Gains from Trade Under Uncertainty: The Case of Electric Power Markets*

Hendrik Bessembinder
Eccles School of Business
University of Utah
E-mail: finhb@business.utah.edu

and

Michael L. Lemmon
Eccles School of Business
University of Utah
E-mail: finmll@business.utah.edu

Current Draft: June 2004

Comments Welcome

* This paper is derived from an earlier manuscript titled "Pricing, Risk Sharing, and Profitability in the Deregulated Electric Power Industry". The authors thank Paul Joskow, Catherine Wolfram, Douglas Cochran, John Dalle Molle, and an anonymous referee for comments on earlier drafts of this paper, and Jean Gray of the Western Area Power Administration and Steve Norris of Arizona Public Service for valuable discussions regarding the evolution of the wholesale power markets. Thanks are due also to seminar participants at Arizona State University, Emory University, Rice University, the University of Texas at Austin, Washington University, and Instituto Technologico Autonomo De Mexico.

Gains from Trade Under Uncertainty: The Case of Electric Power Markets

ABSTRACT

The rapid growth in energy trading and movement towards deregulation of electricity markets have come to a halt in the wake of assertions that western U.S. energy markets were manipulated. This paper refocuses attention on the potential efficiency gains from competitive wholesale power trading, showing that for any given level of average demand, retail electricity prices will be lower if electricity is traded in competitive wholesale markets than if electricity is delivered by integrated producer-retailers. Wholesale power trading allows for the diversification of demand risk, and the greatest efficiency gains accrue when power demand is least correlated across markets and when there is substantial geographic variation in expected demand. Simulation evidence indicates that real time power trading could reduce retail prices by conservative estimates of 3 to 4% on average in the U.S., and that the combination of forward and real time trading could reduce prices by 6 to 10% or more. This analysis indicates that economic efficiency would be best served by policy aimed at ensuring that power markets are indeed competitive, and that sufficient transmission capacity exists for profitable power trades to be completed.

I. Introduction

While power companies have bought and sold power using bilateral forward contracts for some time, recent years have seen the rapid development of active short-term markets for next-day and next-hour delivery. The Energy Policy Act of 1992 was the catalyst for the development of wholesale power trading in the United States. Wholesale energy trading increased throughout the 1990's, and the U. S. Department of Energy reports that U.S. wholesale power purchases by utilities reached 2,976 billion kilowatt hours (or about $89 billion) in 2001, a 32-percent increase over 2000 purchases.[1] Simultaneously, most U.S. states were in the process of deregulating power markets.

The general trend towards increased power trading and deregulation of power markets was altered markedly by recent events in western U.S. markets, particularly in California between the summer of 2000 and summer of 2001. Spot prices in California's wholesale markets rose markedly, to levels that economic analyses (e.g. Borenstein, Bushnell, and Wolak, 2002 and Joskow and Kahn, 2002) indicate exceeded marginal production costs by a significant margin. Power retailing firms, including Southern California Edison and Pacific Gas and Electric, who were obligated to deliver power at fixed retail prices while discouraged from using forward contracts to lock in power acquisition costs, defaulted on financial obligations and face potential bankruptcy. Still-regulated retail power prices in California have been adjusted upward to partially reflect higher wholesale prices. The Federal Energy Regulatory Commission (FERC) investigated events in the western energy markets, reporting that it found "evidence of manipulation of both electricity and natural gas markets".[2]

These events have led many observers (see, for example, Krugman, 2001 and Ciccone, 2001) to question the efficiency of electricity market deregulation entirely. The U.S. Department of Energy reports that electricity market restructuring has either been delayed or suspended in 8 states that had

---

[1] This data is contained in the "Electric Power Annual 2001", a summary of which can be downloaded from http://www.eia.doe.gov/cneaf/electricity/epa/epa_sum.html.

2 See "Final Report on Price Manipulation in Western Markets", which can be downloaded from http://www.ferc.gov/electric/bulkpower/pa02-2/pa02-2.htm. The first author was employed by FERC as a consultant in this investigation.

previously begun to implement deregulation, and that eighteen other states that had previously been studying deregulation now report no ongoing efforts towards deregulation.[3] In California, a bill (State Senate Bill 888) was introduced in February 2003 that proposes to phase out competition and to allow state regulators to again set profit margins and prices for power industry participants. Coleman (2003) quotes the bill's author, with respect to the state's experiment in electricity market deregulation, as saying: "We're not mending it, we're ending it".

The purpose of this paper is to refocus attention on the potential efficiency gains that can be obtained if electricity is traded in a competitive market place. More generally, we seek to quantify the economic efficiency gains that arise from intra-industry trade under uncertainty. We show that concepts developed in the theory of investment portfolios can be applied to better understand these efficiency gains. The analysis presented here affirms that sustainable retail electricity prices will indeed be lower if electricity is traded in competitive wholesale markets than if electricity is delivered by integrated producer-retailers who are allowed by regulators to recover their average costs, and provides insights regarding cross-sectional variation in the efficiency gain from power trading. The key implications of the analysis hold for any given level of expected power demand.

Of course, either competitive *or* regulated retail prices will increase when demand rises, if production costs increase with output. Ciccone (2001) reports that overall U.S. power demand has grown by two to three percent per year in the preceding decade, while California experienced even higher demand growth of four to six percent per year. In the meantime, almost no new California generating capacity was added. These demand increases without capacity enhancements would have led to retail price increases even if power were still delivered by integrated producer-retailer firms subject to government regulation. The analysis presented here shows that competitive power trading mitigates the price rise. More broadly, to assess whether consumer welfare has been improved by deregulation and the

3 This data is also contained in the "Electric Power Annual 2001", as described in footnote 1. Deregulation plans differed somewhat across states. The most common scenarios involving the de-coupling of power production from

development of power markets, it is crucial to separate the effects of increasing power demand relative to system capacity from that of competition and real time power trading.

This paper shows that wholesale power trade improves efficiency because it allows for the diversification of demand risk across producers. In the absence of trading opportunities the risk of local demand fluctuations must be borne by local producers. With wholesale trade, demand risk is aggregated, and the volatility of system-wide rather than of the various local demands becomes relevant. While all producers gain from wholesale trading, the largest gains from trade accrue to producers whose local demand is least correlated with system-wide demand, ceteris paribus. If variation in local power demand is not highly correlated with system demand, i.e. is diversifiable, the local producer is likely to be able to sell into the wholesale markets at times when aggregate demand and prices are high.

A portion of the efficiency gain from wholesale power trading is attributable to forecastable (e.g. seasonal) variation in demand across areas. Since this demand variation is predictable, the accompanying efficiency gains can be captured using bilateral forward contracts.[4] An additional efficiency gain occurs from trading in response to surprise demand shocks. To capture these efficiency gains requires reactive trading, approximated in power markets by trading in day-ahead and hour-ahead contracts, and by "spot" markets in general. To capture the efficiency gains from either forward or real time trades obviously requires an interconnected power grid with sufficient transmission capacity.

We consider first a market structure in which several risk-neutral producers supply electricity, each in a designated area, meeting local demand from native production capacity. Retail prices are set in advance by regulators to provide a fair expected return on invested capital. We then consider a market structure where each producer maintains an exclusive retail franchise, but can trade in a competitive wholesale spot market for power. To assess the degree of efficiency improvement attributable to the ability to trade power in wholesale markets we examine the fixed retail price that yields zero expected

---

retailing, and the opening of both power generation and retailing to competition.

economic profits under this market structure, and compare it to the retail price required under the first structure.

To provide some indication of the magnitude of the potential efficiency gains we conduct a simulation analysis that relies on actual data on power demand and production capacity from the California, PJM, and New York power markets, which in aggregate serve approximately 17% of total U.S. electricity demand, and on bootstrapped data intended to proxy for the remaining U.S. power markets. The results indicate price reductions in specific areas ranging from near zero to over twenty percent. The simulations illustrate that the efficiency gains attributable to trading power in competitive wholesale markets will be greatest (i) when systemwide demand increases relative to production capacity, (ii) when demand is less correlated across areas, (iii) when there is more cross-sectional variation in mean demand, and (iv) when there is less base load capacity (implying greater reliance on gas and oil-fired peaking plants). Calibrating the results to actual data, we obtain conservative estimates indicating that real time power trading could reduce retail prices by 3 to 4% on average in the U.S., and that the combination of forward and real time trading could reduce nation-wide prices by 6 to 10% or more. Of course these gains would need to be balanced against both the direct and indirect costs associated with transmitting power across regions and interconnecting the power grid.

In considering relations between power trading, demand risk, and pricing issues, we focus on a different set of issues than others who have recently studied the electricity markets. Several researchers, including Wolfram (1999), Borenstein and Bushnell (1999), Green and McDaniel (1998), Green (1998), Green and Newbery (1992), Newbery (1995), and Joskow and Schmalensee (1983) focus on the possibility that power markets may not be fully competitive.[5] Numerous assertions that power producers

4 Bessembinder and Lemmon (2002) examine the equilibrium pricing of electricity forward contracts, and also consider optimal forward positions for power producing and power marketing companies.

5 Another group of authors, including Bailey (1998), Harvey, Hogan, and Pope (1996), Hogan (1993), and Tabors (1996) has considered issues related to pricing of power transmission capacity. We also abstract from these interesting and complex questions, by assuming that power transmission is free. Bailey (1998) considers the power grid in the western United States, and concludes that prices at various geographic locations are effectively

exercise a degree of market power have arisen,[6] and careful empirical analyses support the conclusion that wholesale prices in California have exceeded marginal production costs (see Borenstein, Bushnell, and Wolak (2002) and Joskow and Kahn (2002)). In contrast, we assume in our analysis that wholesale power markets are competitive. It is not our intent to argue that the actual power markets have been competitive. Rather, our goal is to refocus attention on the possible efficiency gains that competitive markets could engender. If actual power markets are not effectively competitive then the efficiency gains described here need not be realized. The solution, in this case, is to take steps to improve the competitiveness of the markets, as argued by Harvey and Hogan (2001). Our analysis can be viewed as a reminder that competitive trading in electricity has potential to improve efficiency, as providing insights as to cross-sectional variation in the efficiency improvement, and as quantifying the efficiency gains that might be attained if mechanisms are developed to allow markets to reach competitive equilibrium.

