Efficiency of the California Electricity Reserves Markets

Konstantinos Metaxoglou   Aaron Smith*

September 9, 2005

Abstract

We test the efficiency of the California electricity reserves markets by examining systematic differences between its day- and hour-ahead prices. This efficiency hypothesis test requires estimation of a high-dimensional vector moving average model. We employ a particular state-space representation that allows the EM algorithm to obtain exact maximum likelihood estimates using analytical expressions. In contrast, existing exact maximum likelihood methods require numerical computation of the scores or the Hessian, which is infeasible in our high-dimensional framework. We uncover a significant premium in the day- over the hour-ahead market, which created profitable trading schemes, such as Enron's infamous "Get Shorty" strategy. We attribute this inefficiency to two features of the market design. (*JEL* C22, C12, C13, G14).

Keywords: Electricity, Efficiency Tests, State Space, Vector Moving Averages.

1 Introduction

During the last 15 years, a massive wave of restructuring has transformed the wholesale electricity sector in the United States. This reform created markets that determine prices and help ensure a continuous balance between supply and demand. In California, these markets appeared to operate well from their inception in the spring of 1998 until the summer of 2000, when prices skyrocketed and rolling blackouts ensued. The crisis attracted increased attention on the efficiency of the restructured markets. In this article, we study the state's reserves markets, which

*Metaxoglou: Department of Economics, University of California Davis, One Shields Ave., Davis, CA 95616 (email: kmetaxoglou@ucdavis.edu); Smith: Department of Agricultural and Resource Economics, University of California Davis, One Shields Ave., Davis, CA 95616. (email: adsmith@ucdavis.edu); We would like to thank especially Chris Knittel for invaluable comments and suggestions and Jamal Batakji from the California ISO, as well as Frank Wolak for sharing their knowledge of the markets' operations. Tim Cogley made insightful suggestions in the preliminary steps of the draft. We thank Colin Cameron, Oscar Jorda, Prasad Naik, Robert Shumway and the participants of the Econometrics Seminar at UC Davis for their comments. All remaining errors are ours.

are designed to ensure the availability of sufficient generating capacity at all times. We show that these markets operated inefficiently, not only during the crisis, but also after the crisis.

A trading day in the restructured California electricity sector was divided into 24 intervals, one for each hour. For each of these 24 intervals there was a single day-ahead (DA) energy auction run by the California Power Exchange (PX) before January 2001. By single, we mean that bids to buy or sell for any of the 24 hours of the trading day had to be submitted by the market participants once; by 7:00 the day before. For each of the same intervals, the state's Independent System Operator (ISO) ran 24 separate auctions for real-time energy; participants could submit bids up to 45 minutes before the beginning of each relevant hour. In addition, the ISO procured energy reserves, also known as ancillary services, which gave the ISO the option to call for energy to be delivered at short notice. As in PX, participants submitted bids in the DA ancillary services markets once for any of the 24 hours of the trading day; by 12:00 the day before. Similarly to the ISO real-time (spot) market, bids in the hour-ahead (HA) reserves markets had to be received two hours prior to the beginning of the relevant trading interval; say by 10:00 for the trading interval starting at 12:00 or hour ending 13:00. Generally, the DA markets for energy and reserves served as preliminary estimates of the resource availability in the system; the HA and spot markets provided adjustments to DA commitments.

Under the efficient markets hypothesis, DA prices for energy and reserves should be unbiased predictors of the respective spot and HA prices. Otherwise, buyers would transfer their demand to the lower-priced market and sellers would shift their offers to the higher-priced market until equilibrium is restored. This article presents the first test of an efficient market hypothesis of reserves markets, which have several distinctive characteristics. First, sellers in reserves markets provide reserved capacity rather than energy, i.e., they are paid to keep some generation capacity ready to provide energy if needed. Second, the ISO determines the proportion of reserves to be procured in each of the DA and HA markets, rather than the buyer of the reserves. Therefore, the ISO is likely to procure a greater proportion of reserves in advance using the DA market than the buyers would, given its increased concerns for the reliability of the grid. Third, sellers of reserves can bid to provide capacity only if they intend to physically provide the capacity. At any time after the close of the DA or the HA markets, the ISO maintained the right to dispatch energy from a resource attached to a winning bid for all types of reserves. Therefore, pure speculative ("paper") trading was not allowed, as in other markets in the country. We argue that the second and third of these features caused the DA price for reserves to significantly exceed the HA price, particularly in the periods during and since the crisis. This DA premium created profitable trading schemes, such as Enron's infamous "Get Shorty" strategy, in which traders sold *non-existent* capacity in the DA market and bought it back in the HA market.

For energy markets in California and elsewhere, several authors have tested and rejected the efficient market hypothesis. Longstaff and Wang (2004) tested the Pennsylvania, New Jersey, Maryland (PJM) energy market and attributed the persistent differences between DA and spot prices to risk aversion. They argue that buyers of energy were prepared to pay a premium to secure energy a DA. Borenstein et. al. (2004) and Saravia (2003) studied the California and New York markets, respectively, providing evidence that anti-competitive behavior by market participants caused DA prices to differ substantially from spot prices. In addition, Borenstein at. al. also highlight that uncertainty regarding regulatory penalties for spot trading led the majority of the market participants to avoid arbitrage between the spot and day-ahaed markets.

We estimate separate DA premia for each of the 24 trading hours. Given the market structure, we model the difference (spread) between DA and HA prices as a high-dimensional vector moving average (VMA) process. We employ a particular state-space representation of the VMA that enables exact maximum likelihood (ML) estimation by the EM algorithm using analytical expressions. In contrast, existing ML methods require numerical differentiation, which is computationally infeasible in our high-dimensional framework.

Our state space representation capitalizes on the fact that a VMA model possesses the same Wold representation as a VMA plus white noise. Therefore, by introducing white noise, we do not disturb the series' structure, and we obtain a state space representation with nontrivial noise in its observation equation. In its turn, this observation noise allows implementation of the EM algorithm of Dempster, Laird and Rubin (1977), as in Shumway and Stoffer (1982) and Engle and Watson (1983). Our EM algorithm requires a pass of the Kalman filter along with a fixed-interval smoother in the E step and least squares-type regressions in the M step. Moreover, it can easily incorporate missing observations that are abundant in our data set. In Section 3, we present in detail our EM algorithm for estimation of VMA models following our discussion of the mechanics of the reserves markets in Section 2. Our empirical results appear in Section 4 and Section 5 concludes the paper.

2 The ISO Reserves Markets

The state's ISO is responsible for the reliability of the high-voltage transmission grid within its control area by maintaining a continuous balance between generation and load. Among other things, the ISO ensures that sufficient reserves of unloaded power are maintained on an hourly basis every day, as required by criteria and standards that its operations conform with. These reserves include the following five products procured in DA and/or HA markets: regulation up, regulation down, spinning, non-spinning and replacement.

Flexible generating units that can increase and decrease their output instantly under automatic generation control provide regulation up and regulation down respectively. A synchronized generating unit, already operating at its minimum level, can potentially provide spinning reserve. This provision depends on its ability to convert reserved capacity to energy within 10 minutes and maintain that output for at least 2 hours. The same response time and availability requirements hold for curtailable load and off-line generating units that supply non-spinning reserve. Generation that is capable of starting up, if not already operating, and ramping to a specified level within 60 minutes may provide replacement reserves. Curtailable demand with 60-minute response time is also a potential source of replacement reserves.

The reserve market participants, known as scheduling coordinators (SCs), were required by the ISO to contribute their share to the ISO total reserve requirements. Only those SCs that represented load serving entities, such as utilities, beyond generating units had to fulfill this requirement. The ISO made available a preliminary estimate for these requirement at 18:00 two days ahead of the relevant day. Those SCs that represented various generators in addition to load serving entities could self-provide the total or a fraction of their assigned shares. The ISO then purchased any difference between their total and self-provided shares from other SCs in its DA and HA reserve markets. Hence, the ISO's role was to make sure that a transaction between a SC that was required to buy and a SC that was willing to sell reserves took place. The former received a bill and the later received a payment. Around 80 percent of these transactions took place DA with the rest deferred HA. These important institutional details provide an explanation for our empirical findings below.

For the DA market, SCs had to bid their reserved capacity by 12:00 noon on the day prior to the relevant trading day for all of its 24 hours. The SCs had to submit their bids for the HA reserve market two hours prior to each trading interval (e.g. by 10:00 for the interval starting at 12:00 or hour ending 13:00). In both the DA and HA markets, bids had to fully identify the name and location of the specific load or generating unit (if within the ISO control area) or the point of entry to the ISO control area (if outside the ISO control area), as well as its technical specifications (ramp rates, upward and downward ranges of generating capacity etc.).

