ON THE EXERCISE OF MARKET POWER THROUGH STRATEGIC WITHHOLDING IN CALIFORNIA

Scott M. Harvey* and William W. Hogan**

*LECG, LLC.
Cambridge, MA

**Center for Business and Government
John F. Kennedy School of Government
Harvard University
Cambridge, MA

April 24, 2001

TABLE OF CONTENTS

Page

ON THE EXERCISE OF MARKET POWER THROUGH STRATEGIC WITHHOLDING IN CALIFORNIA

Scott Harvey and William W. Hogan

EXECUTIVE SUMMARY

Beginning in June of 2000, the shock of unexpectedly high prices in the California electricity market convinced everyone of the need for policies to correct the apparent market failures. The public debate and policy discussions have been dominated by a focus on market power as a principal problem amenable to regulatory solution. However, design of effective policies to moderate prices or mitigate their effects depends on the diagnosis of the underlying causes. High prices attributable largely to an exercise of market power in electric generation would point to particular market participants and behaviors that could be targeted for regulatory action. By contrast, high prices attributable to bad electricity market design would indicate a need for changes in the design. High prices attributable to higher fuel prices, environmental constraints and capacity shortages, on the other hand, would prompt actions to address the cost of fuel and environmental limitations and indicate that retail loads should receive the appropriate price signal for conservation.

Suppliers could affect market prices by strategically withholding some capacity in order to profit on the capacity actually sold in the market. But charging high prices during periods of scarcity is not classified as exercising market power if there is no strategic withholding of supply. Likewise, refusing to supply without being paid is not an exercise of market power. Although the potential for withholding exists for many suppliers, the focus of attention has been on the exercise of market power by thermal generators in California.

On its face, the experience of extremely high prices suggests that the exercise of market power could be important. But at the same time the data show that there have been profound changes in the California market such that the thermal generators have actually increased their production more than demand has grown. If anything, thermal generators that hit annual output limits produced too much rather than too little in the summer of 2000. Furthermore, the widespread impacts of higher electricity prices throughout the western market, both on and off peak, indicate that if the exercise of market power is important it is occurring to an extent and through channels unprecedented in this or other electricity markets. In short, this is a complicated story, and there is ample room for further investigation of the data and diagnosis of causes.

Examination of the major analyses of the exercise of market power reveals that the estimated magnitude of the possible strategic withholding of electric generation is small enough to make it important to verify the simplifying assumptions. If strategic withholding were large and pervasive, then the real details of the California electricity market could be ignored. But it is by now apparent that the evidence is not clear, and any finding of the presence or absence of strategic withholding of generation in the California electricity market could turn on the simplifying assumptions used in the analysis of the data. For example, annual limits on production dictate that plants should not run in many hours when prices are higher than direct incremental costs; hence, examinations of output decisions for individual hours or months are necessarily incomplete. The variation in real time conditions is large enough to produce

significant reductions in output compared to the expectations given day-ahead prices; hence, with capacity constraints average optimal production is necessarily less than optimal production at average prices. Limits on the ramping rate of generation units, start-up costs, minimum load costs and other operational inflexibilities imply that a dispatch day is not just twenty-four separate hours and must be analyzed chronologically, recognizing these factors. And so on. Accounting for such effects can reverse the implications of the previous evidence. Unfortunately, the real details are neither simple nor incidental.

It is difficult to conduct a study of market power based solely on publicly available data. A fuller analysis would require data available only to the California Independent System Operator, and has not been done. Many factors contributed to higher electricity prices in California, and the market power theme is only, at most, part of the story. The import of the previous analyses is not to prove that market power has been exercised in the California electricity market but, rather, to suggest that it might be important. The import of the sensitivity analysis here is not to prove that market power has not been exercised in the electricity market but, rather, to suggest that it is unlikely to be the dominant factor and may not even be significant. With the available data in the public domain, and the special complications introduced by the California market design, the margin of error in estimating the extent of the possible exercise of market power through strategic withholding of electric generation is of the same order of magnitude as the effect being measured. On balance, to date the publicly available data provides no reason for the Federal Energy Regulatory Commission to change its conclusion that there is no evidence of strategic withholding nor any proof that no strategic withholding has occurred.

By contrast, there is general agreement that the California electricity market design is "seriously flawed." Furthermore, there is evidence that the policy responses that have been adopted in California have accelerated an already serious market collapse. Hence, without dismissing the possibility of the exercise of market power, the principal policy focus should be on fashioning workable solutions for the other more serious problems in market design that relate to the underlying causes of the market meltdown.

Separate from market power mitigation, California should pay its bills, raise incremental prices to retail customers, and move as quickly as possible to operating a coordinated and efficient market with consistent pricing for all that includes unit commitment, day-ahead scheduling, and real-time balancing. Although not a panacea, these steps would address immediate problems and set the stage for longer-term initiatives to expand generation capacity, transmission infrastructure, and the reach of an efficient market to the western interconnected grid.

ON THE EXERCISE OF MARKET POWER THROUGH STRATEGIC WITHHOLDING IN CALIFORNIA

Scott Harvey and William W. Hogan[1]

April 24, 2001

I. OVERVIEW

The continuing high prices in the California electricity market have elicited a sustained interest in understanding the reasons for the high prices. Design of effective policies to moderate prices or mitigate their effects depends on the diagnosis of the causes. For example, high prices attributable to bad market design would indicate a need for changes in the design. High prices attributable largely to an exercise of market power in electric generation would point to particular market participants and behaviors that could be targeted for regulatory action. High prices attributable to higher fuel prices, environmental constraints and capacity shortages, on the other hand, would prompt actions to address the cost of fuel and environmental limitations and indicate that retail loads should receive the appropriate price signal for conservation.

Suppliers could affect market prices by strategically withholding some capacity in order to profit on the capacity actually sold in the market. Hence, market power is defined as the ability to withhold production on some units in order to increase market prices and profit more from production on other units. By contrast, simply charging high prices during periods of scarcity is not classified as exercising market power if there is no strategic withholding of supply.[2] Likewise, refusing to supply without being paid is not an exercise of market power.

---

[1] Scott Harvey is a Director with LECG, LLC, an economic and management consulting company. William Hogan is the Lucius N. Littauer Professor of Public Policy and Administration at the John F. Kennedy School of Government, Harvard University. This study was supported by Mirant. The authors are or have been consultants on electricity market design and transmission pricing, market power or generation valuation issues for American National Power; Brazil Power Exchange Administrator (ASMAE); British National Grid Company; Calpine Corporation; Commonwealth Edison; Constellation Power Source; Detroit Edison; Duquesne Light Company; Electricity Corporation of New Zealand; General Electric Capital; GPU, Inc. (and the Supporting Companies of PJM); GPU PowerNet Pty Ltd.; ISO New England; Midwest ISO; National Independent Energy Producers; New England Power; New York Energy Association; New York ISO; New York Power Pool; New York Utilities Collaborative; Niagara Mohawk Corporation; Ontario IMO; PJM Office of the Interconnection; Reliant Energy; San Diego Gas & Electric; Sempra Energy; Mirant/Southern Energy; TransÉnergie; Transpower of New Zealand Ltd.; Westbrook Power; Williams Energy Group; and Wisconsin Electric Power Company. Jasper Camacho, John Jankowski, Matthew Kunkle, Ana Palomares, Joel Sachar and Jennifer Vanderhart provided research assistance. The views presented here are not necessarily attributable to any of those mentioned, and any errors are solely the responsibility of the authors.

[2] Another complication is the possibility of exercising market power by producing too much in order to create system bottlenecks. This is not addressed here. See for instance, Judith B. Cardell, Carrie Cullen Hitt and

It is widely recognized that the potential for exercising market power through strategic withholding exists in electricity markets in many regions of the country, and California is no exception. As a result, policies are already in place to mitigate market power, and market monitoring to detect the exercise of market power is a major and growing activity in restructured electricity markets. For example, in California a number of generating plants have been under Reliability Must Run (RMR) contracts in part to mitigate the exercise of market power. Both the California Independent System Operator (CAISO) and Power Exchange (PX) have had independent external monitoring committees and internal staff devoted to the analysis of market performance.

Of course, no market can be perfectly competitive or perfectly free from some exercise of market power. But there is a legitimate policy concern if strategic behavior by market participants becomes pervasive and significant. With prices exploding starting in about June of 2000, the concern in California has been that market power beyond that resident in owners of RMR plants was both available and exploited to a degree that explained all or a large part of the price increase and apparent supply shortage in the latter half of 2000. Although the potential for withholding exists for many suppliers, the focus of attention has been on the exercise of market power by thermal generators in California. If there has been significant strategic withholding by anyone, including thermal generators, the analysis should identify the responsible parties and the appropriate remedies would be clearer.

On its face, the experience of extremely high prices suggests that the exercise of market power could be important. But at the same time the data show that there have been profound changes in the California market such that the thermal generators have actually increased their production more than demand has grown. Furthermore, the widespread impacts of higher prices throughout the western market, both on and off peak, suggest that if the exercise of market power is important it is occurring to an extent and through channels unprecedented in this or other electricity markets. In short, this is a complicated story, and there is ample room for further investigation of the data and diagnosis of causes.

It is important to understand the reason for the high electricity prices in California, both for developing policies in California and so that the rest of the nation can avoid similar outcomes. It is therefore important to identify the exercise of market power, if it has been an important contributor to these high prices. At the same time, it is important to critically examine the implication that exercise of market power is the dominant explanation for outcomes in California. A conviction that the major problems in the California market arise largely from the exercise of market power in electricity generation could distract from consideration of more fundamental problems. However, if the principal sources of high prices in California lie elsewhere, then a policy preoccupation with market power will lead to choices that exacerbate and prolong the period of high wholesale market prices in California or a broader region.

From this perspective, an important public policy question arises in understanding and testing the data and analyses directed at the estimation of the extent and importance of the exercise of

---

William W. Hogan, "Market Power and Strategic Interaction in Electricity Networks," Resource and Energy Economics., Vol. 19, 1997, pp. 109-137 market power. If the data indicate that the exercise of market power is a substantial problem in California, then policies to mitigate the exercise of market power in electric generation should be a priority, and may be all that is needed. If, however, the evidence points to the exercise of market power as a secondary concern compared to perhaps market design, environmental limits, gas and generation supply, then the focus of public policy should be different.

The flow of market power analyses is increasing, and the debate is unsettled. The principal conclusion here is that with the available data in the public domain, and the special complications introduced by the California market design, the margin of error in estimating possible exercise of market power in electric generation through strategic withholding is of the same order of magnitude as the effect being measured.[3] In this regard, it is unlikely that the exercise of market power is extensive, easy to detect, or easy to correct. By contrast, there is general agreement that the California market design is "seriously flawed." Hence, without dismissing the possibility of the exercise of market power, the principal policy focus should be on fashioning workable solutions for the other more serious problems in market design.

II. ANALYSES OF MARKET POWER

The accumulation of market power studies is expanding at a rate that is not surprising given the importance of the shock to the California market and, as we shall see, the difficulty of untangling a complicated story. The various studies directed at assessing the role of generator market power in elevating California electricity prices during the year 2000 have reached conflicting conclusions. In some cases, indirect evidence has been found to support an interpretation that there must have been an exercise of market power through withholding; however, the very nature of such indirect analysis precludes the identification of the particular parties who were responsible.[4] In other instances the analyses found the story to be more complicated, with explanations of the data that either did not indicate the exercise of market power or found the evidence unable to support any conclusion.[5]

As time passes, the horizon for studies assessing the impact of price increases extends correspondingly.[6] As we move beyond the end of year 2000, there is a need to account for an

---

3 We have not examined the possible exercise of market power in gas supply, transmission, storage or distribution.

4 Frank Wolak, Robert Nordhaus, and Carl Shapiro, "An Analysis of the June 2000 Price Spikes in the California ISO's Energy and Ancillary Service Markets," September 6, 2000 (hereafter MSC). Eric Hildebrandt, "Declaration of Eric Hildebrandt," October 2000 (hereafter Hildebrandt Oct), and Eric Hildebrandt "Analysis of Market Power in California's Wholesale Energy Markets," November 21, 2000 (hereafter Hildebrandt Nov).

5 Staff Report to the Federal Energy Regulatory Commission on Western Markets and the Causes of the Summer 2000 Price Abnormalities, Part I, November 1, 2000 (hereafter Staff Report); Scott Harvey and William W. Hogan, "Issues in the Analysis of Market Power in California," October 27, 2000 (hereafter Harvey-Hogan); California Power Exchange Corp, Compliance Unit, "Price Movements in California Electricity Markets," September 29, 2000 (hereafter CALPX); Report on Plant Outages in the State of California, FERC Office of the General Counsel Market Oversight & Enforcement and Office of Markets, Tariffs and Rates, Division of Energy Markets, February 1, 2001 (hereafter FERC Outage Report).

6 For example, Eric Hildebrandt, "Further Analyses of the Exercise and Cost Impacts of Market Power in California's Wholesale Energy Market," California Independent System Operator, March 2001; Anjali Sheffrin,

increasing accumulation of market failures, including the introduction of credit problems and new price regulations, both of which further complicate the analysis of the market. Nonetheless, as of the end of the first quarter 2001, the more recent analyses had not resolved the uncertainty.[7] Hence, the focus of the analysis below is restricted to the year 2000, with an emphasis on the data for the Summer of 2000 that has received the most attention.

For the period covering the Summer of 2000, a paper by Paul Joskow and Edward Kahn carries the analysis further than the previous studies in an attempt to move beyond the limitations of indirect detection of an exercise of market power by some anonymous market participant to more directly identify thermal generators that might be withholding output in a strategic exercise of market power that contributed to the high wholesale market prices for electricity in California during the summer of 2000.[8] The purpose of the analysis below is to carry forward the examination of the market, further discussing the simulation studies and extending Joskow and Kahn's analysis of real-time operating data. The identification of strategic withholding in real time is an important question, and also a difficult question to address, particularly for someone lacking access to the dispatch data. Here we expand our previous discussion to identify some of the issues to be addressed in distinguishing strategic withholding from other causes of high prices, examine further some of the evidence, and suggest some of the policy implications of our current understanding of the state of the market.

It is difficult to conduct a study of market power based solely on publicly available data. On balance, however, we do not see that the evidence to date provides reason for the Federal Energy Regulatory Commission (FERC) to change its conclusion that there is no evidence of withholding nor any proof that no withholding has occurred.[9] The Joskow-Kahn simulation study improves on certain elements of the earlier MSC study that were likely to understate the competitive price level, but relies on approximations of demand that are likely to understate the competitive price relative to the prior MSC study, and the proxies for hydro and geothermal supply may have the same effect. In addition, the Joskow-Kahn simulation shares limitations with the MSC study regarding the treatment of start-up and no-load costs, non-allowance environmental restrictions, outage rates, uncertainty, and market inefficiency that are likely to cause the simulation to understate the competitive price level.

As Joskow and Kahn note, the data analyzed in their innovative study of real-time withholding does not provide evidence that can distinguish between strategic withholding and other market factors. If the data used in their analysis are adjusted for the hours in which real-time prices were low or units were likely to be ramping, their measure of an output gap is far less, rather than more, than ancillary service procurement, reversing the sign of their principal indicator of

---

"Empirical Evidence of Strategic Bidding in California ISO Real Time Market," California Independent System Operator, March 21, 2001.

7 California Independent System Operator, "Response of the California Independent System Opertor (sic) Corporation to Letter Order of March 30, 2001," Submission to Federal Energy Regulatory Commission, April 6, 2001, pp. 2-3.

8 Paul L. Joskow and Edward Kahn, "A Quantitative Analysis of Pricing Behavior in California's Electricity Market During Summer 2000," January 2001 (hereafter Joskow-Kahn).

9 Staff Report, pp. 1-4, 5-16. FERC Outage Report, pp. 1, 52.

possible strategic withholding. This measure of an output gap is overstated by an indeterminate amount because it does not account for real-time deratings, the impact of intra-zonal congestion, environmental output limits, uncertainty and market inefficiency and may be based on overstated capacities. On the other hand, the relationship between the estimated output gap and ancillary service procurement does not establish that there was no withholding because it is not known how much ancillary services were procured from fossil units not included in the real-time output data (particularly gas turbines) or from hydro or geothermal units. Moreover, a large element of uncertainty is introduced into these comparisons by a lack of information on how replacement and other reserves were actually dispatched by the system operator in real time.

The point of these comments is not that it is impossible to detect economic withholding. Most of the ambiguity arises from a lack of public information on factors that ISOs would routinely take into account in the real-time operation of the transmission system. While it will at times be difficult to distinguish between economic withholding and market inefficiency, or pricing intended to manage limited energy units or environmental constraints, most of the other uncertainties plaguing the analysis of the publicly available data should not hinder identification of withholding from ISO dispatch data.

The import of the Joskow-Kahn analysis is not to prove that market power has been exercised but, rather, to suggest that it might be important. The import of the sensitivity analysis here is not to prove that market power has not been exercised but, rather, to suggest that it is unlikely to be the dominant factor and may not even be significant. By contrast, there appears to be little disagreement that other problems of shortage and bad market design are at least large enough to dictate that the solution requires more than just market power mitigation devices.

An important feature of the Joskow-Kahn study is that it came at a time when it had become reasonably clear that a simple theory of anticompetitive withholding of thermal generating capacity from the California market is not consistent with the events of the year 2000. Since at least November, with the benefit of 20-20 hindsight, it has been apparent that some of the thermal generators suspected of withholding output in the earlier part of the year had actually supplied too much output in the earlier part of the year, causing them to reach annual run time or other environmental limits and to be shut down completely by environmental regulations or restricted to running only in emergency conditions. Generators that have been shut down by environmental regulations for running too long and too hard during this period seem not to have engaged in simple anti-competitive withholding of output. In hindsight, these units were supplying too much output, not too little, during much of the year and should have offered their output at higher prices than they did. In light of this evidence, it is not sufficient in establishing the existence of anticompetitive withholding to demonstrate that particular generators were providing less output than their design capacity in high-priced hours. We are reminded that when cumulative constraints are binding, were it not for the constraint, generators would have an interest in producing more at lower prices. A theory of anticompetitive withholding must demonstrate that thermal generators with cumulative output constraints offered too much output at too low prices in some hours and too little output at too high prices in other hours, and did so in a manner that predictably increased profits. This is a much more complex analysis than any that has been undertaken.

III. ANALYTICAL ISSUES

It is difficult in many industries to clearly distinguish price increases arising largely from the exercise of market power from price increases arising from other changes in market conditions. A number of unique features of the California electricity market, such as one-part bids, and separate non-cooptimized markets for energy, ancillary services and congestion management, greatly complicate assessment of competitive behavior in the California electricity market, making the task more difficult than usual.

Moreover, in analyzing performance in the California electricity market it is important to recognize that a demonstration that the observed outcomes are inconsistent with those in an efficient competitive market does not necessarily imply that the market is not operating competitively; the market may simply be clearing very inefficiently as a result of the market design. The California market design explicitly prevented the central market operator, the CAISO, from coordinating efficient markets for energy, ancillary services or congestion management, and it has been widely predicted that the result would be inefficient market outcomes and higher costs to consumers.[10] While the large increases in wholesale market prices are a subject of public policy concern regardless of the source of these increases, in prescribing remedies it is important to understand whether the principal cause of price increases has been the exercise of market power, poor market design, changes in market conditions or other elements of public policy. Prescribing remedies based on the presumption that the price increases arise solely or largely from the exercise of market power will only prolong the agony if the actual source of the problem lies elsewhere.

The marginal cost of electric generators in California depend on several factors, including heat rates (Btu/kWh), the cost of fuel (e.g., oil or gas), NOx allowance costs and variable O&M costs. As any of these factors increase, the marginal cost of the generator will increase.[11] At historic fuel and allowance cost levels the incremental cost of a GT of the type assumed by FERC in its proxy price analysis (18,073 Btu/kWh, 2 lbs. NOx/MWh, $2/MWh O&M) would have

---

[10] See, for example, William W. Hogan, "A Wholesale Pool Spot Market Must Be Administered by the Independent System Operator: Avoiding the Separation Fallacy," *Electricity Journal*, December 1995; Steven Stoft, "Analysis of the California WEPEX Applications to FERC," October 8, 1996; Steven Stoft, "California's ISO: Why Not Clear the Market," *Electricity Journal*, December 1996; Eric Woychik, "California's Schedule Coordinator: Market Maker with Advantage," November 26, 1997; William W. Hogan, "Rethinking WEPEX: What's Wrong with Least Cost?" *Public Utilities Fortnightly*, January 1, 1998; Steven Stoft, "Gaming Intra-Zonal Congestion in California," March 6, 1998; Larry Ruff, "Separation of the ISO from the Power Exchange: Some Structural and Operational Implications," October 25, 1995. Larry Ruff, "The California PX Auction: Whatever Happened to the ISO and Why Should Anybody Care," UC Power Conference, March 10, 1997; Richard O'Neill, "Rules and Institutions for Imperfect Markets," Harvard Electricity Policy Group, January 10, 1997; Charles R. Imbrecht, Statement before the Senate Select Committee on Business Development, November 29, 1995; and Charles R. Imbrecht, Statement before the FERC, Technical Conference on the WEPEX Applications, August 1, 1996.

[11] FERC uses these factors in setting its California proxy price, see FERC, Order Directing Sellers to Provide Refunds of Excess Amounts Charged for Certain Electric Energy Sales During January 2001 or Alternatively, to Provide Further Cost or Other Justification for Such Charges, March 9, 2001, Docket No. EL00-95-017 et al.

incremental costs in the range of $58/MWh).[12] At prices of $10/Mmbtu for gas and $40 for allowances, the incremental cost of electricity would be in the vicinity of $262/MWh.[13]

There are at least four general approaches that could be taken to address questions of withholding and market power in the California electricity market. One approach would be to test directly whether particular thermal generating capacity has been anticompetitively withheld from the day-ahead market. A second more indirect approach would be to simulate the competitive level of day-ahead prices and compare them to actual day-ahead prices. A third direct approach would be to test whether particular thermal generating capacity has been anticompetitively withheld from the real-time market. A fourth and again indirect approach would be to simulate the competitive level of real-time prices and compare them to actual real-time prices. Within these general approaches, there are additional choices in how these tests would be applied. A fundamental problem in addressing market power issues with any or all of these approaches is that the margin of error may well be larger than the magnitude of the effects that one is attempting to measure. Hence, even with the best of analysis, the policy conclusions may not be clear. Estimates of a small amount of strategic withholding may not be statistically significant, and correcting behavior that produced a small amount of withholding may not have a material effect on prices. What would be relevant from a public policy perspective would be reasonably unambiguous evidence of a large amount of strategic withholding. The difficulties are not insurmountable in principle, and the importance of the questions is sufficient to justify the effort to untangle the various contributions to the observed result of high prices.

