability of winning none of the bid FTRs:
\[ Pr\ (CP > P_{X^+}) = Pr(Z\gamma + \varepsilon_{CP} > P_{X^+}) = Pr(\varepsilon_{CP} < Z\gamma - P_{X^+}) = F\left(\frac{Z\gamma - P_{X^+}}{\sigma_{CP}}\right) \quad (78) \]

The probability of winning 100% of the quantity bid is:

\[ Pr\ (CP < P_{X^-}) = Pr(Z\gamma + \varepsilon_{CP} < P_{X^-}) = Pr(\varepsilon_{CP} < P_{X^-} - Z\gamma) = F\left(\frac{P_{X^-} - Z\gamma}{\sigma_{CP}}\right) \quad (79) \]

The likelihood of observing \( 0 < WP < 1 \) is:

\[ LH\left[ WP = WP_i \right] = LH\left[ WP = \frac{P_{X^+} - (Z\gamma + \varepsilon_{CP})}{P_{X^+} - P_{X^-}} \right] = LH\left[ \varepsilon_{CP} = P_{X^+} - Z\gamma - WP(P_{X^+} - P_{X^-}) \right] = \frac{1}{\sigma_{CP}} f\left(\frac{P_{X^+} - Z\gamma - WP(P_{X^+} - P_{X^-})}{\sigma_{CP}}\right) \quad (80) \]

An indicator variable set is defined as follows:

\[ J_1 = 1 \text{ if } WP = 1, \text{ else } 0 \]
\[ J_2 = 1 \text{ if } WP = 0, \text{ else } 0 \]
\[ J_3 = 1 \text{ if } 0 < WP < 1, \text{ else } 0 \]

And the log-likelihood function (LLF) is described by:

\[ LLF = J_1 \sum_{WP=0} \ln Pr(CP > P_{X^+}) + J_2 \sum_{WP=1} \ln Pr(CP < P_{X^-}) + J_3 \sum_{WP \in (0,1)} \ln LH(WP = WP_i) \quad (81) \]

Equation (81) is maximized with respect to \( \gamma \) and \( \sigma_{CP} \), yielding the best unbiased estimate of those parameters.

**Profits estimator**

The inputs into the estimator for profit are the number of FTRs awarded, the clearing price and the ex-post value of the FTR. The profit from owning an FTR is defined as:
\[
\Pi = WP \cdot Q_x (V - CP)
\]  \hspace{1cm} (82)

Where

- \( P \) = profit, $/on-peak hour
- \( WP \) = winning percentage
- \( Q_x \) = size of bid, in MW
- \( V \) = congestion rent accruing to owner of FTR
- \( CP \) = clearing price of FTR path

To obtain a measure of profit that is more common across all awards, both sides of Equation (82) are divided by \( Q_x \cdot WP \), which leaves us with a result that is measured in $/MWh of owned FTRs. Every component of Equation (82) has been previously estimated in an earlier stage of the model except for the value of the FTR. Therefore, this stage of the model tests how good a bidder is at “picking winners.” There are two skills being tested in this essay. One is the ability to make a bid along a certain path that will maximize expected profit along that path. The second skill is selecting which of the near-infinite (> 7,000!) number of paths to bid upon, i.e., choosing to go after particularly profitable FTRs. This can be boiled down to forecasting the value and clearing price along different paths in the grid, and as was shown in the previous chapter, this is not a trivial exercise. Furthermore, some firms do not choose their paths: a hedger is constrained to bidding along paths that he will expect to deliver power along in the day-ahead market.

The point being that the only variable in this stage of the model will be the \textit{ex-post} value of the FTR, or the amount of congestion rent collected along the path. As previously discussed in this chapter, it is not unreasonable to assume that the congestion is exogenous to the owner of the FTR – he cannot change it. Thus, this stage of the model is a test of picking winners in the same way one might pick a racehorse to bet upon.
The differentiation between earning zero profit (i.e., not winning any FTRs) and earning some (positive or negative) profit is contained in the winning-percentage estimation. This estimation provides evaluation of only non-zero profits.

Let the per-MWh value of an FTR be estimated by:

\[ V = W\delta + \epsilon_v, \; \epsilon_v \sim N(0, \sigma_v) \]  \hspace{1cm} (83)

\( W \) is a vector of exogenous regressors. Thus, Equation (83) can be written as:

\[ \frac{\Pi}{WP \cdot Q_x} = (W\delta + \epsilon_v - CP) \]  \hspace{1cm} (84)

And written in the form of a likelihood function we obtain:

\[ LH[\epsilon_v] = \frac{\Pi}{WP \cdot Q_x} + CP - W\delta \]

\[ = \frac{1}{\sigma_v} f \left( \frac{\Pi}{WP \cdot Q_x} + CP - W\delta}{\sigma_v} \]  \hspace{1cm} (85)
2.7 Econometric model specification

Given the estimators as specified in Section 2.6, consideration must now be made for which variables to use as exogenous regressors. The models define upper and lower price limits, but in reality many bids are single component bids. Therefore, the upper and lower prices are the same. It is not valid to perform two estimations for a single variable. For this reason, the bid data set was segregated into two subsets, single-component bids and multiple-component bids. For single-component bids, a single price regression was run, but for multiple-component bids the upper and lower limits were modeled separately.

**Independent variables for single-price estimator**

Based upon this information, the single-component price was modeled using the following variables.

AVGVAL: this is the average value of the FTR over the past three months of observed data; that is, using data from two, three and four months before the month of the FTR in question. The value for the previous month is not available at the time of the auction, since the auction takes place half-way through the period in question. Thus, use of this variable tests the bidders’ response versus changes in the absolute value of the return from the FTR over the three previous months. We would expect the parameter on this variable to be positive: as the value increases, then it is “safer” for a firm to make a higher bid. As mentioned in the previous section, we expect the upper end of the bid range to be the upper end of the distribution of the values of the FTR, and a higher mean value will translate to higher upper limit, *ceteris paribus*.

SDVAL: the uncertainty of the upper price is affected by the uncertainty of the value of the FTR. This was captured by using the standard deviation of the value of the FTR over the previous three months as a variable. We would expect the parameter on this variable to be negative: a greater variance means greater uncertainty on the value of the FTR, and greater uncertainty in an outcome creates more conservative bidding responses.
AVGCP: this is the average value of the clearing price of the FTR over the past three auctions. This serves as a lower limit on the value of the bid. We would expect the parameter on this variable to be positive: as the clearing price increases, all bids are shifted upwards. As the clearing price mostly affects the lower part of the demand curve, we would expect this variable to be less important for the maximum price than for the minimum price.

SDCP: this is the standard deviation of the clearing price over the past three months. A higher number means less certainty in the value of the clearing price. We would expect the parameter on this variable to be negative, once again implying a conservative response to increased uncertainty.

PJM: this is a binary indicator variable that take a value of one if a bidding firm is a PJM native utilities, zero otherwise. This is to test whether PJM firms modify their bids in a significant and consistent manner that would reflect a difference in the quality of information held by these firms versus “outsider” firms. The parameter value will be based upon the behavior of all firms: if outsiders tend to be risk averse and underbid, we expect this variable to be positive if PJM firms are more certain about the true values of the clearing price. If other firms tend to be over-optimistic and bid too high for the FTR, raising the clearing price above the expected value of the FTR, we would expect to see PJM firms bidding below the other firms’ average. Given that clearing prices are more often than not below the final values, we would expect this parameter to be predominantly positive.

EXP: this variable measures the experience of the firm in the FTR market, specified by how many previous FTR auctions the firm has taken part in. As a firm obtains more experience in the FTR market, we would expect it to gain skill in predicting the clearing price. If we assume that PJM firms have better information, then we would expect the parameter to have the same sign as the PJM variable. The idea behind this is that over time, outsiders become more like PJM firms in their knowledge base. One possible confounding factor is that a lot of outsider firms have short stays in the market, whereas
the PJM firms are long term players, and therefore many of the firms with lots of experience are the PJM firms.

