American Economic Review: Papers & Proceedings 2011, 101:3, 247–252
http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.3.247

Measuring the Benefits of Greater Spatial Granularity in Short-Term Pricing in Wholesale Electricity Markets

By Frank A. Wolak*

This paper quantifies the economic benefits associated with the introduction of greater spatial granularity in short-term pricing in the California wholesale electricity market. On April 1, 2009, California switched to a nodal-pricing market with an integrated day-ahead forward market and real-time imbalance market from its original zonal market design with bilateral day-ahead scheduling, a real-time imbalance market, and an intrazonal congestion management process. Semiparametric conditional mean functions that control for observable differences in hourly system conditions are estimated for three hourly market performance measures from January 1, 2008 to June 30, 2010 to quantify the change in the conditional means of these three measures before versus after the implementation of the nodal-pricing market.

The market performance measures focus on quantifying the economic benefits associated with the operation of the 258 natural gas-fired generation units located in the California Independent System Operator (ISO) control area.[1] All but a very small fraction of the fossil fuel–fired generation capacity in California consumes natural gas. Approximately 60 percent of the installed generation capacity in the California ISO control area is natural gas–fired. The remaining capacity is nuclear, hydroelectric, wind, or solar powered. These generation facilities have substantially lower variable costs of production than natural gas units and can therefore be expected to operate if they are available under either market design. The hourly market performance measures are: (i) total amount of energy consumed by natural gas–fired generation units; (ii) total number of starts by natural gas–fired generation units; and (iii) total hourly variable costs for natural gas-fired generation units.

After controlling for the hourly output of all natural gas–fired generation units and daily prices of natural gas in Northern and Southern California, the conditional mean of the total amount of energy consumed each hour (including the energy consumed to start up generation units in that hour) and the conditional mean of the total variable costs (including the costs to start up generation units in that hour) for all natural gas–fired units are both precisely estimated to be lower after the introduction of nodal pricing. The conditional mean of the total amount of energy consumed each hour is 2.5 percent lower, and the conditional mean of the total variable cost is 2.1 percent lower. Estimating a separate hour-of-the-day coefficient for the conditional mean difference finds the largest percentage reductions for total energy consumed and total variable costs in the morning hours of the day, between 7 and 9 a.m. For total hourly starts, the conditional mean is precisely estimated to be 0.18 starts higher after the introduction of the nodal pricing. Separate hour-of-the-day coefficient estimates reveal no difference in the conditional mean number of starts before versus after the introduction of nodal pricing for most of the hours of the day except for the hours between 6 a.m. and 8 a.m.

These empirical results argue in favor of existing zonal short-term markets adopting nodal pricing, because the switch to nodal pricing in California appears to have resulted in net economic benefits in the form of less energy and variable costs being required to produce the same level of total hourly output from all natural gas–fired generation units in the California ISO control area. Less granular spatial-pricing markets existed during the early stages of electricity industry restructuring in the United States. All US markets have now adopted nodal-pricing designs, with Texas being the last to do so in 2010.

* Program on Energy and Sustainable Development and Department of Economics, Stanford University, Stanford, CA 94305, and NBER (e-mail: wolak@stanford.edu). EunYi Chung and Laurence Wong provided outstanding research assistance.

[1] Electricity generating plants are typically composed of a number of generating units. For example, a 1,500 megawatt (MW) facility may be composed of three 500 MW units.

I. Nodal versus Zonal Pricing

Short-term wholesale electricity markets differ from markets for other products because the electricity produced by a generation unit at one location and sold to a customer at another location is not actually delivered to that location in the same sense that an automobile produced in Detroit is delivered to the customer that purchased it in San Francisco. Energy injected into the transmission network flows according to Kirchhoff's laws, rather than from the seller to the buyer. The capacity of the transmission network often limits the amount of energy generation units at certain locations can inject and the amount that consumers at certain locations can withdraw. This circumstance is referred to as transmission congestion, and it can cause the market to become segmented, meaning that some generation units cannot compete to sell energy at certain locations because of the configuration of the transmission network, the locations and outputs of other generation units, and the locations and levels of final demand. Under these circumstances, a market mechanism that assumes that all generation units in the geographic region covered by the wholesale market can compete to sell energy anywhere in that geographic region will likely produce an infeasible dispatch of the available generation units, because capacity constraints in the transmission network and other operating constraints prevent the suppliers that offer the lowest prices for their output from selling all of their available energy.