The application of any of these approaches to the identification of the exercise of market power requires that the analysis:

- Account for fuel, O&M,[14] and NOx emission allowance costs;
- Account for environmental output restrictions other than NOx allowances, such as water temperature restrictions or annual operating hour or capacity factor restrictions;
- Account for the impact of transmission congestion (inter- and intra-zonal congestion and RMR calls in California);
- Account for capacity used to provide ancillary services (regulation, spinning reserves, 10-minute reserves, and potentially replacement reserves);
- Account for capacity not available due to forced or maintenance outages or unit deratings; and

---

12 18.073mmBtu/MW*$3/mmbtu gas + 1$/lb. NOx * 2lb./MW+ $2/MWh O&M

13 18.073mmBtu/MW*$10/mmbtu gas + 30$/lb. NOx * 2lb./MW+ $2/MWh O&M

14 Including any extraordinary O&M costs that might be incurred at very high operating levels as a result of increased risk of triggering a partial or unit wide forced outage, or as a result of continued operation of the unit after mechanical failures have occurred. See, for example, FERC Outage Report, pp. 27, 34 and 49.

- Distinguish between the effects of withholding and imperfect foresight and/or market inefficiency.

Under the first approach, the identification of anticompetitive withholding of capacity in the day-ahead market is further complicated by the need to:

- Account for capacity not sold in the day-ahead market because the day-ahead price was below the expected real-time price;[15]

- Account for capacity not sold in the day-ahead market because prices in the day-ahead market were not expected to be high enough to recover the start-up and/or minimum-load costs of the unit;

- Account for capacity not sold in the California day-ahead market because the output of that capacity was sold in other WSCC markets; and

- Account for bidding strategies intended to hedge outage risk.

While the application of this first approach requires considerable information, much of this information would be routinely available to an ISO coordinating a day-ahead market. The application of this approach to the California market, however, is made particularly difficult, even for the California ISO, by a number of unique features of the California market. In particular; 1) the California ISO does not operate a day-ahead energy market; 2) the bids into the California day-ahead energy market used for most analyses (i.e., the PX market) are in the form of portfolio bids. These supply offers are not tied to the capacity of any specific unit, and do not include capacity intended to be used to provide ancillary services (which would be separately bid into markets coordinated by the CAISO); 3) both the ISO and PX markets are based on one-part bids; 4) the CAISO does not manage intra-zonal transmission congestion in the day-ahead market and accepts infeasible transmission schedules in the day-ahead market; and 5) the day-ahead markets operated by the PX and ISO have a number of pay-as-bid elements that introduce inefficiencies and bidding incentives that must be distinguished from the exercise of market power.[16]

Tests for the exercise of market power based on the simulation of day-ahead prices simplify the task by avoiding the need to account for arbitrage of day-ahead and real-time prices (because the approach attempts to estimate equilibrium day-ahead and real-time prices), but must:

---

15 In other markets, such as PJM and New York, suppliers would probably be more likely to use demand bids (internal or external virtual demand bids in PJM and external virtual demand bids in New York) to arbitrage day-ahead and real-time prices than generation bids, to avoid profit reducing distortions in the unit commitment. There is no such incentive to use demand bids for arbitrage under the California market design and the separation of the ISO and PX, and the PX cost allocation rules appear to make it more expensive for generators to use load, rather than supply bids, to arbitrage day-ahead and real-time prices.

16 Thus, if energy prices in the PX are lower than expected, the day-ahead ancillary services offers may be based on different assumptions regarding the configuration of on-line capacity than those which underlay the portfolio offers into the PX.

- Account for the impact on commitment decisions of start-up and minimum-load costs and other unit characteristics;

- Account for the impact of real-time events (generation and transmission outages, erroneous load forecasts) that raise day-ahead and expected real-time prices relative to simulated prices; and

- Account for the supply and bidding behavior of non-thermal generation, import supply and export demand.

Even with good data and models, simulation studies have a number of limitations in estimating competitive price levels that might be used to identify the exercise of market power. First, it is in practice difficult to calibrate a simulation model to replicate real-world supply and demand patterns (i.e., load variations, generation outages and transmission outages), yet failure to do so can result in differences between real-world and simulated prices that are unrelated to competitive behavior. Second, it is important, but difficult, to accurately model reserve requirements and identify reserve shortages in the simulation. Third, most simulations do not directly account for the effect of supply and demand shocks that occur following the unit commitment decision. Thus, simulation approaches that determine load and outages prior to the unit commitment step and then dispatch the committed units to meet the load will not fully capture the volatility of real-time prices or the impact of expected real-time prices on day-ahead bids and prices. Fourth, it is difficult to model import supply and export demand so as to accurately reflect the prices at which imports are available in relation to internal supply and demand conditions. Fifth, it is difficult to include the effects of inefficient market design in a simulation. Sixth, a full year or more of data would be needed to address the impact of cumulative annual output limits.

The third approach, tests of withholding from the real-time market, can avoid some of the complexities associated with the simulation approach by focusing on real-time economic withholding by on-line units, and can take the unit commitment as given. Still these tests cannot always easily identify physical withholding of the capacity of off-line units.

Finally, simulation models could be applied to the real-time dispatch to test whether actual real-time prices are consistent with the simulated level of prices. Simulating real-time prices is simpler than simulating day-ahead prices because the unit commitment can be taken as given, but otherwise must address essentially the same difficulties noted above with regard to simulations of day-ahead prices.

While all of these tests are at best difficult to apply, ISOs in the course of their real-time dispatch acquire most of the information needed to assess whether there has been substantial economic withholding in real time. Thus, the ISO dispatch data reveals which units were on-line in real-time, which on-line units and off-line quick-start units were using undispatched capacity to provide ancillary services (and particularly which units were being dispatched out of merit to maintain 10-minute reserves), and which capacity could not be dispatched because of transmission congestion, generation outages or generation deratings. Moreover, many environmental restrictions on output are in part implemented by the ISO, and are thus accounted for in the dispatch (such as restrictions permitting the use of some capacity only during various stages of emergency conditions).

Review of the dispatch data would not avoid all complications in identifying economic withholding in real time. First, it would still be necessary to account for the impact of any environmental output restrictions not administered by the ISO. Second, it would be necessary to account for intra-day gas prices and O&M costs associated with operating at high output rates.[17] Third, it would be necessary to distinguish between the effect of market inefficiency and economic withholding.[18] Fourth, other information would be needed to determine whether real-time prices were elevated by physical, rather than economic, withholding.

The task of identifying real-time withholding is much more difficult for those lacking access to real-time dispatch information. Absent such information, it cannot be determined whether undispatched capacity was economically withheld, providing ancillary services, constrained down due to transmission congestion, or unavailable due to deratings or environmental output restrictions. Limited to public data, the analyses so far have been incomplete and unable to identify which, if any, of the thermal generators have been consistently withholding capacity in a strategic effort to raise prices above competitive market clearing levels.

IV. EMPIRICAL ANALYSIS OF MARKET POWER

A. Overview

The paper by Paul Joskow and Edward Kahn goes further with public data than any before and attempts to empirically assess whether the exercise of market power contributed materially to the high wholesale market prices for electricity in California during the summer of 2000, using variations on both the second and third approaches discussed above. First, to motivate a more detailed examination of the data they employ a simulation methodology similar to that previously utilized by Borenstein, Bushnell, and Wolak[19] and the MSC, to simulate the competitive level of wholesale prices in California.[20] Second, they utilize EPA and WSCC hour by hour output data for individual thermal generators to assess whether economic capacity was withheld from the market in real-time during high priced hours in the June to September 2000 period.[21]

Although the Joskow-Kahn paper takes better account in its simulation analysis of NOx emission costs than either the BB&W or MSC papers, the analysis faces many of the same limitations as

---

[17] Some units may incur increased outage risks or additional O&M costs at very high operating levels. Because units rarely operate at these levels, their existence may be uncertain and at least hard to verify.

[18] This would in principle be possible by comparing the degree of economic withholding across otherwise similarly situated firms having large and small shares of the real-time market.

[19] Severin Borenstein, James Bushnell and Frank Wolak, "Diagnosing Market Power in California's Restructured Wholesale Electricity Market," August 2000 (hereafter BB&W).

[20] Joskow-Kahn, pp. 10-20.

[21] Joskow-Kahn, pp. 22-33.

the earlier studies, as well as having some new problems arising from differences in data and methodology. In particular, the Joskow-Kahn study has limitations with respect to accounting for: environmental restrictions other than NOx allowances; intra-zonal congestion and RMR calls; capacity not available due to forced or maintenance outages or deratings; the impact of start-up and no-load costs and operating inflexibilities; the supply of non-thermal generation and imports; the impact of reserve requirements and the distinction between market inefficiency and anticompetitive withholding. The limitations of the data and methodology generally have the effect that the simulation is likely to understate the competitive level of prices, potentially materially so, and thus suggest the presence of anticompetitive economic withholding, even if the market were operating competitively, but inefficiently.

Joskow and Kahn break new ground in attempting to use EPA and WSCC data to go beyond the simulation and assess whether output was economically withheld during June 2000 by examining actual real-time output data rather than simulation results. Again, however, it is seen that the data and methodological approach has a number of limitations that may cause the analysis to identify economic withholding when none is occurring. In particular, the Joskow-Kahn study has limitations with respect to accounting for environmental restrictions other than NOx allowances, capacity not available due to deratings, capacity used to provide ancillary services, and capacity not dispatched because of real-time congestion or balancing requirements. Moreover, the hours in which withholding is identified include hours in which real-time prices were low for all or portions of the hour, accounting for lower output than in hours with high real-time prices. The present paper replicates the Joskow-Kahn analysis of real-time withholding and extends it to address some of these limitations. The revised analysis suggests that there might, or might not, have been material real-time withholding; there are too many unknowns to draw conclusions.

A fundamental problem with both approaches employed by Joskow and Kahn to identify the exercise of market power is that the margin of error in the methodology and data may be larger than the magnitude of the effects we are trying to identify. This does not mean that withholding cannot be identified, the point is rather that these very indirect methods of inferring the existence of economic or physical withholding have a large measure of error relative to the size of the effect. As suggested above, economic withholding that is very difficult to indirectly identify under the Joskow-Kahn, MSC and BB&W methodologies, will often be readily apparent in the dispatch data available to the system operator, particularly in real-time. The dispatch data will reveal which units: were providing reserves or regulation with capacity not dispatched to meet load; were constrained down by the ISO due to congestion; were dispatched down in real-time by the ISO to balance load and generation; or were output limited because they had not been released from environmental output restrictions. Moreover, the dispatch data will reveal what capacity was not dispatched because the capacity was not offered to the system operator (which could reflect outages, environmental restrictions, or physical withholding) and what capacity was not dispatched because it was economically withheld from providing either energy or ancillary services.[22]

---

[22] The dispatch data alone, however, will not reveal whether capacity is being physically withheld from the real-time market to exercise market power or because it is not available due to outages; whether capacity is being

B. Price Simulation Analyses

1. Overview

The Joskow-Kahn simulation of day-ahead prices follows the methodology of the BB&W and MSC papers, but provides a much improved simulation of the impact of NO_x allowance costs. The Joskow-Kahn simulation also takes better account of the full ancillary service requirements of the California ISO.[23]

Prices in the California electricity market since June 2000 appear to have been importantly affected by the prices for NOx emission allowances. The South Coast Air Quality Management District (SCAQMD) RTC program caps and reduces NOx allowances in the Los Angeles region by allocating emission allowances to historical pollution sources and permitting these allowances to be bought and sold in a free market. Although individual allowances are allocated for a one-year period and cannot be carried forward, the RTC program allocates allowances on a January 1 to December 31 cycle to some firms and a July 1 to June 30 cycle to other firms. This provides a degree of inter-temporal flexibility as emission sources on the July 1 to June 30 cycle can buy allowances from emission sources on the January 1 to December 31 cycle if supply is tight for the cycle ending June 30 and sell allowances for use by those on the later cycle if they are long.

A central fact in understanding California electricity prices during 2000 is that due to a combination of ratcheting down of emission levels (reducing the supply of emission allowances as shown in Figure 1 below) and low hydro conditions resulting in unusually high output levels by thermal electricity generators (see Figure 5 below), emission allowances prices in the SCAWMD region rose far above historical levels.

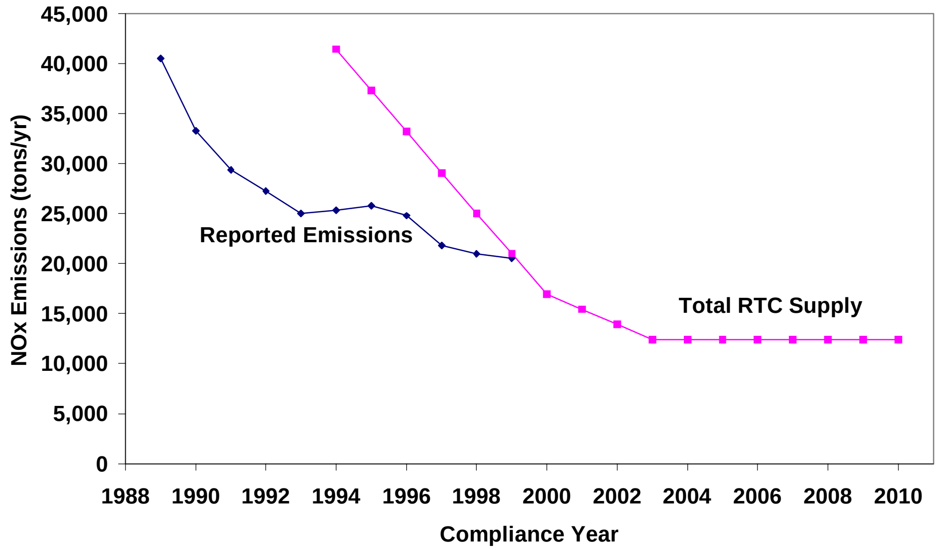
---

economically withheld from the real-time market to exercise market power or to allocate limited energy to the highest valued hours; or whether capacity is offered at very high prices because operation of the unit at such high levels carries a significant risk of triggering a forced outage.

[23] The Joskow-Kahn simulation includes in its measure of demand the CAISO demand for capacity to provide all ancillary services, estimated by Joskow and Kahn to be roughly 10 percent of load, rather than only including the capacity used for regulation as in the BB&W and MSC studies (Joskow-Kahn, p. 12). This methodological change would, other things equal, cause the study to identify more hours of shortage during the summer of 2000 than did the MSC study.

Figure 1

RECLAIM NOx Emissions and RTC Supply (tons/year)



Source: Reproduced from: South Coast Air Quality Management District, Review of RECLAIM Findings, October 20, 2000, p. 1-6 (hereafter RECLAIM Findings).

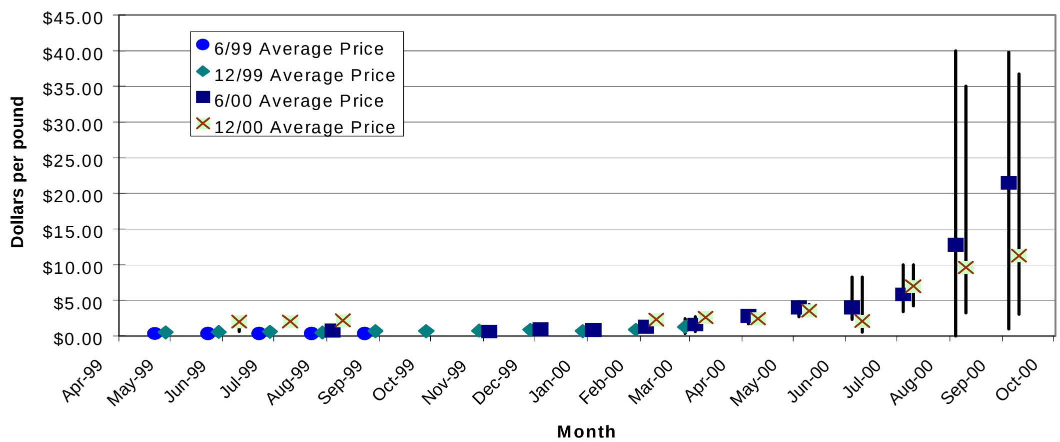
Unfortunately, while it is clear that emission allowance prices rose (see Figure 2), the nature of the process by which these prices are reported and available to electricity market analysts gives rise to some ambiguity in exactly at what point in time allowance prices rose to exactly what level. The Joskow-Kahn simulation analysis both takes these allowance costs into account and addresses the uncertainty in the time path of allowance prices by simulating electricity prices for a variety of allowance price levels.[24]

---

[24] Joskow-Kahn, Table 2, p. 16. It is, however, somewhat unclear what emission rates were assumed in calculating these prices. Joskow and Kahn note that the emission rates utilized in their analysis are based on publicly available data and regulatory filings but the data and filings are not identified.

Figure 2

NOx RTC Monthly Average Price Trends



Source: Reproduced from RECLAIM Findings, p. 6-4.

The Joskow-Kahn simulation study also provides an improved modeling of California reserve requirements. Previous simulation studies of day-ahead prices (in particular those of BB&W and the MSC) dispatched generation resources to meet load plus upward regulation requirements. This approach of the previous studies has three limitations. First, this approach would sometimes fail to indicate hours of capacity shortage because the measure of required capacity does not include reserves. Thus, a capacity shortage would exist whenever the available generation resources were insufficient to meet load (including losses) plus upward regulation requirements plus reserve requirements even if there were sufficient capacity available to meet load and provide upward regulation.

Second, this previous approach would also have the potential to understate the cost of meeting load in both shortage and non-shortage conditions because it implicitly assumes that CAISO reserve requirements can be met entirely with undispatched extra-marginal capacity. In practice, however, spinning and 10-minute reserve requirements must be carried on capacity capable of increasing output by the amount of the reserve requirement within ten minutes.[25] System operators therefore often find it necessary to reduce output on infra-marginal generators in order to maintain sufficient ramping capability on the units providing reserves such that the reserves could be dispatched within ten minutes in the event of a contingency.[26] This reduction in the output of infra-marginal generators correspondingly forces the dispatch of higher-cost extra-marginal generation to meet load. This is illustrated for a very simple market in Table 3. The example assumes that the system operator seeks to carry 60 MW of reserves, that 25 MW can be carried on high-cost hydro and GT units that are not generating and that the remaining 35 MW

---

25 See WSCC Minimum Operating Reliability Criteria, p. 2.

26 It is important to keep in mind that the main purpose of 10-minute reserves is to maintain the reliability of the transmission system in the event of the sudden loss of a large generator or transmission line.

must be carried on gas units. Column (B) sets forth assumed ramp rates for the gas units and it is seen that the 35 MW of reserves must be carried both on the marginal and infra-marginal cost units. As a result, there is 1,040 MW of capacity on-line to meet 990 MW of load and the market price of energy is $70/MWh.

Table 3
Schedule of Reserves and Ramp Constraints

			Schedules	
	Supply (A)	10-Minute Ramp Rate (B)	Energy (B)	Reserves (C)
Nuke 1	100 @ 0	0	100	0
Nuke 2	100 @ 0	0	100	0
Cogen 3	100 @ 5	0	100	0
Coal 4	100 @ 17	5	100	0
Gas 5	100 @ 30	10	100	0
Gas 6	100 @ 35	10	100	0
Gas 7	100 @ 35	10	90	0
Import 8	100 @ 40	0	100	
Gas 9	100 @ 5	10	90	10
Gas 10	50 @ 50	5	45	5
Gas 11	50 @ 60	5	45	5
Gas 12	40 @ 70	5	20	5
Gas 13	40 @ 80	5		
GT	10 @ 75	10	0	10
Hydro	15 @ 80	15	0	15
Load = 990	Reserves = 60			

If the prices for this electric system were simulated taking account only of the need to meet load, ignoring reserve requirements, it would be possible to meet load with only 1,000 MW of capacity, as shown in Table 4, and Gas 12 would not be dispatched. The simulated market price of energy would be only $60/MWh compared to the actual price of $70/MWh. We do not know the extent to which the need to maintain 10-minute reserve levels in practice requires the CAISO to dispatch higher-cost generation than would otherwise be the case. It is our experience, however, that it would be an unusual on-peak hour in which the need to maintain 10-minute and spinning reserves did not raise the market price of energy in New York. The omission of reserve requirements from the BB&W and MSC analyses is therefore likely to have depressed the simulated prices compared to market prices, at least during peak hours.

Table 4 Schedule with No Reserve Constraint			
		Schedules	
	Supply (A)	Energy (B)	Reserves (C)
Nuke 1	100 @ 0	100	0
Nuke 2	100 @ 0	100	0
Cogen 3	100 @ 5	100	0
Coal 4	100 @ 17	100	0
Gas 5	100 @ 30	100	0
Gas 6	100 @ 35	100	0
Gas 7	100 @ 35	100	0
Import 8	100 @ 40	100	
Gas 9	100 @ 45	100	0
Gas 10	50 @ 50	50	0
Gas 11	50 @ 60	40	0
Gas 12	40 @ 70	0	0
Gas 13	40 @ 80	0	
GT	10 @ 75	0	0
Hydro	15 @ 80	0	0
Load = 990			

The Joskow-Kahn study corrects for these downward biases by modeling load as actual load plus 10 percent to account for reserves plus upward regulation.[27] This procedure should correct for the considerations discussed above, permitting the analysis to identify capacity shortage hours and avoid the potential understatement of the market price of energy. In fact, this approach would have the potential to cause the simulation to overstate the competitive market price of energy because it would implicitly assume that all reserves are provided by infra-marginal generation which would also be an inaccurate assumption. Thus, in the example portrayed in Table 4, such an assumption would entail meeting load of 1,050 MW, which would require dispatching the GT at $75/MWh.[28] If we correctly understand the Joskow-Kahn methodology for accounting for reserves, it would eliminate this source of downward bias in the simulation results

---

27 Joskow-Kahn, p. 12. "We add 10% to each demand level reflecting the CAISO's demand for ancillary services capacity."

28 It is possible that we have not correctly understood the manner in which Joskow and Kahn have adjusted load to account for reserves but, other things equal, their methodology would tend to cause their simulation to identify more shortage hours than the MSC simulation.

of BB&W and the MSC. The Joskow-Kahn methodology provides a conservative measure of the impact of reserves on the competitive price level. Indeed, this methodology would in general be too conservative.

Joskow and Kahn also utilize a more realistic fuel cost figure ($5.58/mmBtu or roughly $32.50/bbl) for oil-fired generation than that apparently utilized in the BB&W and MSC studies.[29] This fuel cost is still somewhat on the low side during much of 2000, but is likely within 10 percent of the actual cost during the April to June 2000 period.

The Joskow-Kahn price simulation does not, however, appear to address a number of other limitations of the BB&W and MSC simulation methodology that have been described in a previous paper.[30]

- Accounting for environmental restrictions, other than NOx allowance costs;
- Treatment of hydro power as price taking;
- Accounting for the impact of transmission congestion;
- Treatment of start-up and minimum load costs;
- Assumed outage rates; and
- Impact of the inefficiency of the California market structure.