SHORT: this is a binary indicator variable that takes the value one if the bid in question is a short bid, zero otherwise. A short bid is one in which the upper limit is less than or equal to zero. This is a bid on a contra-flow FTR, or a case of somebody betting that there will be less congestion than the aggregate market consensus. This variable is included to model any structural differences between long bids (bets on more congestion) versus short bids. There is no obvious ex-ante expectation on this variable. Obviously, the upper limit is upper-constrained at zero, but the absolute position of the bid on the number line is should be captured by the AVGVAL and AVGCP variables, which can be seen as “goalposts” in between which a bidder is attempting to place his bid. The SHORT variable tells us whether the position of the bid, relative to these goalposts, is consistently structurally different. Employing the SHORT variable assumes that all of the differences in behavior can be captured in a constant term. This is what is referred to as a “fixed-effects” model: whether a bid is long or short, it exhibits the same response to the other exogenous variables: the slope of the line stays the same with respect to all variables, all that changes is the intercept. However, if the two forms of bids are in fact structurally different, then simply moving the intercept may not provide us with sufficient differentiation. Instead, the long and short bids can be segregated and modeled as separate populations. This provides us with a variable-effects model. In this study, both fixed and variable effects models were estimated.

SUMMER: this is a binary indicator variable that takes the value of one for auctions held for the months of June, July, and August, zero otherwise. It is designed to capture different firm behavior during the high-load summer period. If there is more congestion during high-load periods, we should expect this variable to be positive.

WINTER: this is a binary indicator variable that is similar to the SUMMER variable, except it takes a value of one for the months of December, January, and February. Once again, this is designed to capture behavior changes in the high-load winter months.
During cold weather, heating-related loads increase, but the transmission capacity of the grid also increases. Therefore, the expectation on this variable is not *ex-ante* intuitively obvious. If congestion increases in the winter, it will be positive, but if the additional network capacity outweighs the congestion effects of higher loads, it will be negative.

**Independent variables for upper price limit estimator**

The lower price limit was estimated using the same variables as the single-component price. Expectation on all variables is the same as those in the upper price distribution.

**Independent variables for lower price limit estimator**

The lower price limit was estimated using the same variables as the single-component price. As mentioned above, we can compare the responses for the lower price and upper price to the AVGVAL and AVGCP variables. We expect the upper limit to be more strongly influenced by the AVGVAL variable, the lower price to be more strongly influenced by the AVGCP variable. Expectation on all variables is the same as those in the upper price distribution.

A summary of the independent variables used in the estimators for the upper and lower bounds on bid prices, as well as their expectations, is contained in Table 2.1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Expectation on parameter sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>PJM</td>
<td>Indeterminate</td>
</tr>
<tr>
<td>AVGVAL</td>
<td>Positive</td>
</tr>
<tr>
<td>SDVAL</td>
<td>Negative</td>
</tr>
<tr>
<td>AVGCP</td>
<td>Positive</td>
</tr>
<tr>
<td>SDCP</td>
<td>Negative</td>
</tr>
<tr>
<td>EXP</td>
<td>Same as PJM parameter</td>
</tr>
<tr>
<td>SHORT</td>
<td>Indeterminate</td>
</tr>
<tr>
<td>SUMMER</td>
<td>Positive</td>
</tr>
<tr>
<td>WINTER</td>
<td>Indeterminate</td>
</tr>
</tbody>
</table>
Independent variables for bid size estimator

As mentioned above, it makes no sense for a bidder to bid a quantity that is less than the quantity of all FTRs awarded along a certain path. However, the number of FTRs available is not certain. The quantity is theoretically infinite if enough bidders are willing to take short positions on the corresponding contra-flow FTR. Thus, the bidder needs to estimate the expected size of the market. One question that needs to be asked: is it more dangerous to bid a quantity that is too high or too low? Bidding a high quantity can only raise the clearing price or leave it unaffected, because any part of a bid above the available quantity is cut-off, and the rest of the bid rescaled from a range of [0, 1] times the quantity available. This effectively raises a person’s minimum and average bid prices. Bidding a quantity that is smaller than the quantity cannot raise the clearing price, but instead can decrease it. However, it also reduces the quantity that can be won. Bidding too much can result in negative profit (winning an over-priced FTR), whereas bidding too little can result in not gaining access to a winning FTR. For most bidders, we can assume that the risks from bidding for too great a quantity are far greater than the risks of not bidding enough, and thus in situations of increased uncertainty we can expect a bidder to bid for less, and not more FTRs.

AVGCP: a high clearing price implies that an FTR is taking a larger share of the capacity of constrained lines. Thus, an increase in the past clearing price should lead to an increase in the expected size of the FTR available. We expect the parameter on this variable to be positive.

SDCP: the less certain the clearing price, the less certain the size of the FTR. If the quantity of FTR is smaller than that that the bidder bids on, he is raising the clearing price (his portion of his demand curve beyond the maximum quantity available will be removed.) Thus, with increased uncertainty we would expect to see a decrease in quantity bid.

AVGVAL: while the relationship between the past value and the current size of an FTR is not obvious, we might hypothesize that a higher-valued FTR will attract more bidders.
The presence of more bidders along a certain path raises the clearing price and therefore raises the quantity awarded. Therefore, we would expect the parameter on this variable to be positive.

SDVAL: increasing uncertainty in the past value creates greater uncertainty on the size of the FTR awarded. We would therefore expect the parameter on this variable to be negative.

PJM: this is a indicator variable designed to measure the differences in quality of knowledge held by PJM firms. Once again, the sign is predicated upon the behavior of other firms: if the other firms are systematically bidding for too much FTR, we expect this to be negative. If it is positive, it means that the outside firms tend to bid for too little FTR capacity, so long as we hold the assumption that PJM firms are better informed “insiders.”

EXP: this is the experience indicator as described above. We expect it to have the same sign as the PJM indicator, with the same caveat about confounding with the PJM variable.

SHORT: this is a fixed-effects indicator variable to measure the difference in bids into the long and short markets. We have no prior expectation on this variable.

Past market size: clearly, the size of FTR available is affected by the behavior of other bidders. For this reason, we need some measures of other market activity. For this purpose, two extra variables were employed.

LAG_SHORT: this is the total MW quantity of short (negatively-priced) FTRs awarded in the previous month’s auction.

LAG_LONG:  this is the total MW quantity of long FTRs awarded in the previous month’s auction. It is rare to see more than one firm bidding on the same FTR path, thus
these variables are measures of the competition for FTRs. An increase in the number of long FTRs awarded will decrease the availability along any one path; therefore we expect the parameter on LAG_LONG to be negative. Conversely, more short FTR awards increases the availability of long FTRs. Thus, for a long FTR bid we would expect the parameter on this LAG_SHORT to be positive.

SUMMER: this is a binary indicator variable that assumes the value one for the months of June, July, and August, zero otherwise. It is designed to capture the effects of the grid capacity being reduced by high temperatures in the summer. We expect this variable to take a negative-valued parameter.

WINTER: this is a binary indicator variable that assumes the value one for the months of December, January, and February, zero otherwise. It is designed to capture the effects of the grid capacity being increased by low temperatures in the winter. We expect this variable to take a positive-valued parameter.

A summary of the independent variables used in the estimators for bid size, as well as the expectation of the signs of their parameter, is contained in Table 2.2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Expectation on parameter sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVGCP</td>
<td>Positive</td>
</tr>
<tr>
<td>SDCP</td>
<td>Negative</td>
</tr>
<tr>
<td>AVGVAL</td>
<td>Positive</td>
</tr>
<tr>
<td>SDVAL</td>
<td>Negative</td>
</tr>
<tr>
<td>PJM</td>
<td>Indeterminate</td>
</tr>
<tr>
<td>EXP</td>
<td>Same as PJM parameter</td>
</tr>
<tr>
<td>SHORT</td>
<td>Indeterminate</td>
</tr>
<tr>
<td>LAG_LONG</td>
<td>Negative</td>
</tr>
<tr>
<td>LAG_SHORT</td>
<td>Positive</td>
</tr>
<tr>
<td>SUMMER</td>
<td>Negative</td>
</tr>
<tr>
<td>WINTER</td>
<td>Positive</td>
</tr>
</tbody>
</table>
Independent variables for winning percentage

As can be seen from Equation (76), when calculating the winning percentage we are actually estimating the clearing price of an FTR. The clearing price along an FTR is a function of the bids placed on that path by the bidder in question, as well as the competing bids for all other FTRs in the market. The bidder’s own bid is defined by three parameters as described above: maximum price, minimum price, and quantity. However, it is usually safe to assume that any one bid is a very small fraction of the total number of bids into the monthly FTR market. Therefore, we need to weight the prices by the quantity of the bid. For this reason, confounded variables were formed by multiplying the upper and lower price bounds of a bid by the MW reservation size of the bid.