ISO software created a step supply curve by stacking megawatt (MW) offers in ascending order of their prices ($/MW). This market supply curve crossed with a vertical demand representing the SCs' total assigned minus self-provided amounts produced the market clearing price. This price was the base for the bills and payments described above. All the resources providing reserves in the ISO control area came from two major zones, areas within the state connected by Path 15, a major transmission line. The first was north of Path 15 (NP15) and the second south of Path 15 (SP15). If the zonal mix of the resources providing the reserves above crossed a 55/45 percent threshold for one or all the reserves, the ISO had the option to procure reserves zonally leading to zonal market clearing prices.

Additionally, before August 1999, the ISO cleared its DA and HA markets sequentially starting from regulation and then moving to the lower quality reserves. This quality ranking was based on the reserves' response times to a dispatch instruction. Furthermore, due to the fact that a generating unit providing a higher quality reserve, say spinning, is usually able to provide a lower quality reserve, say non-spinning or replacement, any extra-marginal bids in a DA or HA higher-quality market were rolled over by the ISO software to the lower quality one. Although this sequential market clearing resulted in the lowest possible cost for each reserve individually, it did not necessarily minimize the overall procurement costs. Beginning August 1999, the ISO also implemented a Rational Buyer's algorithm on its purchases, in which amounts of a higher quality reserve could substitute amounts for any of the lower quality reserves if doing so reduced the overall procurement costs. However, the implementation of this algorithm did not take place unless the ISO purchased all or none of its reserves zonally.

According to our discussion, SC submitted their final DA bids by 12:00 the day before the trading day and the ISO cleared its HA markets one hour prior to the beginning of each trading interval (e.g. 11:00 for the trading interval starting at 12:00 or hour ending 13:00). Hence, there was a window of between 11 and 34 hours from the time that the SCs submitted their DA bids to the time that the ISO cleared the corresponding HA market. It follows that the difference between DA and HA prices depends on accumulated information in the intervening 11 to 34 hours, even in an efficient market. Denoting the information arriving in hour ending h and day d by u_{hd}, we can express the difference between DA and HA prices as a moving average (MA) of between 11 and 34 u_{hd} terms (see Borenstein et. al.). Defining an observation in our sample as $y_{hd} = PHA_{hd} - PDA_{hd}$, where PHA_{hd} and PDA_{hd} are the HA and DA prices for hour ending $h = 1, \ldots .24$ and day d, we write:

$$
\begin{array}{lcccccccccccc}
y_{1d} = \beta_1 & & & & & + & u_{1d} & + & \theta_{1,1}u_{24,d-1} & +...+ & \theta_{1,18}u_{7,d-1} \\
y_{2d} = \beta_2 & & & & u_{2d} & + & \theta_{2,1}u_{1d} & + & \theta_{2,2}u_{24,d-1} & +...+ & \theta_{2,19}u_{7,d-1} \\
\vdots & & & & \vdots & & \vdots & & \vdots & & \vdots \\
y_{24d} = \beta_{24} & + & u_{24d} & +...+ & \theta_{24,22}u_{2d} & + & \theta_{24,23}u_{1d} & + & \theta_{24,24}u_{24,d-1} & +...+ & \theta_{24,41}u_{7,d-1}
\end{array}
$$

with an equivalent VMA(1) representation of dimension 24:

$$Y_d = B + \Theta_0 U_d + \Theta_1 U_{d-1}, \tag{1}$$

where $Y_d = PHA_d - PDA_d$, for appropriately defined matrices Θ_0 and Θ_1. The serially uncorrelated zero mean Gaussian error vector U_d has diagonal covariance matrix Σ_u and $B = (\beta_1, \ldots, \beta_{24})'$. Because our hypothesis is that the DA prices are unbiased estimates of the HA prices, we test whether B is statistically significantly different from zero. Taking into account the MA structure is important for correct inference about B. In the next section, we develop our

EM algorithm for ML estimation of VMA(1) models.

3 Estimation Method

The available methods for exact ML estimation of Gaussian MA processes employ numerical techniques, which are not only computationally intensive, but also numerically unstable, particularly for high-dimensional vector processes. Approximate ML techniques based on analytical expressions alleviate the problems of numerical imprecision and computational burden. Their caveat is that they perform poorly compared to their exact analogs in small samples when the roots of the determinantal polynomial in the MA coefficient lie close to the unit circle. In this section, we provide a state space representation that allows the EM algorithm to produce exact ML estimates using analytical expressions.

We capitalize on the fact that a VMA process has the same representation as a VMA plus white noise. We utilize the "additional" noise in the observation equation to employ the EM algorithm as in Shumway and Stoffer and Engle and Watson. Our result is very intuitive: the E step requires a pass of the Kalman filter and a fixed interval smoother that also handle missing observations; the M-step collapses to least-squares type regressions.

Most of the difficulty associated with the evaluation and the numerical maximization of the likelihood function for Gaussian MA processes, particularly in the vector case, is due to two facts. The inverse and the determinant of the covariance matrix of the observed series are not only difficult to compute, but also highly non-linear functions of the parameters. Whittle (1951) proposes approximations for the inverse and the covariance matrix of univariate MA models. The approximation error is negligible in large samples and his estimation procedure is asymptotically equivalent to ML. Walker (1961) simplifies the method suggested by Whittle by approximating the asymptotic distribution of the sample correlations and using the Wold decomposition. Godolphin (1977) demonstrates the asymptotic ML equivalence of the estimator by Walker for univariate MA processes. Additionally, he shows that the estimates of the MA parameters can be derived from Walker's procedure without the Wold decomposition due to the non-singular transformation from the autocorrelations to the MA parameters.

Durbin (1959) proposes an asymptotically efficient iterative least squares method for estimating a univariate MA processes using an initial autoregressive approximation. The validity of his approximation hinges on the underlying assumption of invertibility. de Frutos and Serrano (1997) propose a generalized least squares procedure for estimating VMA models because they assume that the approximation errors when lagged VMA residuals are replaced with lagged residuals from a long vector autoregression follow a MA process. In the same spirit of autoregressive approximations, Galbraith and Zinde-Walsh (1994) and Galbraith et. al. (2002) examine non-iterative estimators for the univariate and multivariate MA cases respectively.

Box and Jenkins (1970) avoid operations on the determinant and the inverse of the covariance matrix in their derivation of an approximation to the exact likelihood of univariate MA processes. To do so, they use the sum of squares of the residuals' expectations conditional on the observed data. They obtain the conditional expectations of the initial residuals by "back-forecasting". Osborn (1977) provides the exact likelihood for VMA processes using generalized least squares estimates of starting residuals. Hilmer and Tiao (1979) also derive an expression for the VMA exact likelihood that is equivalent to Osborne's. Phadke and Kedem (1978) simplify the evaluation of exact VMA likelihood functions by means of the Cholesky decomposition of the covariance matrix.

We consider the following VMA(1):

$$Y_t = B^\top D_t + \Theta(L)u_t, \qquad t = 1, \ldots, n, \tag{2}$$

where $\Theta(L) = I + \Theta L$,Θ is a $d \times d$ matrix, L denotes the lag operator and u_t is a Gaussian vector white noise process with zero mean and covariance Σ_u. D_t is a matrix of non-stochastic observable covariates. We also define $\Omega = \Sigma_u \otimes I_{n+1}$ with I_{n+1} being a $(n+1) \times (n+1)$ identity matrix and $\otimes$ the Kronecker product. Ignoring the constant term, we write the exact log-likelihood for (2) as:

$$l\left(\theta|\mathbf{Y}\right) = -\frac{1}{2}\left(ln\left|\Omega\right| + (\mathbf{Y} - \mathbf{D}B)^\top \Omega^{-1}(\mathbf{Y} - \mathbf{D}B\right), \tag{3}$$

where $\theta = vec(\Theta, \Sigma_u)$ and $Y - DB = U = (U_0, \ldots, U_n)^\top$ for $^\top$ and $|\cdot|$ being the transpose and determinant symbols. The task of evaluating the expression in (3) is complicated because it is difficult to compute the inverse and the determinant of Ω. We can obtain a simpler form for (3) if we assume $u_0 = 0$ (see section 7.2.1 in Lütkepohl 1993), but he have already highlighted the associated side effects above. Moreover, the equations obtained by setting the derivatives of (3) equal to zero are highly non-linear and, therefore, to estimate θ we need to maximize the log-likelihood numerically.