This paper is organized as follows. Section II introduces the model. Section III analytically evaluates the efficiency gains obtained from wholesale trading. Section IV reports simulation-based evidence that helps to quantity the potential efficiency gains, while Section V concludes.

II. Overview of the Model

There are N power producers, each of which initially has an exclusive franchise on power demand in its region. The demand for producer i in a given period is exogenous, and is denoted by the random variable Q_i. The exogenous stochastic demand of power producers is assumed to follow a joint probability distribution function, denoted by $H(Q_1, Q_2, ..., Q_N)$. We assume that the retail price per unit is fixed in advance to ensure expected economic profits of zero. Retail customers consume as much power as they desire at that price. The retail price, denoted P_i, can vary across producers.

---

determined within an integrated market on about 80% of all days. However, on 20% of days prices display evidence of geographic segmentation, indicative of transmission congestion.

6 See, for example, "Five Power Generators Sued in California", New York Times, May 3, 2001 and "Power Politics: A Failed Energy Plan Catches Up to New York", New York Times, June 1, 2001.

Fixed price retail contracts are inefficient in that they fail to provide an incentive to limit power consumption when supply is scarce relative to demand. Of course, any efficiency gains from contracts that allow retail prices to vary in real time would have to be balanced against the likelihood that retail customers do not have the appropriate comparative advantage in risk bearing. In any case, fixed retail prices are commonly observed, particularly for residential customers.[7]

We assume that the cost function of each producer is an increasing convex function of the amount of electricity produced, denoted $F\alpha_i + g(Q_i/\alpha_i)$, where α_i is a variable that reflects the production capacity scale at firm i, and where $g' > 0$ and $g'' > 0$. Note that fixed costs $F_i = F\alpha_i$, are increasing while variable costs, $g(Q_i/\alpha_i)$, are decreasing as a function of plant scale. The intent is to capture the most important features of the power supply function with the simplest possible approach. Increasing convex production costs give rise to increasing marginal production costs and upward sloping supply curves, which are consistent with empirical observations such as higher wholesale power prices during the day than at night. Increasing marginal costs arise due to the industry's reliance on an array of production technologies (including hydro, nuclear, coal, and natural gas) with differing marginal costs, and on specialized "peaking capacity" that is used during periods of high demand. As Joskow and Schmalensee (1983, chapter 5) note, fuel efficiency is less important for peaking than for base-load plants, implying higher marginal costs at times when peaking plants are used. To simplify the analysis, all producers are assumed to have identical cost functions, but are allowed to have different scale.

We assume that there is no risk in real time. That is, power companies are able to forecast demand in the immediate future (e.g. over the next hour) with precision, and can either adjust their own

---

7 Note that the fixed retail price assumption does not rule out seasonal variation, e.g. higher prices during the summer than during the spring. The assumption is that retail prices do not react to wholesale market shocks. Hybrid "interruptible" contracts, where producers reserve the right to interrupt power supply in times of scarcity have evolved, as have contracts that allow for limited adjustments in retail prices as a function of wholesale prices. To date it is mainly larger, more sophisticated, retail customers that use these contracts. It seems likely that new contracts providing customers with a degree of protection against price volatility, but providing incentives to reduce usage at times of peak demand will evolve.

output or enter contracts in the wholesale market at competitive prices. However, when viewed from the earlier point in time when retail prices are set or when production capacity decisions are made, future demand is uncertain. We focus on the market for power at a single future date. Given that electricity is essentially not storable, this simple single-period analysis appears to be capable of capturing many important features of the actual markets.

III. Power Pricing and Trading with Regulated Retail Markets.

A. The Initial Market Setting.

Initially, each power company has an exclusive franchise on demand in its region, serves that demand from its own production capacity, and retail prices are fixed (via regulation for example) to yield zero expected economic profits. The main purpose in examining this market setting is to obtain a basis for comparison of the results obtained once competitive power trading is introduced.

The expected profits to firm i for producing during a future period are given by:

$$E(\pi_i) = P_i E[Q_i] - F_i - E[g(Q_i / \alpha_i)] \; . \tag{1}$$

Setting this expression equal to zero, the zero-expected economic profit retail price can be stated as:

$$P_i^* = \frac{F_i + E[g(Q_i / \alpha_i)]}{E[Q_i]} \; . \tag{2}$$

B. Wholesale Power Markets

We next introduce a competitive wholesale spot market for power, where producers can exchange power amongst themselves, while each producer maintains an exclusive franchise on retail demand in its region. Let Q_i^W denote the quantity of energy sold (purchased if negative) by producer i in the spot market, and P^W denote the market clearing wholesale price. Each firm's energy production is the sum of its retail demand and its spot market sale, $Q_i + Q_i^W$. Each producer selects the quantity sold in the wholesale market to maximize real-time profits, which are given as:

$$\pi_i = P_i Q_i + P^W Q_i^W - F_i - g\left(\frac{Q_i + Q_i^W}{\alpha_i}\right). \tag{3}$$

The optimal quantity sold in the spot market by producer i is:

$$Q_i^{W^*} = \alpha_i g'^{-1}(P^W) - Q_i \ , \tag{4}$$

where $g'^{-1}(\bullet)$ denotes the inverse function operator with respect to the first derivative of the cost function.

The market clearing wholesale price can be expressed as:

$$P^{W^*} = g'\left(\frac{Q^D}{\sum_{i=1}^{N}\alpha_i}\right), \tag{5}$$

where $Q^D \equiv \sum_{i=1}^{N} Q_i$ is aggregate electricity demand.

Substituting this expression into equation (4), each producer's sale in the wholesale market can be rewritten as:

$$Q_i^{W^*} = S_i Q^D - Q_i \ , \tag{6}$$

where $S_i \equiv \alpha_i / \sum_{i=1}^{N} \alpha_i$ is firm i's pro-rata share of total system production capacity. The first term on the right side of (6) is the optimal physical production by firm i. It sells in the wholesale market the excess of its physical production over its own demand. With wholesale markets, each firm's optimal physical production is simply its share of system production capacity times total system demand. As a consequence, expected production costs with wholesale trading depend on the variation in system demand, rather than on the variation in local demand. This gives rise to the diversification-induced reduction in average production costs.

Before the demand realizations are known, wholesale prices and profits are random variables.

The firm's expected profits as of an earlier date are:

$$E(\pi_i) = P_i E[Q_i] + E[P^W Q_i^W] - F_i - E\left[g\left(\frac{Q_i + Q_i^W}{\alpha_i}\right)\right] \ . \tag{7}$$

Substituting from the expressions developed above (to reflect that each firm will optimize in the real-time markets), setting expected profits equal to zero, and solving for the fixed retail price gives:

$$P_i^{**} = \frac{F_i + E\left[g\left(Q^D / \sum_{i=1}^{N} \alpha_i\right)\right] - E\left[g'\left(Q^D / \sum_{i=1}^{N} \alpha_i\right)\left(S_i Q^D - Q_i\right)\right]}{E[Q_i]} \ . \tag{8}$$

C. Efficiency Gains From Wholesale Trading.

To quantify the efficiency gains arising due to competitive wholesale trading, we examine the change in the zero-expected-profit retail price due to the introduction of wholesale markets. Comparing expression (8) to expression (2), we obtain the following proposition:

Proposition 1

> The zero-expected-profit retail price with the existence of a competitive spot market for electricity is, for every producer, strictly less than or equal to the zero-expected-profit retail price when there are no wholesale markets. A portion of the efficiency gain can be captured through bilateral forward contracts entered in advance, while the remainder requires real-time trading. There is cross-sectional variation in the efficiency gains from wholesale trading, with the largest efficiency gains realized by producers whose local demand is least correlated with system demand. (Proof provided in Appendix.)

The appendix demonstrates that the difference between P_i^* in equation (2) and P_i^{**} in equation (8) can be expressed as:

$$P_i^* - P_i^{**} = \frac{E\left[g(Q_i / \alpha_i)\right] - E\left[g\left(Q^D / \sum_{i=1}^{N} \alpha_i\right)\right] + E\left[P^W\right]E\left[S_i Q^D - Q_i\right] + Cov\left[P^W, (S_i Q^D - Q_i)\right]}{E[Q_i]} \geq 0 \ . \tag{9}$$

The wholesale spot market allows producers with greater demand relative to production capacity to transfer production to those with relatively less demand, resulting in better risk-sharing across utilities and lower retail prices to consumers. If (i) there is *any* cross-sectional variation in demand surprises across areas, or (ii) forecast relative demand differs across firms, then efficiency is improved and the zero-expected-profit retail price declines with the introduction of the wholesale market. In practice these conditions will surely be met.