QPMAX: this is the quantity times the upper price. As it increases, we expect an increase in the clearing price.

QPMIN: this is the quantity times the lower price. As it increases, we expect to see an increase in the clearing price. That is, we expect the parameters on both QPMAX and QPMIN to be positive.

We also need measures of the competition for FTRs. For these, the total sizes of the short and long bid markets were used as proxies.

Q_LONG: this is the total amount of long bids placed in the FTR monthly market. As it increases, we expect the clearing price to increase, as the presence of more competition can only increase, and not decrease the equilibrium price.

Q_SHORT: this is the total amount of short bids placed into the monthly market. An increase in the number of short bids increases the supply of FTRs, thus decreasing the clearing price of a long FTR.

The clearing price was estimated for both fixed- and variable-effects models, but a SHORT indicator was not required, as that is included in the specification of the bid
components. Including it again would lead to double-counting of the significance of this variable.

A summary of the independent variables used in the estimators for clearing price, as well as the expectation of the signs of their parameter, is contained in Table 2.3.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Expectation on parameter sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>QPMAX</td>
<td>Positive</td>
</tr>
<tr>
<td>QMIN</td>
<td>Positive</td>
</tr>
<tr>
<td>Q_LONG</td>
<td>Positive</td>
</tr>
<tr>
<td>Q_SHORT</td>
<td>Negative</td>
</tr>
</tbody>
</table>

There is an alternative method of examining the winning percentage of a bid. In the above methodology, the effects of the PJM and experience variables are masked by the fact that they are inputs into the first-stage estimators, and those first-stage estimates are then used as inputs into the clearing-price estimators. While this model may be more rigorous and defensible from an “economic principles” standpoint, it would also be instructive to have a direct measure of the differences between insider and outsider firms and the differences between new entrants and experienced bidders as related to winning bids. Another variable, not addressed to date, is the complexity of the bid. Some researchers (Scott and Wolf) claim, unambiguously, that a complicated bid broken into many small steps will outperform a simpler bid with fewer parts. This claim can be tested by including a “PARTS” variable into this simplified regression. An increase in the number of parts should increase the winning percentage. However, this only applies in cases where the clearing price of the bid falls between the upper- and lower-price bounds. If the bid is a 100% or 0% winner, the complexity of the bid is irrelevant. A slightly different estimation approach must be taken here, directly examining winning versus bid complexity, without clearing price as an intermediate parameter. For this reason, a Logit estimation technique was employed. The bids were segregated into two bins: those for which the winning percentage was zero, and those for which it was greater than zero. That is, a binary variable WIN was defined, equaling zero or one depending on the status
of the bid. This formulation enables the use of a direct binary choice estimator. The variable WIN was regressed versus the three independent variables PJM, EXP, and PARTS. We would expect all three variables to be positive: being an insider, being experienced and using complicated strategies should all lead to increases in the probability of winning.

**Independent variables for profits**

As can be seen from Equation (82), estimation of the profits requires the estimation of the *ex-post* value of the FTR, which is an estimate of the congestion. As seen in Chapter One of this dissertation, this is not a trivial estimate. Thus, for simplicity, it was decided that the value shall be modeled only by a constant term and an indicator variable for differentiating short and long paths. From Figures 2.12 and 2.13 we see that the profits in both markets are generally positive, therefore we expect the constant term to be positive. The parameter on the SHORT variable will tell us if short paths, on average, have a higher expected payout.
2.8 Data

All of the data employed in this study were downloaded from publicly available files on the PJM website: www.pjm.com.

Monthly FTR auction bid data

Bid data for each monthly auction from June 2000 to April 2003 were downloaded. All off-peak FTR data were discarded. Each of these files contained the following data points for each bid:

- Company Code
- FTR source node
- FTR sink node
- Bid size
- Bid price

A total of 287,396 on-peak bids were made into the FTR markets over the period of study. The data as listed above contained every component of a composite bid as a separate data line. That is, if a company made a four-stage bid, it was recorded as four separate bids with a common path and company code, but with a different price and (possibly) a different size.

Composite bids, where the same firm made more than one bid on the same path in the same month, were combined into single bids defined by the company, the source and sink nodes, and the three bid components described in the previous section: the upper price limit, the lower price limit and the total MW quantity bid upon. Also listed was the number of parts that a composite bid was broken into. The outcome of this was a database of 113,669 composite bids.

Linearity of composite functions

As mentioned in Section 2.6 and highlighted in Figure 2.17, the above aggregation technique converts each composite bid to a linear function defined by two points. If the bid is linear, meaning that each composite part is the same size and the price increments
are equally spaced, we are not losing any information by modeling the bid as linear. But if the parts are different sizes and spaced differently, then we are losing information. The question is, how much information are we losing? This can be answered by looking at the implied slopes of the individual bid components. Of course, each component has a slope of zero, since it at a single price. But an implied slope can be calculated by simply dividing the price difference between consecutive bids by the quantity of the first of these two bids. For an \( n \)-part bid this yields \( n-1 \) implied slopes. We can then look at the variance of the implied slopes. If a bid function is linear, each slope will be identical, and the variance of the family of implied slopes will be zero. The less linearity a function exhibits, the greater the variance of the implied slopes. Due to the dispersed nature of the data (stored in separate files for each month,) performing such an analysis on all data was onerous. Thus, such an analysis was performed on some randomly selected composite bids. The assumption of linearity holds strictly true in about 25\% of the cases. Linearity held over a significant range of the bid function in about 50\% of the bids, with firms having slightly different behavior at the extreme ends of the bid, and in about a quarter of the cases, a complex structure was observed. One firm consistently employing complicated structures was Morgan Stanley. Many of its bids consisted of four- or five-part bids that resembled a logistic function; that is, the quantities were small at the upper and lower bounds, but large in the middle. Most of the complex structures employed by Morgan Stanley were non-linear but symmetrical, meaning a linear estimator would not be biased. El Paso Merchant Energy employed many four-part bids that were uniform in price difference, but increased in bid size as the price decreased, that is, its bid functions were strongly convex to the origin. Most of the active PJM member firms (Exelon, Allegheny Power) as well as some active power traders employed linear bidding structures. Linearization is an abstraction technique that is necessary to be able to perform analysis on this scale, and like all abstraction techniques, some detail is sacrificed. However, results of the analysis described above lend confidence to the notion that linearization in this study has not caused a damaging amount of data loss: the simplified structures employed are accurate (or at least unbiased) proxies in a clear majority of the employed cases. Another point needs to be made here: the structure of a bid is only important when considering bids that were partially successful, when the
winning percentage was some number between, but not including zero and one hundred percent. If a bid won either zero or the entire amount bid, the only important variable is the price at the extreme, and the distribution of the bid components is irrelevant. Of the 112,610 separate bids that were eventually modeled, only 5,468 were “partial” winners. 11,452 bids were 100% successful, and the other 95,690 were total losers. Therefore, the linearization routine affected only about 4.8% of all bids entered.

**Monthly FTR auction results data**
The auction outcomes for each month from June 2000 to April 2003 were downloaded from the PJM website. These data files are spreadsheets containing two sheets of relevant data. On one sheet, all winning bids were listed, containing the following information:

- Company name
- FTR source node
- FTR sink node
- FTR value
- Size of MW reservation won

A second data sheet contained the virtual locational marginal prices at all allowable nodes in the PJM network. This is the primary output of the simultaneous feasibility model, the simulation of what the prices would be given the theoretical flows composed of the winning FTR bids. Since there is no actual energy being priced in this market, the prices must all be relative to some reference node. During the period under study the reference (price = 0) node was Alburtis, a town about 10 miles southwest of Allentown, PA.

The winning bid data were concatenated with the total bid data. Based upon this combination, names of the bidding company could be assigned to the semi-anonymous company codes contained in the bid data. Winning percentages were also calculated.
The LMP data were transformed into a month-by-node table of values. That is, the LMP at a given node for each month in the period of study will be a row of numbers beside the name of the node in question.