Using the results that a VMA of some order q plus white noise remains a VMA of the same order q (see Theorem 2 in Pieris 1988), we write $\Theta(L)u_t \equiv \Gamma(L)v_t + \varepsilon_t$, where v_t and ε_t denote vector white noise processes. This setup allows us to treat the lag of v_t as observable in the complete-data log-likelihood that underlies the EM algorithm. Therefore, we write (2) in state space form as:

$$\begin{aligned} Y_t &= B'D_t + Za_t + \varepsilon_t \quad & \varepsilon_t &\sim \; N(0, \Sigma_\varepsilon) \\ a_t &= Ta_{t-1} + \eta_t & \eta_t &\sim \; N(0, \Sigma_\eta) \end{aligned} \tag{4}$$

$$Z = \left[\begin{array}{cc} I_d & \Gamma \end{array}\right], \quad \begin{array}{c} a_t^\top = \left[v_t^\top, v_{t-1}^\top\right]^\top \\ \eta_t^\top = \left[v_t^\top, 0\right]^\top \end{array}, T = \left[\begin{array}{cc} 0 & 0 \\ I_d & 0 \end{array}\right], \Sigma_\eta = \left[\begin{array}{cc} \Sigma_v & 0 \\ 0 & 0 \end{array}\right].$$

I_d is an identity matrix of dimension $(d \times d)$. We label the equation for Y_t the observation equation and we refer to the equation for a_t as the state equation. The Gaussian independently distributed disturbance vectors ε_t and η_t are mutually uncorrelated and independent of a_1. The initial state vector is also Gaussian with mean vector $a_{1|0} = E(a_t) = 0$ and covariance $P_{1|0}$, where $vec(P_{1|0}) = (I - T \otimes T)^{-1} vec(\Sigma_\eta)$ and I is an identity matrix of the proper dimension. We maintain diagonality and invertibility for both Σ_ε and Σ_η for the remainder of our discussion in this section.

If $\Sigma_\varepsilon = 0$, then (4) is identical to (2) and $\Gamma \equiv \Theta$. However, if $\Sigma_\varepsilon \neq 0$, then there exists a unique mapping from $(Z, \Sigma_\varepsilon, \Sigma_v)$ to (Θ, Σ_u) by matching the MA parameters Θ to the infinite MA representation of (4). From the Kalman filter, we have

$$Y_t = (I + Z^\top (I + TL + T^2L^2 + \ldots)KL)u_{t+1}, \tag{5}$$

where the matrix K denotes the steady-state value of the Kalman gain (see section 13.5 in Hamilton 1994). Therefore, we obtain the MA coefficient from the expression $\Theta = Z^\top K$. Moreover, Σ_u equals the steady-state value of the covariance of the one step prediction error from the Kalman filter. The Wold representation in (5) illustrates the unique mapping from $(Z, \Sigma_\varepsilon, \Sigma_v)$ to (Θ, Σ_u). However, the reverse mapping is not unique, which implies that the parameters $(Z, \Sigma_\varepsilon, \Sigma_v)$ are not separately identified[1]. Therefore, to identify the parameters in (4), we set Σ_ε to a constant, which allows us to obtain ML estimates of (Z, Σ_v). We then calculate the ML estimates of the parameters of interest (Θ, Σ_u).

Omitting constants, the exact *complete-data log-likelihood* for (4) is:

$$l(\theta|\mathbf{Y}, a) = -\frac{n}{2} ln\, |\Sigma_\varepsilon| - \left(\frac{n+1}{2}\right) ln\, |\Sigma_v| - \frac{1}{2} trace \left(\sum_{t=1}^{n} \left(\Sigma_\varepsilon^{-1} \varepsilon_t \varepsilon_t^\top\right) + \sum_{t=0}^{n} \left(\Sigma_v^{-1} v_t v_t^\top\right) \right) \tag{6}$$

where $\varepsilon_t = Y_t - B^\top D_t - Za_t$, $v_t = a_t - Ta_{t-1}$, To maximize the *incomplete (observed)-data* log-likelihood $l(\theta|\mathbf{Y})$, we apply the EM algorithm of Dempster et. al. using the following decomposition of the complete data log-likelihood:

$$l(\theta|\mathbf{Y}, a) = l(\theta|\mathbf{Y}) + log f(a, |\mathbf{Y}, \theta). \tag{7}$$

[1] To see this, it is sufficient to assume an MA(1) of the form $y_t = \eta_t + \theta \eta_{t-1} + \varepsilon_t$, where η_t and ε_t are Gaussian white noise processes uncorrelated at all lags and leads with zero means, and σ^2, ω^2 variances. In this case, we have that $\gamma_0 = E[y_t^2] = (1+\theta^2)\sigma^2 + \omega^2$ and $\gamma_1 = E[y_t y_{t-1}] = \theta\sigma^2$, while $\gamma_j = E[y_t y_{t-j}] = 0$ for $j > 1$. Hence, we have only two equations from which we need to rertieve θ, σ^2 and ω^2. This is impossible unless a fixed value is assumed for ω^2.

The second term in (7) refers to the logarithm of the density of the *unobserved (missing)* data given the observed data (see Krishnan and McLachlan 1997 and the references therein). The EM algorithm involves an iterative point-to-set map, $M(\theta)$, from the parameter space to itself that finds the zeros for the score of the *expected complete-data* log-likelihood conditional on $\mathbf{Y}$.

From (7), we write $l(\theta|\mathbf{Y}) = l(\theta|\mathbf{Y}, a) - log f(a|\mathbf{Y}, \theta)$, and in the E step we take expectations over the distribution of a given $\mathbf{Y}$ and a current estimate of θ, say $\theta^{(i)}$:

$$l(\theta|\mathbf{Y}_n) = Q\left(\theta|\theta^{(i)}\right) - H\left(\theta|\theta^{(i)}\right) = \int \left(l(\theta|\mathbf{Y}, a)\right) f(a|\mathbf{Y}, \theta^{(i)})da - \int log f(a|\mathbf{Y}, \theta) f(a|\mathbf{Y}, \theta^{(i)})da. \tag{8}$$

The M step solves for $\theta^{(i+1)}$ by maximizing $Q\left(\theta|\theta^{(i)}\right)$ with respect to θ. Hence, for a sequence of iterates $\theta^{(0)}, \theta^{(1)}, \ldots$, with $\theta^{(i+1)} = M(\theta^{(i)})$ the difference in $l(\theta|\mathbf{Y})$ at successive iterates is:

$$l(\theta^{(i+1)}|\mathbf{Y}) - l\left(\theta^{(i)}|\mathbf{Y}\right) = Q\left(\theta^{(i+1)}|\theta^{(i)}\right) - Q\left(\theta^{(i)}|\theta^{(i)}\right) - \left(H\left(\theta^{(i+1)}|\theta^{(i)}\right) - H\left(\theta^{(i)}|\theta^{(i)}\right)\right). \tag{9}$$

The difference of Q functions is positive by construction and the difference in the H functions is negative by the concavity of the logarithmic function and Jensen's inequality (Dempster, Laird and Rubin.). Therefore, each iteration of the algorithm increases $l(\theta|\mathbf{Y})$ and, if $l(\theta|\mathbf{Y})$ is bounded from above, $l(\theta^{(i)}|\mathbf{Y})$ converges to a stationary value of $l\left(\theta|\mathbf{Y}\right)$ (see the conditions of Wu 1983). For the class of well-behaved problems as the one considered here, where $l(\theta|\mathbf{Y})$ is unimodal and concave over the entire parameter space, this stationary point is a global maximum and the EM algorithm yields the unique ML estimate of θ based on the *observed* data.

The E step requires passes of the Kalman filter and a fixed-interval smoother to obtain the conditional mean and variance of the state and disturbance vectors. The Kalman filter recursions provide a convenient prediction error decomposition for (4) through the mean and the covariance of the state vector a_{t+1} conditional on the observations $\mathbf{Y}_t = (Y_1, Y_2, \ldots Y_t)$, i.e. $a_{t+1|t} = E\left[a_{t+1}|\mathbf{Y}_t\right]$ and $P_{t+1|t} = var(a_{t+1}|\mathbf{Y}_t)$.The backward recursions of a fixed-interval smoother then give the estimates of the state and disturbance vectors along with their covariances conditional on $\mathbf{Y}_n = (Y_1, Y_2, \ldots Y_n)$, i.e. $a_{t|n} = E\left[a_t|\mathbf{Y}_n\right]$, $\varepsilon_{t|n} = E\left[\varepsilon_t|\mathbf{Y}_n\right]$, $v_{t|n} = E\left[v_t|\mathbf{Y}_n\right]$, and $P_{t|n} = var(a_t|\mathbf{Y}_n)$, $var(\varepsilon_t|\mathbf{Y}_n)$ $var(v_t|\mathbf{Y}_n)$ respectively. Omitting constants, the *expected complete-data log-likelihood* is:

$$\begin{aligned} Q(\theta|\theta^{(i)}) &= \frac{n}{2} ln\left|\Sigma_\varepsilon^{-1}\right| - \frac{1}{2} trace\left(\sum_{t=1}^{n} \Sigma_\varepsilon^{-1}\left(\varepsilon_t \varepsilon_t^\top + var(\varepsilon_t|\mathbf{Y}_n)\right)\right) \\ & \left(\frac{n+1}{2}\right) ln\left|\Sigma_v^{-1}\right| - -\frac{1}{2} trace\left(\sum_{t=0}^{n}\left(\Sigma_v^{-1}\left(v_t v_t^\top + var(v_t|\mathbf{Y}_n)\right)\right)\right) \end{aligned} \tag{10}$$