Efficiency is improved and retail prices are decreased by wholesale power trading, which obviously requires an interconnected power grid. Power markets in the United States are currently served by three (Western, Eastern, and Texas) grids that are largely not interconnected. This analysis implies that additional efficiency gains could be obtained by allowing for power transactions across grids, and by alleviating transmission bottlenecks within the existing grids, to allow further diversification of demand shocks. The governors of several western U.S. states recently issued a public call for the construction of new power lines to facilitate additional transmission of power across western states (see Janofsky (2001)). The greatest gains might be obtained by connecting the Texas grid, which is currently confined to the state of Texas, to the Western and Eastern grids.[8]

Of course any potential gains from better diversification would have to be balanced against increased line losses and other costs of interconnection, and we do not model these costs here. As an example of indirect costs, some have argued that the interconnected power grid widened the highly publicized power blackout in the northeastern U.S. that occurred on August 14, 2003.[9]

The first three terms in the numerator on the right side of (9) reflect efficiency gains from trades in the wholesale market that can be forecast *a priori*. The first term is the producer's expected variable costs of generating power to cover local demand, while the second term is the producer's expected

8 See, for example, "Texas May Face a Glut of Electricity, but that Won't Aid the Rest of U.S.", *Wall Street Journal*, May 7, 2001, page A1.

variable costs from producing its pro rata share of system demand. The differential between the first term and the second term will be positive on average because production costs are convex in output, and in a direct corollary to results obtained in the theory of investment portfolios, system demand is on average less volatile than local demand. The third term in the numerator on the right side of (9) reflects trades undertaken in response to predictable demand variation. Firms that expect their own demand to exceed their pro-rata share of system demand will plan to purchase power, and vice versa. For example, firms in the Southwestern United States will sell power in the wholesale markets in the winter and purchase wholesale power during the summer, while producers in the Northwestern United States optimally take the opposite side of the contracts. Note that, since the first three terms in the numerator of (9) are based on ex ante expectations, the efficiency improvements attributable to these terms could be captured with bilateral forward contracts negotiated in advance.

The final term in the numerator on the right side of (9) reflects efficiency gains from the reactive trades that will take place in spot markets in response to *ex ante* unpredictable demand shocks. These efficiency gains cannot be captured by forward contracts, as they comprise optimal responses to demand surprises. Although this term is strictly non-negative (and is zero only if local demand never differs from system average demand), implying efficiency gains for all producers, there is cross-sectional variation in the efficiency gains. Letting *Cov* denote covariance, *Corr* denote correlation, and σ_X denote the standard deviation of *X*, the last term in the numerator on the right side of (9) can also be expressed as:

$$Cov\left[P^W, \left(S_i Q^D - Q_i\right)\right] = Corr\left[P^W, S_i Q^D\right]\sigma_{P^W}\sigma_{S_i Q^D} - Corr\left[P^W, Q_i\right]\sigma_{P^W}\sigma_{Q_i} \ . \tag{10}$$

Expression (10) shows that the efficiency gains from wholesale trading will be greatest when the correlation between local demand and wholesale price (which is a monotonic increasing function of system demand) is lowest, *ceteris paribus*. A relatively high correlation of local with system demand

9 For a discussion of system reliability issues in general and the August 2003 blackout in particular, see “Understanding the 2003 Power Outages, Cascading Blackouts, & the Transmission Grid” at the web site http://www.icfconsulting.com/Markets/Energy/northeast-blackout.asp .

implies that the producer needs to satisfy high local demand at the same times that system demand and wholesale prices are high, and vice versa, resulting in relatively little efficiency gain from wholesale trading. In contrast, a producer with a low or negative correlation of local with system demand will be more likely to have low own demand when system demand is high, allowing it to sell excess production when wholesale prices are high, while purchasing to satisfy large local demand at times when wholesale prices are low.

That efficiency gains from wholesale trading are greatest when cross-sectional demand correlation is lowest is analogous to the standard result from financial portfolio theory that the risk reduction due to diversification is greatest when correlations across returns are lowest. Examples of locales with low correlations of local with system power demand might include temperate coastal environments such as the San Diego, San Francisco, and Seattle areas.

IV. Simulation Evidence on the Magnitude of Efficiency Gains From Wholesale Trade.

Expression (9) demonstrates efficiency gains from wholesale power trading for all producers. In this section we attempt to quantify the potential efficiency gains from wholesale trading by use of a simulation that relies on generating cost data from the California market, as well as hourly demand (load) data during calendar year 2000 from the California, New York, and PJM markets.[10] Our intent is to estimate potential efficiency gains for the U.S. as a whole, and to provide insights as to intertemporal and cross-sectional variation in the efficiency gains. We use the California, New York, and PJM data because it is available to us, and rely on bootstrap techniques to extend the analysis to other geographic areas and other production cost and demand scenarios.

California power markets have been the subject of substantial recently scrutiny, much of which has focused on allegations that prices exceeded competitive levels or that the market was subject to

10 The PJM market includes all or parts of Delaware, Illinois, Maryland, New Jersey, Ohio, Pennsylvania, Virginia, West Virginia, and the District of Columbia.

various forms of manipulation. It is not our goal to revisit these controversies. Rather, we use the California generation data because it is available and allows us to estimate marginal production costs based on actual generating plant efficiencies. Data on the output and operating efficiency of California fossil-fuel generating plants was kindly provided to us by Severin Borenstein, and was also used in Borenstein, Bushnell, and Wolak (2002). Similarly, the use of actual demand data from California, New York, and PJM allows us to be realistic in terms of the parameters of the power demand distribution.

The California demand data includes scheduled and actual hourly load for the entire California market, and also scheduled hourly load for five delivery zones within California during calendar year 2000, and is obtained from the web site of the University of California Energy Institute.[11] The New York and PJM data include load measured on an hourly basis during calendar year 2000.[12]

Table 1 reports some summary statistics on the hourly load data. Panel A reports on actual hourly loads for the California, New York and PJM markets. The PJM market is largest, with mean load of 30,114 MW and maximum load of 49,462 MW. The California market is nearly as large, with mean and maximum load of 27,188 and 43,590 MW, respectively. The New York ISO is smaller, with mean load of 17,833 and maximum load of 28,138 MW.

As noted above the generation data available to us is based on California power plants. Power prices will depend on average and maximum demand relative to production capacity. In order to make use of the California production data for other markets we scale each demand observation from the non-California markets, dividing by the ratio of the local to the California mean, so that the mean of each scaled demand series is the same as the mean California load. In assessing the simulation outcomes it is important to keep in mind that all of the demand data, real and simulated, is scaled to match California

11 http://www.ucei.berkeley.edu/datamine/datamine.htm. The five California delivery zones are NP15 (North of Path 15), SP15 (South of Path 15), Humboldt, SF (a portion of the San Francisco area), and ZP 26 (Zonal Path 26, an area of west-central California north of Santa Barbara and south of Monterey). We exclude the Humboldt data for brevity, as it represents less than 0.2% of California demand.

12 The New York and PJM demand data are obtained from the websites http://www.nyiso.com/markets/index.html#DAM and http://www.pjm.com/markets/jsp/loadhryr.jsp , respectively.

demand.

Panel B of Table 1 reports correlations in hourly demand across the three markets for which hourly load data is available, and also for the average demand across the three markets. These correlations are relatively high. The correlation between California and PJM demand is 0.735, while the correlation between California and New York demand is 0.828. PJM and New York demand are highly correlated, at 0.955.

Actual hourly load data is only available for the three markets mentioned. To obtain some additional indication as to plausible correlations in demand we examine also scheduled hourly load data for four delivery zones within California. Panel C of Table 1 reports on average scheduled hourly loads for these delivery zones, while Panel D reports on correlations in scheduled hourly load. The correlation between NP15 (North of Path 15) and SP15 (South of Path 15) load is 0.548, and the correlation between SF and NP15 load is 0.232. The average pairwise correlation reported on Panel D of Table 1 is 0.447. This data is suggestive that considering simulations incorporating demand correlations at or below 0.5 is likely to be instructive.

A. Electricity Demand

The underlying state variable in our model is the demand for electricity. To assess the potential efficiency gains from wholesale trade we first compute the retail price required by a power company that serves only one of the observed markets. Results reported are based on California as the single, local, market. Results are qualitatively similar if we take either New York or PJM as the local market. We then assess the retail price required to cover expected costs if the power company produces to cover a pro rata share of system demand. System demand is defined as the average demand observed across California, New York, and PJM, as well from 0 to 10 simulated markets described below.

Let $\tilde{Q}_i^A$ denote the average demand in hour i across the California, New York, and PJM markets. We create a bootstrap sample of 20,000 observations on $\tilde{Q}_i^A$ by drawing at random and with replacement from the set of 8,784 actual hourly observations. The summary data in Table 1 imply that total annual output by producers in the California, New York, and PJM markets is 658 million MW. This represents approximately 17% of total 2002 U.S. power production.[13] We wish to assess the possible efficiency gains from nationwide (or international) wholesale power trading. In lieu of actual hourly load data for the remaining markets we construct simulated demand data for ten additional markets, each with the same mean demand as the California market. These additional demand series are constructed as weighted averages of $\tilde{Q}_i^A$, and new series denoted $\tilde{Q}_i^R$, that have the same probability distribution as $\tilde{Q}_i^A$ but are uncorrelated with the observed demand. Each $\tilde{Q}_i^R$, series is constructed by drawing 20,000 observations at random and with replacement from $\tilde{Q}_i^A$.

Demand in hour i for simulated market j is then computed as:

$$\tilde{Q}_i^j = \mu + \left[(1-\omega)\tilde{Q}_i^{R*} + \omega\tilde{Q}_i^{A*}\right], \tag{11}$$

where μ is the mean of $\tilde{Q}_i^A$, an asterisk indicates a data series stated as deviations from mean, and $0 \le \omega \le 1$ is a weighting variable used to ensure a desired correlation between simulated and observed demand. Since $E[\tilde{Q}_i^{R*}] = E[\tilde{Q}_i^{A*}] = 0$, $\sigma^2_{Q^R} = \sigma^2_{Q^D}$, and $E[\tilde{Q}^{R*}\tilde{Q}^D] = 0$. The portion of the simulated demand series that is uncorrelated with observed demand is $(1-\omega)\tilde{Q}^{R*}$, while $\omega\tilde{Q}^{D*}$ is the portion of simulated demand that is perfectly correlated with observed electricity demand. Using (11) and the properties of variances and covariances, the correlation between the simulated demand series and observed demand $\tilde{Q}_i^A$ will be given as:

$$\rho = \frac{\omega}{\sqrt{1-2\omega+2\omega^2}} \tag{12}$$

---

13 The sum of mean hourly load across the three markets is 75,135 MW. Multiplying by 8760 hours per year gives 658 million MW. The national production data for 2002 is from the web site http://www.eia.doe.gov/neic/quickfacts/quickelectric.htm

To ensure that the simulated demand series has a specified correlation, ρ, with observed demand $\tilde{Q}_i^A$, requires that the weighting variable be selected as:

$$\omega = \frac{\rho^2 - \rho\sqrt{1-\rho^2}}{2\rho^2 - 1}. \tag{13}$$

Using the bootstrapped actual demand data $\tilde{Q}_i^A$ and the simulated demand data $\tilde{Q}_i^j$ we simulate system demand, $\tilde{Q}_i^D$ as

$$Q_i^D = \left[3Q_i^A + \sum_{j=0}^{N} Q_i^j\right]/(N+3). \tag{14}$$

In expression (14) the constant 3 appears because $\tilde{Q}_i^A$ reflects mean demand across the California, New York, and PJM markets.