It was necessary to calculate the values of the FTRs, both in the present and past months, as these are used as variables in the price estimators. The hourly day-ahead LMPs for every day from June 1, 2000 to April 30, 2003 were downloaded, one data file for each day. These daily files were concatenated into monthly files. Off-peak data were filtered out, and the resulting hourly on-peak DA-LMPs were averaged over the month in question. These monthly average-hourly DA-LMPs were then arranged into a node-by-month table, similar to that for the FTR LMPs. This format is necessary for use of the table lookup functions in Microsoft Excel. Thus, the end result was a matrix of average hourly nodal LMPs arranged by node (in the rows) and months (in the columns.)

From these tables, a table of complete bid data was constructed, containing all the necessary information to perform the estimations. Each line of this table contained the following data:

- Month
- Company name (or alphanumeric PJM company code, if name was unknown)
- Source node
- Sink node
- Maximum price of bid
- Minimum price of bid
- Total MW bid upon
- Total MW awarded
- PJM indicator variable
- Short FTR indicator variable
- Average value of the FTR along that path for the second to fourth past months
- Standard deviation of the FTR value over the second to fourth past months
- Average clearing price of the FTR along that path for the past three months
- Standard deviation of the clearing price over the past three months
• Number of parts in the bid
• The number of previous monthly FTR auctions the firm has participated in
• The MW total of long FTRs bid upon in the month in question
• The MW total of short FTRs bid upon in the month in question
• The MW total of long FTRs awarded in the previous month
• The MW total of short FTRs awarded in the previous month

Summary statistics
Table 2.4 contains summary statistics of the observed data. In each case there were 112,610 observations.

Table 2.4: Summary statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
<th>Min</th>
<th>Median</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of FTR, $/MWh</td>
<td>1.6</td>
<td>8.1</td>
<td>-111.7</td>
<td>0.6</td>
<td>115.7</td>
</tr>
<tr>
<td>Profit</td>
<td>-0.1</td>
<td>2.8</td>
<td>-199.2</td>
<td>0.0</td>
<td>90.1</td>
</tr>
<tr>
<td>Winning percentage</td>
<td>11.7%</td>
<td>30.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PJM indicator variable</td>
<td>0.10</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHORT indicator variable</td>
<td>0.71</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of parts in bid</td>
<td>2.5</td>
<td>4.4</td>
<td>1</td>
<td>1</td>
<td>180</td>
</tr>
<tr>
<td>Average past value of FTR</td>
<td>1.8</td>
<td>13.1</td>
<td>-95.7</td>
<td>1.6</td>
<td>100.1</td>
</tr>
<tr>
<td>SD of past value of FTR</td>
<td>4.8</td>
<td>9.5</td>
<td>0</td>
<td>1.6</td>
<td>60.6</td>
</tr>
<tr>
<td>Experience</td>
<td>5.8</td>
<td>6.5</td>
<td>0</td>
<td>3</td>
<td>34</td>
</tr>
<tr>
<td>Average past clearing price</td>
<td>1.2</td>
<td>6.8</td>
<td>-131.7</td>
<td>1.3</td>
<td>118.0</td>
</tr>
<tr>
<td>SD of past clearing price</td>
<td>2.6</td>
<td>5.2</td>
<td>0</td>
<td>1.0</td>
<td>137.5</td>
</tr>
<tr>
<td>Bid size, MW</td>
<td>30.1</td>
<td>8.1</td>
<td>0.3</td>
<td>7.5</td>
<td>2000</td>
</tr>
<tr>
<td>Short FTR awarded, GW</td>
<td>1.6</td>
<td>0.9</td>
<td>0.0</td>
<td>1.7</td>
<td>2.9</td>
</tr>
<tr>
<td>Long FTR awarded, GW</td>
<td>5.6</td>
<td>2.7</td>
<td>1.3</td>
<td>4.1</td>
<td>10.8</td>
</tr>
<tr>
<td>Total short bids, GW</td>
<td>24.2</td>
<td>14.6</td>
<td>1.5</td>
<td>26.6</td>
<td>52.3</td>
</tr>
<tr>
<td>Total long bids, GW</td>
<td>34.1</td>
<td>11.0</td>
<td>7.5</td>
<td>34.0</td>
<td>56.6</td>
</tr>
</tbody>
</table>

Variable correlation

The Pearson correlation coefficients between the variables employed in each of the three data sets are shown in Tables 2.5, 2.6 and 2.7.
Table 2.5: Variable correlations, bid price regressions

<table>
<thead>
<tr>
<th></th>
<th>SHORT</th>
<th>AVGVAL</th>
<th>SDVAL</th>
<th>AVGCP</th>
<th>SDCP</th>
<th>EXP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PJM</td>
<td>-0.41</td>
<td>0.04</td>
<td>-0.05</td>
<td>0.02</td>
<td>-0.07</td>
<td>0.66</td>
</tr>
<tr>
<td>SHORT</td>
<td>-0.18</td>
<td>-0.03</td>
<td>-0.13</td>
<td>0.12</td>
<td>-0.38</td>
<td></td>
</tr>
<tr>
<td>AVGVAL</td>
<td></td>
<td>-0.15</td>
<td>0.72</td>
<td></td>
<td>-0.41</td>
<td>-0.11</td>
</tr>
<tr>
<td>SDVAL</td>
<td></td>
<td>-0.31</td>
<td>0.53</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVGCP</td>
<td></td>
<td></td>
<td>-0.46</td>
<td>-0.16</td>
<td></td>
<td>0.03</td>
</tr>
<tr>
<td>SDCP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.6: Variable correlations, bid quantity regression  
(note: all other correlations as in Table 2.5)

<table>
<thead>
<tr>
<th></th>
<th>LAG_SHORT</th>
<th>LAG_LONG</th>
</tr>
</thead>
<tbody>
<tr>
<td>PJM</td>
<td>-0.06</td>
<td>-0.03</td>
</tr>
<tr>
<td>SHORT</td>
<td>0.19</td>
<td>0.20</td>
</tr>
<tr>
<td>AVGVAL</td>
<td>0.16</td>
<td>0.28</td>
</tr>
<tr>
<td>SDVAL</td>
<td>-0.02</td>
<td>-0.22</td>
</tr>
<tr>
<td>AVGCP</td>
<td>0.17</td>
<td>0.32</td>
</tr>
<tr>
<td>SDCP</td>
<td>-0.06</td>
<td>-0.17</td>
</tr>
<tr>
<td>EXP</td>
<td>-0.13</td>
<td>-0.18</td>
</tr>
<tr>
<td>LAG_SHORT</td>
<td></td>
<td>0.75</td>
</tr>
</tbody>
</table>

Table 2.7: Variable correlations, clearing price regression

<table>
<thead>
<tr>
<th></th>
<th>QPMIN</th>
<th>Q_SHORT</th>
<th>Q_LONG</th>
</tr>
</thead>
<tbody>
<tr>
<td>QPMAX</td>
<td>0.46</td>
<td>-0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>QMIN</td>
<td></td>
<td>-0.10</td>
<td>0.16</td>
</tr>
<tr>
<td>Q_SHORT</td>
<td></td>
<td></td>
<td>0.23</td>
</tr>
</tbody>
</table>

Some observations can be made from observations of the correlation coefficients. From Table 2.5 we can see that the correlation between PJM and SHORT is negative, which tells us that most of the short bids are being entered by non-PJM firms. This lends credence to the idea that PJM native firms are more likely to be hedgers than speculators.
Also note that the PJM and EXP variables are fairly strongly correlated ($\rho = 0.66$). This tells us that the PJM firms are the ones that stay in the market the longest, and thus tend to be among the more experienced. This tells us that the PJM and EXP variables are confounded to some degree, thus reducing the effectiveness of the EXP variable in measuring learning effects over time. The starkest observation from Table 2.5 is the correlation coefficient between the average value, lagged over months $t-4$ to $t-2$, and the average clearing price, lagged over months $t-3$ to $t-1$. Essentially, this statistic measures the correlation between the clearing price of an FTR and the value of that FTR in the previous month. One of the hypotheses being tested here is that the firm predicates its bids upon, among other things, the value of the FTR in the past few months. Thus, as the value increases, we would expect the firm’s bids to increase, which should lead to an increase in the clearing price. The correlation, highlighted in Table 2.5, is an early validation of this hypothesis. That the two variables are correlated means that the estimator sacrifices some efficiency, but as these two variables define the upper and lower limits of the rational bidding space, both must be included in the price regressions.