Wholesale electricity markets that do not respect the configuration of the transmission network in the dispatch and pricing process must have mechanisms that alter the results of the market to achieve a physically feasible dispatch. The former California market provides one example of this sort of mechanism. All market participants were required to submit balanced schedules in the day-ahead scheduling process that were physically feasible within California's three-zone market, but not necessarily physically feasible for the entire transmission network. A balanced schedule meant that the total amount of energy scheduled from generation units under the control of that market participant equaled the amount of load scheduled by that market participant. Physical feasibility for the zonal market meant that the schedules submitted by all market participants did not violate any of the transmission capacity constraints across the boundaries of the three zones in California. Following the close of the day-ahead market and before real-time operation, the California ISO moved generation units up and down relative to their day-ahead schedules along the offer curves they submitted to the real-time market to achieve a physically feasible dispatch that respected all transmission and other relevant operating constraints. Originally, a supplier that had to move up relative to its day-ahead schedule would be paid as offered for the additional energy and a supplier that had to move down would buy back this energy as offered. Suppliers quickly figured out when their generation units were likely to be called "out of merit order" in a zone to provide more or less energy and would alter their supply offers to take advantage of that fact. This led to the implementation of mechanisms that paid or charged these suppliers regulated prices to manage what was called intrazonal congestion.

It is important to emphasize that if a transmission network within a zone has sufficient capacity to ensure that output levels that are physically feasible at the zonal level are also physically feasible for the entire transmission network, then ignoring the intrazonal details of the transmission network in the day-ahead market may introduce few, if any, market inefficiencies and have the benefit of a significantly simplified market mechanism. This may explain why one-zone or multiple-zone-pricing markets are common outside of the United States, where the electricity industry often started the restructuring process composed of state-owned vertically-integrated entities with transmission networks that were built to eliminate or minimize the incidence of transmission congestion. In contrast, at the start of the restructuring process in the United States in the late 1990s, investor-owned, vertically integrated utilities subject to more than 60 years of state-level regulation of retail prices had transmission networks that had not seen significant capacity additions since the 1970s. In spite of this, the markets in New England (ISO-NE), New York (NYISO), and Texas (ERCOT) also began operation with zonal pricing. ISO-NE had a one-zone market, and NYISO and ERCOT had multiple zones. Increasing reliance on "intrazonal congestion" management processes to achieve a feasible dispatch was a significant contributing factor to these markets adopting nodal pricing.

Under a nodal-pricing market, all relevant operating constraints are modeled to determine the output levels for all generation units and the prices that they receive. This implies that if system conditions remain the same between the close of the day-ahead market and real-time system operation, the results of the day-ahead scheduling process are physically feasible. All US nodal-pricing markets allow suppliers to submit two-part offers composed of a start-up cost offer and energy offer curve, instead of an energy offer curve, as is typically the case for zonal markets. Consequently, another argument in favor of nodal pricing is that start-up costs are explicitly taken into account in the decision to turn on generation units for the day.

By definition, nodal pricing is the least-cost approach to operating an electricity market under the assumption that all suppliers submit their minimum variable cost curve as their offer curve for energy. The nodal price for each location is the shadow price associated with withdrawing one more unit of energy at that location, or equivalently, the change in the minimized as-offered total cost of serving load at all locations in the transmission network associated with withdrawing one more MWh of energy at that location.