We assess the retail price required for a firm producing a pro rata share of $\tilde{Q}_i^D$ and compare it to the retail price required if the same firm produces only for its home market, while varying ρ, the correlation of simulated market demand with observed demand, from 0 to 1.0, and varying N, the number of simulated markets, from zero to 10. We later investigate the effect on gains from trade of altering production capacity and mean demand in the local and the simulated markets.

B. Generating Costs

To simulate the efficiency gains from wholesale power trading we require estimates of the cost function, $g(Q_i)$, that relates total production costs to electricity demand. The California electricity market is served by a variety of generation, including hydro, nuclear, coal, natural gas, and fuel oil. The marginal costs of these different types of generation vary dramatically, with hydro and nuclear being relatively inexpensive and natural gas and fuel oil being more expensive. Base load demand (load generally served on an around-the-clock basis) is primarily served by hydro, nuclear, and coal fired generation, while demand fluctuations above the base level are most often served by the more expensive generating sources, which can be brought on line relatively quickly.

Data on the natural gas and oil-fired generators in California as of the end of 1998 is obtained from Borenstein, Bushnell, and Wolak (2002), and includes 91 generating plants spread across seven unique owners. The total capacity of these 91 generators is 17,610 Megawatts (MW). We assume that the majority of power demand is covered from base load capacity (e.g. hydro, nuclear, and coal) at a constant marginal cost, while peak demand (i.e. that beyond base load) is covered from the gas and oil-fired generators. Data supplied by the California Energy Commission indicates that 33.4% of California power production is supplied by gas and oil generation, while data provided by the United States Department of Energy indicates that 20% of total U.S. power generation is fueled by gas and oil.[14] We use an empirical approach to assess the amount of base load capacity that is consistent with the observed production from gas and oil peaking generation.

More specifically, we initially set base load capacity at an arbitrary level, *B*. We then take each outcome, $\tilde{Q}_i^A$ from the empirical demand distribution and measure the portion covered by peaking plants (i.e. not from base load) as max(0, $\tilde{Q}_i^A$ - *B*). This provides 20,000 outcomes on production from gas and oil plants. We then vary B until average production from gas and oil plants is a specified percentage of average total production. The result is that base load production capacity of 22,050 MW is consistent with 20% of $\tilde{Q}_i^A$ being generated by gas and oil, while base load capacity of 18,150 MW is consistent with 33.4% of $\tilde{Q}_i^A$ production from gas and oil-powered plants. The initial simulations assume base load capacity of 18,000 MW. We subsequently report simulations where base load capacity varies from 17,000 to 27,000 MW. In each simulation demand in excess of base load is assumed to be covered by the oil and gas-fired plants for which we have explicit data.

The data on the generating plants include the fuel source, the heat rate of the generator (i.e., the

.14 These data are obtained from the websites http://www.energy.ca.gov/html/energysources.html and http://www.eia.doe.gov/neic/quickfacts/quickelectric.htm respectively.

efficiency with which fuel is converted into electricity), operating and maintenance costs, and data on outage rates, which indicate, on average, how often each generator is out of service. We obtain monthly prices for natural gas, fuel oil, and jet fuel for delivery in California from the Website of the Energy Information Administration at the Department of Energy.[15]

Data on plant efficiency, outage rates, fuel costs, and bootstrapped demand outcomes are used to simulate generating costs. For each of the 20,000 demand realizations we first assess whether a specific generator is in service. Individual generation units are removed at random from a percentage of individual outcomes that corresponds to the units overall outage rate. We record the actual date for each bootstrapped outcome on system demand, and use actual fuel prices in effect on that date. Our bootstrapped data thus endogenously account for any correlation between demand and input prices.[16] We then use data on plant efficiency and fuel prices, and aggregate across the generators that are in service for the specific demand realization, to form the "supply stack", from which we can assess total variable costs and marginal production costs for each demand outcome.

The procedure described above is computationally expensive. For computational efficiency in our price simulations we fit a simple piecewise quadratic cost function that closely matches the precise cost function. For demand realizations up to the base load (B) level generating costs are set to $42.35 per MWh. For demand realizations from base load to 15,000 MW over base load generating costs are estimated as $42.35+0.000223*(Q^D- B)+4.105e-8*(Q^D-B)^2$. For demand realizations greater that B + 15,000 MW generating costs estimated as $54.9327+0.027509*(Q^D-(B+15{,}000))+0.000013*(Q^D-(B+15{,}000)^2$. A comparison of the actual and fitted cost functions for production above base load is

15 http://www.eia.doe.gov/emeu/states/main_ca.html.

16 As it turns out, the correlation between demand and fuel prices is not empirically important during calendar year 2000. At a monthly frequency, the correlation between power demand and the California gas prices used in the simulations is -0.034. We also re-estimated the cost function using daily NYMEX settlement prices for the nearest to delivery Henry Hub gas contract. The correlation between daily electricity demand and the Henry Hub price is just 0.036. The cost function obtained when using the Henry Hub daily data is shifted downward (as average Henry Hub prices were lower than California prices), but the shape of the cost function does not change in any meaningful way.

shown in Figure 1, and reveals that the fitted cost function closely approximates the actual function. This fitted cost function is used for the simulations described below.

C. Simulation Outcomes

To explore the magnitude of the possible gains from wholesale trading we conduct a series of simulations based on the bootstrapped data on electricity demand and the fitted cost function. We compute the retail prices that yield zero expected profits to producers in both the initial case where each producer has an exclusive obligation to serve demand in its region from its own production (expression 2), and in the case where all producers are allowed to participate in a deregulated wholesale market (expression 8).

The most basic simulation uses data from the California, PJM, and New York, electricity markets. As noted above, the fraction of output produced by natural gas and oil in California is 33.4%, which is consistent in the demand data with base load capacity of approximately 18,000 MW. Fixed generation costs are set at $500,000.[17] System demand initially includes only the California, New York and PJM markets. The most basic simulation outcomes therefore model the efficiency gains that could arise to a firm that participates in the combined California, New York and PJM markets (ignoring the effects of transmission losses, which we do not model), instead of just the California markets. In addition to this most basic case, we examine the effects of adding additional markets to the system, where we vary the number of additional markets and the correlation coefficient between these simulated markets and the three existing markets.

For each parameterization, we generate 20,000 outcomes on demand and production costs. The parameters of (2) and (8) are then computed across the 20,000 outcomes to provide zero-expected profit retail prices without and with wholesale trading. In the first two sets of simulations we report, all firms are identical and face the same demand probability distribution, so that the first three terms in the numerator of expression (9) are zero. These simulations capture the efficiency gains resulting from real time trading in response to demand shocks only, and therefore estimate the *lower bound* of the efficiency gains from wholesale power trading (gross of line losses). In the third set of simulations we allow for heterogeneity in the mean level of demand across markets. In addition to gains resulting from real-time trading in response to demand shocks, these simulations also capture potential efficiency gains from trades in the wholesale market that can be forecast *a priori* (i.e., efficiency gains reflected in the first three terms in the numerator on the right side of (9)). Note that, since the first three terms in the numerator of (9) are based on ex ante expectations, the efficiency improvements attributable to these terms in theory be captured with bilateral forward contracts negotiated in advance, but only if the markets are physically interconnected.

C.1. Base case simulations

Given the assumptions described above, the zero expected profit retail price for a firm serving only the California market is \$84.40 per MWh. The gains from trade when wholesale trading is allowed are illustrated by Figures 2 and 3. Figure 2 plots the zero-expected profit retail price *with* wholesale trading, as a function of the number of additional markets (beyond New York and PJM) and the correlation between these additional markets and observed demand, while Figure 3 plots the percentage reduction in retail price attributable to wholesale trading.

Consistent with the analytical results presented in Section III.C, retail prices decrease with lower correlation in demand and with an increase in the number of markets included in the system. A large number of individual markets and/or lower demand correlations imply greater diversification-induced risk reduction due to wholesale trade. This result is directly analogous to the insight from financial portfolio theory that risk reduction from diversification is greater if correlations are lower and when more assets (or markets in this case) are included in the system.

---

[17] The level of fixed costs assumed for the simulation does not materially affect estimates of price reductions due to trade. It does, however, affect levels of prices with or without trade. Fixed costs were selected to obtain minimum

Note that the decrease in retail price is significant, even with no additional markets beyond New York and PJM in the system. The retail price in this case falls to $75.91, a decrease of 10.1%. If additional markets are added to the system the retail price falls further (unless demand in the new markets is perfectly correlated with observed demand). In the most extreme case considered, with 10 additional markets and zero correlation between the simulated and the observed markets, the retail price falls to $63.30, a decrease of 25.0% relative to that obtained without wholesale trading. While this final case may be unrealistic due to the assumption of zero correlation, these results illustrate the degree to which the number of power markets in the integrated system and their correlation structure affect the potential trading gains.

We view the inclusion of ten simulated markets in addition to the three observed markets as conservative in light of the observation that the three observed markets account for about 17% of national power production. Which correlations are most reasonable is more difficult to say. If the correlation of California with PJM demand of 0.735 is representative of correlations between unobserved markets and observed markets, then Figure 1 implies efficiency gains to California producers from nationwide wholesale trading of approximately 16%, gross of line losses.