A second test of the relationship between past value and clearing price was employed. A data set was developed consisting of 667 nodes from the DA-LMP and FTR virtual LMP tables described above. These 667 nodes comprised the non-redundant set of all nodes that had observed price data for the full period of study. Thus, in this case, we are comparing LMPs calculated from the FTR auction with those observed in actual DA-LMP markets. For each node, the clearing price LMP was regressed on various ages of the value DA-LMP. The minimum lag used was two months, because at the time of the auction, the values for the previous month are not known. That is, the auction for, say, April FTRs take place on March 15th, and at this point the aggregate values of the FTRs for entire month of March are yet to be determined. Thus, the February values are the most recent ones that can be used by bidders to inform their decisions.

The mean-value correlation function between the clearing price and the lag of the FTR values is shown in Figure 2.17. The term “mean-value correlation” refers to the mean value of the correlation of each of the 667 observed nodes.
Figure 2.19: Mean correlation function of clearing price versus lagged FTR value

If we were discussing autocorrelation, Figure 2.17 would describe a classic autoregressive (AR) model: one where the value at time $t$ is a function of the value at time $t-k$. The behavior of autocorrelation functions is described in more detail in Shumway and Stoffer (2000). The AR behavior can be analogized to a cross-correlation: Figure 2.17 tells us that there is a functional relationship between the clearing price at time $t$ and the value of the FTR at time $t-k$. In this study the average value over the past three months has been employed, which is akin to an AR(3) model with equal parameters on each of the three lags. A possible extension of the work here may involve using empirically-determined vector-ARIMA functions to model the time-series behavior of clearing price versus past values of both clearing price and FTR value. This is the approach taken by Cammack, although she had only one set of prices to examine, not many thousands as in this study.

**Experimental procedure**

The OLS regressions were performed using built-in regression functions in MINITAB v.14 statistical software. The maximum likelihood estimation models were performed using custom-built models in MATLAB program language. The Newton Raphson root-solving technique was employed to find the roots of the first derivatives of the maximum
likelihood functions. The first and second derivatives were calculated using a central-differencing numerical estimation technique, as described in Chapter One of this study. In order to converge, the Newton-Raphson technique requires an initial estimate of the parameter vector that is reasonably close to the actual value. An OLS estimation of the clearing prices was used as a first estimate, but this proved unstable in the model. Therefore, it was necessary to use a crude grid-search routine to home in on the true value. In this model, the value of the log-likelihood function (LLF) was calculated at the initial (default) value of the parameters, and then calculated after adding some small increment to one of the parameters. If the incremented estimate of the LLF was greater than the default value, then the incremented value of the parameter in question was taken as the default, and the process repeated. If the positively-incremented value of the LLF was less than the default, then an increment was subtracted from the parameter in question, and the process repeated. When a point was reached where neither the positively- or negatively-incremented LLF was greater than the default, a local maximum was obtained. This process was performed one variable at a time. A large increment, typically 0.5, was first employed, and all of the variables cycled through repeatedly until all exhibited a maximum behavior. The process then began anew with a smaller increment, and was repeated several times until maxima using an increment of 0.0001 were located. At this point, the parameter values were close enough to the true optimum to ensure convergence in the Newton-Raphson routine. The information matrix was modeled as the inverse of the Hessian matrix, and the parameter variances were taken as the square roots of the diagonal terms of the information matrix.
2.9 Results and discussion

Single-component-bid-price regression results

Table 2.8 contains the results of the estimation of the parameters for the single-component bid price, employing both the fixed-effects and variable-effects models. The \( t \)-statistics for each parameter estimate are shown in parentheses below the parameter, and parameters that are statistically significant to a 95% confidence level (\( i.e., |t\text{-stat}| > 1.96 \)) are highlighted in bold-face type. The variable effects columns show the effects as isolated for long and then short FTR bids, respectively. Some statistical qualities of the results are shown at the bottom of the tables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fixed effects parameter</th>
<th>Variable effects parameter (long)</th>
<th>Variable effects parameter (short)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.780 (35.33)</td>
<td>0.464 (13.74)</td>
<td>-0.021 (-0.59)</td>
</tr>
<tr>
<td>PJM</td>
<td>2.986 (23.18)</td>
<td>1.115 (13.05)</td>
<td>4.163 (16.69)</td>
</tr>
<tr>
<td>AVGVAL</td>
<td>-0.053 (-13.85)</td>
<td>0.071 (20.98)</td>
<td>0.005 (1.06)</td>
</tr>
<tr>
<td>SDVAL</td>
<td>-0.026 (-7.71)</td>
<td>0.002 (13.74)</td>
<td>-0.287 (-44.84)</td>
</tr>
<tr>
<td>AVGCP</td>
<td>0.099 (19.35)</td>
<td>0.113 (24.71)</td>
<td>0.106 (16.49)</td>
</tr>
<tr>
<td>SDCP</td>
<td>-0.119 (-14.54)</td>
<td>-0.002 (-0.23)</td>
<td>-0.061 (-6.03)</td>
</tr>
<tr>
<td>EXP</td>
<td>-0.149 (-27.99)</td>
<td>-0.038 (-9.73)</td>
<td>-0.249 (-34.88)</td>
</tr>
<tr>
<td>SUMMER</td>
<td>0.474 (10.39)</td>
<td>0.087 (2.73)</td>
<td>1.017 (17.52)</td>
</tr>
<tr>
<td>WINTER</td>
<td>0.969 (27.69)</td>
<td>0.064 (1.42)</td>
<td>1.302 (32.91)</td>
</tr>
<tr>
<td>SHORT</td>
<td>-2.345 (-53.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>81,109</td>
<td>16,170</td>
<td>64,939</td>
</tr>
<tr>
<td>( R^2 ) statistic</td>
<td>7.1%</td>
<td>18.7%</td>
<td>11.2%</td>
</tr>
</tbody>
</table>
Upper- and lower-price-limit regressions results

Tables 2.9 and 2.10 contain the results of the estimation of the parameters for upper and lower bid price in multiple-component bids, employing both the fixed-effects and variable-effects models. The variable effects columns show the effects as isolated for long and then short FTR bids, respectively. Some statistical qualities of the results are shown at the bottom of the tables.

Table 2.9: Parameter estimates, upper limit of bid price regressions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fixed effects parameter</th>
<th>Variable effects parameter (long)</th>
<th>Variable effects parameter (short)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-4.728 (-18.99)</td>
<td>0.217 (5.62)</td>
<td>-8.319 (-32.13)</td>
</tr>
<tr>
<td>PJM</td>
<td>1.746 (5.89)</td>
<td>0.698 (15.54)</td>
<td>-5.909 (-14.71)</td>
</tr>
<tr>
<td>AVGVAL</td>
<td>0.011 (1.42)</td>
<td>0.112 (23.44)</td>
<td>-0.041 (-25.15)</td>
</tr>
<tr>
<td>SDVAL</td>
<td>0.077 (7.87)</td>
<td>-0.047 (-8.69)</td>
<td>0.028 (3.39)</td>
</tr>
<tr>
<td>AVGCP</td>
<td>0.926 (46.30)</td>
<td>0.133 (23.94)</td>
<td>0.914 (38.11)</td>
</tr>
<tr>
<td>SDCP</td>
<td>-0.323 (-18.19)</td>
<td>-0.040 (-6.29)</td>
<td>0.172 (8.59)</td>
</tr>
<tr>
<td>EXP</td>
<td>0.204 (11.93)</td>
<td>0.033 (12.55)</td>
<td>0.403 (17.35)</td>
</tr>
<tr>
<td>SUMMER</td>
<td>-3.733 (-11.77)</td>
<td>0.388 (6.87)</td>
<td>-0.587 (-1.89)</td>
</tr>
<tr>
<td>WINTER</td>
<td>2.646 (9.54)</td>
<td>0.415 (8.66)</td>
<td>-0.255 (-0.89)</td>
</tr>
<tr>
<td>SHORT</td>
<td>-16.303 (-62.63)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>31,501</td>
<td>16,759</td>
<td>14,742</td>
</tr>
<tr>
<td>R² statistic</td>
<td>50.8%</td>
<td>18.6%</td>
<td>33.9%</td>
</tr>
</tbody>
</table>
Table 2.10: Parameter estimates, lower limit of bid price regressions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fixed effects parameter</th>
<th>Variable effects parameter (long)</th>
<th>Variable effects parameter (short)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.643 (-5.73)</td>
<td>-0.580 (-22.60)</td>
<td>-23.825 (-41.71)</td>
</tr>
<tr>
<td>PJM</td>
<td>-0.594 (-4.45)</td>
<td>0.777 (26.09)</td>
<td>3.439 (3.88)</td>
</tr>
<tr>
<td>AVGVAL</td>
<td>0.008 (2.12)</td>
<td>0.089 (28.16)</td>
<td>-0.137 (-9.01)</td>
</tr>
<tr>
<td>SDVAL</td>
<td>0.042 (9.52)</td>
<td>-0.057 (-16.01)</td>
<td>0.165 (8.97)</td>
</tr>
<tr>
<td>AVGCP</td>
<td>0.463 (51.33)</td>
<td>0.025 (6.78)</td>
<td>2.367 (44.67)</td>
</tr>
<tr>
<td>SDCP</td>
<td>-0.187 (-23.41)</td>
<td>-0.012 (-2.77)</td>
<td>0.907 (20.54)</td>
</tr>
<tr>
<td>EXP</td>
<td>0.114 (14.75)</td>
<td>0.023 (13.26)</td>
<td>0.702 (13.69)</td>
</tr>
<tr>
<td>SUMMER</td>
<td>-0.047 (-0.33)</td>
<td>0.385 (10.28)</td>
<td>-11.138 (-16.27)</td>
</tr>
<tr>
<td>WINTER</td>
<td>0.361 (2.89)</td>
<td>0.589 (18.51)</td>
<td>4.614 (7.32)</td>
</tr>
<tr>
<td>SHORT</td>
<td>-6.222 (-53.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>31,501</td>
<td>16,759</td>
<td>14,742</td>
</tr>
<tr>
<td>R² statistic</td>
<td>49.1%</td>
<td>15.6%</td>
<td>33.4%</td>
</tr>
</tbody>
</table>