Moreover, the Joskow-Kahn price simulation methodology introduces some new sources of error that were not present in the earlier analyses. In particular, the Joskow-Kahn simulation analysis is based on ten monthly load deciles, rather than actual hourly supply and demand, which appears likely to result in downward biased simulations of the average competitive price by reducing the number of shortage hours. The Joskow-Kahn study also allocates each month's internal hydro power to meet load over time based on a set of assumptions rather than the actual supply of hydro power and assumes that geothermal generation is not energy limited. Depending on whether these assumptions under or overstate the supply of hydro power in shortage hours and the magnitude of the overstatement of geothermal energy supply, the methodology may over or understate the competitive price level.

2. Previously Discussed Methodological Problems

Methodological limitations of the BB&W and MSC studies that also appear to apply to the Joskow-Kahn simulation study are briefly summarized below. For a few topics for which the Joskow-Kahn methodology is somewhat different than that applied in the earlier studies, somewhat more discussion is provided.

---

29 See Harvey-Hogan, pp. 41-43.

30 See Harvey-Hogan.

a) Environmental Output Restrictions (other than NOx allowance costs)

The Joskow-Kahn paper provides important improvements on the BB&W and MSC papers in its treatment of NOx allowance costs. Moreover, improved accounting for allowance costs results in simulated prices that are much closer to actual prices in August and September than those simulated by the DMA.[31] On the other hand, however, the Joskow-Kahn simulation does not account for any of the other environmental limitations that affected output, nor does it account for the impact of NOx limitations in areas without markets for allowances. Two types of environmental restrictions that need to be taken into account in simulating generation output are NOx limitations imposed through annual operating hour or capacity factor limits and water temperature restrictions.

Other NOx Limitations

It is widely known by now that a number of California generating plants are subject to annual restrictions on their hours of operation or their annual capacity factor in order to limit NOx emissions. Plants subject to such restrictions and reaching their limits in 2000 include at least the Reliant's Coolwater 1 unit; Ellwood plant and Mandalay 3 unit; and Mirant's (Southern Energy)[32] Potrero 4, 5 and 6 units. AES's Los Alamitos, Redondo Beach and Huntington Beach units were also shut down during November by environmental limits.[33] These units account for roughly 4,000-5,000 MW of capacity.

If the plants subject to these kinds of annual operating restrictions are expected to have little chance of reaching their annual operating limit, the limits would be treated by the unit owner as non-binding and the limits would not affect the unit owner's bidding strategy or California electricity prices. If, however, it is expected that these plants will reach their annual operating limits, then a perfectly competitive firm bidding the output of such plants into the market would include in its bids a premium over the other incremental costs of operating the plants. The purpose of such a premium would be to allocate the output of such output-limited plants to the highest-priced hours (i.e., the hours in which their energy was most valuable).

The clear reality for California in 2000 is that a number of the plants subject to such annual operating limits used up their annual quota of operating hours prior to the end of the year. In some cases, such as Reliant's Mandalay 3 and Ellwood units**,** mechanisms were eventually worked out to enable the plants to continue operating, for at least a limited number of hours.[34] In other cases, plants ceased operation.[35]

---

31 Joskow-Kahn Table 2, p. 16; Hildebrandt Oct, p. 6 exhibit 1; Hildebrandt Nov, p.6, Table 1.

32 Since Joskow and Kahn completed their study, Southern Energy has changed its name to Mirant. We generally refer to the company as Mirant in this paper, except in the discussion and presentation of Joskow-Kahn output gap findings where we refer to the company as Southern for consistency.

33 See FERC Outage Report, pp. 8, 19, 32; *MW Daily*, November 22, 2000, p. 2; December 1, 2000, p. 7; and December 8, 2000, p. 8.

34 See, for example, FERC Outage Report, pp. 8, 19, 32.

35 See, for example, *MW Daily*, November 22, 2000, p. 2; December 1, 2000, p. 7; and December 8, 2000, p. 8; South Coast Air Quality Management District Website (aqmd.gov/news).

The Joskow-Kahn study assumption that additional NOx allowances were available to Los Angeles area generators at the South Coast market price would not necessarily be applicable to generators in other regions, which could have either lower or higher incremental costs of acquiring NOx allowances. For example, generating plants in San Diego are also subject to NOx restrictions but there is no established market in which additional allowances can be acquired once the annual allowance is exhausted.[36]

It is clear in the case of plants that were shut down as a result of exhausting their annual quotas that the competitive offer curve for these plants should include an opportunity cost for using up an annual hour or MW-hour of output. In addition to accounting for the output restrictions on the units that are known to have actually reached their environmental operating limits, a simulation analysis must account for any other environmental limits that would have been reached had the unit operator not included an environmental opportunity cost in its bids so as to conserve the unit's available output. Neither the BB&W, MSC nor Joskow-Kahn studies, however, include such an opportunity cost in their simulations. As a result, these studies would overdispatch the plants subject to such restrictions, resulting in the plants exhausting their annual operating allowance at an even earlier point in the year than was actually the case. This overdispatch of these plants would tend to reduce the price simulated by BB&W, MSC and Joskow-Kahn relative to the actual market-clearing prices,[37] but the price decrease would reflect the effect of violating environmental limits, not the impact of competition. Because these studies do not model environmental limits, the effects of environmental limits cannot be distinguished from the exercise of market power.

While it is clearly necessary to account for these environmental output restrictions in a simulation analysis, it is recognized that this is not straightforward in circumstances such as prevailed in California in 2000. First, in a number of instances, plants which reached their limits were allowed to continue to operate, at least to some degree, and sometimes in conjunction with additional payments. In these circumstances, the expectations that would have been held by competitive suppliers regarding the consequences of reaching their environmental operating limits are somewhat uncertain. Should plant owners have acted on the assumption that the operating limits would be relaxed in this manner or should they have anticipated that they would be shut down?

Second, the year 2000 in California was almost certainly one in which thermal plants were utilized more heavily than was anticipated early in the year. California ISO data show that thermal units were not utilized more heavily in aggregate in 2000 than in 1999 until May (see Figure 5 below). Absent foresight of how the year would end, thermal plants subject to environmental operating limits may have been bid into the market with no environmental opportunity cost early in the year, under the expectation that their environmental constraints would be non-binding. Such a misforecast would have required higher environmental adders to be included in bids in later months as the continued low hydro and high-load conditions made it

---

36 See Comments of Duke Energy North America LLC, November 22, 2000, p. 21.

37 Meaningful simulations would necessarily encompass the period covered by the run-time limitations, which would permit verification that operating the generation facilities as assumed by the simulation would not entail violating environmental run time or capacity factor limits.

apparent that these units would in fact reach their environmental limits unless their running time was carefully allocated.

There is more than one way to model these environmental limits, and reasonable people will likely differ on the most appropriate approach. It may in fact be desirable to simulate prices using a range of approaches to understand the sensitivity of the simulation results to these assumptions. It does seem clear, however, that it is not sufficient to simulate electricity prices for California in 2000 under the assumption that none of these environmental limits existed.

The fact that some of the plants subject to such environmental operating restrictions were actually shutdown during 2000 because they reached their annual operating limit is particularly significant in evaluating the possible exercise of market power. These units reaching their operating limits implies that, rather than anticompetitively withholding their output from the market to raise prices as implied by the BB&W, MSC and Joskow-Kahn methodology, these units actually offered themselves into the market at prices that with the benefit of perfect hindsight we now know without question were too low. Perfectly competitive firms with perfect foresight would have offered the output of these plants into the California market during 2000 in many hours at *higher* prices than those at which the output of these plants was actually offered. The reason for such higher offer prices would be to conserve the annual operating hours of these plants so that the units would not exhaust their ability to run prior to the end of the year. A theory of anticompetitive withholding of capacity that is applied to the capacity that reached its environmental limits must therefore be premised on inefficiently low-priced supply offers in some hours and anticompetitively high-priced supply offers in other hours, resulting in a net increase in profits over the year. The conditions required for such a strategy to be profitable are complex, particularly when there are other energy-limited suppliers in the market who would divert energy supplies from the low-priced hours to the high-priced hours. These conditions are not taken into account in the Joskow-Kahn, BB&W or MSC studies.[38]

The impact of the annual run time and capacity factor limitations is also important in assessing the impact on California prices of the reduction in imports. Joskow and Kahn analyze the impact on their simulated California electricity prices of the reduction in imports between 1999 and 2000, and include the combined effect of the reduction in imports and the increase in allowance prices.[39] The Joskow-Kahn analysis of the impact of reduced imports on electricity prices does not, however, take account of the cumulative impact of sustained reductions in imports on the hours of operation of units subject to environmental restrictions on their annual output.

Thus, reduced imports in a particular hour raise electricity prices by forcing the use of higher-cost generation, including generation with higher NOx allowance costs. Reduced electricity imports over the course of the summer also exhaust NOx allowances, driving up the price of

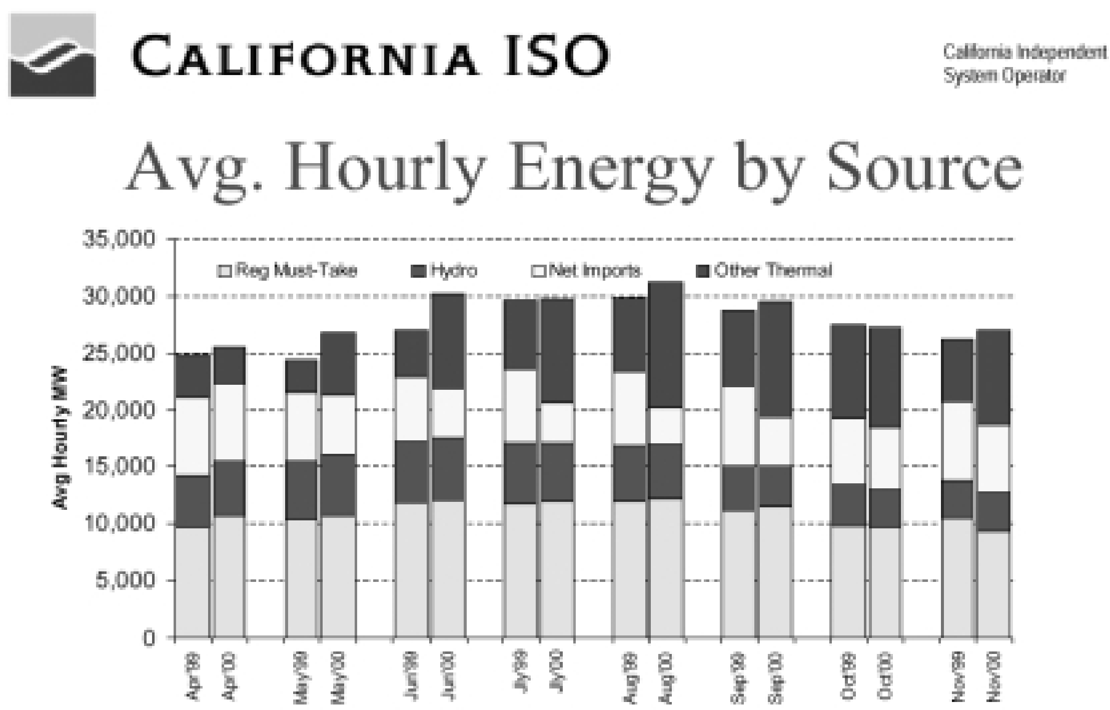
---

38 Indeed, such a theory does not appear to have been advanced by any study of the exercise of market power in California. Moreover, such a theory could not be tested with publicly available data as a test would at least require access to data on RMR calls, the dispatch of reserves by the ISO, and unit-specific real-time adjustment bid data. Even with such data it would be necessary to somehow distinguish between the exercise of market power, imperfect foresight and the effects of the price caps in accounting for why some units reached their environmental operating limits prior to the end of the year.

39 Joskow-Kahn, pp. 17-18.

NOx allowances, and exhaust the annual operating hours of restricted units, forcing load to be met with higher-cost units, raising prices and inducing conservation or creating shortages. It is clear from CAISO data that the combination of increased demand (insulated from wholesale electricity prices) and reduced imports led to a sustained increase in the output of California's thermal generation that was much larger than the increase in demand alone (see Figures 5 and 24).

Figure 5



November average based on data for November 1-15.

Source: Eric Hildebrandt, "Market Analysis Report," November 30, 2000.

Similar data showing substantially increased output in 2000 on a unit-by-unit basis have also been reported by FERC for a subset of these thermal units.[40] In considering the effects of the aggregate increase in utilization, it should be kept in mind that some units included in the 1999 output data were not operating at all during June 2000 (such as PG&E's Hunters Point units),[41] increasing the load on the remaining units.

It is also noteworthy in this regard that even with the low import supply elasticity assumed by Joskow and Kahn, their simulation analysis predicts much lower levels of imports,[42] and thus

40 FERC Outage Report, pp. 13, 20, 24, 29, 39, 46.

41 It should be noted that PG&E as a large net buyer presumably had a substantial financial incentive to maximize the output of its thermal units.

42 Joskow-Kahn, pp. 13-14, 37.

much higher operating levels for thermal generators, than those that prevailed in the real world. The supply offers in the simulation would likely therefore cause thermal generators to reach their run time limits even sooner in the simulation than they did in the real world. If this kind of simulation approach is to be applied to the California electricity industry, it needs to be applied over an annual period and the total output in the simulation of each unit subject to environmental restrictions needs to be compared to and validated against the actual environmental limits.[43]

Water Temperature

Duke's South Bay plant in San Diego is subject to cooling water thermal discharge limits that can limit output.[44] Mirant's Contra Costa and Pittsburgh plants are subject to environmental restrictions, sometimes referred to as the Delta Dispatch. These restrictions tend to be active from May through about July 15. The restrictions predate Mirant's acquisition of these generating plants and are intended to protect the Delta Smelt, a fish species in the area. The restrictions limit the output and ramp of the Mirant units, with the limits depending on water temperature in the river. The full capacity of these units is only available during this period during local or system emergencies. None of these limits are taken into account in the simulations and the Delta restrictions in particular are in effect during the June period.

Non-allowance environmental constraints affecting the output of the thermal units are a potentially significant factor affecting the supply of energy, and thus California electricity prices, during the summer of 2000. If conclusions are to be drawn regarding the exercise of market power, these limitations need to be taken into account. If these limitations are not taken into account, then the analysis cannot distinguish whether high prices are attributable to the exercise of market power or to the high environmental cost of incremental output.

b) Price Taking Hydro Power

Like the Borenstein, Bushnell & Wolak and MSC studies, the Joskow-Kahn study takes the output of hydro resources as given in defining the competitive supply curve and estimating the competitive market price.[45] Thus, Joskow and Kahn in effect place the output of hydro resources at the bottom of the supply curve they use to simulate competitive prices. This methodology, however, will cause the simulation to understate the actual competitive price level in any hour in which hydro units were on the margin (i.e., setting prices) in the real world.[46] Information regarding the frequency of the time that such hydro units were on the margin is apparently not

---

43 One way to validate the results of such a simulation against environmental output limits would be to compare monthly, and cumulative annual output and run times in the simulation to real-world monthly output and run times from the CEMS data. Such a comparison needs to be made, however, in the context of a chronological simulation model that recognizes generator characteristics such as start-up costs, minimum load levels, minimum down times and minimum run times, as discussed below in Section III.B.2.d.

44 See Comments of Duke Energy North America LLC, November 22, 2000, p. 21.

45 There are, however, important differences in how Joskow and Kahn and the MSC implement this approach, which are discussed in Section III.C.2 below.

46 This potential for understating the competitive price level is discussed at greater length in Harvey-Hogan, pp. 34-35.

publicly available, nor is it known to what extent energy-limited hydro (and geothermal) suppliers used their bids to allocate their energy output to the highest valued hours, but data reported by the California PX showing that many of the high-price bids in the day-ahead market during June 2000 were submitted for IOU resources suggest that this may be a material consideration.[47]

There is also no discussion in the Joskow-Kahn paper of how they handled the output of geothermal units (which are also energy-limited resources and may have used bids to allocate their energy output to shortage hours). If Joskow and Kahn followed the BB&W and MSC methodology, geothermal units would have been treated as price taking as well, which may or may not accurately characterize their bidding strategy and could also cause the simulation to understate the competitive price level.

c) *Transmission Congestion*

Like BB&W and the MSC, Joskow and Kahn do not attempt to model the impact of transmission congestion on the cost of meeting California load in a competitive market. Instead, they compare the market-clearing price that they estimate, ignoring the effects of congestion, with the hypothetical unconstrained PX price, which is also calculated without regard to the impact of transmission congestion.[48] While this approach greatly simplifies the analysis, it has implications for the conclusions that can be drawn from the Joskow-Kahn analysis. First, if transmission congestion were expected to exist in real time, then inter-temporal arbitrage by generators, in conjunction with the large amount of load often buying energy in the California real-time market and the segmented structure of California energy and ancillary services markets, would cause thermal generators lacking market power but located in the constrained-up regions to submit PX bids that exceed the unconstrained PX price, but instead approximate the actual zonal PX price and expected real-time zonal price. Thus, if generators in the constrained-up zone bid the expected real-time market-clearing price for that zone into the day-ahead market and their costs into the real-time market, the day-ahead zonal price would be equal to the real-time zonal price in the constrained-up zone. The hypothetical unconstrained price calculated from the day-ahead bids, however, could be higher than a hypothetical unconstrained price

---

47 See CalPX Compliance Unit, Price Movements in California Electricity Markets, August 31, 2000, p. 11 (hereafter CalPXI), and CalPXII, pp. 55-57. It should be kept in mind that because of the structure of the California market, one might actually observe much of the generation offered into the PX market at the expected market-clearing price. It would then be difficult to determine which units were marginal in a cost sense and which units are simply bidding to ensure that they are paid the market-clearing price.

Because most of the internal California hydro resources were owned by utilities that were net buyers during the summer of 2000, one might anticipate that the hydro resource owners would want to avoid submitting bids that would raise the market-clearing price. From such a perspective, the ideal course of action might be to allocate their output to the hours with the highest expected price, bidding it into the market at zero. To the extent, however, that pondage hydro or geothermal resources were bid in to provide reserves and the owners sought to conserve their available energy, the owners would be likely to bid the energy in at high prices, such as $750/MWh, which would then set the market-clearing price if these resources were dispatched to meet load during shortage conditions.

48 Joskow and Kahn do not discuss the impact of transmission congestion or RMR contracts on their analysis. This issue is discussed in BB&W, p. 18.

calculated from the cost-based real-time bids. Recall that sellers are not paid the hypothetical unconstrained price; they are paid the zonal price (or their bid, if dispatched to manage intra-zonal congestion). Thus, in a world with perfect competition and perfect foresight, the unconstrained PX price could exceed the price simulated by Joskow and Kahn in any period in which the NP-15 or SP-15 price exceeded the unconstrained PX price.

A rough assessment of the potential magnitude of the effect of congestion on the unconstrained PX price can be made by calculating the average difference between the unconstrained PX price during June and the price in the constrained-down zone. This difference averaged $4.37/MWh over June.[49] This is an upper bound on the impact of the congestion on the calculated unconstrained price and it is clearly both lower than the total price difference simulated by Joskow and Kahn and non-trivial.

A second potential effect of congestion relates to intra-zonal congestion. Until the implementation of Amendment 26 in early 2000, much intra-zonal congestion was managed through the call of RMR contracts, after the close of the PX market. In this non-locational bidding structure, firms lacking market power but having perfect foresight should be expected to bid their capacity into the market at the locational market-clearing price. Firms with locational market power that is mitigated through RMR contracts would, again with perfect foresight, offer their capacity into the market in periods with transmission congestion at the lower of the locational market-clearing price or the mitigated price under their RMR contracts, and would be called by the ISO out of merit at the RMR contract price to meet real-time load. Finally, loads would, with perfect foresight, take these RMR calls into account in bidding into the PX. These effects are recognized by BB&W who note the potential for RMR contract calls to cause the actual PX price to be lower than the simulated PX price curve used in their analysis. The same potential exists in the Joskow-Kahn study.[50]

In practice, neither loads nor suppliers have perfect foresight, but they should be expected to bid so as to attempt to capture the market-clearing price at their location, whether that is the zonal price or a price reflecting the impact of intra-zonal congestion.[51]

Although the Joskow-Kahn, BB&W and MSC papers do not discuss the implications of prescheduling of RMR generation subsequent to the implementation of Amendment 26, this system would cause the unconstrained PX supply curve to be lower and to the right of the supply curve estimated by Joskow-Kahn, BB&W and the MSC. Like the other impacts of congestion, the magnitude of the impact of RMR generation on the difference between the actual and simulated price is variable and hard to quantify without reference to data on actual RMR calls.

While it is possible to use publicly available data to place an upper bound on the effect of ignoring inter-zonal congestion as discussed above, no publicly available data exist for intra-zonal congestion or the unit-by-unit level of hourly RMR calls. As a result, neither the

---

[49] The difference averaged $13.06 in congested hours.

[50] BB&W, p. 29.

[51] Moreover, mistakes arising from the imperfect foresight of real-world suppliers would contribute to the market inefficiency that also raises prices as discussed below.

magnitude nor the direction of the impact of excluding transmission congestion from the analysis can be reliably predicted from publicly available data without abandoning the Joskow-Kahn simulation model and adopting a model that explicitly takes account of transmission congestion.

d) Start-Up and Minimum-Load Costs

Like BB&W and the MSC, the Joskow-Kahn analysis excludes start-up costs, minimum-load costs, and operating parameters such as minimum down times and run times[52] in defining the simulated competitive supply curve and in simulating the estimated market-clearing price.[53] It is a straightforward result of unit commitment logic that when these costs and restrictions exist, it will at times be more efficient to meet load with high incremental cost output from a unit that is already on-line or a high-cost but quick-start unit, than to meet that load by starting a unit with low incremental energy costs but a long start-up time or high start-up costs. While taking account of these costs greatly complicates the analysis, these are real costs that generators must recover in the market price of energy if they are to operate and that, under a one-part bidding system such as in California, will be reflected in the bids of competitive generators.

The impact of ignoring start-up and minimum-load costs would very likely be small on days in which the market was in a shortage situation and prices reached $750/MWh for many hours, but it could be appreciable over the study period as a whole and even during June on days when prices were well below the price cap. For example, in hour 17 on June 17 the unconstrained PX price was $125, which according to Joskow and Kahn would have exceeded the incremental running costs of all of the steam units in the CEMS database.[54] The Joskow-Kahn simulation model would find that all of this generation should therefore have been dispatched to meet load before the price could rise to $125. Even if their inferences regarding incremental costs were correct, however, it is likely that it would not have been economic to use all of the steam generation to meet load in this hour. The $125/MWh price was the highest hourly unconstrained PX price in the day-ahead market for June 17. While there were six hours that day during which the unconstrained PX price exceeded $100, it is unlikely that a steam unit with average running costs of $110 would have expected to recover its start-up costs (or losses from remaining on overnight at prices that were less than $40 for hours 2 through 8) by operating at prices only slightly above its running costs for a few hours. Only a unit with average running costs materially below $120 would have been economic to bring on-line to meet load during June 17.