As a basis for comparison, White (1995) has also provided an estimate of the gains from trade that would arise in a fully competitive power market. Using actual trading data from four of the largest utilities in California, he estimates gains from trade of over 4%. This estimate is near the lower end of the range of the estimates we present, but is not inconsistent with our estimates. A key distinction is that the White estimates by construction apply to intra-California trading. Our simulations examine the potential gains that would be expected if the California markets could be interconnected with the PJM, New York, and other national markets. As noted earlier, power markets in the United States are currently served by three (Western, Eastern, and Texas) grids that are largely not interconnected. This analysis implies that additional efficiency gains could be obtained by allowing for power transactions across grids, which

---

retail prices of about $60 per MWh, which seems empirically reasonable.

would allow further diversification of demand shocks. Of course any potential gains from better diversification would have to be balanced against increased line losses and other costs of interconnection.

C.2. The effect of altering base load capacity

To this point the simulations have assumed base load capacity of 18,000 MW, which is consistent in the observed demand data with 66.6% of total production being covered from base load capacity, while the remainder is covered by natural gas and oil-fired peaking plants, as actually observed in California. On a national basis, however, only about 20% of production is serviced by natural gas and oil-fired generation. The base load capacity that corresponds, given the observed demand distribution, to 20% of production from peaking capacity is 22,000 MWh. The amount of base load capacity is relevant in assessing gains from trade, because the gains from trade arise from the convexity of the marginal cost function, and greater base load capacity implies more demand realizations in the flat portion of the cost curve. In this section we examine how changes in base load capacity affect the potential gains from wholesale trading. The simulations in this subsection fix the number of additional markets at 10, while varying base load capacity and demand correlation.

Figure 4 reports the percentage reduction in the retail price as a result of wholesale power trading as a function of base load capacity. As illustrated in the Figure, higher base load capacity lowers the potential gains from wholesale trading. The intuition for this result is that with greater base load capacity, fewer of the demand realizations fall into the convex portion of the cost curve, so there is less gain to sharing demand shocks across regions. Focusing, for example, on demand correlations of 0.8, efficiency gains from wholesale power trading range from 16.9% when base load capacity is 17,000 MWh to only 0.1% when base load capacity is 27,000 MWh.[18] An implication is that gains from trade will be greater

18 The simulations indicate lower retail prices (not reported) both with and without wholesale trading, with greater base load capacity. This is mainly an artifact of our assumption that fixed generating costs are invariant regardless of base load, which reflects our focus on quantifying the determinants of gains from trade rather than assessing the optimal amount of base load capacity. To determine optimal base load capacity would require assessment of the tradeoffs of greater fixed costs against lower marginal costs on plants used sparingly, a task beyond the scope of this paper.

during periods when base load capacity is reduced, e.g. when hydro output declines due to drought. As expected, the price reductions from wholesale trading continue to be decreasing in the correlation across markets, regardless of base load capacity.

Results obtained with a base load capacity of 22,000 MWh are of particular interest, since this base load capacity corresponds to 20% of production from gas and oil peaking plants, the U.S. national average. In this case the decrease in retail prices due to wholesale trading ranges from 2.5% to 4.0%, depending on the correlation in demand across markets. At a conservative correlation estimate of 0.7, the gain from trade is 3.6%.

We view the 3 to 4% range to be a conservative estimate of the gains from trade arising from real time trading on a nationwide basis. The source of the conservatism is that the gains from trade are convex in base load capacity, so that the average gain from trade across various amounts of base load exceeds the gain from trade evaluated at average base load capacity. Focusing, for example, on the case where the demand correlation is 0.7, the average efficiency gain from trade across all base load capacities considered is 6.5%, while the gain evaluated at average base load capacity is 3.6%.

Each percentage point reduction in power prices equates to a large change in total dollar expenditures. The United States Department of Energy reports national electricity consumption of 3463 MWh during 2002, at a weighted average price of $72.10 per MWh, giving total electricity expenditures of $249.7 billion.[19] A four percent reduction in this total, for example, equates to $10.0 billion.

C.3. The effect of demand variation

The simulations conducted to this point have incorporated the assumption that each market is characterized by constant and equal mean demand. The final set of simulations we report consider the effect of variation in mean demand. The simulations are conducted while assuming base load capacity of 22,000 MWh (corresponding to the actual national average usage of peaking plants) and demand

---

19 http://www.eia.doe.gov/neic/quickfacts/quickelectric.htm .

correlation between the simulated and actual markets of 0.8. In these simulations we shift each individual demand outcome by an indicated percentage of mean demand, varying from a 20% decrease to a 20% increase.

These simulations provide insights regarding two distinct but related issues. Simulations that shift both local and system demand by the same percentage illustrate the effect of demand variation on the efficiency gains from real time trading. Simulations that alter local and system demand by differing percentages quantify also the efficiency gains arising from the first three terms in the numerator of equation (9), which depend on cross-sectional variation in expected demand, and that can be captured by forward contracting.

Figure 5 reports the percentage reduction in retail prices attributable to wholesale trading for varying levels of average local and system demand. To assess the effect of variation in average local demand, consider outcomes when system average demand is held constant at its base case value (0% shift in the figure). Then the efficiency gain from wholesale trading is monotone increasing in local demand, ranging from a 0.6% price improvement if average local demand is 20% lower than in the base case, to a 3.5% reduction if local demand is also held at the base level, to a 19.8% efficiency improvement if average local demand increases by 20%. To interpret this result it is important to recall that we are comparing the retail price with wholesale trading to the retail price without trading. Given fixed production capacity, retail prices increase with average demand, with *or* without wholesale trading. These results indicate that retail prices increase *less* with average local demand if there is wholesale trading, and that the percentage improvement due to wholesale trading is greatest when average demand is highest.

Figure 5 also illustrates that the gains from real time trading increase with overall demand. Focusing on the diagonal of the Figure, starting at the nearest point, if both local and system mean demand are decreased by 20% the efficiency gain is only 0.2%. This occurs because at low levels of system and local demand all firms are less likely to face demand realizations in the convex part of the cost function. The efficiency gain from real time trading increases monotonically with average demand, to 7.7% if both local and system demand are increased by 10%, and to 12.2% if both local and system demand rise by 20%.

Finally, Figure 5 shows that the efficiency gains from wholesale trade are greatest when mean demand diverges most across markets. If average local demand is decreased by 20% while average system demand is increased by 20% the price reduction from wholesale trade is 19.3%, and if local demand increases by 20% while system demand falls by 20% the retail price reduction from wholesale trade is 20.9%. In either of these divergent situations firms in the low mean demand market can efficiently service load in the high mean demand market.

Gains from trade arising due to differences in mean demand are likely to be important in practice due to seasonal effects in demand. Power companies can respond to higher average annual demand by placing more production capacity in the vicinity of the high demand areas, but cannot readily make a similar locational response to varying seasonal demands. The well-documented higher summer demand in the U.S. southwest and the lower summer demand in the U.S. northwest provide an example of seasonal demand variation that can be exploited through wholesale power trade. While seasonal variation in average demand can be accommodated with forward rather than real time contracts, sufficient transmission capacity is required to capture the efficiencies.

Table 2 reports some data useful in assessing the extent to which predictable seasonal demand variation is relevant in determining efficiency gains from wholesale power trade. Each cell of Table 2 reports mean hourly demand for the month indicated relative to mean hourly demand in the same market for the full year. Efficiency gains from predictable demand variation arise when these data series differ across markets in a given month. For example, in January, California demand is 94% of normal while PJM demand is 107% of normal. Let $|107/94 - 1| = 0.138$ denote the magnitude of the January demand variation between California and PJM. By this measure, seasonal demand variation between California and PJM averages 6.1% across months. Data from the California delivery zones suggests that larger seasonal demand variations are plausible. For example, the seasonal demand variation between the SP15 delivery zone and the ZP26 delivery zone averages 12.0% across the twelve months.

Any summary statement made in the absence of actual demand and production cost data for all parts of the country must be approximate and conditional. However, we note that Figure 5 displays efficiency gains in the vicinity of six to ten percent when cross-sectional mean demand varies by approximately four to twelve percent. The simulations underlying Figure 5 also incorporate realistic but arguably conservative assumptions as to additional markets (10 additional markets, implying that observed demand is 23% of total), correlations across markets (0.8), and constant base load capacity (20% of production from peaking plants). We therefore conclude that a conservative range of estimates for efficiency gains due to competitive wholesale power trading lies in the vicinity of six to 10 percent per year. These estimates equate on a national basis to savings of approximately $15 to $25 billion per year.

V. Conclusions

This paper applies concepts from financial portfolio theory to analyze prices, profitability, and risk sharing in deregulated power markets. Wholesale power trading improves efficiency by smoothing production across the system. Trades arranged in advance can capture a portion of the efficiency gains, that attributable to predictable differences in demand across locales. The rest of the efficiency gains are attributable to diversifying demand shocks across the system. Capturing these efficiencies requires real-time trades made in response to surprises in demand. While all producers gain from wholesale trading, the largest gains accrue to when the correlation of local with system demand is low, and when there is cross-sectional variation in mean demand.

We provide simulation evidence that relies on actual production cost data from California and actual demand data from the California, New York, and PJM markets. Since we are interested in nationwide efficiency gains, we also construct bootstrapped demand data for up to ten hypothetical markets to assess the potential efficiency gains from wholesale trading. Realized efficiency gains will depend on numerous factors, including base load (constant marginal cost) production capacity, the number of markets integrated into the system, the correlation of power demand across markets, the amount of seasonal variation in average demand, and line losses (which we do not model). Results of the simulation indicate that power prices in California (which relies more on peaking plants and less on base load) could be decreased by 10 to 20% if all efficiency gains from trade were realized. For the U.S. as a whole, efficiency gains from real time trading are likely to be in the vicinity of 3 to 4%. Considering also the use of forward contracts to exploit predictable variation in mean demand, we estimate nationwide efficiency gains of 6 to 10%.