A number of studies—Catherine D. Wolfram (1998) for the United Kingdom, Frank A. Wolak (2000) for Australia, Wolak (2003) for California—have shown that large suppliers in wholesale electricity markets have a substantial ability (which is often accompanied by a strong incentive) to exercise unilateral market power which causes them to submit offer curves that are higher than their minimum variable cost curve. Severin Borenstein, James B. Bushnell, and Wolak (2002) demonstrate that this exercise of unilateral market power can lead to significant deviations from least-cost system operation. Consequently, whether nodal pricing yields a more efficient dispatch in terms of total energy consumed or total operating costs is an empirical question. It depends on the extent to which the opportunities for suppliers to exercise unilateral market power are limited by the zonal market design more than enough to compensate for deviations from least cost system operation caused by not dispatching generation units to minimize the as-bid cost of operating the system subject to all relevant operating constraints.

II. Empirical Framework and Implementation Details

This section describes the data used to compute the hourly market performance measures and the econometric modeling framework used to assess the differences in the conditional means of the three market performance measures before versus after the implementation of nodal pricing in California. To control for the hourly differences in observable system conditions as flexibly as possible, Peter M. Robinson's (1988) semiparametric partially linear model is used to estimate the conditional mean function for each market performance measure.

The total hourly amount of energy consumed by all natural gas–fired generation units includes both the energy consumed to produce electricity during that hour and the energy used to start units during that hour. To calculate the first component, the hourly metered output of each natural gas–fired generation unit is obtained from the California ISO's settlement system. This information is combined with the generation unit–level heat rate curve that all natural gas–fired generation unit owners are required to submit as part of the California ISO's local market power mitigation mechanism. This curve is a piecewise linear function that can have up to ten heat rate level and output quantity pairs up to the full capacity of the generation unit. The vertical axis gives the heat rate denominated in millions of British Thermal Units (MMBTUs) of natural gas burned to produce each additional MWh for the level of output from that generation unit on the horizontal axis. The heat rate value on this piecewise linear curve times the generation unit's metered output for that hour is the first component of the total amount of energy consumed by that generation unit during that hour.

The second component of hourly total energy consumed is the total energy used to start up any generation units turned on during the hour. A unit is defined as starting in hour t if its output in hour $t - 1$ is zero and its output in hour t is positive. There are three types of starts, each of which can require a different amount of energy. A hot start is preceded by less than or equal to two hours of zero output. A warm start is preceded by more than two hours and less than six hours of zero output. A cold start is preceded by more than six hours of zero output.

Natural gas–fired generation unit owners are required to file information with the California ISO on the total amount of energy required for each type of start for each of their units. For each start that occurred in an hour of the sample, the type of start was determined and the amount of energy appropriate for that type of start was added to the energy consumed during that hour to produce that unit's metered output level. Summing the total MMBTUs of energy consumed in that hour to start all units turned on during that hour and the total MMBTUs of energy required to produce electricity during the hour over all units producing during the hour yields $TOTAL\_ENERGY(t)$.

The total number of generation units started in hour t, $STARTS(t)$, is the total number of units in hour t that have zero metered output in hour $t - 1$ and positive output in hour t. The final market performance measure, $TOTAL\_VC(t)$, the total variable costs of all natural gas–fired generation units in hour t is calculated as follows. The marginal cost for each generation unit is computed by multiplying the heat rate associated with the unit's metered output for that hour times the daily price of natural gas for that unit plus the variable operating and maintenance costs that the unit's owner submits to the California ISO for its local market power mitigation mechanism. Information on daily natural gas prices for all generation units is also collected as part of the California ISO's market monitoring process. The total variable cost for the unit is computed as the product of its marginal cost and metered output for the hour. For units that start up in hour t, the total energy to start the unit is converted to start-up costs by multiplying the MMBTUs of energy consumed by the unit's daily price of natural gas. Summing the variable operating costs over all generation units operating in hour t and the start-up costs for all units starting in hour t yields $TOTAL\_VC(t)$.