It is important to understand that these minimum-load costs can be large and can change the apparent profitability of unit operation. For example, Table 6 portrays the apparent profitability of Alamitos 2 on June 17 (one of the days during which unconstrained PX prices at times

---

52 Start-up costs are the fuel costs of bringing a generation unit online to meet load. In the case of large gas-fired steam generators, this can be $100,000 or more. Minimum load costs are the cost of keeping a steam generating unit operating at its minimum operating level in order to have it available to operate at a higher level later in the day.

53 BB&W provide various rationales for their failure to take account of start-up and minimum load costs (BB&W, p. 22), which are discussed in Harvey-Hogan, pp. 14-16 and 38-39. Joskow and Kahn do not provide any rationale for their decision not to take account of these costs.

54 Joskow-Kahn, p. 25.

exceeded $120/MWh), calculated based on the unconstrained PX price, the actual CEMS output,[55] the incremental heat rate reported in Klein,[56] and a Southern California gas price of $4.99 per mmBtu. These approximate the assumptions apparently used in the various simulation analyses. With these assumptions, the unit would appear to earn a margin of nearly $45,573 for the day, a gross profit margin of nearly 46 percent of its gross revenues.

[55] It should be kept in mind that the CEMS data reports the gross output of each unit. Some of this electricity output is consumed by the plant itself and therefore does not generate revenues. Tables 6 to 9 below therefore overstate unit margins because they do not include those costs.

[56] Joel Klein, "The Use of Heat Rates in Production Cost Modeling and Market Modeling," April 17, 1998. 10056 is the incremental heat rate at full output. The incremental heat rate is lower at lower output levels.

Table 6 Alamitos 2 Profitability Incremental Heat Rate and Unconstrained PX Price June 17, 2000						
Hour	**Heat Rate (A)**	**Gross Output MWh (B)**	**Unconstrained PX Price (C)**	**Incremental Cost $/MWh (D)**	**Margin $/MWh (E)**	**Margin $/Hour (F)**
0	10056	10	50.00	50.18	-0.18	-1.79
1	10056	9	38.99	50.18	-11.19	-100.71
2	10056	9	35.76	50.18	-14.42	-129.78
3	10056	9	35.75	50.18	-14.43	-129.87
4	10056	9	35.01	50.18	-15.17	-136.53
5	10056	29	28.01	50.18	-22.17	-642.91
6	10056	37	32.00	50.18	-18.18	-672.64
7	10056	25	37.90	50.18	-12.28	-306.99
8	10056	34	45.01	50.18	-5.17	-175.76
9	10056	53	72.57	50.18	22.39	1186.70
10	10056	56	76.52	50.18	26.34	1475.07
11	10056	56	90.00	50.18	39.82	2229.95
12	10056	57	105.00	50.18	54.82	3124.77
13	10056	94	116.13	50.18	65.95	6199.35
14	10056	122	115.00	50.18	64.82	7908.11
15	10056	124	119.98	50.18	69.80	8655.27
16	10056	107	125.00	50.18	74.82	8005.80
17	10056	63	115.00	50.18	64.82	4083.70
18	10056	38	93.92	50.18	43.74	1662.14
19	10056	34	80.01	50.18	29.83	1014.24
20	10056	48	80.01	50.18	29.83	1431.87
21	10056	19	75.61	50.18	25.43	483.18
22	10056	24	65.00	50.18	14.82	355.69
23	10056	15	53.76	50.18	3.58	53.71
Total All Hours						45572.59
Total Profitable Hours						47869.55

No allowance is made for variable O&M, emission allowance or station costs.
Gas Price = 4.99
Sources:
(A) Klein
(B) CEMS
(C) PX
(D) (A) * $4.99/1000
(E) (C) – (D)
(F) (E) * (B)

However, the assumptions of the simulation analyses simplify the problem in important ways. By contrast, Table 7 recalculates the profitability of Alamitos 2 using the actual average heat rate reported in the CEMS data for each hour and the SP-15 PX price. Rather than earning a significant profit, it can be seen that the unit would have lost money had it sold its real-time output in the PX market.[57] Part of the difference is attributable to the difference between the actual SP-15 price and the hypothetical unconstrained price, but most of the difference is attributable to the fact that the unit's actual average heat rate was much higher than the incremental heat rate reported in Klein for most of the hours of the day. The unit's output weighted average heat rate was almost 17,000, compared to the incremental heat rate of 10,056 in Klein.

[57] It should be kept in mind that these tables do not reflect the actual day-ahead revenues of the unit. Profits are calculated in these tables based on real-time output and day-ahead prices. The actual day-ahead schedule of the unit is not known. The tables also do not include any ancillary service revenues that the unit might have earned in the day-ahead or real-time markets.

Table 7 Alamitos 2 Margins – PX Prices and Actual Heat Rate June 17, 2000						
Hour	**Heat Rate (A)**	**Gross Output MWh (B)**	**SP-15 PX Price (C)**	**Incremental Cost $/MWh (D)**	**Margin $/MWh (E)**	**Margin $/Hour (F)**
0	33832.31	10	35.00	168.82	-133.82	-1338.23
1	36213.68	9	30.00	180.71	-150.71	-1356.36
2	36480.34	9	24.02	182.04	-158.02	-1422.15
3	35558.12	9	20.51	177.44	-156.92	-1412.33
4	36258.12	9	19.99	180.93	-160.94	-1448.44
5	20490.45	29	15.00	102.25	-87.25	-2530.17
6	17751.98	37	24.99	88.58	-63.59	-2352.92
7	20660.92	25	22.90	103.10	-80.20	-2004.95
8	18735.97	34	32.02	93.49	-61.47	-2090.07
9	16404.21	53	38.01	81.86	-43.85	-2323.89
10	15736.13	56	70.00	78.52	-8.52	-477.30
11	15666.48	56	74.23	78.18	-3.95	-220.96
12	15810.93	57	82.00	78.90	3.10	176.90
13	14684.29	94	101.13	73.27	27.86	2618.41
14	14206.75	122	113.60	70.89	42.71	5210.42
15	14204.22	124	119.98	70.88	49.10	6088.52
16	14272.18	107	123.60	71.22	52.38	5604.86
17	15224.18	63	113.60	75.97	37.63	2370.78
18	16774.29	38	85.00	83.70	1.30	49.26
19	18068.33	34	75.00	90.16	-15.16	-515.47
20	16900.48	48	74.99	84.33	-9.34	-448.48
21	22101.21	19	52.00	110.29	-58.29	-1107.42
22	22146.79	24	51.30	110.51	-59.21	-1421.10
23	25961.54	15	38.01	129.55	-91.54	-1373.07
Total All Hours		1081				-1724.19
Total Profitable Hours						22119.12
No allowance is made for variable O&M, emission allowance or station costs. Gas Price = 4.99 Sources: (A), (B). CEMS (C) PX (D) (A) * $4.99/1000 (E) (C) – (D) (F) (E) * (B)						

While incremental heat rates are appropriate for calculating incremental dispatch costs, they are not appropriate for evaluating unit commitment decisions. The unit commitment decision must take into account the average heat rate and costs of a unit, as well as its incremental costs. Bidding strategy is greatly complicated for California generators by the CAISO and PX reliance on one-part bids, but the decision to prevent generators from submitting multi-part bids does not change the underlying economics that will govern the bids of competitive generators. Generators in competitive markets will submit one-part bids that limit the financial losses arising from uneconomic PX positions. This will entail submitting one-part bids that exceed incremental costs on potentially marginal or extra-marginal units.

The Joskow-Kahn methodology not only likely understates the bid price at which capacity that actually operated in the real-world would be offered to the day-ahead market (because this capacity would lose money on marginal operations in the Joskow-Kahn model) but also likely assumes that capacity that did not operate in the real world because of high start-up and minimum-load costs, would have operated in many additional hours in which the units would have lost money. The Joskow-Kahn simulation methodology implicitly assumes that a steam unit would be available to meet load at incremental cost in each and every hour, even if it was only needed to meet load for a single hour and that the units could be turned off whenever prices made their operation uneconomic, even for a single hour.

Load could indeed be met more cheaply if start-up and minimum-load costs did not exist and units did not have inflexible operating characteristics, but real generators dispatched to meet California load did incur start-up and minimum-load costs and did have operating constraints. These costs and operating constraints raise the cost of meeting load in a competitive market. The effect of these assumptions can also be illustrated using the actual output of Alamitos 2 on June 17. It can be seen in Table 8 that even if average heat rates are used to evaluate operating profitability, the operation of Alamitos 2 appears highly profitable if it is modeled as being able to turn on whenever prices are high and turn off whenever prices are low. Thus, with this flexibility, it would appear highly profitable to use Alamitos to meet load on June 17, earning $22,119, when in fact the day-ahead prices for energy alone were not high enough to justify operation of Alamitos 2 at the levels to which it was actually dispatched.

Thus, if Alamitos 2 were not providing ancillary services and submitted cost-based energy bids reflecting the CEMS heat rate data and the Joskow-Kahn fuel price assumptions to a centralized day-ahead unit commitment process such as those coordinated by the PJM-ISO and NYISO, Alamitos 2 would not have been committed to meet these schedules at these prices because operation of the unit would not have been economic. In the real world, even price-taking competitive suppliers should bid their units into the CAP PX and ISO markets so as to recover their variable operating costs. The Joskow-Kahn simulation model would commit Alamitos 2 and other units to operate in circumstances in which it is not efficient for the plant to run and it would be unprofitable for price-taking suppliers to operate in the actual dispatch.

Table 8 Alamitos 2 Profitability June 17, 2000			
	SP-15 PX Prices	**Unconstrained PX Prices**	**SP-15 Real-Time Prices**
Actual Heat Rate			
Profitable Hours	22119		3744
All Hours	-1724		-29080
Incremental Heat Rate			
Profitable Hours	40723	47870	19449
All Hours	35537	45573	8180
Calculations do not include variable O&M, emission allowances or station costs.			

Since the Joskow-Kahn simulation results, like those of the BB&W and the MSC, take no account of start-up and minimum load costs, the impact of minimum-load and start-up costs and operating constraints would show up in the Joskow-Kahn study as indistinguishable from anticompetitive withholding.

The impact of these implicit assumptions regarding minimum-load costs and operating constraints is potentially magnified by the increase in emission allowance costs during 2000. If emissions are a function of energy inputs, total daily emissions per MW will be higher than implied by incremental heat rates alone and these costs will be much higher per MW during the hours with poorer heat rate performance. Table 9 illustrates this impact for Alamitos 2 on June 17, using the CEMS emission rate for 1Q 2000 and an assumed NOx allowance price of $10/lb. It can be seen that the operation of Alamitos 2 would have been dramatically unprofitable at day-ahead PX prices on this day, losing slightly more than $33/MWh.[58] At the same time, a calculation based on incremental heat rates, and assuming that the unit would have operated only during the hours in which it was profitable, would imply profits of $28/MWh.[59]

[58] -$36,382/1081 MWh.

[59] $30,574/1081 MWh.

Table 9 Alamitos 2 Profitability June 17,2000 With and Without Emissions Allowance Costs				
	All Hours		**Profitable Hours**	
	No Allowance Costs	**Allowance Costs**	**No Allowance Costs**	**Allowance Costs**
Actual Heat Rate	-1724	-36382	22119	5299
Incremental Heat Rate	45573	25027	47870	30574
Emission rate .189 pounds per mmBtu per CEMS. Assumed allowance costs $10/pound.				
Calculated profitability does not include variable O&M or station costs.				

These calculations for Alamitos 2 are only illustrative and we have not repeated this calculation for every unit for every day of June. The point of these calculations is that the financial impact of minimum load costs and operating inflexibilities is not necessarily insignificant. Simulation models that implicitly or explicitly make unit commitment decisions without regard to these costs have the potential to misstate the competitive level of electricity prices.

A failure to include either start-up and minimum-load costs or operating parameters in simulation models will also hinder the ability of these models to accurately account for the environmental constraints discussed above. In simulations lacking these costs and constraints, environmental constraints would be less limiting because units subject to such limits could be dispatched only in high-priced hours, conserving their hours on-line or operating factor. In the real world, however, some of the units subject to these environmental limits were steam units with start-up times, minimum-load levels and minimum down times. They could not be simply turned on and off hour by hour as needed, and having them available to meet load in high-priced hours could entail running them and using up limited operating time or output in other lower-price hours. These effects would be entirely missing in non-chronological simulation models that do not account for these kinds of constraints.

Simple non-chronological models that do not account for start-up and no-load costs cannot accurately simulate the competitive level of prices. As illustrated above, this is not a detail. It can be a fundamental limitation. A better alternative would be to assume that market participants were able to replicate the results of a centralized unit commitment process through their individual one-part bids and self-commitment and thus to estimate market prices based on the estimated multi-part cost functions of the various generators.[60] This method of simulating

[60] Simulation models such as GE MAPS could do this, as well as taking account of transmission constraints.

market prices would still be biased toward finding market power, because it would tend to attribute to market power the inefficiency arising from the California market design,[61] but the estimates would be less biased than those provided by a methodology which simply assumes that start-up and no-load costs and other operating inflexibilities are not material.[62]

e) *Outage Assumptions*

Like the BB&W and MSC analyses, the Joskow-Kahn study bases its simulation on assumed outage rates rather than actual outage data for the period studied. The Joskow-Kahn methodology for taking account of forced outages differs, however, from that of BB&W and the MSC. It assumes that all California generators (both fossil and must-take) suffer outages in each period in proportion to the forced outage rate in the Henwood database.[63] Like the methodology in the BB&W and MSC studies, this approach will overstate the available generation to the extent that the actual outage rate differed from that in the Henwood database, either due to fortuitous events or due to hard use.[64] The methodology therefore cannot distinguish between high prices due to the exercise of market power or to higher than assumed outage rates.

Moreover, it would be inappropriate to assume that increased levels of forced outages necessarily reflect the exercise of market power. For example, AES encountered substantially higher outage rates on all of its units during 2000 than in 1998 or 1999,[65] yet AES had sold its output forward and did not benefit from higher prices.[66] Indeed, AES lost money in California during 2000.[67] Similarly, the summer-long outage of PG&E's Hunters Point units 2 and 3 and the outage of Hunters Point 4 during May and June presumably raised prices in the real world relative to the prices simulated by Joskow and Kahn and this effect would be included in the price difference they attribute to market power but, as a substantial net buyer, PG&E presumably had no incentive to withdraw output from the market to raise prices.[68]

---

[61] See the discussion in Harvey-Hogan, pp. 8-14, 31-34, and item 6 below.

[62] While it is more difficult to model the electric industry if account is taken of start-up and minimum-load costs, these costs are sufficiently important that the standard industry simulation tools take them into account. GE-MAPS, for example, could be used to simulate unit commitment and prices taking account of start-up and no-load costs. Since MAPS also models the transmission system, this approach would also have enabled simulation of the effect of locational constraints and RMR calls, modeling of reserve requirements, and evaluation of environmental output limits.

[63] Joskow-Kahn, p. 14.

[64] Thermal generators in California operated at higher levels than in previous years throughout the period May through December, so historical outage rates may be not accurate for 2000 operating conditions (see Figure 5 above and Figure 24 below).

[65] Stu Ryan, AES Pacific, February 1, 2001, Analyst Presentation (hereafter Ryan).

[66] A further complication, not addressed here, would be the full analysis of the forward contracting position of the thermal generators, which could affect their incentives for strategic withholding.

[67] AES January 29, 2001, Press Release re annual earnings. Aesc.com/investor/press/index.html

[68] Joskow and Kahn state that they relied on the commercially available Henwood database for capacities (p. 14) but it is not entirely clear which units were included from the database. In particular, it is not clear whether the capacities of units owned by entities other than the five large non-utility generators (AES, Duke, Dynegy/NRG, Mirant and Reliant) that were not available in June 2000(such as 377 MW of PG&E's Hunters Point steam

In addition, by smoothing the incidence of outages over time, the Joskow-Kahn approach to outage modeling introduces a new problem not found in the BB&W or MSC studies.[69] The Joskow-Kahn approach of proportionally derating every unit in every hour is easy to implement but has the limitation that to the extent that real-world outages of either fossil generation or must-take generation were lumpy, this methodology would tend to reduce the number of shortage hours by smoothing out the impact of outages.[70] This limitation of the Joskow-Kahn methodology can be illustrated with a simple example. Suppose that there were two 1,000 MW units in the market with a 5 percent outage probability. The Joskow-Kahn methodology would assume that the units would provide 1,900 MW of capacity with a probability equal to one, while in the real world there would be a 90.25 percent probability of having 2,000 MW of capacity, a 9.5 percent probability of having 1,000 MW of capacity and a .25 percent probability of having no capacity. Obviously, if load exceeds 1,000 MW, shortages would be much less likely if outages occurred in the manner assumed by Joskow and Kahn than they would be in the real world. If real-world units suffered outages in the smooth predictable manner assumed by the Joskow-Kahn model, less available capacity would be needed to meet load, supply would be larger relative to demand and prices would be lower. Such an assumption can obscure the reasons for high prices in California and the distinction between competitive behavior and the exercise of market power, because real-world units do not suffer outages in this nice, smooth predictable manner.

The actual impact of the Joskow-Kahn assumptions regarding outage rates and incidence are difficult to assess without knowing the actual pattern of real-world outages. Data showing atypically high outage rates for AES units in the summer of 2000[71] and knowledge of PG&E's Hunter Point outages, in combination with the information in the FERC Outage Report, however, suggest that understatement of real-world forced outage rates have depressed the simulated competitive prices in the Joskow-Kahn study.

---

units and the 140 MW of Thermo-Ecotek's Riverside units) were treated as available in the simulation. We have confirmed that the Hunters Point units were treated as available in the simulation, which would tend to cause the simulation to understate the competitive price.

69 BB&W and the MSC went to great lengths to avoid this potential source of bias, using actual output data for must-take and geothermal generation and using Monte-Carlo simulation methods to account for the impact of random forced outages of thermal units on prices, see BB&W, pp. 28, 41-43.

70 We have confirmed that the unit capacities used by Joskow and Kahn in the simulation analysis were net output capacities (i.e., net of the electricity output consumed by the generating unit).

71 See Ryan.

f) Inefficiency of California Market Structure

The California electricity markets have a number of design features that predictably cause suppliers entirely lacking market power to bid their resources into the market at the expected market-clearing price rather than at their costs.[72] In addition, the one-part bidding mechanism requires that market participants make guesses about market-clearing prices in order to determine the offer price of potentially extra-marginal units with start-up or minimum-load costs.[73] Even if these guesses do not change the identity of the resources used to meet load, errors in predicting the market-clearing price could raise the prices paid by loads (if generators guess that the market-clearing price will be higher that it actually turns out to be) or could reduce the prices paid by loads (if generators guess that the market-clearing price will be lower than it actually turns out to be). If there were no change in the actual resource cost of meeting load, the errors might cancel out and have no impact on market efficiency or real-world market prices.

In practice, however, it is unlikely in our view that load will be met at least cost in markets that clear based on generators guessing the market-clearing price. Instead, it is likely that some of the generators that overestimate the market-clearing price would be lower cost than some of the generators whose bids clear in the market. In this circumstance, the bidding errors attributable to non-uniform pricing would not cancel out, because the guesses would change the resources used to meet load and the supply curve would in effect be shifted in as a result of some supply resources not clearing in the market despite being infra-marginal on a cost basis.

The inefficiency of the California market might also exacerbate the impact of the annual run time and capacity factor limits by making it more difficult for California generators to bid their units into the market to provide reserves but with sufficiently high energy bids that the units would rarely be dispatched for energy. The lack of cooptimization of the energy and reserve markets, the pay-as-bid structure of the ancillary services market arising from the rational buyer protocol, and the non-least-cost dispatch of reserves and other energy to meet load in real time create a market structure in which energy-limited resources may not be efficiently allocated between reserves and energy. This effect would potentially be exacerbated by the lowering of the price caps in July and August, which would cause many suppliers to offer energy at the price cap, increasing the likelihood that generators trying to ration their hours of operation would be dispatched for energy if they offered their capacity in the market. These kinds of inefficiency in the California market structure have been long recognized, but their costs may have been magnified by the energy shortage created by low hydro conditions and hot weather in the west.

The design of the California market was premised on there being other advantages that compensate for this inefficiency in the short-term market. Even if that is the case, however, the inefficiency of the short-term market must be taken into account in empirical analyses of market performance. While reasonable people may disagree over the magnitude of the short-run inefficiency arising from the non-uniform, pay-as-bid pricing elements of the California market design, simulations of competitive prices intended to identify the exercise of market power must

---

72 These features, as well as evidence of the inefficiency to which they lead, are discussed in Harvey-Hogan, pp. 6-14.

73 See Harvey-Hogan, pp. 14-16, 38-39.

recognize that this inefficiency exists and account for it in the analysis. Like BBW and the MSC, the Joskow-Kahn analysis of supply, in effect, simulates the market-clearing prices that would arise under a centralized market-clearing process for energy and ancillary services, rather than simulating likely competitive outcomes under the actual California market structure.

The quantitative impact of the inefficiency introduced by the California market design is hard to assess. Indeed, it is not apparent that there is any simple, reliable method of assessing the impact of this inefficiency.[74] It should be apparent, however, that it is not appropriate to simply attribute to the exercise of market power all differences between the price levels that might prevail in centrally coordinated day-ahead and real-time markets[75] and the price levels prevailing under the California market design.

3. New Problems

While the Joskow-Kahn study improves on some areas of the BB&W and MSC studies, there are also two areas (the method of measuring load and the method of determining hydro output) in which the methodology and data employed in the Joskow-Kahn study could introduce new problems.[76] The method of measuring load is likely to cause the simulation to understate the competitive price level while the impact of the hydro methodology cannot be determined without access to data on actual hydro output in high-load hours.

a) Monthly Load Aggregation

The Joskow-Kahn simulation analyzes supply and demand for monthly rather than hourly periods, and analyzes load within each month by deciles.[77] This approach tends to understate the impact of peak load demand on prices and is likely to understate the number of shortage hours.

One source of the bias from this approach is that monthly load is highly variable, and there is considerable variation in loads within the top decile of load hours. Thus, as seen in Figures 10, 11 and 12, the actual loads in the top decile exceeded the average load for the decile in June by 1,000 MW or more in 18 of the 72 hours, in July for 19 of the 74 hours, and in August for 12 of

---

[74] One might attempt to develop measures of market inefficiency from data on the frequency of irrational price patterns in PX and CAISO markets that might be compared to similar statistics for PJM and NY. Even with such measures, however, it is doubtful whether there is a reliable criteria for translating indicators or relative market efficiency into price impacts.