From these results, it is clear that the long and short markets are distinctly different. In each fixed-effects model the SHORT variable was large and strongly significant, and the parameters in the long and short models in all three cases showed significant differences. (The two sets of variable-effects parameters were subjected to “difference of two means” tests and the results were significant with t-statistics of 500 to 3,000.) This is not surprising: the two different markets represent two fundamentally different belief sets about the expected movement of the market. In each case, the fixed effects model was a mixture of two different behavior modes that masked the true effects of either market. For this reason, it is not meaningful to examine the fixed-effects parameters any further, instead concentrating on the variable-effects models when analyzing the results.
The count of single-component bids is more than double that of the multi-part bids, but this is because of the large number of single-component short bids placed by one company. The numbers of single- and multi-part long bids are approximately equal. Considering first the long market: the response to changing past values and past clearing prices is consistent with the theory: as the prices and values increase, firms increase their bids. These parameters were positive and significant in six of six cases. Oddly, the responses were more significant for the upper price limit than the lower – it was expected that the lower price would be more strongly affected by changes in the clearing price. In most of the cases, the response to the variance of the prices and values was as expected: as uncertainty increased, prices decreased. The SD parameters were negative in five of the six cases and not significant in the instance where the parameter was positive.

Firms consistently bid higher in both the summer and winter in the long market. The parameters on these variables were positive and statistically significant in six of six cases.

In all three cases, PJM firms made significantly higher bids. For the multi-part bids, the experience variable was consistent with this: both upper and lower bounds on multi-part bids increased with experience. However, the opposite was true of single component bids: the parameter on the EXP variable was negative. This finding might imply some difference in sophistication between firms using single- and multi-part bids: Scott and Wolf state that multi-part bids strongly dominate simple bids, and it might be assumed here that PJM firms have better information than non-PJM firms, thus, as the sophisticated non-PJM firms gain experience, they gravitate towards the behavior of PJM firms, whereas non-sophisticated firms, using dominated single bids, do not learn as well. There is, of course, possible confounding in that PJM firms may be more frequent suppliers of multi-part bids. Two PJM firms, Exelon and Allegheny Energy, are by far the most frequent suppliers of bids with many (>10) parts. On the other hand, PJM firms such as PP&L and PSEG frequently employ single-part bids.

The short market does not behave as tidily as the long market. This is partly due to the fact that certain firms (El Paso, FPL) entered the market for short periods of time with
large quantities of bids employing novel strategies. This causes the estimates over an entire sample to be less well-behaved, given the strong between-firm heterogeneity. Short single-component bids provide well over half of all the bids studied in this essay (about 64,000 of 112,000 in total) and almost all of these come from three firms. Multi-part short bids number about 15,000 and come from a distinctly different set of firms. This helps explain some of the observed differences in the multi- and single-part bids: they are provided by two separate populations, firm-wise.

The single-part results are generally in agreement with expectations, except that the PJM and EXP variables have different signs. However, only a very small number of bids come from PJM firms. The responses in the multi-part bid regressions are less consistent. Firms do not respond to past FTR values, clearing prices and variances in the way we expect, often increasing the price bid in response to increased uncertainty. The summer and winter responses are also either non-significant or non-intuitive. The reasons for these results are not obviously explained.

**Bid quantity regression results**

Table 2.11 contains the results of the estimation of the bid-quantity parameters, employing both the fixed-effects and variable-effects models. The same description for \( t \)-stats and significance as applied to Tables 2.8 to 2.10 applies here.
Once again, the value of the SHORT indicator variable is significant, and the differences between the parameter estimates for the long and short variable-effects models are all very significantly different to zero. This tells us that the variable effects models yield a more accurate reflection of two very different behavior regimes. The greatest difference here is in the constant term: average bids in the long market are far bigger than those in the short market, consistent again with the hedger-versus-speculator model. The next most interesting result is the response to the size of the market in past months: the long market behaves counter-intuitively, responding to an increase in short FTRs awarded by decreasing bid size. The short market is also reversed: we would expect the short market
parameter to be negative and the long market parameter to be positive. This is difficult to explain. One possible cause is that the firm population does not remain constant, but changes significantly from month-to-month. It is possible that within a firm the bid sizes do not vary greatly from month to month, but the entry and exit of different firms with different bid-size strategies is the root cause of the variability from month to month. This hypothesis could be tested by either doing firm-level regressions or a fixed-effects panel model with indicator variables for each firm.

Responses to the SUMMER and WINTER variables are consistent with the physics of power-flows: in all cases the firms make larger bids in the winter and smaller ones in the summer, responding to the increased amount of capacity available in the winter and vice-versa in the summer.

Clearing-price regression results

Table 2.12 contains the results of the estimation of the clearing-price parameters, employing both the fixed-effects and variable-effects models.

Table 2.12: Parameter estimates, clearing price regressions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fixed effects parameter</th>
<th>Variable effects parameter (long)</th>
<th>Variable effects parameter (short)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>9.776 (36.44)</td>
<td>20.453 (36.23)</td>
<td>17.128 (20.12)</td>
</tr>
<tr>
<td>QMAX</td>
<td>-6.747 (-17.25)</td>
<td>3.445 (6.02)</td>
<td>-5.872 (-4.69)</td>
</tr>
<tr>
<td>QMIN</td>
<td>6.595 (37.07)</td>
<td>2.323 (39.25)</td>
<td>5.504 (9.55)</td>
</tr>
<tr>
<td>Q_LONG</td>
<td>0.154 (30.94)</td>
<td>-0.103 (-10.56)</td>
<td>0.890 (37.87)</td>
</tr>
<tr>
<td>Q_SHORT</td>
<td>0.039 (6.70)</td>
<td>-0.018 (-1.39)</td>
<td>0.030 (1.52)</td>
</tr>
<tr>
<td>S.D. of error term</td>
<td>13.428 (82.60)</td>
<td>18.337 (66.69)</td>
<td>33.740 (46.69)</td>
</tr>
</tbody>
</table>

As mentioned in Section 2.6, the clearing price is determined by all bids into the FTR market, and it is estimated here by the bid made by the bidder in question on the path in
question and the sum total of all other bids. The QP MAX and QP MIN variables are size-weighted upper and lower prices on the bidder’s own bid, estimated from the first-stage OLS regressions detailed in Tables 2.8 - 2.11. Once again, the fixed effects model shows the mix of two very different regimes, so it is again relevant to focus on the variable-effects results. In the long market, the clearing price increases with increases in both the upper and lower quantity-weighted price variables. However, the responses to the market sizes were either counterintuitive or insignificant. As the rest of the long market gets larger, the clearing price decreases. This defies logic: more bidders should mean higher clearing prices. All else staying equal, the entrance of new bidders should drive up prices if the existing bidders retain their bidding strategies. If established bidders lower their prices over time, while the market grows, then this result may be logical, but there is evidence that PJM firms raise their bids over time, and less experienced firms lower theirs. However, PJM bidders do not supply enough bids to purchase the full quantity of FTRs, thus they are less likely to be the marginal bidder (since it has been shown that PJM firms tend to make higher bids), and it is the marginal bidder who establishes the clearing price. Therefore, the response seen in Table 2.12 is in fact consistent with previous observed behavior: new bidders entering the market bid lower, and those bidders are the marginal, price-setting bidders.