The conditional mean functions estimated can all be written as $y_t = W_t'\alpha + X_t'\beta + \theta(Z_t) + \varepsilon_t$, with $E(\varepsilon_t | X_t, W_t, Z_t) = 0$, for different definitions of y_t. The function $\theta(Z)$ is an unknown function of the vector Z, W is a vector of 23 hour-of-day and 11 month-of-the-year dummy variables, Z_t is a three-dimensional vector composed of the natural logarithm of total hourly output in MWh of all natural gas–fired generation units, the natural logarithm of the daily price of natural gas in northern California in \$/MMBTU, and the natural logarithm of the daily natural gas price in southern California in \$/MMBTU, and α and β are unknown parameter vectors. For the total energy conditional mean function, y_t equals the natural logarithm of $TOTAL\_ENERGY(t)$, X_t is a single dummy variable that takes on the value 1 for all hours after midnight March 31, 2009 and zero otherwise. For the total starts conditional mean function, y_t equals $STARTS(t)$, and for the total variable costs conditional mean function, y_t equals the natural logarithm of $TOTAL\_VC(t)$, and X_t and Z_t are as defined above. For the models that estimate 24 hour-of-day coefficients for the nodal pricing period, X_t is a (24×1) vector with kth element X_{tk}, which equals one during hour-of-the-day k for all days from April 1, 2009 until the end of the sample.

The two-step estimation procedure first subtracts $E(y_t | Z_t) = E(W_t' | Z_t)\alpha + E(X_t' | Z_t)\beta + \theta(Z_t)$ from the conditional mean function to yield $(y_t - E(y_t | Z_t)) = (W_t' - E(W_t' | Z_t))\alpha + (X_t' - E(X_t' | Z_t))\beta + \varepsilon_t$. Each of the conditional expectations is estimated using a kernel regression, $E_h(y | z) = \left[\sum_{s=1}^{T} y_s K((z - Z_s)/h)\right] / \left[\sum_{s=1}^{T} K((z - Z_s)/h)\right]$, with Gaussian kernel $K(t)$, and smoothing parameter h. The value of h is chosen by minimizing $\sum_{t=1}^{T} [(y_t - E_h^{-t}(y_t | Z_t)) - (W_t' - E_h^{-t}(W_t' | Z_t))\alpha - (X_t' - E_h^{-t}(X_t' | Z_t))\beta]^2$ with respect to h, α, and β, where $E_h^{-t}(y | z)$ equals $E_h(y | z)$ with the tth observation excluded from the sums in the numerator and denominator. Applying ordinary least squares to the transformed model with h fixed at the optimized value yields root-T-consistent and asymptotically normal estimates of α and β. Standard error estimates are constructed using the expression given in Robinson (1988).

III. Empirical Results

Table 1 reports the results of estimating the conditional mean function for each market performance measure when X_t is a single dummy variable that equals 1 for all hours with nodal pricing and zero otherwise. These estimates imply that the conditional mean of total hourly energy is 2.5 percent lower after March 31, 2009. The conditional mean of total hourly starts is 0.18 starts higher and the conditional mean of total variable costs is 2.1 percent lower after March 31, 2009.

Figures 1 to 3 plot the estimates of hour-of-the-day change in the conditional mean of the

Table 1—Change in Hourly Conditional Mean After Nodal-Pricing Market Implemented

Dependent variable	$\ln(TOTAL\_ENERGY(t))$	$STARTS(t)$	$\ln(TOTAL\_VC(t))$
β	−0.0246	0.178	−0.0207
Standard error	0.0020	0.056	0.0020