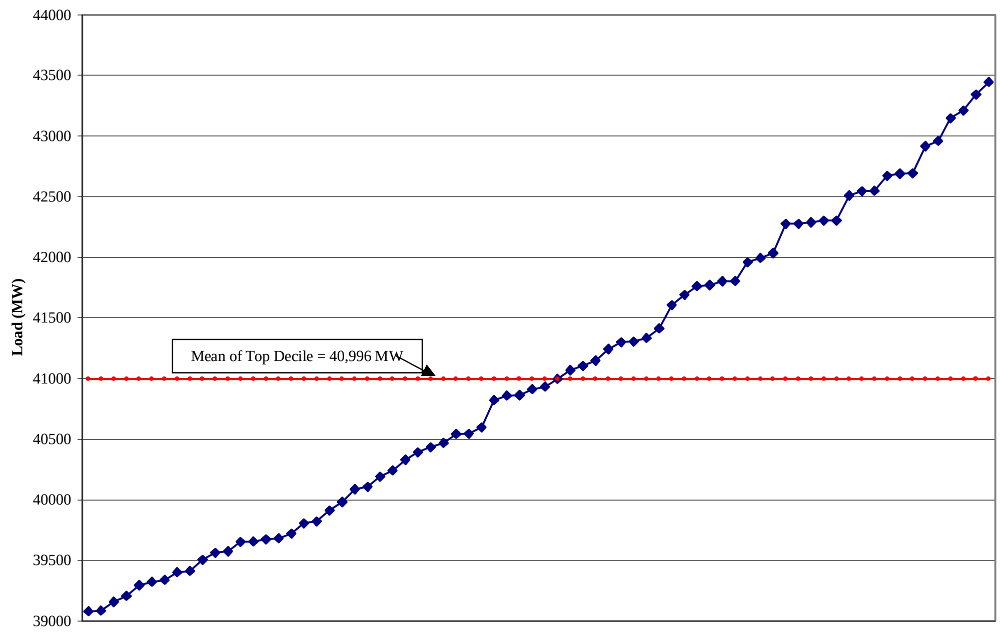
[75] Such as those coordinated by the PJM ISO and NY ISO.

[76] An additional methodological issue that is probably not material in June and July, but may be more important in the August through December 2000 period (with increased gas price volatility), is the use of monthly average gas prices rather than daily gas prices. If gas prices are correlated with the amount of gas-fired generation running, and thus with the heat rate of the marginal unit, the use of monthly averages may cause simulation results to understate the competitive level of electricity prices.

[77] Joskow-Kahn, p. 12. Joskow-Kahn state that they take this approach because they lack access to publicly available hourly data on hydro power production inside California. It is not apparent, however, how averaging loads in this manner reduces the error associated with their procedures for modeling hydro power. It appears, on the contrary, that this procedure serves only to introduce error into the demand measures used in their simulation, in addition to the potentially poor measurement of hydro supply.

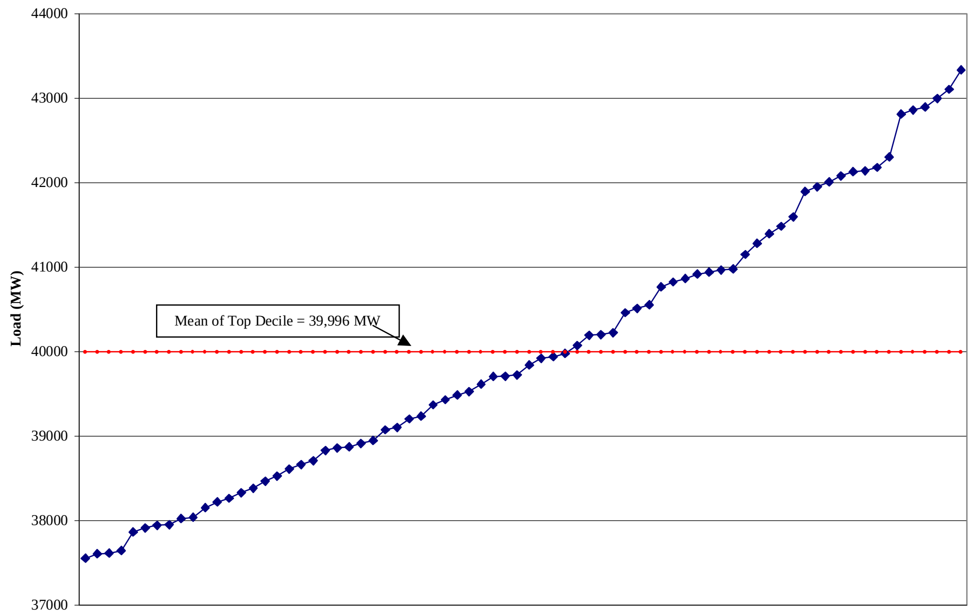
the 74 hours. Using the average decile load rather than actual loads in the Joskow-Kahn simulation therefore materially understates load in many of the high-priced hours. This understating of load in the highest price periods has the potential to materially understate the simulated prices in these hours, especially by eliminating shortage hours.

Figure 10
Top Decile of Actual Load, June 2000



If the relationship between prices and loads were linear (i.e., a 500 MW increase in loads resulted in an X dollar increase in prices at all points along the supply curve), then this averaging of loads would not affect the average prices simulated by the study, as if loads were 500 MW higher than the monthly average in some hours and 500 MW lower than the monthly average in other hours, the prices would also be correspondingly higher and lower and the simulated average monthly price would not be affected by the intra-month variation in loads.

Figure 11
Top Decile of Actual Load, July 2000

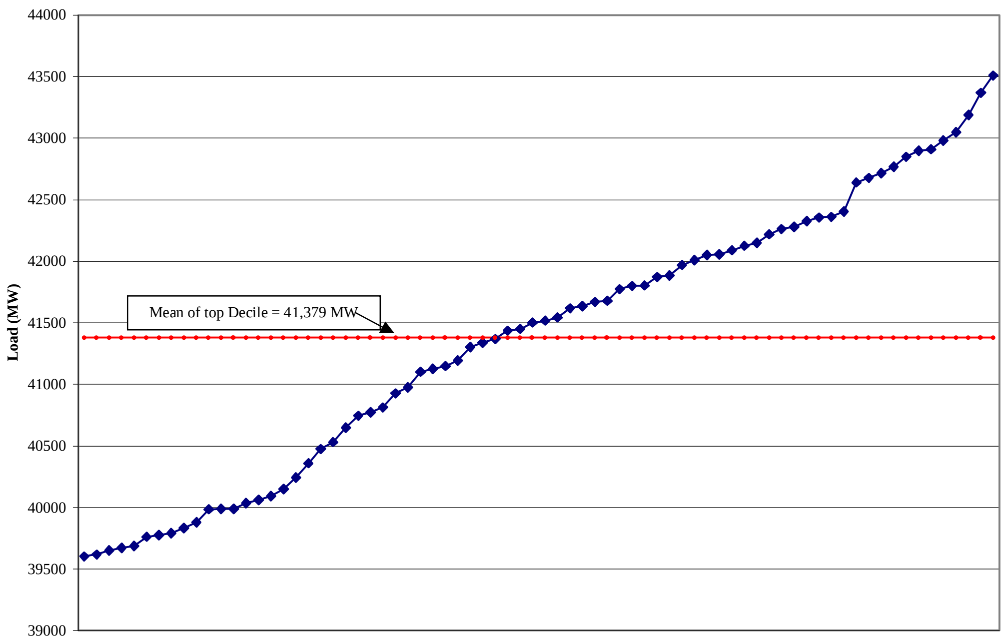


The relationship between loads and prices in a competitive California electricity market should, however, be expected to be non-linear for two reasons.[78] First, as capacity is exhausted, the running and opportunity costs of the marginal units probably rise much more rapidly than load. Second, the system is much more likely to be in a state of shortage, with prices rising to the price cap in very high load conditions.

---

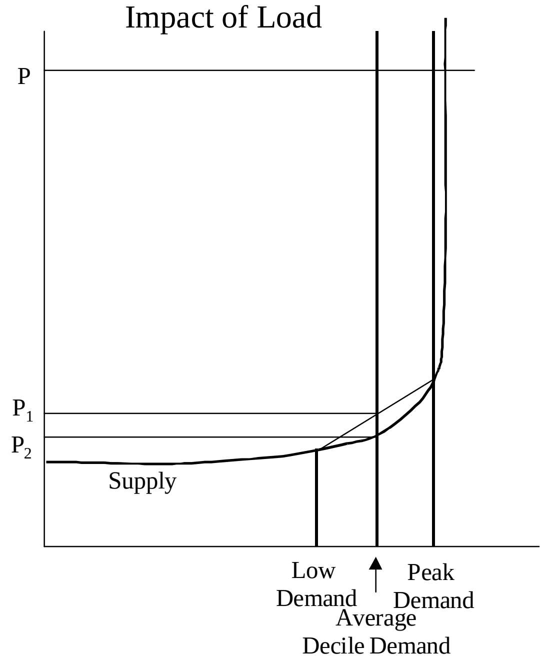
78 BB&W also assume such a convex relationship, BB&W p. 20.

Figure 12
Top Decile of Actual Load, August 2000



In these circumstances, the load averaging procedure of Joskow and Kahn is likely to reduce the number of hours that the simulated market clears on the very steep portion of the supply curve or that the price rises to the price cap as a result of the market failing to clear in a shortage. Because the actual competitive market price rises more when load is above average within the highest decile than it falls when load is below average in this decile, the use of deciles, rather than hour-by-hour data, to simulate competitive prices tends to bias the simulated price downward, relative to actual average prices, as illustrated in Figure 13. P_1 is the average of the prices when load is low and high, while P_2 is the price corresponding to the average load. If the supply curve is convex, P_1 will lie above P_2. Joskow and Kahn in effect simulate P_2.

Figure 13



A related issue is how the Joskow-Kahn study treated shortage hours. There is no discussion in Joskow-Kahn of how prices were determined in hours in which the simulated market failed to clear, a topic that was explicitly discussed in BB&W.[79] Having accounted for ancillary services demand equal to 10 percent of load, one would expect that the Joskow-Kahn study would have identified many more hours of shortage than did BB&W or the MSC.[80] The implication of this lack of discussion of shortage hours is that the Joskow-Kahn procedure for estimating peak loads, perhaps in combination with the assumptions regarding hydro power availability discussed below, eliminated shortage hours in the simulation of June prices. If this is the case, it may not be surprising that the Joskow-Kahn simulation would fail to replicate real-world prices since shortage conditions appear to have driven most of the high priced hours during June.[81]

b) Hydro and Geothermal Power Allocation

Joskow and Kahn apparently allocate monthly hydro power equal to 60 percent of the historical monthly hourly average supply to each hour in the month, and then allocate the rest to the high-

---

79 BB&W, p. 21.

80 The BB&W and MSC analysis of demand did not take account of the capacity required to provide reserves, BB&W, p. 19 and Harvey-Hogan, pp. 30-31.

81 See Harvey-Hogan, pp. 16 to 25, for a discussion of the relationship between high prices and shortage conditions during June. A California ISO Department of Market Analysis Study, “Report on California Energy Market Issues and Performance: May-June 2000,” August 10, 2000, argues that many of the high-priced hours did not reflect shortages (pp. 47-51). This analysis, however, is not based on actual real-world unit availability, but assumes no thermal units were derated, no outages of thermal units occurred during these days, and that any units that came on-line were immediately available at full output.

priced hours, up to 8,000 MW.[82] This procedure allocates 8,000 MW of hydro power to the top four load decile hours, and 3,155 MW to the lowest five decile hours.[83] If the full 8,000 MW of hydro capacity were not actually available in all of the high-priced hours, the Joskow-Kahn methodology would overstate the capacity available in the peak hours in which prices were high, eliminating shortage hours and downward biasing simulated prices in these hours.[84]

Like the load aggregation procedure discussed above, the potential impact of the hydro power allocation rule on the results of the price simulation arises from the non-linear impact on prices of supply reductions when the market is tight. The hydro power allocation rule would not make much difference if the supply curve had a constant slope over its entire length. In this situation, an increase in load of 1,000 MW would have the same impact on price whether it happened on a high-load hour or a low-load hour and, similarly, a change in hydro power supply would have the same impact on price if it happened on a high-load hour or a low-load hour. If, on the other hand, the supply curve becomes increasingly steep at high-load levels and eventually becomes nearly vertical, then these assumptions could cause the simulation to materially understate the competitive level of prices. It cannot be determined whether the Joskow-Kahn hydro power assumption causes their simulation results to overstate or understate the competitive price level without analyzing the actual pattern of hydro power availability, but the impact could be material.[85]

Although not explicitly described in the paper, we have confirmed with the authors that unlike the MSC and BB&W analyses, the Joskow-Kahn study did not impose limits on the energy output of California geothermal units. This is also a potentially significant assumption whose impact cannot be evaluated without comparing the simulated output with the actual output of the geothermal units. It is noteworthy, however, that the California ISO's recent CAISO 2001 Summer Assessment specifically mentioned the impact of declining steam field pressure on the dependable output of geothermal capacity.[86]

While assumptions like these might be appropriate in a demonstration that it is conceivable that there could have been economic withholding, a study based on such assumptions cannot provide an indication that output withholding by thermal generators occurred because it cannot

---

82 The BB&W and MSC studies made use of confidential hour-by-hour data on actual hydro and geothermal output that was not available to Joskow-Kahn.

83 Joskow-Kahn, p. 38.

84 The potential for overstated hydro capacity arises not only from the possibility that the 8,000 MW capacity figure might be high for peak availability but also from the possibility that the number of hours per day that this peak capacity was available was more limited than implied by the non-chronological modeling of supply and demand in the Joskow-Kahn model. The actual hydro generation data apparently utilized by BB&W and the MSC would not only better account for total hydro capacity, but would also better account for limitations on hydro power availability both within a day and over a multi-day hot spell.

85 Joskow-Kahn maintain that the assumptions are conservative but no basis is provided for this view (Joskow-Kahn, pp. 12-13).

86 California ISO, "CAISO 2001 Summer Assessment," March 22, 2001; "declining steam field pressure has affected the power output of geothermal units within the CAISO Control Area, thereby reducing the overall maximum "Dependable generating capability," p. 5. A similar comment is repeated on page 16. The assessment also refers to the "reduced capacity levels of geothermal resources," p. 8

distinguish economic withholding from incorrect assumptions regarding the level of hydro and geothermal output in high-priced hours.

4. Conclusions

While the Joskow-Kahn simulation study improves on the BB&W and MSC studies in its treatment of NOx allowance costs and ancillary services requirements, it confronts many of the same limitations of the earlier studies. Particularly important in this regard are likely to be the Joskow-Kahn study's treatment of environmental restrictions other than NOx allowance costs, hydro power offer prices, start-up and minimum-load costs, and outage assumptions. In addition, the treatment of load, hydro and geothermal power availability and outage probabilities create sources of imprecision not present in the BB&W and MSC studies.

Overall, the simulation results do not provide evidence that the high prices in California have been attributable to economic withholding. While the study shows that California prices have been higher than they might have been in a world without environmental output limitations, start-up and minimum load costs of thermal generation, energy limitations on hydro and geothermal output and with outage and load patterns matching those assumed in the study, the difference cannot be attributed to market power, as it might be attributable to some or all of these other differences between the simulation and actual power markets.

C. Withholding Analysis

1. Overview

A pathbreaking contribution of the Joskow-Kahn analysis is the use of CEMS data to attempt to test whether available thermal generating capacity was withheld from the real-time market during high-priced hours in June 2000. The CEMS data compiled by the EPA portrays the hour-by-hour gross output of the covered thermal units.[87]

Joskow and Kahn study the hours in June for which the PX day-ahead price exceeded $120 and compare the actual output of the units studied to the estimated capacity of these units. The analysis posits that any difference between actual and potential output can be attributed only to: 1) capacity used to meet ancillary service requirements; 2) capacity out of service due to forced outages; 3) capacity not available due to inter-zonal transmission constraints; and 4) economic or physical withholding. Joskow and Kahn therefore seek to account for the impact of the first three factors, with any unexplained difference between output and capacity being attributed to withholding and other factors.[88]

This gap or difference between the energy output of a unit and the maximum capacity becomes the focus for examining the possibility of strategic withholding. The extra capacity could be

---

[87] http://www.epa.gov/airmarkets/reporting/edr21/edrinst.june00.pdf, http://www.epa.gov/ardpublic/acidrain/ftp/rawhles.html.

[88] Joskow-Kahn, p. 23.

used to supply other services, such as reserves, or result from factors that limit the ability of generators to produce their full output. The gap could also reflect intentional withholding of supply in order to influence market prices. Hence, the size of the gap and the likely magnitudes of these various contributing factors become important in determining the strength of the evidence for any particular interpretation.

This gap analysis faces many of the complications described above for the simulation studies. For example, limitations on ramping of generators could contribute to the gap, but the magnitude is difficult to estimate without detailed dispatch data. Here we examine the analysis as best we can with the available data to indicate the sensitivity of the results and the likely order of magnitude of the potential errors of estimation.

Joskow and Kahn find that the difference between the output and capacity of the plants they study is less than the ISO's ancillary service requirements in NP-15 and greater than the ISO's ancillary service requirements in SP-15.[89] They further attempt to allow for the impact of forced outages by excluding the capacity of units not operating in the hour, day or prior day from the calculation of the output gap. If the output gap is calculated based on the output of units operating in a given hour, they find a mean output gap of 690 MW in NP-15 compared to total ancillary service requirements of 1,510 MW, and a mean output gap of 1,954 MW in SP-15, compared to total ancillary service requirements of 1,672 MW.[90] For SP-15 they also compare the estimated output gap to ancillary service requirements and find 49 hours in which there is no congestion from SP-15 to NP-15 in which the estimated output gap in SP-15 exceeds the estimated ancillary service requirements by 500 MW or more, using the best of their measures of the impact of forced outages.[91] Joskow and Kahn then repeat their calculation of the output gap for June using the WSCC EHV data, finding a smaller output gap in NP-15 and a larger output gap in SP15 than in their analysis based on EPA data. They attribute the larger output gap in SP-15 to the inclusion of additional generation capacity.[92] This analysis does not, however, include any adjustment for the impact of unit outages. Joskow and Kahn then use the EHV data to extend their analysis of the output gap through September, but without any adjustment to available capacity for the impact of unit outages. They again find a large output gap in SP-15, which they caveat by noting that the gap probably includes a fairly large amount of capacity whose operation was not economic under the price cap during August and September, and that there is no allowance for the impact of forced outages.[93] Joskow and Kahn conclude that despite these limitations, their analysis reveals unexplained output gaps that are "sufficiently large to suggest that power supplies were withheld in the zone during the June through September period."[94]

---

[89] Joskow-Kahn, Table 8, p. 26.

[90] Joskow-Kahn, pp. 25-30.

[91] Joskow-Kahn, Table 11, p. 30.

[92] Joskow-Kahn, Table 12, p. 31.

[93] Joskow-Kahn, pp. 31-33.

[94] Joskow-Kahn, p. 33.

The Joskow-Kahn analysis is a thoughtful effort to quantitatively assess a difficult public policy question. Nevertheless, the analysis has a number of acknowledged and unacknowledged data and methodological limitations, most of which tend to overstate the capacity available to meet load in these high-priced hours or to otherwise overstate the output gap. These limitations include: basing the comparison on hours in which day-ahead, rather than real-time prices, were high, omission of partial unit outages or environmental output limitations; incomplete accounting for ancillary service demand; omission of start-up and minimum load costs; assuming equality between peak hourly demand and average hourly demand; inability to account for the impact of the CAISO's dispatch instructions; potentially overstated unit capacities; an assumption that cumulative annual output limitations were not binding; and assuming that market inefficiency had no impact on available supply. In addition, our replication of their work identified a few data problems.

2. Day-Ahead Prices and Real-Time Output

Joskow and Kahn identify the high priced hours in which they study withholding based on the day-ahead unconstrained PX price. In June the threshold they use for a high-priced hour is a day-ahead price of \$120, \$90 in July, \$130 in August and \$110 in September. Importantly, Joskow and Kahn select the hours they analyze based on high day-ahead PX prices, but then test for withholding based on real-time output. This tends to bias the study towards finding evidence of withholding, even if none existed because the real-time output of a competitive firm depends on real-time prices, not just day-ahead prices. Even if output is sold forward at day-ahead prices, this is only a financial commitment. The ultimate opportunity cost for real-time production is the real-time price that would apply to any differences between day-ahead schedules and actual production.

In particular, the selection of high-priced hours restricts attention to the supply region where Joskow and Kahn presume that competitively offered plants would all be at their maximum output. Higher real-time prices should not produce higher output, but lower real-time prices could result in lower output. The effects do not average out. Therefore, with any significant volatility in real-time prices relative to day-ahead prices, this asymmetry would tend to overstate the implied withholding of supply.

The day-ahead prices used in the Joskow-Kahn study were determined in the PX and reflected day-ahead supply and demand conditions. The day-ahead prices are related to real-time prices because day-ahead demand and supply offers will be affected by expected real-time demand, supply and prices. Nevertheless, it needs to be recognized in empirical analysis that actual real-time prices will often differ from expected real-time prices. In particular, day-ahead prices are sometimes high in expectation of high loads that fail to materialize in real time. When this occurs and real-time load is lower than anticipated, real-time utilization and prices will also be reduced.

Examples of such unfulfilled expectations can be seen on June 16, 23 and 30 in the DMA report on May-June 2000, and for a number of hours on June 15.[95] Because the Joskow-Kahn methodology is based on day-ahead prices and real-time output, it treats output reductions associated with real-time load and price declines as anticompetitive withholding, when it could reflect an efficient and competitive response to lower demand and prices.

In order to assess the impact of the Joskow-Kahn study's reliance on day-ahead, rather than real-time prices, the Joskow-Kahn withholding analysis has been replicated for the companies and units whose output was analyzed by Joskow and Kahn. First, Joskow and Kahn's finding that the output gap for these units averaged 983 MW in the North and 3,351 MW in the South in the 137 June hours in which the PX day-ahead price was $120/MWh or more was largely replicated. The results of this replication are portrayed in Table 14. One difference that should be noted as it will affect some of the discussion that follows is that the generators labeled by Joskow and Kahn as NP-15 generators are actually located in three zones: NP-15, Z-26 and SF, as noted on Table 14. This difference is important later, as their analysis of ancillary service supply did not include the CAISO data for the SF and Z-26 zones, although their calculation of an output gap included generators in these zones. In order to maintain consistency with the Joskow-Kahn calculations and facilitate comparisons, Z-26 is treated below as if it is part of NP-15, which is not accurate.

---

95 California ISO, Department of Market Analysis, "Report on California Energy Market Issues and Performance: May-June 2000," August 10, 2000, p. 13.

Table 14 Output Gap (MW), Not Adjusted for Outages June 2000, PX Price > $120/MWh						
	Joskow-Kahn Calculation			Replication		
Owner	Maximum Output (A)	Mean Output (B)	Output Gap (C)	Maximum Output (D)	Mean Output (E)	Output Gap (F)
NP-15, SF and Z-26						
Duke	2,563	2,422	141	2,563	2,422	141
Southern	2,932	2,090	842	2,932	2,090	842
Total	5,495	4,512	983	5,495	4,512	983
SF and Humboldt						
PG&E	NA	NA	NA	279	93	186
SP-15 (excluding Z-26)						
AES	3,681	2,542	1,139	3,681	2,542	1,139
Duke	733	643	90	733	643	90
Dynegy	2,000	1,014	986	2,000	1,014	986
Reliant	3,487	2,351	1,136	3,487	2,351	1,136
Total	9,901	6,550	3,351	9,901	6,550	3,351

Source:
(A) – (C): Joskow-Kahn, Table 8
(D) – (E): CEMS Data
(F): Col. (D) – Col. (E)

Table 14 also differs from the figures reported by Joskow and Kahn in that it includes a similarly calculated output gap for PG&E, which continued to own two plants in the north during the study period. It can be seen that using the Joskow-Kahn methodology there is a large implied output gap for the PG&E plants as well as for the plants of the new generators.