Using standard matrix calculus results, the M step, which solves for $\theta^{(i+1)}$ by maximizing $Q\left(\theta|\theta^{(i)}\right)$ with respect to θ, implies the following analytical expressions:

$$\begin{bmatrix} Z^{(i+1)} \\ B^{(i+1)} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{n}\left(a_{t|n}a'_{t|n}+P_{t|n}\right) & \sum_{t=1}^{n}a_{t|n}D'_t \\ \sum_{t=1}^{n}D_t a'_{t|n} & \sum_{t=1}^{n}D_tD'_t \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{n}a_{t|n}Y'_t \\ \sum_{t=1}^{n}D_tY'_t \end{bmatrix}, \tag{11}$$

$$\Sigma_v^{(i+1)} = \left(\frac{1}{n+1}\right)\sum_{t=0}^{n}\left((\eta_{t|n})(\eta_{t|n})' + var(\eta_t|\mathbf{Y}_n)\right). \tag{12}$$

Finally, because Y_t is of dimension $d \times 1$ and Σ_ε is diagonal, (11) is equivalent to d univariate regressions. We conclude this section with the several remarks about our estimation method:

Remark 1: With Σ_ε being diagonal, we apply the univariate filtering and smoothing algorithm of Durbin and Koopman (2000) directly. In the smoother pass, we use the iterations summarized in Koopman (2001, Section 4.3.1), which avoid the inversion of the contemporaneous covariance matrix $P_{t|t}$. This univariate algorithm saves computation time and simplifies the calculation of the exact log-likelihood because it requires no matrix inversion in either the filter or the smoother pass.

Remark 2: We performed numerous simulations for a wide range of values of Σ_ε to assess the impact of noise in the observation equation on the global convergence rate of the EM algorithm, which we denote by $r = lim_{i\to\infty}\left(\left\|\theta^{(i+1)}-\theta^*\right\| / \left\|\theta^{(i)}-\theta^*\right\|\right)$, where $\theta^{(i)} \to \theta^*$. Following Meng and Rubin (1994) we used $r = max\ r_j$, $r_j = lim_{i\to\infty}\left(\left|\theta_j^{(i+1)}-\theta_j^{(i)}\right| / \left|\theta_j^{(i)}-\theta_j^{(i-1)}\right|\right)$, where $\theta_j^{(i)}$ is the jth element of θ in the ith iteration. Our simulations showed that r is decreasing in Σ_ε, which implies a faster rate of convergence for larger Σ_ε.

Remark 3: Several methods exist for estimating the covariance of MLEs obtained from the EM algorithm. A closed form expression for the asymptotic covariance of Θ is given by $var(\Theta) = (1/n)\,\Sigma_U {\Sigma_Y}^{-1}$, $vec\left(\Sigma_Y\right) = \left(I-(\Theta\otimes\Theta)\right)^{-1} vec\left(\Sigma_U\right)$ (see section 6.3 in Lütkepohl and section 4.3.1. in Reinsel (1993)). This expression is particularly useful for high-dimensional models, where numerical differentiation is computationally demanding. For low-dimensional models, Meilijson (1989) shows that numerical computation of the empirical observed information matrix consistently estimates the observed information matrix. An alternative procedure for a numerically stable asymptotic covariance matrix of the MLEs is the Supplemented EM of Meng and Rubin (1991). Another option for consistent standard error estimation is to use the bootstrapping method for state space models of Stoffer and Wall (1991).

4 Empirical Analysis

In our analysis, we use DA and HA market clearing prices publicly available from the Open Access Same Time Information System (OASIS) of the California ISO, between August 1999 and August 2002. Although the reserve markets started their operations on April 1, 1998, the ISO did not record HA prices until June 1, 1999 and did not distinguish regulation up from regulation down prices before August 18, 1999. It is also widely accepted that the entire restructured wholesale electricity sector in California performed without any major problems from its inception up to May 2000. It then entered an almost year-long period of severe crisis that ended in late June 2001. Because we are also interested in the reserve markets' performance in the aftermath of the crisis, we analyze three distinct periods: 1) pre-crisis (August 1999-April 2000), 2) crisis (May 2000-June 2001) and 3) post-crisis (July 2001-August 2002). *We provide all our tables and figures in the end of the paper.*

4.1 Summary Statistics

On average the ISO procures about 80 percent of capacity in the DA market. Table 1 shows the mean DA and HA purchases for each of the five types of reserves. We collected these means from the ISO Department of Market Analysis (DMA) monthly reports between October 1999 and August 2002. Total procurements were largest during the crisis period and smallest in the post-crisis period, but the percentage procured DA is relatively constant. An increase in the amount of self-provided reserves, especially from units operated by the state's utilities, explains the decrease in both DA and HA purchases after the crisis. Both DA and HA prices were substantially greater during the crisis period than before and after the crisis. Tables 5 and 6 show summary statistics for the price data.

Ideally, the ISO procures its various reserve requirements from SCs that represent units from both NP15 and SP15 areas for reliability purposes. However,the ISO may make its purchases for its system-wide requirements from units in either of the two areas. In many cases, we have price but not quantity (MW) information available in our data set for the DA and HA markets. We treat any HA or DA price that is not accompanied by MW procurement in either SP15 or NP15 as missing. For spinning, non-spinning and replacement, a zero MW procurement is not only due to sufficient self-provision, but also due to the implementation of the Rational Buyer's algorithm. Because we define an observation as the spread between the HA and DA price, we treat it as missing if either the DA or the HA price is missing. As a result, the number of hourly non-missing observations in the three periods lies between 24 (post-crisis, spinning, hour 1) and 400 (crisis, regulation up, hour 16) with an average of 263. Therefore, we expect our estimates for some hours to be more precise than others. Furthermore, due to the large number of hours with missing observations for replacement, we exclude it from our efficiency tests below.

Moreover, we focus our analysis on NP15 prices for the following two reasons. First, during the post-crisis period, all 4 products, for which we perform our efficiency tests, have the same NP15 and SP15 prices both DA and HA (except for 42 non-spinning hours). Second, only 8 percent of the NP15 and SP15 pre-crisis and crisis DA or HA prices were different for the 4 highest quality products on average.

The mean DA and HA regulation up and down prices are the highest among the products for both peak (7:00-22:00) and off-peak (23:00-6:00) hours in all three periods. The only exception are the mean peak HA and DA replacement prices during crisis that exceed their regulation down analogues. This is not surprising since regulation is the highest quality product and according to ISO staff the number of units providing regulation is limited. Mean DA and HA regulation down prices are higher in off-peak than peak hours for all the periods. Mean DA and HA peak prices for regulation up (except for post-crisis HA and pre-crisis DA), operating reserves (spinning and non-spinning) and replacement are higher than the corresponding off-peak ones.

Our findings are compatible with the following facts. In general, the ISO uses regulation as a load following product. In order to meet the peak morning loads, imports and in-state generation begin ramping up several hours in advance, in the early morning hours. This causes an over-generation condition, which the ISO mitigates by backing down these units to provide downward regulation. Similarly, in the evening hours, generation and imports start ramping down several hours prior to the sharp drop in load that occurs in the last few hours of the day. For these hours, an under-generation condition exists which the ISO mitigates by ramping up units that provide regulation.

The reason for higher operating reserves DA and HA peak prices is that their required amounts are determined as a fraction of about 6 percent of the load that is served by generation within the ISO control area. Wolak et. al. (1998) also argue that regulation up and spinning prices should be somewhat related to the energy prices, although the relation is not clearly defined. Regulation units can suffer considerably more wear and tear than they would providing energy at a constant level, and in many cases units providing spinning and non-spinning operate below their most efficient point. We note that the mean ISO NP15 spot prices for the three periods were: \$37/MWh (pre-crisis), \$133/MWh (crisis) and \$28/MWh (post-crisis). The mean PX NP15 DA prices were \$34/MWh (pre-crisis) and \$155/MWh (crisis). The vast majority of negative mean spreads (HA-DA) reflects the difference in the volume of procurements in the two markets of Table 1. Interestingly enough, the mean pre- and post-crisis prices for both DA and HA are comparable, but their counterparts during crisis seem to be from a different world (notice the \$89.14/MW for peak replacement DA!).

The non-spinning and replacement medians are very close to zero during the pre- and post-crisis periods. The underlying cause is the large number of zero DA and HA observations that also contribute to the high kurtosis for the prices of the two products. Our discussions with ISO staff indicate that non-spinning and replacement prices less than $1 are not rare and are usually attributed to a robust supplemental energy market, which operated 45 minutes before the ISO real-time market. Wolak et. al. (1998) also provide an analogous comment for the latter product.