In both markets, the response to the lower bid limit is stronger than the response to the upper bid limit, as expected, and in both markets the clearing price clearly increases with an increase in the quantity-weighted lower-price limit. In the short example, the QP MAX variable is acting as a sort of proxy constant: we expect the clearing price of a short-bid path to be negative, which means we would expect a negative constant term. In our case, the constant term is positive, but the QP MAX variable provides the negative weighting on the estimate. It appears that the QP MIN variable is establishing the response in clearing price to the bid, whereas the constant term and QP MAX term are defining the average level of the clearing price.
FTR value regression results

The value of the FTR was defined as part of the profit model. It was decided to model this as a constant, with distinction between long and short markets. Since there are no variable effects, using a fixed-effects model with a SHRT indicator variable does not result in the loss of any information. The results of this regression are shown in Table 2.13, below.

Table 2.13: Parameter estimates, value of FTR regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter estimate</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>11.533</td>
<td>95.00</td>
</tr>
<tr>
<td>SHORT</td>
<td>-1.607</td>
<td>9.77</td>
</tr>
<tr>
<td>S.D. of error term</td>
<td>10.653</td>
<td>183.97</td>
</tr>
</tbody>
</table>

These numbers seem strange when compared to the actual data. For example, the non-weighted mean value of all winning FTRs was -0.522. So how can we get a mean estimate of about 10.3 (the average of the constant minus SHORT)? Remember that this is the third stage of a three-stage estimator: first, we are estimating price limits and bid sizes. We are then using those estimates in an estimate for clearing prices. Those second stage estimates for clearing prices are then used as inputs into the model for FTR value. Hence, we have three layers of error, and as shown above, a standard deviation of the error term that is greater than the mean value of the calculated variable. This is a warning against using complicated, multi-stage estimation techniques. If we decide to do a little cheating and use the observed values of clearing price in the FTR value model, what are the results? A constant term of 1.66, a SHORT parameter value of -4.43 and an error S.D. of 7.74. These numbers are wildly different from the three-stage estimator output.

Winning percentage estimations

Winning percentage was estimated directly from three firm-specific variables: PJM status, firm experience and number of parts to a bid. Fixed and variable effects models were calculated. The results are shown in Table 2.14, below.
Table 2.14: Parameter estimates, winning percentage regressions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fixed effects parameter</th>
<th>Variable effects parameter (long)</th>
<th>Variable effects parameter (short)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2.421 (-174.47)</td>
<td>-1.985 (-79.06)</td>
<td>-2.526 (-148.78)</td>
</tr>
<tr>
<td>PJM</td>
<td>-0.216 (-6.45)</td>
<td>0.070 (1.78)</td>
<td>-0.855 (-11.64)</td>
</tr>
<tr>
<td>EXP</td>
<td>0.060 (40.35)</td>
<td>0.045 (23.24)</td>
<td>0.088 (36.04)</td>
</tr>
<tr>
<td>PARTS</td>
<td>0.064 (32.48)</td>
<td>0.065 (25.29)</td>
<td>0.054 (18.22)</td>
</tr>
<tr>
<td>SHORT</td>
<td>0.064 (-17.39)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.14 helps us answer several questions. Firstly, are PJM firms better bidders? The results indicate that in the long market, PJM firms are neither better nor worse than the average firm, but that in the short market, they do worse than other firms. However, it was mentioned above that PJM firms only make 1.8% of the bids in the short market, so this result can be considered not meaningful. So, the hypothesis that PJM firms are better informed than other firms does not appear to be borne out by the data. PJM firms consistently bid higher than other firms, but do not significantly win more FTRs. This seems like a counter-intuitive result – how can firms that bid higher on average win less? The most obvious answer to this would be that PJM firms are bidding on a generally different subset of FTRs than other firms.

The second result is that firms do appear to get better at bidding with experience. This is not surprising – we expect learning effects to have positive results. The last result is that more complicated bid structures do better on average than less complicated ones. This validates the theoretical findings of Scott and Wolf.
2.10 Conclusion

Many of the questions posited in the introduction to this essay can now be answered. Do firms adjust to information from past auctions? Yes: the levels of bid prices generally move in concert with recent auction clearing prices and recent realized values of FTR ownership. This is not surprising given that we have a somewhat competitive market (HHI of about 1,500 in recent months) populated by profit-maximizing firms adopting both hedger and speculator roles.

The second question posed was whether firms consistently bid below the value of the good or not. The theory-based literature (primarily Wilson) claims that in a uniform-price divisible goods auction, firms will employ bid-shading as a response to uncertainty about the value of an auctioned good. This result is empirically validated in this work. There are two ways of examining this question: firstly: how do firms bid with respect to previous realized data? It has been shown that bids are consistently below past values, and that bid shading increases with increased uncertainty in past values. Secondly, how do bids correlate with the \textit{ex-post} realized value of the good being bid upon? The result is less clear: in some months, firms overbid and in others they underbid; in some months, firms lose money from FTR ownership and in others they profit from such ownership. This is a result of the difficulties in short-term forecasting of congestion amounts. As shown in Chapter One of this essay, predicting congestion, even in the near-term, can be very difficult. However, viewed over time in its entirety, analysis of the data set tells us that in the long run firms profit from owning FTRs. On a month-over-month basis they are very volatile, but have a net positive value. Indeed, over the period of study here PJM paid out $48 million more than they took in from monthly FTR auctions. This existence of persistent profits over time is, at first glance, inconsistent with efficient market theories dictating that over time markets should move towards zero economic profit status. Some have cited these profits as evidence of collusion, even coining the term “collusive-seeming equilibria.” However, the results of this study corroborate the theoretical statements made in the past about bid-shading being an efficient equilibrium strategy. Bid-shading in a divisible goods auction arises out of uncertainty about the
future value of the auctioned good, and the persistent profits that are exhibited from bid-shading are perhaps the “risk-premium” that firms earn for participating in such an uncertain (high-risk) marketplace. Stated otherwise, these profits are consistent with efficient market hypotheses when they are viewed as part of the risk premium that firms must receive in order to enter these markets.

The question of bidder symmetry was frequently addressed in the literature. The general consensus was that bidder asymmetry complicates analysis, but definitely leads to significantly greater profits for better-informed firms. The question of complexity of analysis has not been addressed by this study, but the presence of asymmetry was tested by looking at profitability. Some firms are native to these markets, have operated in them for far longer, and presumably have better historical data than outside firms. On the other hand, as entrenched firms with non-competitive utility arms, PJM firms may be more prone to complacency and less likely to employ sophisticated bid strategies. The results were inconclusive. While PJM firms consistently bid higher than non-PJM firms, their profits were not significantly higher. Thus, there was asymmetry in the inputs, but not in the results. Furthermore, if the PJM firms are better informed we would expect other firms to converge towards PJM firms in their bidding strategies as those firms gained more experience in the market. Once again, this was not shown to be the case. PJM behavior and other firm behavior did not appear to be convergent over time. In concert with this question, the idea of simple and sophisticated bid strategies was examined. Some authors have made theory-based (but not empirically-validated) claims that multi-part bids strongly dominate simple bids. Results from this study show that the winning percentage increases with bid complexity in a statistically significant fashion, thus providing some rare empirical validation for the claims of Scott and Wolf. However, this finding seemed to have little effect on the behavior of firms: we would expect to see bid complexity increasing over time, with firm behavior converging to the efficient optimum. But individual firm behavior did not appear to exhibit this behavior. This finding flies in the face of efficient market theories: if there is a practice that can be shown to be unambiguously better, we expect all firms to move towards this behavior, regardless of their starting point. This was not observed. The simplest explanation that comes readily
to mind is that of institutional inertia: when a bidding strategy becomes entrenched as “standard practice” within a firm, it is not prone to spontaneous change.