We also assess the effect of increasing power demand. Efficiency is improved by wholesale trading for any given level of expected demand. Although retail prices increase with greater expected demand with or without wholesale power trading, the efficiency gains due to wholesale trading are actually larger when average demand is greater. Thus, in assessing whether consumer welfare has been improved by deregulation and the development of power markets it is crucial to separate the effects of increasing power demand from that of wholesale power trading.

The analysis presented here has policy implications. In particular, the desirability of wholesale power trading and power industry deregulation cannot be judged wholly on the basis of the recent increases in retail prices observed in California and other locations. Increased power demand without increased production capacity would have led to higher prices in any case.

Empirical evidence indicates that wholesale electricity markets, particularly in California, have been characterized by a degree of monopoly power. It is not our intent to dispute this evidence, but rather to reaffirm that the trading of power in *competitive* wholesale markets reduces retail power prices. To realize the efficiency gains that are theoretically attainable requires competitive power markets and a complete power grid. Economic efficiency would be best served by policy aimed at ensuring that power markets are indeed competitive, and that sufficient transmission capacity exists for profitable power trades to be completed.

Appendix:

Proof of Proposition 1:

The difference between $\mathrm{P}_{\mathrm{i}}^{*}$ described by equation (2) and $\mathrm{P}_{\mathrm{i}}^{**}$ described by equation (8) can be stated as:

$$P_i^* - P_i^{**} = \frac{E[g(Q_i/\alpha_i)] - E\left[g\left(Q^D / \sum_{i=1}^{N} \alpha_i\right)\right] - E\left[g'\left(Q^D / \sum_{i=1}^{N} \alpha_i\right)\left(Q_i - S_i Q^D\right)\right]}{E[Q_i]} \ . \tag{15}$$

Given that the denominator in equation (15) is positive, the proof requires that the numerator in equation (15) is greater than or equal to zero.

By the mean value theorem:

$$g(Q_i/\alpha_i) = g\left(Q^D / \sum_{i=1}^{N} \alpha_i\right) + g'(x)\left(Q_i/\alpha_i - Q^D / \sum_{i=1}^{N} \alpha_i\right), \tag{16}$$

for some $x \in \left[Q_i/\alpha_i, Q^D / \sum_{i=1}^{N} \alpha_i\right]$. Now suppose that $Q_i > S_i Q^D$, then by definition,

$Q_i/\alpha_i > Q^D / \sum_{i=1}^{N} \alpha_i$, and it follows that:

$$g(Q_i/\alpha_i) > g\left(Q^D / \sum_{i=1}^{N} \alpha_i\right) + g'\left(Q^D / \sum_{i=1}^{N} \alpha_i\right)\left(Q_i/\alpha_i - Q^D / \sum_{i=1}^{N} \alpha_i\right). \tag{17}$$

Now suppose that $Q_i < S_i Q^D$, then by definition, $Q_i/\alpha_i < Q^D / \sum_{i=1}^{N} \alpha_i$, and again it follows that:

$$g(Q_i/\alpha_i) > g\left(Q^D / \sum_{i=1}^{N} \alpha_i\right) + g'\left(Q^D / \sum_{i=1}^{N} \alpha_i\right)\left(Q_i/\alpha_i - Q^D / \sum_{i=1}^{N} \alpha_i\right). \tag{18}$$

Combining the inequalities in equations (17) and (18) and taking expectations on both sides of the resulting inequality with respect to the joint probability distribution of electricity demand, $H(Q_1, Q_2, ..., Q_N)$, yields:

$$\int_N g(Q_i / \alpha_i)\, dH(Q_1, Q_2, ..., Q_N) > \int_N \left[g\left(Q^D / \sum_{i=1}^{N} \alpha_i \right) + g'\left(Q^D / \sum_{i=1}^{N} \alpha_i \right)\left(Q_i - S_i Q^D\right) \right] dH(Q_1, Q_2, ..., Q_N)\ , \tag{19}$$

which proves propositon 1. Using the definition of the market clearing wholesale price from text equation (5) and expanding the expectation in the third term in the numerator in equation (15) yields text equation (9).

References

Bailey, E., 1998, "The Geographic Expanse of the Market for Wholesale Electricity", working paper, Massachusetts Institute of Technology, Boston, MA.

Bessembinder, H. and M. Lemmon, 2002, "Equilibrium Pricing and Optimal Hedging in Electricity Forward Markets", *Journal of Finance*, 57,1347-1382.

Borenstein, S., J. Bushnell, and F. Wolak, 2002, "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market", *American Economic Review*, 92, 1376-1405.

Borenstein, S., and J. Bushnell, 1999, "An Empirical Analysis of the Potential for Market Power in California's Electricity Industry", *Journal of Industrial Economics*, 47.

Coleman, J. 2003, "Senators Unveil Bill to Repeal Energy Regulation", Associated Press wire story. The story can be downloaded from the web site: http://www.consumerwatchdog.org/utilities/nw/nw003259.php3 .

Ciccone, J., 2001, "Time to Build" *Electric Perspectives*, Jan/Feb. Washington D.C.

Green, R., and T. McDaniel, 1998, "Competition in Electricity Supply: will '1998' be worth it?", working paper, University of Cambridge.

Green, R., 1998, "England and Wales – A Competitive Electricity Market?", working paper, University of Cambridge.

Green, R., and D. Newbery, 1992, "Competition in the British Electricity Spot Market", *Journal of Political Economy*, 100, 929-953.

Harvey, S., and W. Hogan, 2001, "On the exercise of Market Power Through Strategic Withholding in California", Working paper, Harvard University.

Harvey, S., W. Hogan, and S. Pope, 1996, "Transmission Capacity Reservations Implemented Through a Spot Market with Transmission Congestion Contracts", *The Electricity Journal*, 42-55.

Hogan, W., 1993, "Markets in Real Electric Networks Require Reactive Prices", *The Energy Journal*, 171-200.

Janofsky, M. 2001, "Western Governors Turn Focus to Need for More Power Lines", *New York Times*, May 9.

Joskow, P., 1997, "Restructuring, Competition, and Regulatory Reform in the U.S. Electricity Sector", *Journal of Economic Perspectives*, 11, 119-138.

Joskow, P. and R. Schmalensee, 1986, "Incentive Regulation for Electric Utilities" *Yale Journal on Regulation*, 4, 1-49.

Joskow, P. and R. Schmalensee, 1983, *Markets for Power*, MIT Press, Cambridge, Massachusetts.

Joskow, P. and E. Kahn, 2002, “A Quantitative Analysis of Pricing Behavior In California’s Wholesale Electricity Market During Summer 2000: The Final Word”, working paper, Massachusetts Institute of Technology.

Krugman, P. 2001, “Enron Goes Overboard”, New York Times (August 17).

Newbery, D., 1995, “Power Markets and Market Power”, *The Energy Journal*, 39-66.

Tabors, R. 1996, “A Market-Based Proposal for Transmission Pricing” *The Electricity Journal*, 61-67.

Wolfram, C. 1999, “Measuring Duopoly Power in the British Electricity Spot Market”, *American Economic Review*,

White, M., 1995, “Dynamic Efficiency and the Regulated Firm: Evidence from Interfirm Trade in Electricity Markets”, working paper, Stanford University, Stanford, CA.

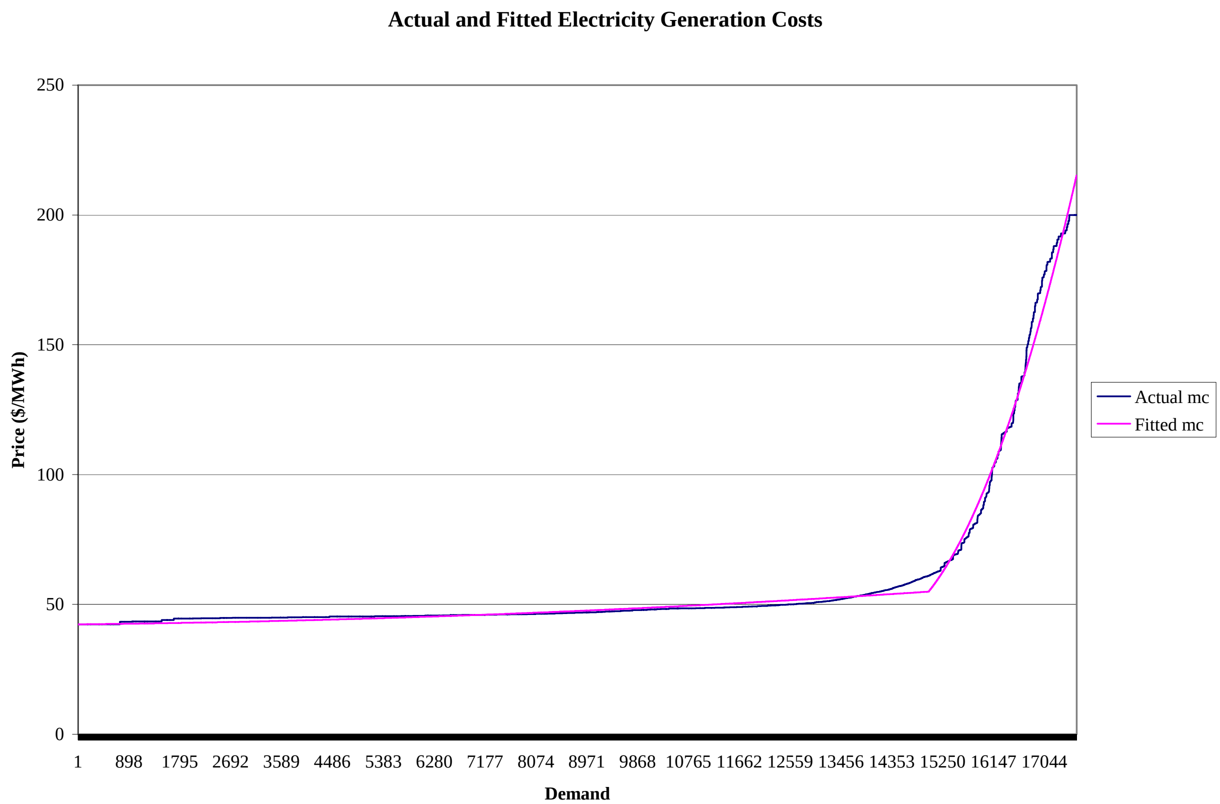

Figure 1. Actual and fitted marginal electricity generation costs. The figure is based on actual heat rate and plant outage data for 91 oil and gas-fired California power plants.