The values of kurtosis for all products are striking. We attribute this pattern to a combination of kinked supply curves and changes in price caps over the three periods. Summing horizontally SC bids often produces a supply curve that has a small slope at low prices, where most of the bids lie, and a large slope at high prices. We present examples of such supply curves in Figure 1. In most hours, the market clears to the left of the kink and the price is moderate. However, occasionally the market clears to the right of the kink and the price increases dramatically. The notable difference in the size of the kurtosis for the three periods is also related to different FERC imposed price caps (see Table 2). These caps appear to serve as a focal point for some bidders, as there was often at least one bid at the price cap in any given hour. After December 8, 2000, all caps reported were soft. Thus, a market participant could receive a payment that exceeded the cap upon cost justification, but could not set the market clearing price. Lower caps led to reasonable kurtosis levels in the crisis and post-crisis periods.

4.2 Efficiency Tests

The model in (1) contains 516 MA parameters, so numerical differentiation even by gradient-based methods is computationally infeasible. We estimate the model by applying our EM algorithm of Section 3. We iterate the filtering and smoothing recursions of Durbin and Koopman (2000) in the E-step and 24 univariate regressions in the M-step until convergence. We calculate the standard errors for B using the empirical observed information matrix of Meilijson. We derive the standard errors associated with the MA parameters in Θ using the analytical expression in Remark 3.

We report our spread estimates from our VMA(1) model and their associated standard errors by reserve type for all three periods in Figure 2. Although not reported, we obtained analogous spread estimates using 24 univariate regressions and constructed confidence intervals using Newey-West standard errors (1 lag). As expected, our VMA(1) spread estimates are very close to the ones from the univariate regressions, but their associated confidence intervals are tighter reflecting the superiority in terms of efficiency of our approach (see Table 3).

Using our VMA(1) estimates, for the pre-crisis period, most of the 24 hourly spreads do not differ significantly from zero for regulation down. Regulation up is the product with the highest number of significant hourly negative spreads (21). For the same period, we don't see any significant spreads for 9 hours of spinning. For non-spinning, there are only 4 significant spreads and are all positive. The significant negative spreads imply an additional cost of up to $6.7/MW (regulation up, hour 23) for buying 1MW in the DA instead of the HA market. If we use the mean regulation up HA off peak pre-crisis price ($11.3/MW), the spread translates to an additional cost of 60 percent for the purchase of 1 MW in the DA.

The overwhelming majority of the hourly spreads are statistically significant and negative for both the crisis and the post-crisis periods. The additional cost of 1MW procured DA instead of HA is as high as $26.9/MW (regulation up, hour 12) and $28.3/MW (spinning, hour 14) during the crisis. These translate to 32 and 53 percent premia for DA purchases respectively; again using the mean peak HA regulation up ($82.4/MW) and spinning ($54.9/MW) prices. In post-crisis period, we see additional costs that are as high as $8.2/MW (regulation down, hour 6; spinning, hour 16). The implied premia are 49 percent and 160 percent for regulation down and spinning respectively, evaluated at the mean HA peak price for the period ($16.8/MW and $5.1/MW). Regulation down has the highest number of significant negative post-crisis spreads for the off-peak hours (23:00 to 6:00). The regulation down spread seems to be driven by a substantial difference between the mean DA ($21.5/MW) and HA ($16.8/MW) prices for this part of the day.

Our premia estimates well exceed those of earlier studies for energy markets in California and PJM. In Borenstein et. al., after averaging over their significant price spreads (ISO-PX) between August 1999 and November 2000 we get premia of 30 and 17 percent for spot purchases in NP15 and SP15 for hours 1:00-6:00. Repeating analogous calculations for hours 8:00-24:00, the implied premia for NP15 is 24 percent for spot purchases. Averaging over the significant unconditional hourly forward premia in Longstaff and Wang for the PJM eastern hub between June 2000 and November 2002 we get 4 percent. Moreover, In Saravia, the implied DA transmission premia for western and central New York are about 12 and 3 percent before the introduction of speculative trading in November 2001. After the introduction of speculative trading, the same premia reduce to 4.9 and 2.8 percent.

We estimated 516 MA parameters for each product in every period. To indicate the nature of the estimated coefficients, we use regulation down, hour 18:00, for all periods in Figure 4. The spikes for hours in the window between 13:00 and 18:00 for all periods are compatible with the comments of a member of the ISO staff: "...the phones ring all the time up to late afternoon, then it is quiet..." and are present in almost all hours for all the products. This happens for two reasons. The DA procurement results and the market clearing prices were finalized by 13:00. Changes to final DA schedules due to unit failures or Reliability Must Run requirements with a potential impact on the HA markets are published approximately by 18:00.

Finally, to account for the effect of the excess kurtosis in our sample, we re-estimated our VMA(1) model and the naive 24 univariate regressions treating any observation below the 1 percent and above the 99 percent percentiles (outlier) also as missing (see Figure 3 for the VMA(1) case). We summarize the number of statistically significant spreads for each product by period in Table 4.Our findings are not driven from the excess mass on the tails of the distribution of our sample under any circumstances.

For the pre-crisis period, the VMA(1) estimated spreads treating the outliers as missing imply additional costs as high as \$4.3/MW (regulation down, hour 7), \$6.8/MW (regulation up, hour 16), \$3.10/MW (spinning, hour 16), \$1.6M/W (non-spinning, hour 7). Evaluated at the mean peak HA prices (\$11.5/MW, \$10.0/MW, \$2.6/MW, \$0.9/MW) for the same period, the premia of DA purchases of 1MW are 37, 68, 120 and 180 percent, respectively. For the crisis period, we have additional costs as high as \$21.0/MW (regulation down, hour 23), \$21.5/MW (regulation up, hour 23), \$24.2/MW (spinning, hour 16), \$15.21M/W (non-spinning, hour 19). Evaluated at the corresponding mean and off peak HA prices (\$75.1/MW, \$71.1/MW, \$50.2/MW, \$48.1/MW) for the same period, the premia of DA purchases of 1MW are 28, 30, 48 and 32 percent, respectively. Finally, for the post-crisis periods, the additional costs are as high as \$7.6/MW (regulation down, hour 23), \$5.3/MW (regulation up, hour 16), \$8.7/MW (spinning, hour 14), \$4.8M/W (non-spinning, hour 16). Evaluated at the corresponding mean and off peak HA prices (\$16.3/MW, \$15.6/MW, \$3.7/MW, \$2/MW) for the same period, the premia of DA purchases of 1MW are 47, 33, 230 and 220 percent.

4.3 Why did Large Premia Persist?

The most compelling question raised by this analysis is why SCs paid a premium for DA purchases during the crisis and continued to do so (although in lesser extent) in the post-crisis period. The distinctive feature of the reserves markets is that the ISO determines the proportion of reserve capacity that will be purchased in the DA market. Thus, the buyers of these ancillary services have no choice over what price they pay for this capacity. On the other hand, the ISO does not pay for the capacity, but is responsible if insufficient reserve capacity is supplied. Given the potentially disastrous effect of even a minor flaw in the grid's operation, it seems reasonable to assume that the ISO is a risk-averse agent. Sufficient amounts of the various types of reserves simply reduce the probability of such a catastrophic event. Therefore, it is not surprising over 85 percent of these purchases were made at a premium in the DA market, at least for the 4 highest quality reserves.

The previous paragraph explains why the ISO is willing to pay a premium to buy in the DA market, but it does not explain why sellers did not take full advantage of it and ultimately bid the price down. One answer lies on the fact the sellers were constrained in the amount of capacity they could bid; they could not offer capacity that they did not intend to provide. At any time after the close of the DA or the HA markets, the ISO maintained the right to dispatch energy from a resource attached to a winning bid for all types of reserves. The substantial premia in the energy market documented in Borenstein et. al. also may have attracted SCs with spare capacity away from the reserves and towards the energy markets, at least during the pre-crisis and the first half of the crisis periods.

Did a large number SCs attempt to exploit the trading opportunities presented by the DA premium? Looking for an answer, we first referred to the Market Monitoring and Information Protocol (MMIP) of the ISO Tariff. The MMIP is the code of conduct for SCs upon their participation in any of the ISO markets. It provides a description of what constitutes *gaming* and *anomalous market behavior* on behalf of the SCs. Both are termed as *practices subject to scrutiny*. The description of the first, at least to our understanding, is clear and rather irrelevant to our puzzle. Any attempt to exercise market power through economic or physical withholding of generating capacity almost explicitly falls within its scope. The definition of *gaming* is less clear to us and can be summarized to the following: "...taking unfair advantage of the rules and procedures set forth in the ISO Tariff...", " ...it may also include actions or behaviors that may render the ISO Markets vulnerable to price manipulation to the detriment of their efficiency...". Notice that the ISO tariff did not prevent SCs from selling DA and then cancelling their positions in HA buying back the entire amount of DA MWs. Our conjecture is that the majority of SCs were hesitant to engage extensively in such kind of activity, due to not very explicit description of what constituted *gaming* behavior and could lead to sanctions and penalties.