The single most unprecedented finding of this study is the large divergence of behavior in the hedger and speculator market, or more clearly, the long-seller and short-seller markets. No work in the literature was found addressing this phenomenon in the context of divisible-goods auctions or FTR markets. Simply put, bidders in the long and short markets exhibit vastly different behavior. In the long market, behavior is more consistent with expectations of firm behavior, whereas in the short market a wider variety of unusual bidding strategies was observed from firms that entered the market for short periods of time before departing after having mixed success with their strategies.
References


*Online, [http://www.pjm.com/contributions/pjm-manuals/pdf/m06v08.pdf](http://www.pjm.com/contributions/pjm-manuals/pdf/m06v08.pdf)*


Appendix B: Firm-level profitability rankings

Table B.1 contains information concerning the profitability rankings of the firms that participated in the PJM monthly FTR auction market between June 2000 and April 2003. The column headed “Rank Index” describes the firm’s average monthly rank coefficient, a modified ranking in terms of profitability in the monthly FTR auctions. This number was derived as follows:

1. The firm’s total profit from owning the FTRs it won in a given month was divided by the MW total of all FTRs owned. This yielded an average monthly profit measured in S/MWh.

2. For each month, all of the firms were ranked in descending order of average profit, from 1 to $n$, $n$ being the number of firms participating in the month in question.

3. The ranks were normalized from a range of $[1, n]$ to a range of $[0, 1]$ by subtracting 1 from each rank and dividing by $(n-1)$. This yields a number called the monthly rank coefficient.

4. A firm’s average monthly rank coefficient was calculated by taking the arithmetic mean of all of that firm’s monthly rank coefficients.

Of course, a lower rank index means that a firm typically makes a higher profit than the average firm. A firm that always tops the monthly market will have an index of 0. The average monthly index is 0.5.

The column headed “SD of Rank Index” is the standard deviation of the firm’s monthly rank coefficients. A larger number means that the firm moves about in the monthly rankings more than a firm with a low number. Since at least two monthly observations are required to calculate a standard deviation, only firms that have participated in at least two monthly auctions are listed in Table B.1.
### Table B.1: Firm-wise profitability rank index

<table>
<thead>
<tr>
<th>Firm</th>
<th>Rank Index</th>
<th>SD of Rank Index</th>
<th>Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coral Power</td>
<td>0.045</td>
<td>0.009</td>
<td>3</td>
</tr>
<tr>
<td>Energy America</td>
<td>0.181</td>
<td>0.148</td>
<td>2</td>
</tr>
<tr>
<td>Split Rock Energy</td>
<td>0.208</td>
<td>0.139</td>
<td>4</td>
</tr>
<tr>
<td>Outback Power Marketing</td>
<td>0.226</td>
<td>0.181</td>
<td>8</td>
</tr>
<tr>
<td>DC Energy</td>
<td>0.254</td>
<td>0.063</td>
<td>4</td>
</tr>
<tr>
<td>DTE</td>
<td>0.283</td>
<td>0.328</td>
<td>4</td>
</tr>
<tr>
<td>Constellation PowerSource</td>
<td>0.398</td>
<td>0.205</td>
<td>25</td>
</tr>
<tr>
<td>Edison Mission</td>
<td>0.406</td>
<td>0.275</td>
<td>13</td>
</tr>
<tr>
<td>PP&amp;L EnergyPlus</td>
<td>0.442</td>
<td>0.295</td>
<td>35</td>
</tr>
<tr>
<td>TXU Energy Trading</td>
<td>0.459</td>
<td>0.281</td>
<td>13</td>
</tr>
<tr>
<td>UGI Utilities</td>
<td>0.464</td>
<td>0.347</td>
<td>8</td>
</tr>
<tr>
<td>Morgan Stanley Capital Group</td>
<td>0.477</td>
<td>0.170</td>
<td>35</td>
</tr>
<tr>
<td>Conectiv Energy Supply</td>
<td>0.479</td>
<td>0.321</td>
<td>11</td>
</tr>
<tr>
<td>H.Q. Energy Services</td>
<td>0.500</td>
<td>0.393</td>
<td>2</td>
</tr>
<tr>
<td>Amoco Energy Trading</td>
<td>0.501</td>
<td>0.358</td>
<td>9</td>
</tr>
<tr>
<td>Exelon/Peco/PowerTeam</td>
<td>0.510</td>
<td>0.280</td>
<td>34</td>
</tr>
<tr>
<td>Aquila Power</td>
<td>0.517</td>
<td>0.389</td>
<td>5</td>
</tr>
<tr>
<td>Williams Energy</td>
<td>0.533</td>
<td>0.213</td>
<td>20</td>
</tr>
<tr>
<td>Citadel Energy Products</td>
<td>0.541</td>
<td>0.395</td>
<td>6</td>
</tr>
<tr>
<td>FPL Energy Power Marketing</td>
<td>0.560</td>
<td>0.320</td>
<td>8</td>
</tr>
<tr>
<td>Dominion Energy Marketing</td>
<td>0.566</td>
<td>0.327</td>
<td>4</td>
</tr>
<tr>
<td>New Power Co.</td>
<td>0.580</td>
<td>0.466</td>
<td>2</td>
</tr>
<tr>
<td>PSEG</td>
<td>0.581</td>
<td>0.224</td>
<td>29</td>
</tr>
<tr>
<td>Enron Power Marketing</td>
<td>0.583</td>
<td>0.316</td>
<td>10</td>
</tr>
<tr>
<td>PG&amp;E Energy Trading</td>
<td>0.591</td>
<td>0.238</td>
<td>5</td>
</tr>
<tr>
<td>Delmarva Power &amp; Light</td>
<td>0.603</td>
<td>0.283</td>
<td>7</td>
</tr>
<tr>
<td>New Energy Inc.</td>
<td>0.608</td>
<td>0.321</td>
<td>11</td>
</tr>
<tr>
<td>Allegheny Energy</td>
<td>0.632</td>
<td>0.204</td>
<td>22</td>
</tr>
<tr>
<td>Quark Power</td>
<td>0.633</td>
<td>0.176</td>
<td>3</td>
</tr>
<tr>
<td>El Paso Merchant Energy</td>
<td>0.633</td>
<td>0.325</td>
<td>18</td>
</tr>
<tr>
<td>NRG Power Marketing</td>
<td>0.633</td>
<td>0.241</td>
<td>8</td>
</tr>
<tr>
<td>Merrill Lynch Capital Services</td>
<td>0.700</td>
<td>0.284</td>
<td>6</td>
</tr>
<tr>
<td>FirstEnergy Trading Services</td>
<td>0.700</td>
<td>0.261</td>
<td>21</td>
</tr>
<tr>
<td>Niagara Mohawk Energy Mktg</td>
<td>0.708</td>
<td>0.295</td>
<td>2</td>
</tr>
<tr>
<td>Southern Co.</td>
<td>0.711</td>
<td>0.273</td>
<td>10</td>
</tr>
<tr>
<td>Enron Energy Services</td>
<td>0.713</td>
<td>0.322</td>
<td>4</td>
</tr>
<tr>
<td>PEPCO</td>
<td>0.766</td>
<td>0.280</td>
<td>3</td>
</tr>
<tr>
<td>Cargill-Alliant</td>
<td>0.787</td>
<td>0.291</td>
<td>5</td>
</tr>
<tr>
<td>Old Dominion Electric Co-op</td>
<td>0.793</td>
<td>0.278</td>
<td>16</td>
</tr>
<tr>
<td>Mack Services Group</td>
<td>0.833</td>
<td>0.079</td>
<td>2</td>
</tr>
</tbody>
</table>
VITA

BARRY POSNER

Born: Lowestoft, Suffolk, UK, 1964

Education

High School Diploma, Eckville (AB) Jr/Sr High School, 1983

Diploma in Engineering Drafting Technology, Saskatchewan Technical Institute, 1985

Bachelor of Science in Chemical Engineering, University of Alberta, 1994

Master of Science in Mining Engineering, University of Alberta, 1998


Professional

Syncrude Canada Ltd., Edmonton and Fort McMurray, AB, 1996-99
Rescan Environmental Services, Vancouver, BC, 1999-2000
The Petroleum Institute, Abu Dhabi, UAE, 2004 - present