Figure 2: Retail Price with Wholesale Trading, as Function of Demand Correlation and Number of Markets. Results are based on the actual probability distribution of hourly California, New York, and PJM power demand during calendar year 2000, as well as bootstrapped hourly demand for up to 10 additional markets having the indicated correlation with observed demand. Production costs are based on heat and outage rates for 91 California power plants and assume that 66.6% of production is from base load at a constant marginal cost.

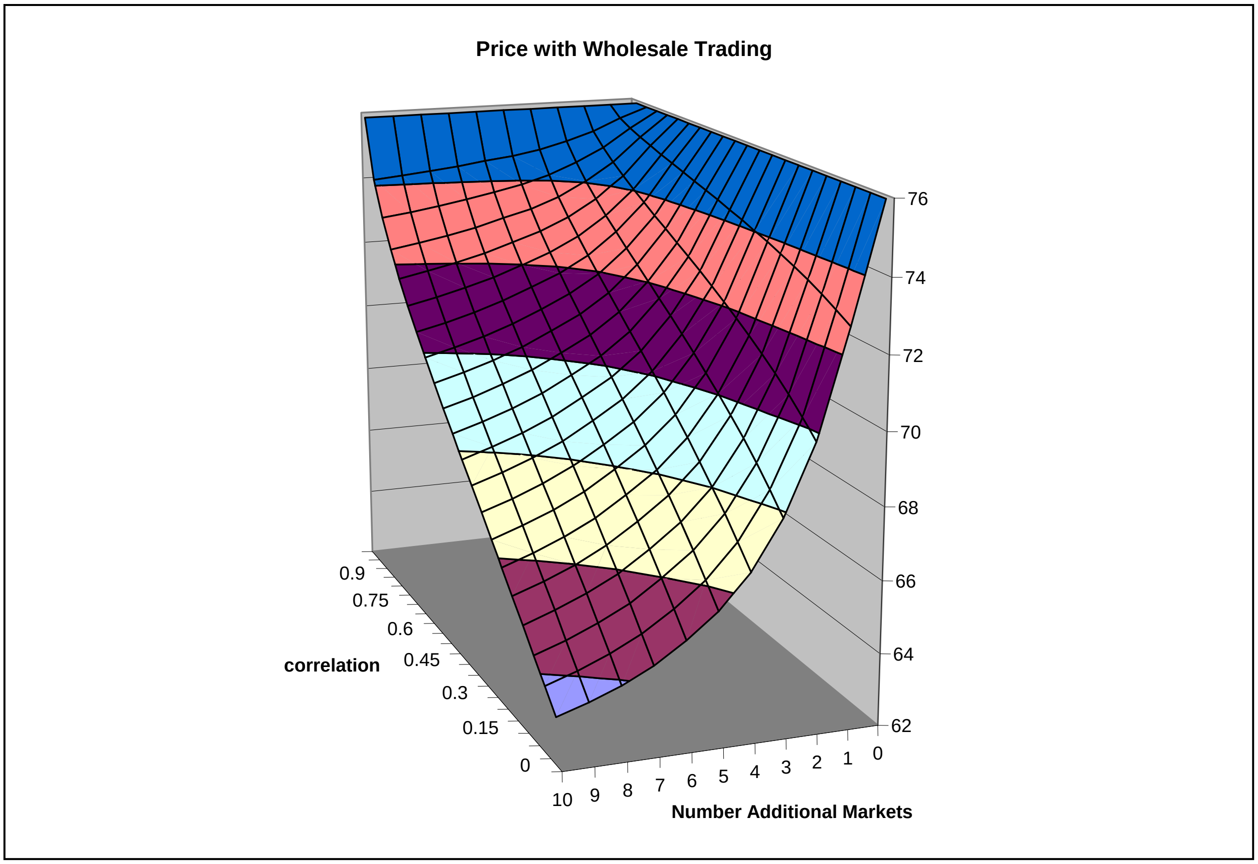

Figure 3: Percentage Reduction in Retail Price due to Wholesale Trading, as Function of Demand Correlation and Number of Markets. Results are based on the actual probability distribution of hourly California, New York, and PJM power demand during calendar year 2000, as well as bootstrapped hourly demand for up to 10 additional markets having the indicated correlation with observed demand. Production costs are based on heat and outage rates for 91 California power plants and assume that 66.6% of production is from base load at a constant marginal cost.

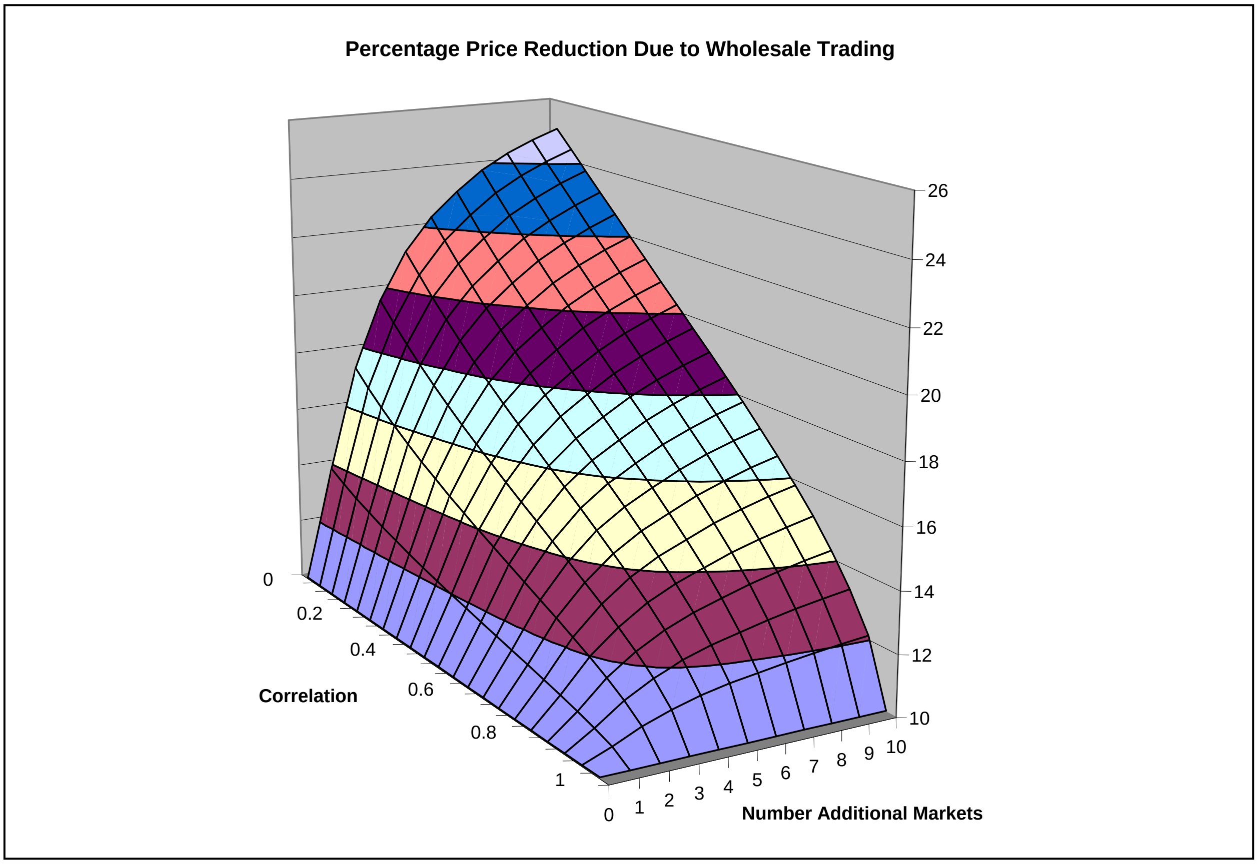

Figure 4: Percentage Reduction in Retail Price due to Wholesale Trading, as Function of Demand Correlation and Base Load Capacity. Results are based on the actual probability distribution of hourly California, New York, and PJM power demand during calendar year 2000, as well as bootstrapped hourly demand for 10 additional markets having the indicated correlation with observed demand. Production costs are based on heat and outage rates for 91 California power plants. Base load capacity represents the portion of power demand served at a constant marginal cost. Given the empirical demand distribution base load capacity of 18,000 MW equates to one-third of production from peaking plants and base load capacity of 22,000 MW equates to one-fifth of production from peaking plants.