However, there is plenty of support for the exploitation of these trading opportunities by a limited number of SCs in a report of the ISO DMA in October 2002. The ISO staff calculated significant net gains for 25 SCs that sold reserves DA and then bought them back HA during 2000 and 2001 (the extent of net gains was limited in 2002 and only through May). The amount of net gains was close to $60 million (California ISO 2002). Since the hours, dates and products for which such buybacks are publicly available, re-estimating our model including appropriate binary variables to account for their presence is possible.

The ISO October 2002 DMA report that we refer to was prepared as part of FERC investigation for price manipulation in electricity and natural gas markets in the West initiated in February 2002. Enron's filing for bankruptcy in December 2001 and the subsequent substantial decline in spot electricity and natural gas prices in the West had triggered allegations for their manipulation by Enron. Three months later, in May 2002, FERC released on its web site two internal Enron memoranda of December 2000 that described various fraudulent trading schemers employed by its traders in the Western markets.

Among these strategies, "Get Shorty" intended to capitalize on the sign of the differences between DA and HA reserve prices reported above. In its final report, FERC staff concluded that "Get Shorty" was within the scope of *anti-gaming prohibition* of the ISO MMIP (FERC 2003). Notice that "Get Shorty" was an *explicit abuse of a market rule* because it was based almost exclusively on the submission of false information. In their written guilty pleas, two chief Enron executives of its West Power Trading Division (West Power) fully acknowledged repeated fraudulent behavior in the ISO reserve markets.

West Power provided scheduling services for both energy and reserve markets in California extensively during the crisis period in 2000 and 2001. Enron traders repeatedly submitted DA bids for reserves they had *no intention* to have standby. Most commonly, they scheduled resources for spinning and replacement outside the ISO control area. They did so because they did not have to identify the resource's exact name, location and technical specifications. The ISO required only the disclosure of the point of entry to its control area of energy dispatched from the reserve of such a resource. The strategy was in most cases successful. They bought back in the HA the entire amount of the DA commitment. Even if their buyback had failed, they had a very good chance to avoid failure of compliance with potential dispatch orders and fines because they could choose among numerous resources to provide the energy required by the ISO. In most cases, they zeroed out their obligations successfully, they eliminated penalties and generated illegitimate profits of millions of dollars.

5 Conclusion

We perform an efficient market hypothesis test of the California electricity reserve markets by examining systematic price differences between their DA and HA prices using ML estimates of a high-dimensional VMA. We avoid the computational burden associated with the exact ML estimates using a state space representation of the VMA process that leads to analytical expressions for the scores if the EM algorithm is employed as in Engle and Watson (1983) and Shumway and Stoffer (1983). A plausible explanation for the signs of the spreads reported lies on the risk-averse behavior of the demand side and on the willingness of a limited number of sellers to game the market rules.

We believe that a policy-evaluation relevance is inherent in our findings, especially in the light of the FERC Standard Market Design efforts that have been put in place since July 2002. First of all, there is a need for explicit market rules. In the case of the California ISO Tariff we found it difficult to explicitly define what constituted *gaming* behavior, a practice that was subject to scrutiny and corrective actions, such as sanctions and penalties. Second, although we do not question the ISO's intention for the smooth operation of the state's transmission grid during the periods we analyzed, we are not convinced that 80 percent of procurements at high DA premia can be fully justified in economic terms. Finally, in terms of policy recommendation, we believe that the introduction of "paper trading" in the reserves market may share the beneficial effects of virtual bidding in the New York market reported in Saravia (2003).

References

[1] Borenstein, S., Bushnell, J., Knittel, C.R., and Wolfram, C., "Inefficiencies and Market Power in Financial Arbitrage: A Study of California's Electricity Markets , *POWER Working Paper PWP-138*, 2004, University of California Energy Institute.

[2] Box, G.E.P., and Jenkins, G.M., *Time Series Analysis: Forecasting and Control*, Holden-Day, Oakland.

[3] California ISO, Department of Market Analysis, "Analysis of Trading and Scheduling Strategies Described in Enron Memos", October 2002.

[4] De Frutos, R.F., and Serrano, G.R. (1997), "A Generalized Least Squares Estimation Method For Invertible Vector Moving Average Models", 1997, *Economics Letters*, 57(2), pp. 149-156.

[5] Dempster, A., Laird, N., and Rubin, "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm", 1977, *Journal of the Royal Statistical Society (Series B)*, 39(1), pp. 1-38.

[6] Durbin, J. "Efficient Estimation of Parameters in Moving-Average Models", 1959, *Biometrika*, 46(3/4). pp. 306-316.

[7] Durbin, J. and Koopman, S.J., *Time Series Analysis by State Space Models*, 2001, Oxford University Press, Oxford.

[8] Durbin, J. and Koopman, S.J., "Fast Filtering and Smoothing for Multivariate State Space Models", 2000, *Journal of Time Series Analysis*, 21(3), pp.281-296.

[9] Engle, R.F., and Watson, M., "Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression Models", 1983, *Journal of Econometrics*, 23(3), pp. 385-400.

[10] Federal Energy Regulatory Commission, "Final Report on Price Manipulation in Western Markets, Fact-Finding Investigation of Potential Manipulation of Electric and Natural Gas Prices", Docket No. PA02-2-000, March 2003.

[11] Galbraith, J.W., Ulah, A., and Zinde-Walsh, V., "Estimation Of The Vector Moving Average Model by Vector Autoregression", 2002, *Econometric Reviews*, 21(2), pp.205-219.

[12] Galbraith, J.W., and Zinde-Walsh, V., "A Simple Non-Iterative Estimator for Moving Average Models", 1994, *Biometrika,* 81(1), pp.143-155.

[13] Godolphin, E.J., "A Direct Representation for the Maximum Likelihood Estimator of a Gaussian Moving Average Process", 1977, *Biometrika*, 64(2), pp. 375-384.

[14] Hamilton, J., *Time Series Analysis,* 1994, Princeton University Press, New Jersey.

[15] Hillmer, S.C., and Tiao, G.C.," Likelihood Function of Stationary Multiple Autoregressive Moving Average Models", 1979, *Journal of the American Statistical Association*, 74(367), pp. 652-660.

[16] Jamshidian, M., and Jennrich, R.I., "Acceleration of the EM Algorithm using Quasi-Newton Methods", 1997, *Journal of the Royal Statistical Society (Series B)*, 59(3), pp. 569-587.

[17] Koopman, S.J., "Disturbance Smoother for State Space Models", 1993, *Biometrika*, 80(1), pp. 117-126.

[18] Krishnan, T., and McLachlan, *The EM Algorithm and Extensions,* 1997, Willey, New York.

[19] Longstaff, F.A., and Wang, A.W., "Electricity Forward Prices: A High-Frequency Empirical Analysis", 2004, *Journal of Finance,* 59(4), pp.1877-1900.

[20] Lüktepohl, H. *Introduction to Multiple Time Series Analysis, 1993,* Springer-Verlag, New-York.

[21] Meng, X.L., and Rubin, D.B., "On the Global and Component-wise Rates of Convergence of the EM Algorithm", 1994, *Linear Algebra and Its Applications*, 199, pp. 413-425.

[22] Meng, X.L., and Rubin, D.B., "Using EM to Obtain Asymptotic Variance-Covariance Matrices: the SEM Algorithm", 1991, *Journal of the American Statistical Association*, 86(416), pp. 899-909.

[23] Meng, X.L., and van Dyk, D., "The EM Algorithm–An Old Folk-Song Sung to a Fast New Tune", 1997, *Journal of the Royal Statistical Society (Series B)*, 59(3), pp. 511-567.

[24] Meilijson, I., "A Fast Improvement to the EM Algorithm on its Own Terms", 1989, *Journal of the Royal Statistical Society (Series B)*, 51(1), pp. 127-138.

[25] Osborn, D.R., "Exact and Approximate Maximum Likelihood Estimators for Vector Moving Average Processes", 1977, *Journal of the Royal Statistical Society (Series B)*, 39(1), pp. 114-118.

[26] Phadke, M.S., and Kedem, G., "Computation of exact likelihood function of multivariate moving average models", 1978, *Biometrika*, 65(3), pp. 511-519.

[27] Pieris, S., "On the Study of Some Functions of Multivariate ARMA Processes", 1988, *Journal of Multivariate Analysis*, 25(1), pp. 146-151.