Having largely replicated the Joskow and Kahn analysis with day-ahead prices, the approach can be applied with real-time price data. The real-time zonal uninstructed energy prices for NP-15 and SP-15[96] were then examined for these same 137 hours, and it was found that they exceeded $120/MWh in only 86 hours in NP-15 and 85 hours in SP-15. Table 15 portrays the result of repeating the Joskow-Kahn output gap calculation for these high-priced real-time hours. It is seen that the average output of both the new generators and PG&E was generally higher in the hours in which the average hourly real-time price was also high. This would be expected in a competitive industry as lower real-time prices should be expected to cause competitive firms to

[96] See caiso.com\marketops\Oasis\pubmkt2.html. Ex Post Market Information (button 29).

reduce output. When the output gap is calculated for these high-priced real-time hours, the gap falls by 204 MW in the north and 251 MW in the south, relative to that calculated by Joskow and Kahn.

Table 15

Output Gap (MW), Not Adjusted for Outages
June 2000

Owner	Maximum Output (A)	PX Price > $120/MWh: Mean Output (B)	PX Price > $120/MWh: Output Gap (C)	PX and Real-Time Price > $120/MWh: Mean Output (D)	PX and Real-Time Price > $120/MWh: Output Gap (E)
NP-15, SF and Z-26					
Duke	2,563	2,422	141	2,416	147
Southern	2,932	2,090	842	2,300	632
Total	5,495	4,512	983	4,716	779
SF and Humboldt					
PG&E	279	93	186	100	179
SP-15					
AES	3,681	2,542	1,139	2,589	1,092
Duke	733	643	90	663	70
Dynegy	2,000	1,014	986	1,125	875
Reliant	3,487	2,351	1,136	2,424	1,063
Total	9,901	6,550	3,351	6,801	3,100

Source:
(A) - (C): Joskow-Kahn, Table 8 and CEMS Replication Table 14, above
(D): CEMS
(E): Col. (A) – Col. (D)

It is possible to use these data to test whether the calculated output gap in the hours in which both the PX and real-time prices exceeded $120/MWh is statistically different from the calculated output gap in the hours in which the PX price exceeded $120/MWh in both NP-15 and SP-15 but the real-time price did not. For the purpose of this test we have dropped the hour included in the Joskow-Kahn analysis in which the hypothetical unconstrained PX price exceeded $120/MWh but the actual SP-15 PX price was less than $120/MWh. [97] For the combined North and South regions the mean output gap was 5,089 MW in the hours in which the real-time price was less than $120/MWh and 3,862 MW in hours in which the real-time price

[97] This is the reason that the mean for NP-15 used in the test differs slightly from the mean in Table 15.

exceeded $120/MWh. The difference is 1,217 MW which is statistically significantly different from zero at above the 99.9 percent confidence level (the t-statistic is 5.70).

For the South alone the difference in means is 662 MW (3,762 to 3,100 MW), which is also statistically significantly different from zero at more than the 99.9 percent confidence level (the t-statistic is 4.37). Finally, for the North alone the difference in means is 555 MW (1,326 to 771 MW), which is statistically significantly different from zero at above the 99.9 percent confidence level (the t-statistic is 6.93).

Joskow and Kahn recognize that their initial comparison of output to estimated maximum unit capacity (Table 8) makes no allowance for capacity not available due to outages. To account for the effect of unit outages on available capacity they develop three alternative measures of available unit capacity: capacity of units on-line in the hour; capacity of units on-line at some time during the day; and capacity of units on-line at some time during that day or at some time during the prior day.[98] While these measures adjust available capacity in the right direction, all are imperfect measures of the amount of capacity not available due to outages, deratings or environmental output restrictions and could materially overstate the available capacity in the California market. Tests 2 and 3 have the limitation that they include the capacity of units that were not actually on-line during the hour and therefore could not generate energy. While such a lack of availability could reflect a unit being physically withheld from the market, it also could reflect a unit that has suffered an outage since the start of the day or the prior day.[99] Moreover, it is evident from the FERC staff discussion of forced and maintenance outages at Reliant and NRG plants that these outages had an important effect on generation availability.[100]

Test 1 excludes from the calculation of the output gap, the capacity of units that were not on-line at the time. This approach avoids including in the calculated output gap the capacity of units that were unavailable due to forced outages. In replicating their calculation of the output gap thus adjusted for outages (Test 1), it became apparent that there were one or more typographical errors in the figures Joskow and Kahn report for Southern Energy/Mirant. Table 8 in Joskow-Kahn reports a mean output of 2,090 for Southern and Table 10 reports a capacity of 2,395 MW, which would imply an output gap of 305 MW rather than the 571 MW reported in Table 10. Our application of the Joskow-Kahn methodology to calculating capacity for Southern Energy yielded a figure of 2,740 MW, however, rather than 2,395 MW for capacity, implying an output gap by their standard of 650 MW.[101] The other capacity figures and output gaps in Joskow-Kahn Table 10 were replicated, as shown in Table 16.

98 Joskow-Kahn, pp. 27-29.

99 As discussed below in item 6, the lack of availability could also reflect a unit whose start-up costs made it uneconomic for the unit to start in order to meet a short term load and price spike.

100 FERC Outage Report, pp. 9, 17-19, 22-23, 27, 41-43, 48-50.

101 We have confirmed that the 571 MW figure is a typographical error and that 650 MW is the correct figure.

Table 16 Output Gap (MW), Adjusted for Outages (Test 1) June 2000, PX Price > $120/MWh						
	Joskow-Kahn Calculation			Replication		
Owner	Maximum Output (A)	Mean Output (B)	Output Gap (C)	Maximum Output (D)	Mean Output (E)	Output Gap (F)
NP-15, SF and Z-26						
Duke	2,541	2,422	119	2,541	2,422	119
Southern	2,395	2,090	571*	2,740	2,090	650
Total	4,936	4,512	690	5,281	4,512	769
SF and Humboldt						
PG&E	NA	NA	NA	109	93	16
SP-15						
AES	2,945	2,542	403	2,945	2,542	403
Duke	723	643	80	723	643	80
Dynegy	1,611	1,014	597	1,611	1,014	597
Reliant	3,225	2,351	874	3,225	2,351	874
Total	8,504	6,550	1,954	8,504	6,550	1,954

* Columns (A), (B) and (C) for Southern do not add due to typographical errors in original.

Sources:

(A), (C): Joskow-Kahn, Table 10
(B): Joskow-Kahn, Table 8
(D), (E): CEMS Replication
(F) (D) – Col. (E)

This calculation of an outage adjusted output gap was then repeated for the 86 hours in the north and 85 hours in the South in which both the PX and real-time prices exceeded $120/MWh. For these hours, the Joskow-Kahn methodology yielded an estimated output gap by their standard of 598 MW in the north and 1,696 MW in the south. This calculation of the real-time output gap is reported in Table 17. The overall output gap of 4,334 MW in Table 8 of Joskow-Kahn thus falls to 2,294 MW if the capacity of off-line units is excluded (to account for outages) and the analysis restricted to hours in which the real-time price exceeded $120/MWh.

Table 17 Output Gap (MW), Adjusted for Outages (Test 1) June 2000						
	PX Price > $120/MWh			PX and Real-Time Price > $120/MWh		
Owner	Maximum Output (A)	Mean Output (B)	Output Gap (C)	Maximum Output (D)	Mean Output (E)	Output Gap (F)
NP-15, SF and Z-26						
Duke	2,541	2,422	119	2,528	2,416	112
Southern	2,740	2,090	650	2,786	2,300	486
Total	5,281	4,512	769	5,314	4,716	598
SF and Humboldt						
PG&E	109	93	16	109	100	9
SP-15						
AES	2,945	2,542	403	2,940	2,589	351
Duke	723	643	80	719	663	56
Dynegy	1,611	1,014	597	1,596	1,125	471
Reliant	3,225	2,351	874	3,242	2,424	818
Total	8,504	6,550	1,954	8,497	6,801	1,696
Sources: (A) - (C): CEMS Replication Table 16, above, and Joskow-Kahn, Tables 8 and 10 (D) – (E): CEMS Replication (F): (D) – (E)						

It is again possible to use these data to test whether the calculated outage adjusted output gap in the hours in which both the PX and real-time prices exceeded $120/MWh is statistically different from the calculated outage adjusted output gap in the hours in which the PX price exceeded $120/MWh in both NP-15 and SP-15 but the real-time price did not. For the purpose of this test we have again dropped the hour included in the Joskow-Kahn analysis in which the hypothetical unconstrained PX price exceeded $120/MWh but the actual SP-15 PX price was less than $120/MWh.[102] For the combined North and South regions the mean outage adjusted output gap was 3,428 MW in the hours in which the real-time price was less than $120/MWh and 2,296 MW in hours in which the real-time price exceeded $120/MWh. The difference is 1,132 MW which is statistically significantly different from zero at above the 99.9 percent confidence level (the t-statistic is 5.49).

---

[102] This is the reason that the mean for NP-15 used in the test differs slightly from the mean in Table 17.

For the South alone the difference in means is 676 MW (2,370 to 1,695 MW),[103] which is also statistically significantly different from zero at more than the 99.9 percent confidence level (the t-statistic is 4.41). Finally, for the North alone the difference in means is 456 MW (1,056 to 601 MW), which is statistically significantly different from zero at above the 99.9 percent confidence level (the t-statistic is 5.77).

3. Outages, Output Restrictions and Deratings

Even the narrowest measure of available capacity employed by Joskow and Kahn (Test 1, which is based on units on-line during the hour) has the potential to overstate available capacity. There are two main reasons for this. First, an important implicit assumption of the Joskow-Kahn withholding analysis is that the impact of forced outages is such that capacity of generating units is either not available at all or available in its entirety. In practice, however, a unit may suffer equipment failures that reduce the unit's capacity, without requiring that the unit be taken off-line.[104] Indeed, it is particularly likely during high-priced hours that competitive unit owners would endeavor to keep units with operating problems on-line but operating at reduced capacity. Moreover, units that have suffered outages and attempt to come back on-line would be treated under Test 1 as fully available during the hour in which they came on-line. This would be true even if the units failed to come back on-line and generated power only for a short period of time.

Second, as discussed above, some units are subject to environmental restrictions that may preclude them from operating at full capacity.

The Joskow-Kahn methodology, however, makes no allowance for environmental restrictions that reduced unit capacity at times during the summer. As discussed in Section IV.B above, these kinds of environmental limits may have affected the capacity of at least Duke's South Bay unit and Mirant's Pittsburgh and Contra Costa units.

In most cases the capacity subject to these environmental restrictions would be available for use in emergency conditions, but not all of the hours analyzed by Joskow and Kahn were emergency conditions. This is particularly true during July and August when rising gas and allowance costs lifted the incremental costs of many units above the price thresholds used to identify high priced hours in the Joskow-Kahn analysis.[105]

It is not known how to assess the impact of deratings and other output limitations on the output gap calculated by Joskow and Kahn using the CEMS data. While a variety of procedures might be used to try to draw inferences regarding such deratings from the CEMS data, it seems to us that these procedures all have the potential to confuse withholding with deratings and vice versa. The reality is that there is no point in devoting a lot of resources to making guesses about the level of deratings and other output limitations in the hours analyzed by Joskow and Kahn, as the

---

103 The 1,695 MW differs slightly from the number in the table due to rounding.

104 A number of apparent examples of such partial outages or deratings are found in the recent FERC Report on California outages, FERC Outage Report, pp. 8-9, 27-28, 43, 49.

105 To the extent that units face incremental costs above the thresholds, they should be removed from the analysis of withholding.

actual data regarding the level of capacity offered to the market by each unit is readily available to the CAISO and thus presumably to FERC.

4. Understated Ancillary Service Requirements.

Joskow and Kahn attempt to account for the CAISO's ancillary service requirements in their assessment of whether economic withholding occurred by accounting for the amount of capacity in SP-15 and NP-15, separately, required for ancillary services. Their approach is to compare their estimate of the total output gap to the total ancillary service "requirements" or "demands" for the zone and to infer the existence of possible withholding if the output gap exceeds the ancillary service "requirement." As they note, such an approach would be conservative because it would implicitly assume that all ancillary services were provided by California thermal units, when in practice some portion of the ancillary services were likely provided by external units, internal hydro or geothermal resources or quick-start units not included in the CEMS data.[106]

Although Joskow and Kahn do not precisely identify the data they utilize to measure ancillary service "demand," their figures can be nearly exactly reproduced by averaging the CAISO published data for hour-ahead ancillary service procurement from generators for NP-15 (excluding SF and Humboldt) and SP-15 (excluding Z-26),[107] as shown in Table 18. It appears from this replication that the Joskow-Kahn methodology is not quite as conservative as their description suggests. First, the data they apparently utilize is for ancillary services procured from generation located within the zone, rather than total demand for ancillary services, and therefore does not include ancillary services scheduled to be provided by load or imports. Second, the figures cited by Joskow and Kahn for ancillary services demand in NP-15 and SP-15 do not include ancillary services procured in the SF or Z-26 zones, but their calculation of the output gap includes thermal generation located in SF and Z-26. Inclusion of the ancillary services procured from generation located in SF and Z-26 increases the ancillary services procurement calculated by Joskow and Kahn by somewhat more than 100 MW, as shown in Table 18.

[106] See Joskow-Kahn, p. 26.

[107] See www.caiso.com/marketops/oasis/pubmkt2.html. Hour-ahead ancillary procurement (button 20).

Table 18 Mean Hour-Ahead Ancillary Service Procurement (MW) from Generation by Zone and Region June 2000, PX Price > $120/MWh						
	NP-15		North[1]	SP-15 (excluding Z-26)		
	Joskow-Kahn (A)	Replication (B)	(C)	Joskow-Kahn (D)	Replication (E)	Humboldt (F)
Regulation	--	407	413	--	519	0
Spin	--	471	496	--	142	13
Non-Spin	--	181	198	--	387	1
Total, excluding replacement	--	1,058	1,107	1,044	1,048	15
Replacement	--	448	502	628	629	2
Total	1,510	1,506	1,610	1,672	1,678	16
Unadjusted Output Gap	983	983	983	3,351	3,351	16

[1] North includes NP-15, SF and Z-26.

Sources:
(A), (D): Joskow-Kahn, Tables 8 and 13
(B), (C), (E) and (F): www.caiso.com\marketops\OASIS\pubmkt2.html

As discussed above, it is also desirable to revise this calculation to exclude the hours in which day-ahead prices were high, but real-time prices were not high and thus generation was dispatched down in real-time. This information is portrayed in Table 19. Perhaps because these hours were somewhat higher load than the hours in which real-time prices were lower, the procurement of ancillary services other than replacement reserves from generators averaged slightly higher in these hours than in the hours analyzed by Joskow and Kahn (2,224 MW versus 2,155 MW). More materially, the CAISO acquired far more replacement reserves on average during these hours than in the 137 hours analyzed by Joskow and Kahn, so that the total ancillary services procurement for generation corresponding to the generation assets included in calculating the output gap was 3,690 MW in the high real-time price hours compared to 3,288 MW in the high priced PX hours (1,610 + 1,678) and the 3,182 MW figure originally calculated by Joskow and Kahn (1,510 + 1,672).[108]

[108] Joskow-Kahn, Table 8, p. 26.

Table 19 **Ancillary Service Procurement (MW) from Generation by Zone and Region June 2000, PX and Real-Time Price > $120/MWh**			
	North[1] (A)	**SP 15 (B)**	**Total**
Regulation	433	535	968
Spin	491	153	645
Non-Spin	201	411	612
Total, excluding replacement	1,125	1,100	2,224
Replacement	621	845	1,466
Total	1,746	1,945	3,690
Output Gap			
No Outage Adjustment	779	3,100	3,879
Adjusted for Outages	598	1,696	2,294

[1] North includes NP-15, SF and Z-26.

Source: www.caiso.com\marketops\OASIS\pubmkt2.html

As Joskow and Kahn note, since replacement reserves were acquired by the CAISO in part to meet load, rather than to provide 10-minute reserves or regulation, replacement reserve procurement does not necessarily give rise to a real-time "output gap."[109] If all of the capacity providing replacement reserves were used to meet real-time load, there would be no output gap in the CEMS data attributable to the replacement reserve capacity. There are, however, several complications in comparing the calculated output gap to ancillary service procurement from generation, excluding replacement reserves.

First, it needs to be recognized that the CAISO procured a substantial proportion of its 10-minute non-spinning reserves from load (133 MW in the north, 287 MW in the south, 420 MW total), or from imports in these high real-time priced hours (see Table 20).[110] We have not been able to identify publicly available information that indicates whether operationally the CAISO interrupted the loads providing non-spinning reserves before dispatching replacement reserves (and thus used the replacement reserves to replace the loads as 10-minute reserves) or after dispatching replacement reserves (and thus dispatched the replacement reserves to meet load and kept the load reserves in the form of reserves). Similarly, it is not known whether the CAISO dispatched imports providing 10-minute reserves before dispatching replacement reserves (and thus used the replacement reserves to replace the imports as 10-minute reserves) or after

[109] Joskow-Kahn, p. 26.

[110] Imports of spin and non-spin average 394 MW in the 86 hours in which both PX and real-time prices exceeded $120/MWh.

dispatching replacement reserves (and thus dispatched the replacement reserves to meet load and kept the import capacity in the form of reserves).

Furthermore, it is not known whether the CAISO sought to maintain the location of its non-replacement reserves in real time or allowed the location of these reserves to shift as it met load, so that the average real-time geographic distribution of reserves might differ from the geographic distribution in the hour-ahead schedules.[111] It can be seen in Table 20 that the total procurement of ancillary services, excluding replacement reserves, is considerably in excess of the procurement from generation. Without knowing how the CAISO used these ancillary service resources in real time it is not possible to determine how much of the capacity included in the calculated real-time output gap for the various subregions was in fact providing ancillary services in real time.

---

[111] Joskow-Kahn consider the impact of transmission congestion on their estimated output gap; however, they only consider the impact of congestion on inter-zonal dispatch of energy. They observe that there was no South to North congestion during most of the hours in which they identify an output gap and apparently draw the conclusion that the dispatch of energy to meet load was not impacted by these constraints (pp. 30-31). These data, however, are also consistent with the possibility that the CAISO met NP-15 load with NP-15 generation, including generation procured as reserves, and carried more of its real-time reserves in SP-15. Thus, part of the reason that the real-time output gap in the north is so much less than ancillary service procurement may be that in real-time during some of these high-priced hours the CAISO disproportionately used its reserves in the north to meet load, resulting in more of its real-time reserves being carried in the south.

Table 20 Ancillary Service Procurement and Requirements (MW) June 2000, PX and Real-Time Price > $120/MWh						
	Ancillary Service Procurement from Generation					Total Ancillary Service Requirements (F)
	North (A)	SP-15 (B)	Imports (C)	Total (D)	Humboldt (E)	
Regulation	433	535	0	968	0	634
Spin	491	153	279	923	9	1,080
Total Non-Spin	334	698	115	1,147	2	1,115
Gen Non-Spin	201	411	115	728	2	
Load Non-Spin	133	287	0	420	0	
Total, excluding replacement	1,258	1,387	394	3,038	10	2,829
Output Gap						
Not Outage Adjusted	779	3,100		3,879		
Outage Adjusted	598	1,696		2,294		
Source: Output Gap, Tables 15 and 17. Ancillary service data: www.caiso.com\marketops\OASIS\pubmkt2.html North includes NP-15, SF and Z-26.						

Overall, the CAISO was maintaining 3,000 MW of regulation room and 10-minute reserves somewhere on the system during the 86 high-priced hours but the data utilized by Joskow and Kahn are not sufficient to resolve the location at which these reserves were carried in real time, particularly in hours in which large amounts of replacement reserves were also available for dispatch. Moreover, since the CAISO apparently does not publish real-time load by zone, it has not been possible to compare the average level of replacement reserves by zone with the amount of load met with reserves in real time in each zone.[112]

Second, the amount of capacity procured as replacement reserves does not necessarily correspond to the amount of capacity needed to meet real-time load. The reported figure for procurement of replacement reserves will understate the amount of reserves potentially used to meet real-time load to the extent that the CAISO purchased capacity to meet its replacement reserve target as regulation, spinning reserves or 10-minute reserves under the rational buyer

[112] Even if such zonal load data were available, it would probably be necessary to analyze the data hour by hour to draw any inferences as to whether real-time load could have been met solely with replacement reserves located in that zone.

protocol. This consideration would cause replacement reserve procurement to understate the amount of reserves actually used to meet load in real time and thus overstate the amount of any output gap that is attributable to ancillary service requirements. An alternative approach to measuring ancillary service requirements that would address this potential overstatement would be to compare the calculated output gap to the total ancillary service requirement (excluding replacement reserves), rather than procurement, as shown in Column F of Table 20. A comparison of Columns (F) and (D) suggest that the figure for ancillary services excluding replacement reserves calculated from hour-ahead ancillary service procurement likely includes, on average, around 200 MW of replacement reserves bought as regulation, spin or non-spin.

In addition, the quantity of reserves procured by the California ISO during June was not always sufficient to meet load and maintain the minimum level of operating reserves.[113] The amount of capacity actually providing reserves in real time would therefore have been less than the requirement identified in Table 20 during these shortage hours, so a comparison of ancillary requirements to the output gap could still overstate the amount of capacity providing ancillary services in real time.

The final reason that the amount of capacity procured as replacement reserves may not correspond to the amount of capacity needed to meet real-time load is that neither data on ancillary service procurements nor requirements informs us as to whether the California ISO's replacement reserve requirement, or target, was on average above or below the amount of additional capacity that would be required to meet load and reserve targets. To the extent that the CAISO systematically over or underestimated load, sought to provide a capacity cushion in excess of expected load,[114] or sought to be very conservative in scheduling replacement reserves, the replacement reserve target would differ from the amount of reserve capacity that was on average used to meet real-time load.

The third complication in comparing the calculated output gap to ancillary service requirements and procurement is that it needs to be kept in mind that an unknown amount of reserves and regulation would have been provided during these hours by units not included in the Joskow-Kahn analysis, including hydro units, geothermal units, and small units not included in the CEMS database. To the extent that these other sources were actually used to meet ancillary services requirements, more of the calculated output gap could reflect withholding or other factors.

Given these considerations, the output gap methodology has a relatively large margin of error in assessing the amount of capacity actually used in real time to provide ancillary services on the units analyzed. While the CAISO dispatch data should reveal which undispatched capacity was providing reserves in real time, there is a rather large margin of error in the Joskow-Kahn approach, ranging from thousands of megawatts of ancillary services that might have been

---

113 Data on reserve shortages do not appear to be publicly available. It is possible to infer from a variety of data that is publicly available that California was short of reserves during many hours in June 2000 (see Harvey-Hogan, pp. 22-25).