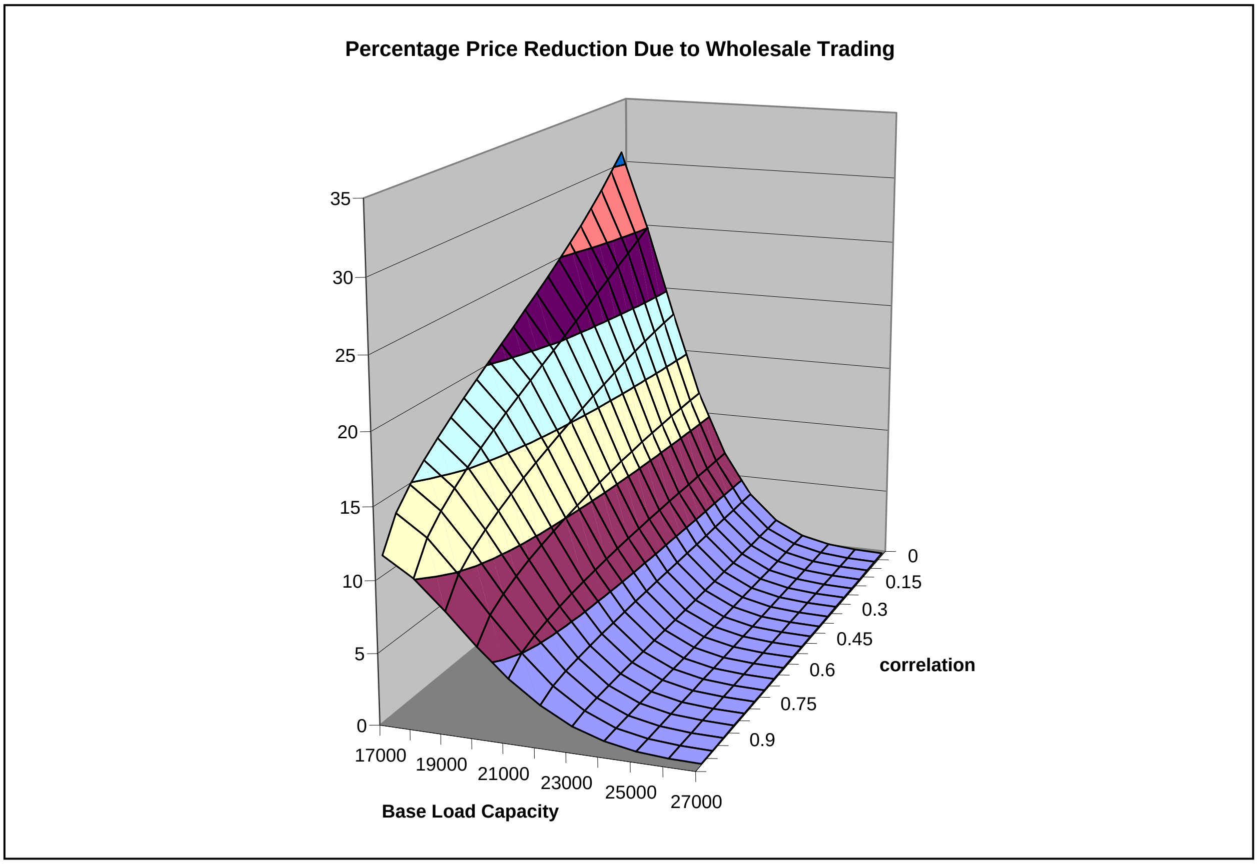

Figure 5: Percentage Reduction in Retail Price due to Wholesale Trading, as Function of Mean Local and System Demand. Local demand is based on the actual probability distribution of hourly California demand, while system demand is based on actual California, New York, and PJM power demand, plus bootstrapped hourly demand for 10 additional markets having a correlation of 0.8 with observed demand. Each demand outcome is increased or decreased by the indicated percentage of mean demand. Production costs are based on heat and outage rates for 91 California power plants. Base load capacity is set at 22,000 MW, which equates to one-fifth of production from peaking plants.

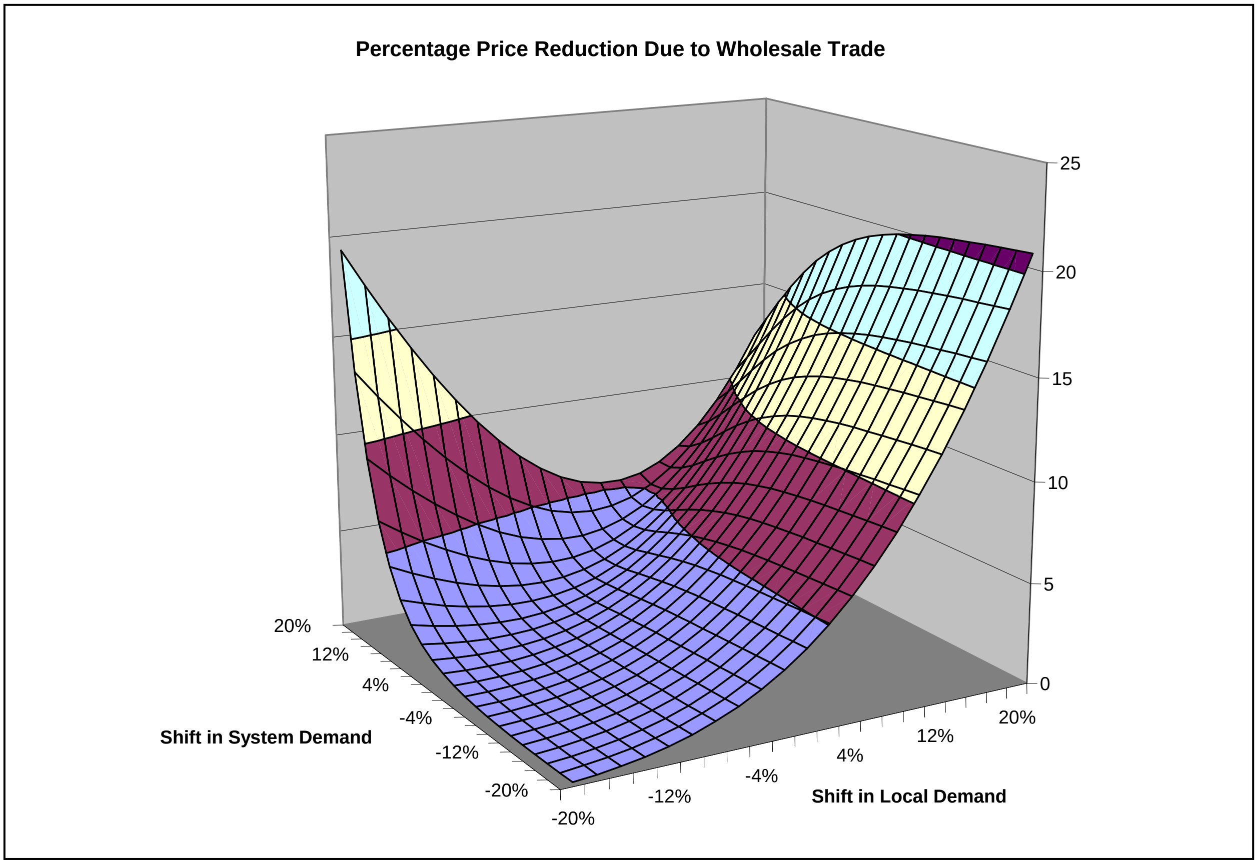

Table 1: Mean Demand and Demand Correlations. Panel A reports on actual hourly load data in megawatts during calendar year 2000, while Panel B reports correlations in actual hourly load. Average represents mean hourly load across the three indicated markets. Panel C reports on scheduled hourly loads during calendar year 2000, for four designated delivery zones within California. Panel D reports correlations among the scheduled California load series. The PJM market includes all or parts of Delaware, Illinois, Maryland, New Jersey, Ohio, Pennsylvania, Virginia, West Virginia, and the District of Columbia. The California delivery zones are NP15 (North of Path 15), SP15 (South of Path 15), SF (a portion of the San Francisco area), and ZP 26 (Zonal Path 26, an area of west-central California north of Santa Barbara and south of Monterey).

Panel A: Actual Hourly Load (MW)				
Market	**Mean**	**Minimum**	**Maximum**	**Standard Dev.**
California	27,188	17,536	43,509	5,193
New York	17,833	11,577	28,138	3,106
PJM	30,114	18,208	49,462	5,529
Average	25,045	16,468	39,996	4,332
Panel B: Correlations in Actual Hourly Load				
	California	**New York**	**PJM**	**Average**
California	1.000	0.828	0.735	0.910
New York		1.000	0.955	0.976
PJM			1.000	0.947
Panel C: Scheduled Hourly Load in California Delivery Zones (MW)				
Zone	**Mean**	**Minimum**	**Maximum**	**Standard Dev.**
SP15	13,859	9,118	19,941	2,365
NP15	9,227	3,585	16,129	1,807
ZP26	1,492	409	2,378	363
SF	123	43	573	49
Panel D: Correlations in Scheduled Hourly Load				
	SP15	**NP15**	**ZP26**	**SF**
SP15	1.000	0.548	0.625	0.563
NP15		1.000	0.280	0.232
ZP26			1.000	0.433

Table 2: Monthly Average Demand Relative to Annual Average Demand. Each cell in the Table reports average hourly electricity demand for the indicated month relative to average hourly demand for the full year in the same market. The PJM market includes all or parts of Delaware, Illinois, Maryland, New Jersey, Ohio, Pennsylvania, Virginia, West Virginia, and the District of Columbia. The California delivery zones are NP15 (North of Path 15), SP15 (South of Path 15), SF (a portion of the San Francisco area), and ZP 26 (Zonal Path 26, an area of west-central California north of Santa Barbara and south of Monterey).

Month	Actual Hourly Demand			Scheduled Hourly Demand			
	California	PJM	NY	SP15	NP15	ZP26	SF
Jan	0.94	1.07	1.04	1.03	1.07	1.00	1.08
Feb	0.94	1.03	1.01	0.93	0.96	0.94	1.10
Mar	0.94	0.93	0.94	0.93	0.88	1.01	1.03
Apr	0.93	0.88	0.91	0.95	0.90	1.16	0.91
May	0.99	0.92	0.93	1.00	0.89	1.14	1.14
Jun	1.10	1.08	1.06	1.07	1.13	1.13	1.29
Jul	1.08	1.06	1.05	1.08	1.08	1.22	1.11
Aug	1.14	1.09	1.09	1.12	1.01	1.19	1.08
Sep	1.05	0.99	1.01	1.06	1.04	1.01	0.93
Oct	0.96	0.90	0.93	0.97	1.12	0.77	0.83
Nov	0.95	0.96	0.97	0.95	1.01	0.75	0.77
Dec	0.96	1.10	1.05	0.90	0.91	0.68	0.72