[28] Reinsel, G.C., *Elements of Multivariate Time Series Analysis,* 1993, Springer-Verlag, New-York.

[29] Saravia, C, "Speculative Trading and Market Performance: The Effect of Arbitrageurs on Efficiency and Market Power in the New York Electricity Market", 2003, *CSEM Working Paper CSEM WP 121,* University of California Energy Institute.

[30] Shumway, R.H., and Stoffer, D.S., "An Approach To Time Series Smoothing and Forecasting using the EM Algorithm", 1982, *Journal of Time Series Analysis,* 3(4), pp. 253-264.

[31] Stoffer, D.S., and Wall, K.D., "Bootstrapping State-Space Models: Gaussian Maximum Likelihood Estimation and the Kalman Filter", 1991, *Journal of the American Statistical Association,* 86(416), pp. 1024-1033.

[32] Walker, A.M., "Large-Sample Estimation of Parameters for Moving-Average Models", 1961, *Biometrika,* 48(3/4), pp. 343-357.

[33] Whittle, P. Hypothesis *Testing in Time Series Analysis*, 1951, Almquist and Wiksell, Upssala.

[34] Wolak, F., Nordhaus, R., and Shapiro, C., "Preliminary Report on the Operation of the Ancillary Services Markets of the California ISO", August 1998.

[35] Wu, C.F.J., "On the convergence properties of the EM algorithm", 1983, *Annals of Statistics* 11(1), pp. 95-103.

Tables and Figures

Table 1: Mean MW Procurements in the California Electricity DA and HA Reserves Markets; pre-crisis: August 1999-April 2000, crisis: May 2000-June 2001, post-crisis: July 2001-August 2002

	pre-crisis			**crisis**			**post-crisis**		
reserve	**DA**	**HA**	%**DA**	**DA**	**HA**	%**DA**	**DA**	**HA**	%**DA**
regulation down	566	60	90	453	68	86	469	49	90
regulation up	525	43	92	539	78	88	466	56	89
spinning	682	120	85	924	127	89	764	49	93
non-spinning	761	126	86	815	87	90	741	46	94
replacement	330	174	66	365	96	77	108	40	77

Table 2: Price Caps in the California Electricity DA and HA Reserves Markets

price cap	**start date**	**end date**
750	1999 − 10 − 01	2000 − 06 − 30
500	2000 − 07 − 01	2000 − 08 − 06
250	2000 − 08 − 07	2000 − 12 − 31
150	2001 − 01 − 01	2001 − 06 − 19
91.87	2001 − 06 − 20	2001 − 12 − 19
108	2001 − 12 − 20	2002 − 04 − 30
91.87	2002 − 05 − 01	2002 − 07 − 08

Table 3: Statistically Signifcant Negative Price Spreads (HA-DA) in the California Electricity Reserves Markets; N-W: 24 univariate regressions, Newey-West (1 lag) std-errors, 0.05 level; VMA: 24-dimensional VMA(1), 0.05 level; pre-crisis: August 1999-April 2000, crisis: May 2000-June 2001, post-crisis: July 2001-August 2002

	pre-crisis		**crisis**		**post-crisis**	
reserve	**N-W**	**VMA**	**N-W**	**VMA**	**N-W**	**VMA**
regulation down	4	10	11	21	9	22
regulation up	7	21	20	23	11	20
spinning	15	15	17	23	18	18
non-spinning	0	4	8	15	12	13

Table 4: Statistically Signifcant Negative Price Spreads (HA-DA) in the California Electricity Reserves Markets; Spreads below 1 and above 99 percentiles treated as missing; N-W: 24 univariate regressions, Newey-West (1 lag) std-errors, 0.05 level; VMA: 24-dimensional VMA(1), 0.05 level; pre-crisis: August 1999-April 2000, crisis: May 2000-June 2001, post-crisis: July 2001-August 2002

	pre-crisis		**crisis**		**post-crisis**	
reserve	**N-W**	**VMA**	**N-W**	**VMA**	**N-W**	**VMA**
regulation down	7	8	12	22	15	24
regulation up	16	24	19	23	13	21
spinning	22	21	21	24	18	18
non-spinning	13	1	15	19	14	12

Table 5: Prices in the California Electricity Reserves Markets; Off-Peak Hours 23:00-6:00; Spread: HA-DA; pre-crisis: August 1999-April 2000, crisis: May 2000-June 2001, post-crisis: July 2001-August 2002

reserve	**pre-crisis**			**crisis**			**post-crisis**		
reg-down	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**
mean	11.33	14.95	−3.62	72.35	85.64	−13.29	18.13	18.54	−0.41
median	9.74	10.25	−0.63	50.00	72.00	−4.00	14.89	13.00	−0.09
std.deviation	13.39	23.99	23.68	63.08	63.52	70.61	16.65	13.68	16.32
kurtosis	167.66	58.88	56.11	2.80	6.23	7.73	8.59	9.14	9.51
reg-up	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**
mean	22.57	19.95	2.62	79.79	85.65	−5.86	16.77	21.51	−4.74
median	14.47	14.48	−0.39	50.12	71.07	−5.58	13.78	18.00	−3.00
std.deviation	32.79	19.12	30.19	75.01	59.51	76.27	15.50	13.72	17.67
kurtosis	144.66	50.78	184.19	8.54	5.84	12.17	7.25	7.19	6.58
spinning	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**
mean	1.01	1.80	−0.80	28.67	34.57	−5.9	1.49	2.41	−0.92
median	0.31	1.00	−0.67	3.17	5.29	−1.00	1.00	1.55	0.00
std.deviation	3.62	4.31	4.52	52.47	57.09	31.18	2.07	4.33	4.80
kurtosis	427.87	72.61	90.63	8.05	5.84	9.70	24.33	50.95	3.97
non-spinning	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**
mean	0.40	0.18	0.23	24.73	25.72	−0.99	0.50	0.20	0.29
median	0.03	0.01	0.00	1.15	3.00	−0.44	0.01	0.01	0.00
std.deviation	4.48	1.59	4.71	48.03	52.98	42.71	3.33	1.57	3.67
kurtosis	894.57	919.50	750.69	8.77	9.52	9.57	515.56	189.14	359.45
replacement	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**
mean	*n/a*	*n/a*	*n/a*	58.37	72.29	−13.92	0.58	0.27	0.31
median	*n/a*	*n/a*	*n/a*	50.00	74.44	−3.97	0.01	0.01	0.00
std.deviation	*n/a*	*n/a*	*n/a*	52.71	46.91	67.32	3.51	1.55	2.34
kurtosis	*n/a*	*n/a*	*n/a*	2.07	1.64	2.63	182.43	149.69	110.68

Table 6: Prices in the California Electricity Reserves Markets; Peak Hours 7:00-22:00; Spread: HA-DA; pre-crisis: August 1999-April 2000, crisis: May 2000-June 2001, post-crisis: July 2001-August 2002

reserve	**pre-crisis**			**crisis**			**post-crisis**		
reg-down	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**
mean	12.89	14.79	−1.90	82.37	97.64	−15.27	16.07	18.76	−2.69
median	8.47	8.70	−0.85	59.97	69.56	−1.43	11.54	12.21	−0.92
std.deviation	33.76	25.97	39.21	88.38	103.82	100.62	17.28	17.29	19.74
kurtosis	315.81	55.33	157.15	23.04	20.72	17.98	8.70	8.38	8.32
reg-up	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**
mean	13.09	12.89	0.21	43.64	50.65	−7.01	12.27	13.03	−0.77
median	9.20	9.19	0.00	21.70	30.00	−3.28	10.00	10.24	−0.46
std.deviation	19.02	16.81	19.96	56.57	56.47	58.93	12.40	7.86	13.13
kurtosis	86.15	86.89	59.29	15.35	9.13	13.38	13.35	13.16	10.05
spinning	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**
mean	5.14	5.14	0.00	54.90	71.47	−16.58	5.12	11.26	−6.14
median	1.00	2.05	−0.67	18.31	28.66	−3.5	1.48	5.00	−2.82
std.deviation	29.51	11.01	26.14	84.96	93.42	83.48	12.78	15.70	16.19
kurtosis	353.63	195.50	310.04	24.84	20.97	24.46	30.47	8.78	12.72
non-spinning	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**
mean	4.25	2.69	1.56	55.51	64.48	−8.97	2.86	5.46	−2.60
median	0.15	0.35	−0.08	13.00	22.94	−1.55	0.69	0.01	0.00
std.deviation	34.69	13.05	32.13	94.33	93.93	98.18	8.84	12.77	11.28
kurtosis	232.49	374.45	195.54	22.79	20.13	24.54	58.45	18.11	19.71
replacement	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**	**HA**	**DA**	**Spread**
mean	1.30	3.61	−2.31	60.21	89.14	−28.93	3.51	6.20	−2.69
median	0.10	0.25	−0.16	30.00	99.00	−1.01	0.01	1.00	0.00
std.deviation	15.71	20.76	21.98	85.30	98.72	88.28	13.12	14.57	15.43
kurtosis	852.15	233.43	209.44	35.04	25.50	28.74	35.53	16.01	17.22