114 In general, the CAISO operating procedures provide that the CAISO will procure 300-700 MW of replacement reserves (in excess of those required to meet load) (see ISO Operating Procedure M-402, July 14, 2000).

provided by off-line thermal units, or other units not included in their analysis (including hydro and geothermal units) to thousands of megawatts of additional ancillary services that might have been provided by the thermal units they analyze, replacing imports or load in providing spinning and non-spinning reserves. It is evident from Tables 17 and 20 that the output gap, adjusted for unit outages, is substantially less than the total reserves including replacement reserves procured from internal generation and substantially less than total ancillary service requirements excluding replacement reserves. The entire output gap could therefore be accounted for by ancillary service requirements (deratings, overstated capacity and the other factors discussed below). Conversely, these data also do not allow one to rule out the possibility that much of the real-time ancillary service requirements were being carried on units not included in the CEMS data and the output gap, in part, reflects economic withholding. Hence, the data could be consistent with either a fully competitive market with no withholding or with a market in which withholding raised prices. Without access to the dispatch data, the margin of error looks to be larger than the effect to be measured.

5. Uneconomic Energy

The Joskow-Kahn analysis does not differentiate between capacity that is economically or physically withheld in order to exercise market power and that which is economically withheld because the market price is less than the unit's incremental costs. Joskow and Kahn argue that all of the units analyzed in the CEMS data had incremental costs in June that were less than $120/MWh, but acknowledge that this was not the case for the units included in their analysis for subsequent months, particularly August and September.[115] Although Joskow and Kahn assert that this limitation does not apply to their results for July,[116] given the low threshold used to define a high priced day in July ($90), it appears likely that many units would have had incremental gas and allowance costs above the threshold. Indeed, one of the peculiarities of the withholding analysis is that lower price thresholds are used in the analysis for the months of July and September, even though gas and NOx allowance costs are both known to be higher than in June.

Even the $120 figure for June needs to be interpreted cautiously. First, it is not clear that the time path of RTC prices can be traced with as much precision as Joskow and Kahn suggest. Most of the transactions are recorded during the two-month settlement period following the end of the cycle. Not only are there relatively few transactions recorded during June, but the link between the record date and the price negotiation date is murky. While it may be reasonable to conclude that arm's-length RTC prices during June were in the range of $5 to $20 per pound, it is not clear that it can be concluded that the RTC price used in formulating bids during the latter part of June was no higher than $10. Second, Joskow and Kahn base their evaluation of fuel costs on a posited marginal heat rate of 12,000 Btu/kWh. It is possible that this conjecture is correct for marginal heat rates, but the CEMS data indicates that it is far from accurate for average heat rates for many of these units. The marginal heat rate is relevant to assessing whether it would be economic to fully dispatch a unit given that it is on-line, but the economics

---

115 Joskow-Kahn, pp. 25, 32-33.

116 Joskow-Kahn, p. 33.

of committing a unit to operate depend on the unit's average or full load heat rate (and the start-up or minimum-load costs incurred in order to have the unit on-line discussed below). It appears that several units included in the Joskow-Kahn withholding analysis, including Redondo 5 and 6, Alamitos 1 and 2, and Alta 3.1 and 3.2 may have had generating costs, including emission allowances but excluding variable O&M, in the range of $120 to $150/MWh. Some of these units may therefore have been off-line during some of the hours analyzed by Joskow and Kahn simply because they were not economic to operate at the expected day-ahead prices for those hours.

Another factor that is not taken into account in the Joskow-Kahn withholding analysis is the increasing impact of environmental restrictions as the summer progressed. The reality is that a number of the units included in the Joskow-Kahn withholding analysis reached environmental operating limits before the end of the year. These units were not engaged in anticompetitive withholding; on the contrary, we know with the benefit of hindsight that they failed to engage in enough economic withholding to stay within their operating limits. These plants should have been bid in at higher prices than was actually the case during the first part of 2000. No allowance is directly made in the Joskow-Kahn analysis for the underpricing of this capacity.

This kind of economically efficient output allocation would be treated the same as market power withholding under the Joskow-Kahn Test 1 approach to analyzing withholding in June 2000, as the units would be counted as withholding if they were on-line but not operating at capacity and the Joskow-Kahn analysis includes no other allowance for the impact of annual capacity factor restrictions.[117] The owners of the units that ultimately reached environmental operating limits would have at some point during the summer begun factoring these environmental restrictions into their scheduling decisions, operating the units only during the highest-priced hours, with large margins.[118] Since the Joskow-Kahn analysis of July-September includes many relatively low-priced hours with low margins, units subject to annual operating restrictions would be expected in a competitive market to be off-line or operating at reduced capacity factors in these hours to conserve their hours of operation for hours of extreme capacity shortage.

6. Start-up and Minimum-Load Costs

Unlike electricity markets in PJM and New York, the California day-ahead markets for energy (PX) and ancillary services (ISO) were based on one-part bids evaluated on an hour-by-hour basis. A day-head PX price of $120 might cover the variable cost of incremental output, yet it might be uneconomic for the unit owner to start that unit or keep it on-line overnight in order to

---

[117] The data source used for analysis of the July-September period includes more units subject to very strict run time limits. Indeed, some of these units had fewer hours of allowed operation per year than the number of high-priced June hours analyzed by Joskow and Kahn.

[118] It is unclear from publicly available data when market participants began to recognize the potential to exhaust their NOx allowances. While AES did not mention the potential to exhaust its NOx allowances in its second quarter report, its third quarter report disclosed this possibility. The AES Corporation, Form 10Q for the period ending September 30, 2000, p. 13. Understanding of AES unit constraints is also complicated by the terms of the forward sale of the capacity to Williams.

sell energy at a price of $120, particularly if the price exceeded $120 for a relatively small number of hours.

Since all of the units included in the CEMS data are steam units with start-up and minimum-load costs, it is incorrect to infer that a competitive firm would always be willing to offer supply from these units at a day-ahead price of $120/MWh, even if $120/MWh accurately reflected the incremental running cost of the unit. As discussed previously in Harvey-Hogan,[119] the day-ahead one-part offer price of a unit lacking market power but having start-up and no-load costs depends in a complex way on expected prices during that hour and other hours of the day. One approach to accounting for start-up and no-load costs would be to use daily strip prices to determine whether a unit would have been economic to operate for the day as a whole,[120] and then analyze real-time output hour by hour for the hours with high real-time prices.

An alternative approach to controlling for start-up and no-load costs would be to restrict the analysis to the output decisions of units that were actually on-line in real-time. Start-up and no-load costs are irrelevant for units that are actually operating, as those costs are sunk in real time. This approach to accounting for start-up and no-load costs corresponds to the Joskow-Kahn Test 1 for measuring available capacity, as this test would only include the capacity of units on-line during the hour. Part of the large difference between the unadjusted output gap Joskow and Kahn calculate and the output gap calculated using Test 1 is, therefore, potentially due to the impact of start-up and no-load costs and high average operating costs, as well as outages. This needs to be kept in mind in evaluating the calculated output gap for the months of July, August and September as Joskow and Kahn do not apply Test 1 to measuring available capacity in these periods, yet the higher operating costs during these periods (due to higher gas and NOx allowance costs) could raise the price at which starting such a unit or keeping it on overnight would be economic for a competitive firm.

7. CAISO Dispatch Instructions

There are several respects in which capacity included in the output gap as calculated by Joskow and Kahn might reflect units dispatched down in real time by the CAISO (aside from the ancillary service requirements discussed above) despite bids less than $120/MWh. First, units submitting $120/MWh or lower adjustment bids would have average output that is less than their capacity in hours in which the average hourly real-time price exceeded $120/MWh if prices varied during the hour and the units operated at different levels during the hour. Second, units submitting $120/MWh or lower adjustment bids could be dispatched down by the California ISO in hours in which real-time prices exceeded $120/MWh as a result of intra-zonal congestion. Third, units that submitted real-time adjustment bids of less than $120/MWh but experienced difficulty ramping or maintaining maximum output or were not dispatched by the ISO due to a slow response time could have average hourly output that is less than their maximum output during hours in which real-time prices exceeded their adjustment bid. Fourth, it is a design

---

[119] See Harvey-Hogan, pp. 14-16.

[120] This evaluation would compare the cost of keeping the unit on overnight or its start-up costs (whichever is lower) to its expected on-peak margin.

feature of the California market that units submitting $120/MWh or lower adjustment bids would not necessarily be dispatched up by the California ISO in an hour in which the real-time price exceeded $120/MWh. There is some information available on real-time price variation; however, there do not appear to be any publicly available data that would enable one to assess directly the practical impact of the second, third and fourth effects during the summer of 2000.

Consider first the impact of price and output averaging. The Joskow-Kahn methodology implicitly analyzes the relationship between unit capacity and average unit utilization in the hour. The average output of the unit, however, is limited by the peak output, and thermal steam units such as those analyzed by Joskow and Kahn would likely have been providing load following capability to the CAISO. Thus, a unit might have been fully dispatched at the intra-hour load peak at which the price exceeded $120, but would be less than fully dispatched on average over the hour if prices fell below $120/MWh during other dispatch intervals. It is apparently not possible to develop estimates of the potential magnitude of the effect by comparing average hourly load to the intra-hour peak load, because the CAISO does not publish data on intra-hour load levels. The magnitude of the change in average loads from hour to hour often exceeds 1,000 MW on peak days, however, suggesting that differences between peak and average hourly load could exceed 500 MW. Moreover, of the 85 or 86 hours in which both the PX and hourly real-time prices exceeded $120, 25 were adjacent to an hour in which the hourly real-time price was less than $120, 15 were adjacent to hours in which the hourly real-time price was less than $80 and 4 were adjacent to hours in which the hourly real-time price was less than $20/MWh, suggesting that steam units likely would have been ramping up or down during a number of these hours.

A non-systematic examination of the CAISO ex-post 10-minute prices on high-priced days in June 2000 (www.caiso.com\cgitbin\pubmrt2.cgi) reveals instances in which hourly real-time prices (the zonal uninstructed energy price) exceeded $120/MWh and the BEEP prices ranged from zero to $400 or more within the hour or within six to eight intervals across hours. It is not surprising that there would be an "output" gap in circumstances in which the dispatch price first falls to zero for 20 to 40 minutes and then suddenly rises.

The significance of this effect has been tested by recalculating the output gap for the hours in which the real-time hourly price was both high and units with bids below $120/MWh likely would not have been ramping up or down (i.e., hours in which the real-time price exceeded $120/MWh in the preceding and following hour as well). The results of this calculation are shown in Table 21, and it is seen that although the available capacity of on-line units is 185 MW higher in these hours, output is 590 MW higher so the apparent output gap is reduced by 405 MW (to 1,889 MW), suggesting that a substantial portion of the apparent output gap may be due to units ramping up or down during hours when prices are rising or falling, rather than due to economic withholding.

Table 21 Output Gap (MW), Adjusted for Outages and Ramping						
	High-Priced RT Hours			Stable High-Priced RT Hours		
Owner	Maximum Output (A)	Mean Output (B)	Output Gap (C)	Maximum Output (D)	Mean Output (E)	Output Gap (F)
NP-15, SF and Z-26						
Duke	2,528	2,416	112	2,535	2,437	98
Southern	2,786	2,300	486	2,816	2,520	296
Total North	5,314	4,716	598	5,351	4,957	394
SF and Humboldt						
PG&E	109	100	9	109	105	4
SP-15						
AES	2,940	2,589	351	2,975	2,643	332
Duke	719	663	56	729	694	35
Dynegy	1,596	1,125	471	1,622	1,235	387
Reliant	3,242	2,424	818	3,319	2,578	741
Total SP-15	8,497	6,801	1,696	8,645	7,150	1,495
Source: (A) – (C) Table 17 (D) – (E) CEMS Data (F) = (D) – (E)						

It is also possible to use these data to test whether the calculated outage adjusted output gap in the hours in which both the PX and real-time prices exceeded $120/MWh is statistically different from the calculated outage adjusted output gap in the hours in which both the PX and real-time prices exceeded $120/MWh and the real-time price was high both before and after the hour in question. Once again, for the purpose of this test we have dropped the hour included in the Joskow-Kahn analysis in which the hypothetical unconstrained PX price exceeded $120/MWh but the actual SP-15 PX price was less than $120/MWh.[121] For the combined North and South regions the mean outage adjusted output gap in the hours in which both the PX and real-time price exceeded $120/MWh but the real-time price was lower in either the preceding or following hour was 3,017 MW but only 1,881 MW in the hours in which the real-time price exceeded $120/MWh both before and after the hour as well. The difference is 1,136 MW which is statistically significantly different from zero at above the 99.9 percent confidence level (the t-statistic is 5.40).

---

[121] This is the reason that the mean for NP-15 used in the test differs slightly from the mean in Table 21.

For the South alone the difference in means is 552 MW (2,045 to 1,493 MW), which is also statistically significantly different from zero at more than the 99 percent confidence level (the t-statistic is 3.35). Finally, for the North alone the difference in means is 584 MW (972 to 388 MW), which is statistically significantly different from zero at above the 99.9 percent confidence level (the t-statistic is 7.54).

Table 22 compares the recalculated output gap in the high-priced non-ramping hours to ancillary service procurement from generation in those same hours. It is seen that the output gap is a few hundred MW less than ancillary service procurement from generation excluding replacement reserves and far less than ancillary service procurement including replacement reserves.[122] As noted above, generation scheduled to provide replacement reserves might have been dispatched to meet load in real time or might have provided reserves while other capacity was dispatched to meet load. Again, it is seen that the potential variation in the size of the factors included in the output gap is too large, relative to the size of the calculated output gap, to permit findings that output either has or has not been economically withheld from the market.

Table 22

Output Gap and Ancillary Service Procurement (MW) from Generation by Zone and Region

June 2000, PX and Real-Time Price > $120/MWh

	Non-Ramping Hours		
	North	**SP-15**	**Total**
Regulation	347	612	959
Spinning Reserves	494	169	664
Non-Spinning Reserves	196	396	593
Total Excluding Replacement	1,038	1,178	2,216
Replacement Reserves	758	1,095	1,854
Total A/S	1,796	2,273	4,069
Outage Adjusted Output Gap	394	1,495	1,889

North includes NP-15, SF and Z-26.

Source: Output Gap Table 21.

Ancillary Service Procurement: www.caiso.com\marketops\OASIS\pubmkt2.html

122 Ramping constraints likely also contributed to an apparent output gap even within the stable high-priced real-time hours. Some of the capacity that is on-line in these hours had only recently come on-line and was likely ramp-constrained in providing energy and reserves. Thus, an average of 150 MW/hour of capacity in these high-priced hours had started up within the prior four hours and roughly 80 MW of the output gap in Table 22 was on these units.

Second, intra-zonal congestion management can require that generators at particular locations be dispatched down by the CAISO despite zonal prices that exceed the unit's bid. In these circumstances, there is a generation pocket in which some resources cannot be utilized to provide either energy or ancillary services. It is therefore possible for generator offers below $120 to not be accepted for dispatch at times that the zonal price exceeds $120.[123] Because of California's zonal pricing system, the published price data do not indicate whether there were any such generation pockets during the high-priced days in June.

The potential for such output-reducing generation pockets can be seen, however, in the publicly available nodal prices of other ISOs. For example, June 26, 2000 was the highest price day in the New York day-ahead market with prices east of Central East exceeding $1,000/MWh, yet there were also a number of generation pockets east of Central East with locational prices below $100/MWh because of transmission constraints that prevented the generation at these locations from being fully scheduled to meet load or to provide reserves.

Third, during high-priced hours in which every unit not providing reserves would ideally be dispatched to its maximum capacity, any inability of units to either ramp quickly up to maximum in the initial high priced hour or to sustain maximum output over a long period of time would translate into an output gap for that unit. Alternatively, a slow responding unit might be skipped by the CAISO and the interval price set by a high-cost unit with a faster response time. There would be an offsetting increase in output by other units including those providing regulation but the source of the replacement output could be units not included in the CEMS data. Fourth, it is a design feature of the California market that the California ISO's real-time dispatch does not clear the market. In particular, the ISO's software has been specifically designed so that it will be the case that if the zonal price is $120 and there is no intra-zonal congestion, it is possible for there to be adjustment Dec bids above $120 that have not been accepted or Inc bids below $120 that have not been accepted. The potential for unaccepted Inc bids below $120 would likely be greatest if prices were falling. Of the 85 and 86 hours with day ahead and average real-time prices exceeding $120, 29 were followed by hours in which the average real-time price was at least $20/MWh lower.

8. Unit Capacity

The Joskow-Kahn study estimates unit capacity based on the maximum output in the quarter. This methodology also introduces imprecision into the calculation of the output gap as it has the potential to either overstate or understate actual capacity. First, it may overstate capacity because it does not account for the impact of ambient temperature on unit capacity. Many of these capacity estimates are based on unit outputs during April and May. Because the capacity of electric generating units can depend on ambient temperature, the Joskow-Kahn methodology

---

[123] Joskow and Kahn assert that "intrazonal constraints would not affect production levels within a zone" (p. 30), but this is incorrect. This might be true under the simplified assumptions of the zonal model, but it is not true in the real dispatch. So-called "intra-zonal" congestion can be managed by interruptible load and extra-zonal generation as well, and both effects would lead to lower total zonal production. This would be more likely to be true during high-priced hours when all thermal generation units are supposedly already at their maximum level of output.

for estimating capacity has the potential to overstate capacity if the maximum output occurred on a day with lower ambient temperatures. The effect of temperature on capacity, however, would typically not be large for the steam units included in the CEMS data for June.[124]

Second, the Joskow-Kahn methodology for determining capacity could also potentially understate capacity, if part of the capacity of the unit were consistently used to provide reserves, not energy. Thus, if a generation owner consistently bid the top range of one of its units into the market for reserves and submitted a very high energy adjustment bid in real time, it is possible that this capacity could go through the quarter without ever being dispatched.

Third, the Joskow-Kahn methodology for determining capacity, particularly the reliance on the highest output achieved by the unit, even if achieved only in a single hour, has the potential for basing the capacity estimate on anomalous data, atypical operating conditions or output levels that can be sustained only for short periods of time.

Fourth, Joskow and Kahn apparently applied their maximum output methodology on a unit-by-unit basis rather than a station or plant basis. In cases in which some or all of the units composing a station share some facilities, some or all of the individual units may be able to achieve higher than nameplate outputs if the other units in the plant are off-line and the plant is operated to maximize the output of the individual unit. The aggregate of these individual unit outputs could then, however, exceed the capacity of the plant, and capacity estimates based on the sum of the maximum hourly output of the individual units, could overstate plant capacity.

An effort was made to assess the potential direction and magnitude of errors in measuring capacity by comparing the unit capacities estimated based on the Joskow-Kahn methodology with other data on unit capacity. Data on the nameplate capacity of the Mirant units are available on the Mirant website as well as in the Klein and EIA reports, permitting a comparison with the Joskow-Kahn figures as shown in Table 23. For these units the Joskow-Kahn estimate of the capacity of Contra Costa 6 and 7 is 691 MW, which exceeds the capacity figures of 676-680 MW in the other sources. On the other hand, it can be seen that these units operated at the rates used by Joskow and Kahn for a non-trivial number of hours. Joskow and Kahn estimate the capacity of Pittsburgh 1-4 at 645 MW compared to a nameplate capacity of 652 MW in the other sources. These output rates, however, appear to have very rarely been achieved. Joskow and Kahn then estimate the capacity of Pittsburgh 5 and 6 at 663 MW compared to a nameplate capacity in other sources of 650 MW. These units appear to have rarely achieved the capacities used by Joskow and Kahn but did operate above their nameplate capacity for a non-trivial number of hours. The Joskow-Kahn capacity estimate for Pittsburgh 7 is in line with other sources but the unit rarely operated at this level. Finally, Joskow and Kahn estimate the capacity of Potrero 3 at 213 MW compared to a nameplate capacity on the Mirant website and in Klein of 207 MW. It is evident that there is a range in capacity estimates and that the highest output a unit has achieved under perhaps ideal conditions is not necessarily a good indicator of the output it can generate on a consistent basis. The effects are likely not large but reflect an additional increment of output gap in Table 22 that is not appropriately attributed to economic withholding.

[124] The effect can be large for combined-cycle and combustion turbines but the EPA/CEMS data do not include combustion turbines.

Table 23 Capacity and Output (MW)								
	Maximum Gross Output 1Half 2000 (A)	**Hours at Maximum 1Half 2000 (B)**	**Maximum Gross Output 2Q 2000 (C)**	**Hours at Maximum 2Q 2000 (D)**	**EIA Net Summer Capability (E)**	**Klein (F)**	**Southern Energy Website (G)**	**Maximum Gross Output Min 1% Hours (H)**
Contra Costa 6	346	227	346	14	339	340	340	346
Contra Costa 7	345	40	345	5	337	340	340	344
Pittsburgh 1	163	1	161	1	163	163	163	160
Pittsburgh 2	163	6	160	8	163	163	163	159
Pittsburgh 3	163	1	163	1	163	163	163	160
Pittsburgh 4	161	8	161	3	163	163	163	157
Pittsburgh 5	332	4	332	2	325	325	325	329
Pittsburgh 6	331	2	331	2	325	325	325	329
Pittsburgh 7	721	1	720	1	682	720	720	716
Potrero 3	213	2	213	2	260	207	207	211
Total	2938		2932		2920	2909	2909	2911

Sources:

(A) CEMS data, maximum gross unit output Jan 1 – June 20, 2000.
(B) CEMS data, maximum number of hours rate in Column (A) was achieved in first half 2000.
(C) CEMS data, maximum gross unit output April 1 – June 30, 2000. Joskow-Kahn.
(D) CEMS data, number of hours rate in Column (C) was achieved in second quarter 2000.
(E) EIA, "Inventory of Nonutility Electric Power Plants in the United States 1999," November 2000. www.eia.doc.gov/cneaf/electricity/ipp/ipp2.pdf.
(F) Joel Klein, "The Use of Heat Rates in Production Cost Modeling and Market Modeling," April 17, 1998.
(G) www.Southernenergy.com
(H) CEMS data, maximum gross output achieved in 1% or more of hours on-line Jan 1 – June 30, 2000.