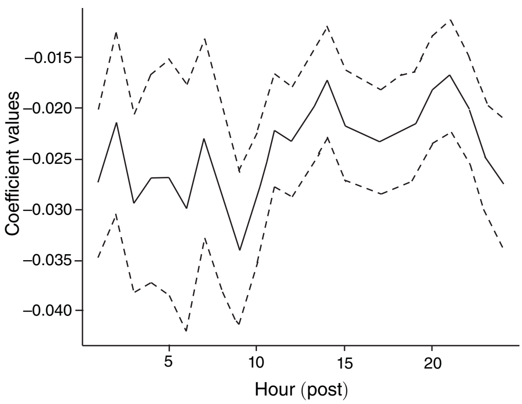

Figure 1. Hour-of-the-Day Percent Change Estimates for Hourly Total Energy

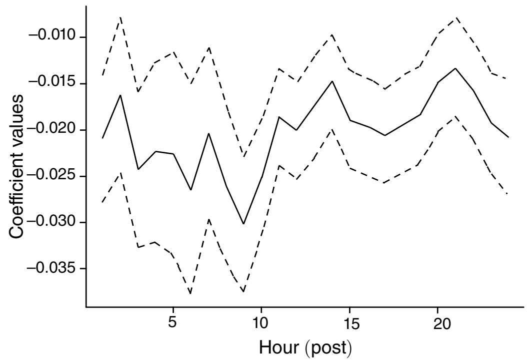

Figure 3. Hour-of-the-Day Percent Change Estimates for Hourly Total Variable Cost

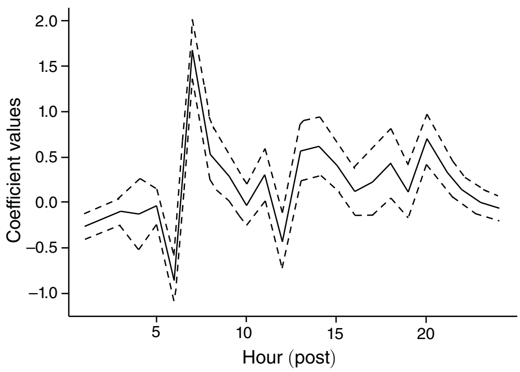

Figure 2. Hour-of-the-Day Change Estimates for Hourly Total Starts

three hourly market performance measures after the implementation of nodal pricing along with the pointwise upper and lower 95 percent confidence intervals for each hour-of-the-day estimate. For the case of total hourly energy, the largest in absolute value reduction is in the morning hours from 7 AM to 9 AM, with remaining hours slightly smaller in absolute value. For total starts, the largest changes are concentrated in the hours from 6 AM to 8 PM, with most of the remaining hours of the day having point estimates that are not statistically different from zero. For total variable costs, the pattern of the absolute value of the hour-of-the-day reductions is similar to that for total hourly energy; the largest in absolute value reductions occur in morning hours from 7 AM to 9 AM.

IV. Conclusions

Multiplying the 2.1 percent average cost reduction by the average hourly value of $TOTAL\_VC(t)$ yields approximately \$12,000 less in total variable costs per hour, which translates into approximately \$105 million in estimated annual cost reductions from the introduction of nodal pricing. Because this analysis focused on the supply side of the market, these results do not address the question of whether California consumers realized any of these benefits. However, these results do indicate that there is a substantial source of net economic benefits that California electricity consumers could access as a result of the introduction of nodal pricing.

REFERENCES

Borenstein, Severin, James B. Bushnell, and Frank A. Wolak. 2002. "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market." *American Economic Review*, 92(5): 1376–1405.

Robinson, Peter M. 1988. "Root- N-Consistent Semiparametric Regression." *Econometrica*, 56(4): 931–54.

Wolak, Frank A. 2000. "An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market." *International Economic Journal*, 14(2): 1–39.

Wolak, Frank A. 2003. "Measuring Unilateral Market Power in Wholesale Electricity Markets: The California Market, 1998–2000." *American Economic Review*, 93(2): 425–30.

Wolfram, Catherine D. 1999. "Measuring Duopoly Power in the British Electricity Spot Market." *American Economic Review*, 89(4): 805–26.