9. Market Inefficiency

The day-ahead and real-time markets coordinated by the CAISO and CA PX operate in important respects as pay-as-bid markets.[125] The need for generators to bid the market-clearing price in order to avoid price discrimination applied through the rational buyer rule and ensure that they are paid the market-clearing price creates a degree of market inefficiency that may appear much the same as anticompetitively motivated withholding.[126] Most of these bidding incentives would not affect clearing prices in a world with perfect foresight. Moreover, these incentives are likely to be more important in the day-ahead markets than in real time. In the real world with imperfect foresight and considerable price volatility, however, the pay-as-bid market

125 These features are discussed in detail in Harvey-Hogan, pp. 4-14.

126 A critical difference is that the market inefficiency affects the incentives and behavior of all firms participating in the market, not merely the largest net supplier, and can be eliminated only by changing the market design.

features have the potential to create inefficiency and raise prices. For instance, in a simulation of the U.K. market, Bower and Bunn found that switching from a market-clearing to a pay-as-bid auction format increased prices from 100 to 200 percent higher during peak periods.[127] This outcome is particularly likely in potential shortage conditions in which the price cap is binding. Moreover, it has long been recognized that the real-time dispatch by the CAISO contains artificial restrictions that are intended to prevent the CAISO from undertaking least-cost dispatch and rationalizing the supply and demand for energy in real-time.[128] The inevitable and intended consequence of these restrictions on least-cost dispatch is to ensure that low-cost generation capacity at times goes utilized in favor of other higher-cost resources. The impact of these restrictions is thereby to restrict output and raise market prices, but this is the intended outcome of the market design, not the result of anticompetitive behavior by market participants.

10. Summary

The output gap calculated by Joskow and Kahn includes the effects of many considerations, including ancillary service requirements, real-time price variations, unit outages, unit deratings, environmental limitations, CAISO dispatch instructions, ramping constraints, effects of minimum load costs, any inability of units to consistently achieve their design capacity and the inefficiency of the California market. All of these effects would show up as economic or physical withholding unless they are accounted for in the analysis. It is seen above that if the output gap is calculated to account for real-time prices, units that are off-line and ramping constraints, the output gap is less than ancillary service procurement from generation other than replacement reserves and far less than ancillary service procurement from generation including replacement reserves or the overall ancillary service requirement, excluding replacement reserves (see Tables 22 and 20). This measure of the output gap still includes the impact of several factors unrelated to economic or physical withholding such as unit deratings, environmental limitations, CAISO dispatch instructions, any inability of units to achieve their design capacity and the inefficiency of the California market. Therefore, the CEMS data do not provide evidence that can distinguish between strategic withholding and other market factors.

At the same time, the calculation of the output gap and ancillary service requirements does not account for the capacity of off-line thermal units such as GTs, reserves provided by hydro or geothermal units or the possibility that the CAISO did not meet its reserve target in all hours. It therefore also cannot be concluded from the demonstration that the output gap is much smaller than reserve requirements, or that all of the capacity included in the output gap was either providing ancillary services, dispatched down by the CAISO, or unable to operate due to unit deratings, environmental limitations or overstated capacities. In other words, the range of error in this approach appears to be larger than the amount of the economic withholding that might have occurred. The result could go either way.

---

[127] John Bower and Derek W. Bunn, "Model-Based Comparisons of Pool and Bilateral Markets for Electricity," *Energy Journal*, Vol. 21, November 3, 2000, pp. 1-29. The higher prices arose from both inefficiencies and from the greater exercise of market power. This huge price increase in the U.K. simulations would be unlikely to result in the California market from inefficiency alone, but it illustrates the problems created by market design.

[128] See footnote 6 above.

V. MARKET DESIGN, PRICES AND INCENTIVES

Joskow and Kahn address a difficult public policy question with publicly available data using two distinct approaches. Both of their approaches face the reality, however, that the range of uncertainty in the publicly available data measuring available capacity, capacity used to provide ancillary services, capacity used to generate energy at the intra-hour peak and capacity dispatched down by the CAISO could exceed the magnitude of any strategic anticompetitive withholding that might have occurred. The CAISO dispatch data would reveal the total available capacity of the on-line units (net of derations, i.e., the upper limit of the units bid in for supplemental energy), and the amount of this capacity that was either used to generate energy, provide ancillary services, or was backed down by the CAISO to manage congestion or balance generation and load. The CAISO dispatch data, moreover, would reveal whether capacity was being economically withheld (i.e., not used to provide energy or ancillary services) in real time during the shortage hours, based on the actual generator schedules and capacity available in each hour. A fundamental limitation of studies based on the CEMS data is that they must be based on very imperfect measures of things the CAISO dispatch data would reveal exactly.

The CAISO dispatch data, however, would not provide the information required to determine whether prices have been affected by physical, rather than economic, withholding. This would require the assessment of the causation of each forced outage or unit derating, assessing the impact of start-up and minimum-load costs on decisions to operate, assessing the impact of annual run restrictions on operating decisions. On this point, the FERC staff has examined outage data for a sample of units and concluded the outages they examined did not reflect physical withholding.

Given the design of the California market, however, even the FERC may not be able to determine from the dispatch data in all instances whether capacity was economically withheld in order to exercise market power, or whether it was economically withheld because of mistaken bids arising from imperfect foresight, which are an inevitable by product of the California market design.

As described by proponents of the California market design: "The ancillary services protocol used in California is strongly dependent on bidders being able to select efficiently between day-ahead energy, ancillary services and real-time markets in selling their scheduled capacity. The primary costs of offering ancillary services to the ISO under the protocol are opportunity costs e.g., the profits foregone in other markets, such as selling into the PX, or offering capacity into the real-time markets without being restricted to providing blocks of reserves. To make efficient ancillary services bids, then, sellers need to be able to gauge their opportunity costs as accurately as possible. Yet this is difficult or impossible given the limited data available."[129] The result of this market design is that ancillary services prices are probably higher than would be the case in an efficient market, as are the costs of providing these services. The obvious reality is that California market participants have never had perfect foresight and never will. Unless they do,

---

[129] Seabron Adamson and Carl Imparato, "Fixing What is Broken: What Steps Are Needed to Complete California's Power Markets?" October 20, 2000, pp. 4-5.

however, the California market will continue to operate the way it does, and action to mitigate market power may be less important than action to fix the faulty market design.

Ultimately, it is impossible to prove the absence of any withholding or any exercise of market power without analyzing the reasons for every outage, derating, and decision not to operate by every supplier, which has not been undertaken by any study. The available information, however, has several elements that suggest that the exercise of market power by California thermal generators is not the primary cause of the high prices in California during 2000-2001. First, electricity prices have been consistently high both inside and outside California, which strongly suggests that the problem is not the exercise of locational market power inside California but a widespread shortage of energy and/or capacity in the WSCC. Indeed, prices have at times been higher outside of California than within California due to transmission constraints on exports. Second, if thermal unit owners were engaged in a simple withholding of generation, then they would not have exceeded the environmental limits on their output. With the benefit of hindsight, it appears indisputable that perfectly competitive thermal generator owners of constrained units blessed with perfect foresight would have offered less capacity into the market from a number of units in many hours during the spring and summer than they actually did, not more capacity, and prices in such a perfectly competitive market would have been higher, not lower, than the actual prices in many of the hours in early 2000. Third, if the high prices in California were attributable to simple withholding by a few thermal generators in California, could these generators be exercising sufficient market power to raise prices off-peak as well as on-peak throughout the entire WSCC?

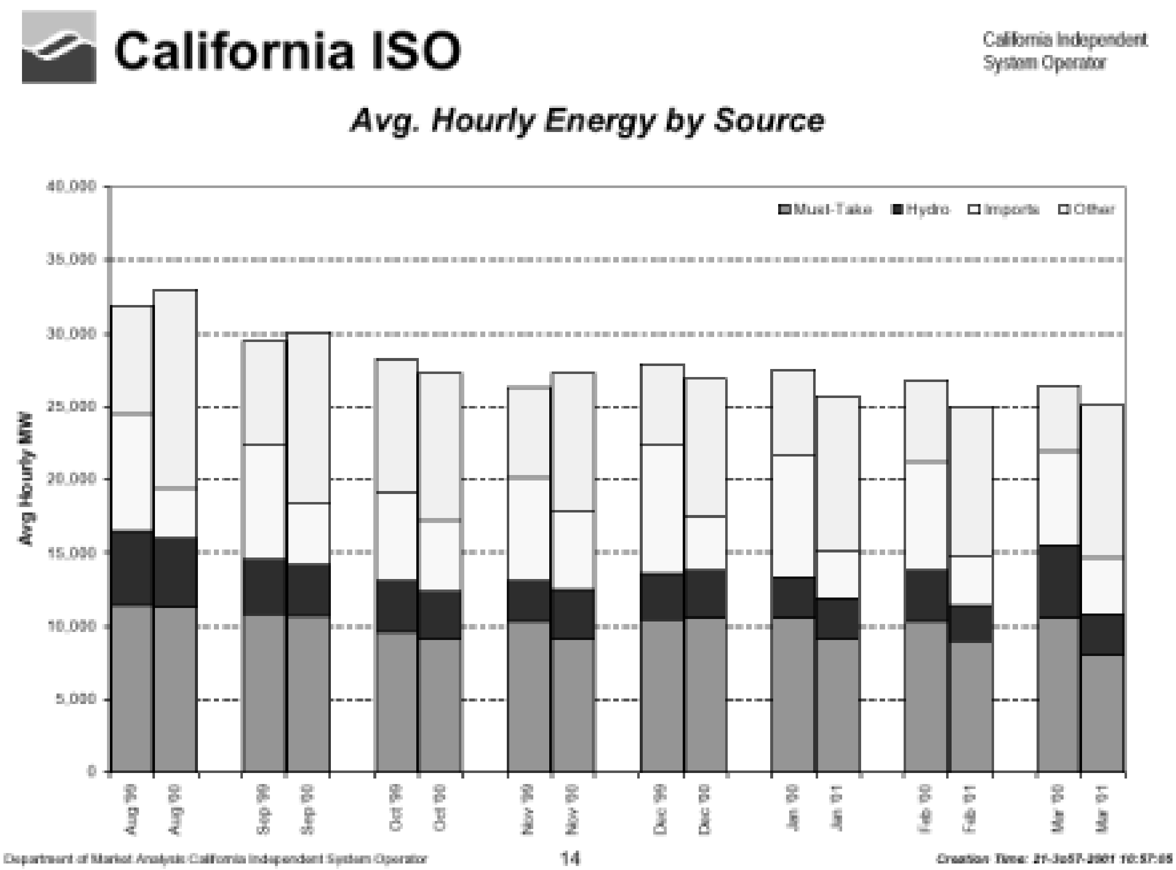
While it is important that allegations of the exercise of market power be carefully investigated, the evidence to date that the high prices in California and the WSCC arise mainly from the exercise of market power by California thermal generators is far from compelling. As Joskow and Kahn have highlighted, many factors contributed to higher prices in California during 2000 and 2001, and the market power theme is only, at most, part of the story. The import of their analysis is not to prove that market power has been exercised but, rather, to suggest that it might be important. The import of the sensitivity analysis here is not to prove that market power has not been exercised but, rather, to suggest that it is unlikely to be the dominant factor and may not even be significant. By contrast, there appears to be little disagreement that other problems of shortage and bad market design are at least large enough to dictate that the solution requires more than just market power mitigation devices. Analysis of the possible exercise of market power should not divert attention from the need for California to:

1. Pay its bills;
2. Raise retail rates, at least at the margin, to reflect the wholesale market price;
3. Assign the financial responsibility for paying electricity bills to someone;
4. Make it clear if and, if so, how, environmental regulations will be modified; and
5. Adopt LMP and reform the wholesale market design.

Similarly, California should provide generators with incentives to contract for gas transportation capacity (spurring the construction of more pipeline capacity), to make investments in emission reductions, to make investments in improved heat rate performance and likely in additional capacity. None of these investments will be made if generators are uncertain of being paid for the electricity they generate or are not permitted to earn a margin above the cost of fuel.

Moreover, generators cannot buy fuel or pay for operations and maintenance if they are not paid for their output. It can be seen in Figure 24 that while the output of thermal generators in California, reflected in the "other" category, has increased dramatically in the winter 2000-2001 relative to the winter of 1999-2000, there has been a corresponding large reduction in the output of California "must-take" generation, as well as hydro and imports. Identifying and reversing the cause of this output reduction ought to be a high priority public policy objective.

Figure 24



Reproduced from Anjali Sheffrin, Market Analysis Report, March 30, 2001.

If the fall in "must-take" generation is largely due to a combination of QF purchase prices that have failed to rise with gas prices and a failure to pay must-take generators for their output, then the recent electricity shortages and accompanying high prices in California (and the WSCC generally) would appear to be in considerable part a consequence of public policy decisions in California.

Even more importantly, it is essential that retail customers be exposed to wholesale market prices at the margin. Despite the rhetoric about inelastic retail demand, San Diego customers responded to high prices last summer by reducing consumption.[130] This behavior indicates that electricity is not worth the cost of production to at least some customers. Retail customers in California have been paying electricity prices that are often less than the cost of the fuel and emission allowances required to generate that electricity. This pricing subsidizes consumption and may create distortions in interfuel consumption decisions.

Although the outcome of the current policy debates in California is uncertain, it is likely that one way or another the customers in California will pay for the power that is consumed; however the payment is being repackaged through asset transfers, bond financing with future surcharges, and use of tax revenue. Importantly, these steps do not protect the customers from the high prices; they only hide and disguise the process of assigning the costs.

Hence, retail customers in California will eventually be asked to make up the difference between the true cost of the electricity they are consuming and the current charges. The current pricing system in effect deprives retail customers of choice and ensures that many retail customers will end up paying for electricity that they would not have consumed if they had the opportunity to avoid these costs by reducing consumption.

Worse, if all loads are insulated from market prices in a shortage situation, the price change required to clear the market will be more extreme, perhaps far more extreme, than if loads were to reduce consumption in response to high prices, even if electricity costs were only passed through on a monthly basis. California's effort to insulate electricity consumers from wholesale market prices has aggravated the increase in wholesale market prices by eliminating the load response that would be required to enable the wholesale market to clear at more moderate prices. Moreover, the retail California electricity prices that are below the fuel cost of generating that electricity create substantial risks of artificially increasing electricity demand on the transmission grid by making it cheaper for customers to buy electricity from the grid than to burn gas or oil to self-generate that electricity.[131]

The most important step that can be taken to moderate electricity prices in the west is to begin flowing wholesale prices through to California end users for their marginal consumption, rather than continuing to subsidize consumption.[132] Moreover, it needs to be recognized that end-users

---

130 James Bushnell and Erin Mansur, "The Impact of Rate Deregulation on Electricity Consumption in San Diego," POWER Working Paper PWP-082, University of California Energy Institute, Berkeley, April, 2001. "A program of real-time applied to even half the customers in the ISO system, if it is credibly committed to and property understood by customers, could produce power savings that allow California to avoid rolling blackouts." (p. 22).

131 The declining supply of energy from regulatory must-take resources since October 2000 may in part reflect this effect.

132 Severin Borenstein, "The Trouble with Electricity Markets (and some solution)," POWER Working Paper, PWP-081, University of California Energy Institute, Berkeley, January 2001. Frank A. Wolak, "A Market (Power) Mitigation Plan for the California Electricity Market," CAISO Market Surveillance Committee Presentation, March 15, 2001.

that reduce consumption and sell that electricity on the spot market are doing exactly what is required to moderate prices in a shortage and should be praised, not criticized and punished.[133]

Less obviously, the California subsidies for end-user consumption cloud the investment outlook for new electric generating capacity, because it is not clear to investors whether retail customers are really willing to pay the cost of adding electric generating capacity to meet their load. As long as customers are paying subsidized prices for electricity, their consumption decisions provide an uncertain guide for investment decisions. Investment decisions that remove consumer preferences from the decision-making process risk repeating the mistakes of the early 1980s when gas pipelines contracted for natural gas at prices that turned out to be higher than consumers were willing to pay.

It is also important that responsibility for paying electricity bills be clearly assigned to someone -- the end-user, the utility, or whoever -- so that those entities can enter into hedging contracts to insulate themselves from price volatility to the degree that is appropriate. This is also important in avoiding the frequent outcome prior to deregulation in which regulators entered into hedging contracts on behalf of customers at prices above what customers were willing to pay.

New investments in emission reduction are also important in reducing the cost of meeting California electricity demand. The incentives to make these investments will be reduced, and some investments deferred, if it is unclear whether current environmental restrictions on emissions will be relaxed. If there will be some relaxation of emission limits, this should be announced as soon as possible and if there will not be any relaxation, this also needs to be clearly announced as soon as possible. Delaying the decision will delay investments and prolong the impact on market prices. The economics of these investments also depend on the consumer response to high prices, as the easiest way of reducing emissions from high emission plants is not to run them because they are not needed. The longer the delay in flowing through wholesale prices into marginal retail prices the longer the likely delay in making environmental investments as well.

Finally, the designed inefficiency of the California market has directly raised prices and also greatly complicated diagnosis of the cause of the high prices. The California ISO should shift immediately to a bid-based least-cost dispatch in real-time and bid-based least-cost congestion management in day-ahead markets.[134] Further, the FERC should recognize that directing the CAISO to maintain minimum WSCC reserve levels at any cost has an important impact on the cost of meeting load in a shortage situation. Rather drive out high-cost supply, it would be better

---

[133] See the discussion of Kaiser, *MW Daily*, December 13, 2000, p. 7.

[134] California Independent System Operator, "Proposed Market Stabilization Plan of the California Independent System Operator Corporation Provided in Response to Letter Order of March 30, 2001," Submission to Federal Energy Regulatory Commission, April 6, 2001. This plan includes some of the suggested reforms, but also introduces other features such as price discrimination between various sources of supply that would likely complicate market operations.

public policy to recognize that every MW of the 3,000 or so MW of regulation and reserves that the CAISO seeks to schedule is not worth $1,000, $750, $500 or $250/MW.[135]

[135] This issue is discussed further in Harvey-Hogan, pp. 25-26. See also John D. Chandley, Scott M. Harvey, William W. Hogan., "Electricity Market Reform in California," Comments in FERC Docket EL00-95-000, Center for Business and Government, Harvard University, November 22, 2000.

BIBLIOGRAPHY

Adamson, Seabron and Carl Imparato, "Fixing What is Broken: What Steps Are Needed to Complete California's Power Markets?" October 20, 2000.

Bower, John and Derek W. Bunn, "Model-Based Comparisons of Pool and Bilateral Markets for Electricity," *Energy Journal*, Vol. 21, November 3, 2000.

Borenstein, Severin, James Bushnell and Frank Wolak, "Diagnosing Market Power in California's Restructured Wholesale Electricity Market," August 2000.

Borenstein, Severin, "The Trouble with Electricity Markets (and some solution)," POWER Working Paper, PWP-081, University of California Energy Institute, Berkeley, January 2001.

Bushnell, James and Erin Mansur, "The Impact of Rate Deregulation on Electricity Consumption in San Diego," POWER Working Paper PWP-082, University of California Energy Institute, Berkeley, April 2001.

California Independent System Operator, Department of Market Analysis Study, "Report on California Energy Market Issues and Performance: May-June 2000," August 10, 2000.

California Independent System Operator, "CAISO 2001 Summer Assessment," March 22, 2001.

California Independent System Operator, "Response of the California Independent System Opertor (sic) Corporation to Letter Order of March 30, 2001," Submission to Federal Energy Regulatory Commission, April 6, 2001.

California Independent System Operator, "Proposed Market Stabilization Plan of the California Independent System Operator Corporation Provided in Response to Letter Order of March 30, 2001," Submission to Federal Energy Regulatory Commission, April 6, 2001.

California Power Exchange Corp, Compliance Unit, "Price Movements in California Electricity Markets," September 29, 2000.

CalPX Compliance Unit, Price Movements in California Electricity Markets, August 31, 2000.

Cardell, Judith B., Carrie Cullen Hitt and William W. Hogan, "Market Power and Strategic Interaction in Electricity Networks," *Resource and Energy Economics*, Vol. 19, 1997.

Chandley, John D., Scott M. Harvey and William W. Hogan., "Electricity Market Reform in California," Comments in FERC Docket EL00-95-000, Center for Business and Government, Harvard University, November 22, 2000.

FERC Office of the General Counsel Market Oversight & Enforcement and Office of Markets, Tariffs and Rates, Division of Energy Markets, Report on Plant Outages in the State of California, February 1, 2001.

FERC, Order Directing Sellers to Provide Refunds of Excess Amounts Charged for Certain Electric Energy Sales During January 2001 or Alternatively, to Provide Further Cost or Other Justification for Such Charges, March 9, 2001.

Harvey, Scott M. and William W. Hogan, “Issues in the Analysis of Market Power in California,” October 27, 2000.

Hildebrandt, Eric, “Declaration of Eric Hildebrandt,” October 2000.

Hildebrandt, Eric, “Analysis of Market Power in California’s Wholesale Energy Markets,” November 21, 2000.

Hildebrandt, Eric, “Market Analysis Report,” November 30, 2000.

Hildebrandt, Eric, "Further Analyses of the Exercise and Cost Impacts of Market Power in California's Wholesale Energy Market," California Independent System Operator, March 2001.

Hogan, William W., “Rethinking WEPEX: What’s Wrong with Least Cost?” *Public Utilities Fortnightly*, January 1, 1998.

Hogan, William W., “A Wholesale Pool Spot Market Must Be Administered by the Independent System Operator: Avoiding the Separation Fallacy,” *Electricity Journal*, December 1995.

Imbrecht, Charles R., Statement before the Senate Select Committee on Business Development, November 29, 1995.

Imbrecht, Charles R., Statement before the FERC, Technical Conference on the WEPEX Applications, August 1996.

Joskow, Paul L. and Edward Kahn, “A Quantitative Analysis of Pricing Behavior in California’s Electricity Market During Summer 2000,” January 2001.

Klein, Joel, “The Use of Heat Rates in Production Cost Modeling and Market Modeling,” April 17, 1998.

O’Neill, Richard, “Rules and Institutions for Imperfect Markets,” Harvard Electricity Policy Group, January 10, 1997.

Ryan, Stu, AES Pacific, Analyst Presentation, February 1, 2001.

Ruff, Larry, “The California PX Auction: Whatever Happened to the ISO and Why Should Anybody Care,” UC Power Conference, March 10, 1997.

Ruff, Larry, “Separation of the ISO from the Power Exchange: Some Structural and Operational Implications,” October 25, 1995.

Sheffrin, Anjali, "Empirical Evidence of Strategic Bidding in California ISO Real Time Market," California Independent System Operator, March 21, 2001.

Sheffrin, Anjali, Market Analysis Report, March 30, 2001.

South Coast Air Quality Management District, Review of RECLAIM Findings, October 20, 2000.

Staff Report to the Federal Energy Regulatory Commission on Western Markets and the Causes of the Summer 2000 Price Abnormalities, Part I, November 1, 2000.

Stoft, Steven, "Analysis of the California WEPEX Applications to FERC," October 8, 1996.

Stoft, Steven, "California's ISO: Why Not Clear the Market," *Electricity Journal*, December 1996.

Stoft, Steven, "Gaming Intra-Zonal Congestion in California," March 6, 1998.

Wolak, Frank, Robert Nordhaus and Carl Shapiro, "An Analysis of the June 2000 Price Spikes in the California ISO's Energy and Ancillary Service Markets," September 6, 2000.

Wolak, Frank A., "A Market (Power) Mitigation Plan for the California Electricity Market," CAISO Market Surveillance Committee Presentation, March 15, 2001.

Woychik, Eric, "California's Schedule Coordinator: Market Maker with Advantage," November 26, 1997.