Figure 1: Representative Steep Supply Curves in the California Electricity Reserves Markets; June 2000

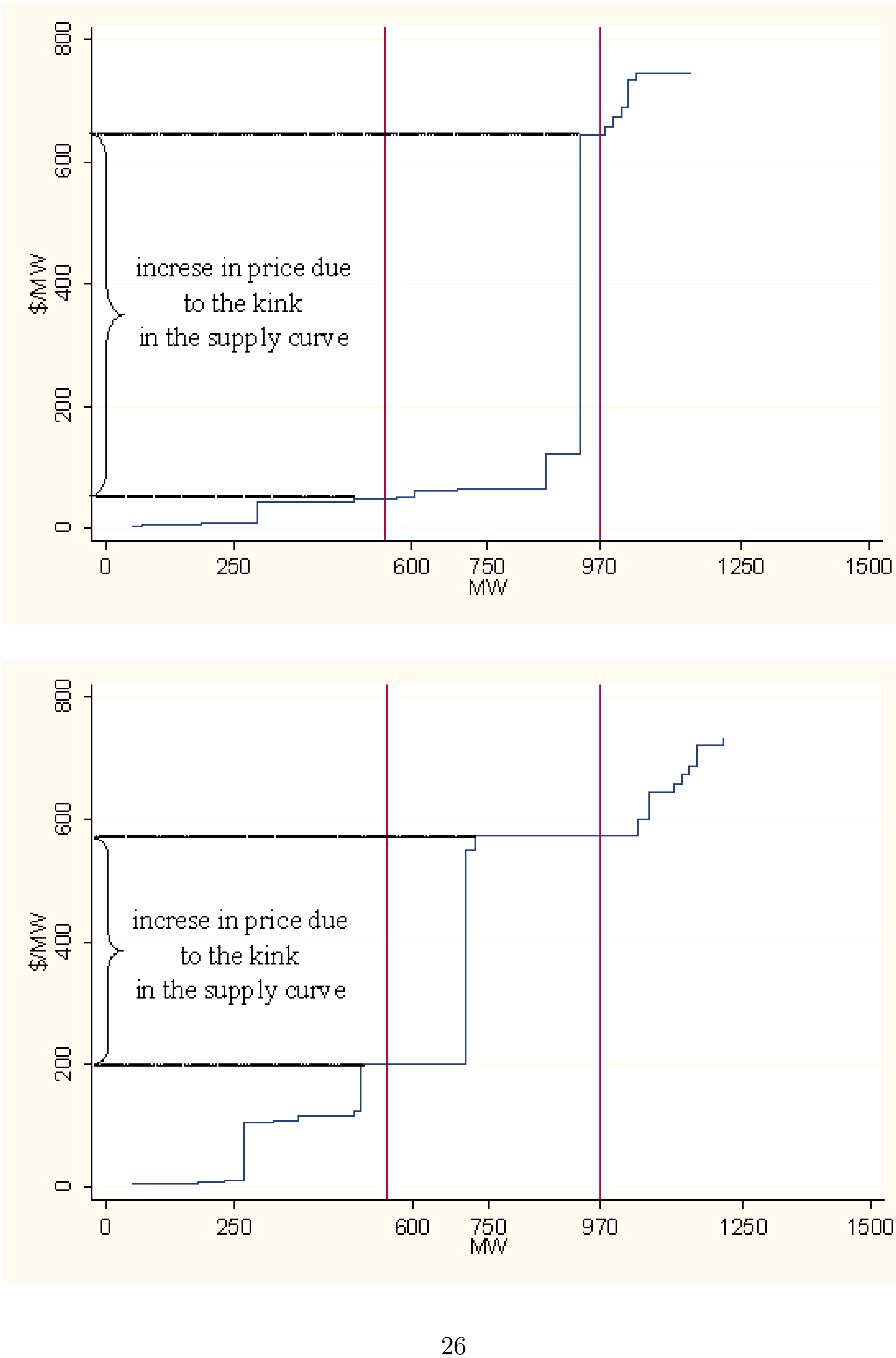

Figure 2: VMA(1) Mean Hourly Price Spread (HA-DA) Estimates in the California Electricity Reserves Markets; dashed lines: 95% confidence intervals; pre-crisis: August 1999-April 2000, crisis: May 2000-June 2001, post-crisis: July 2001-August 2002

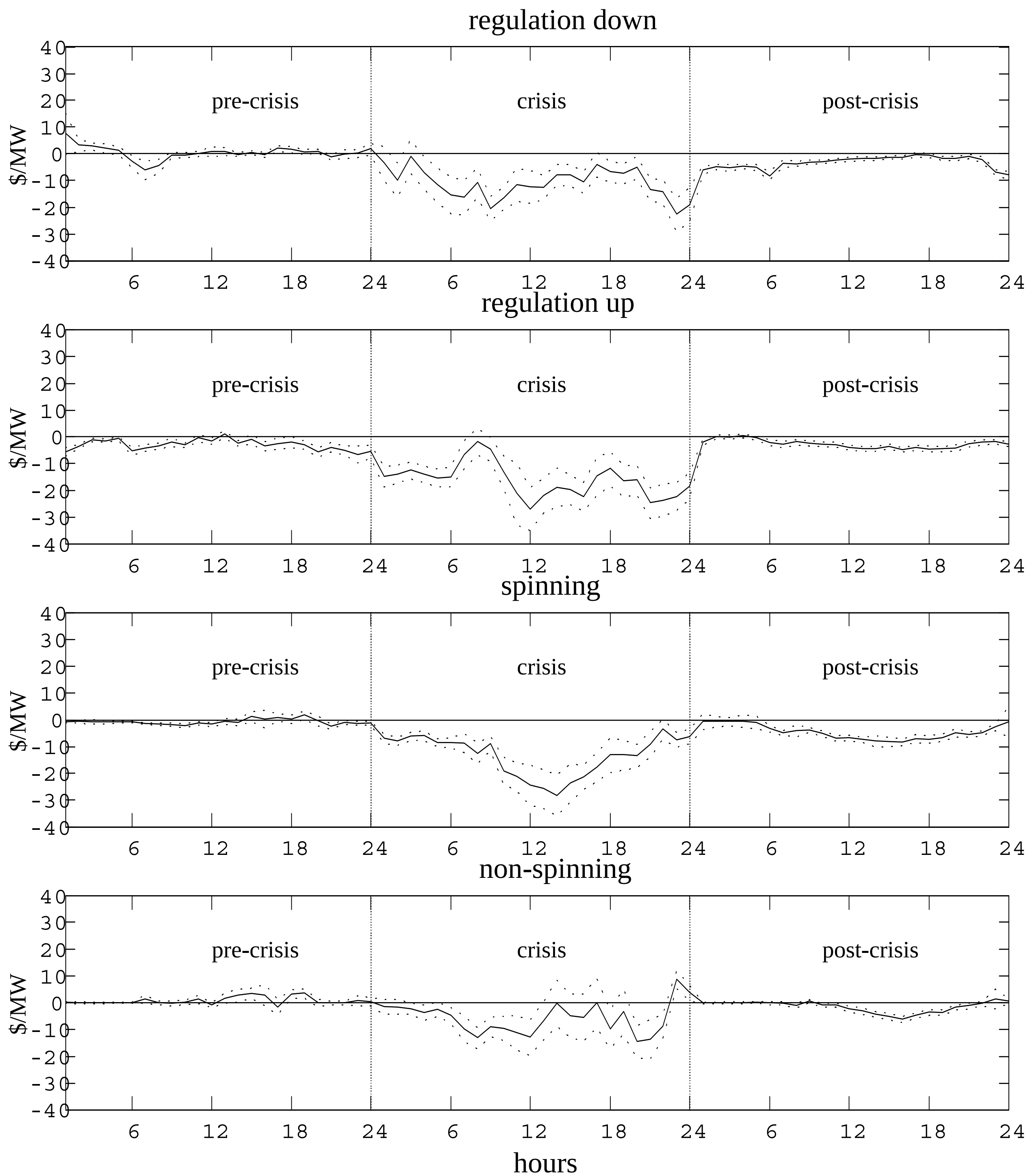

Figure 3: VMA(1) Mean Hourly Price Spread (HA-DA) Estimates in the California Electricity Reserves Markets; dashed lines: 95% confidence intervals; Spreads below 1% and above 99% treated as missing; pre-crisis: August 1999-April 2000, crisis: May 2000-June 2001, post-crisis: July 2001-August 2002

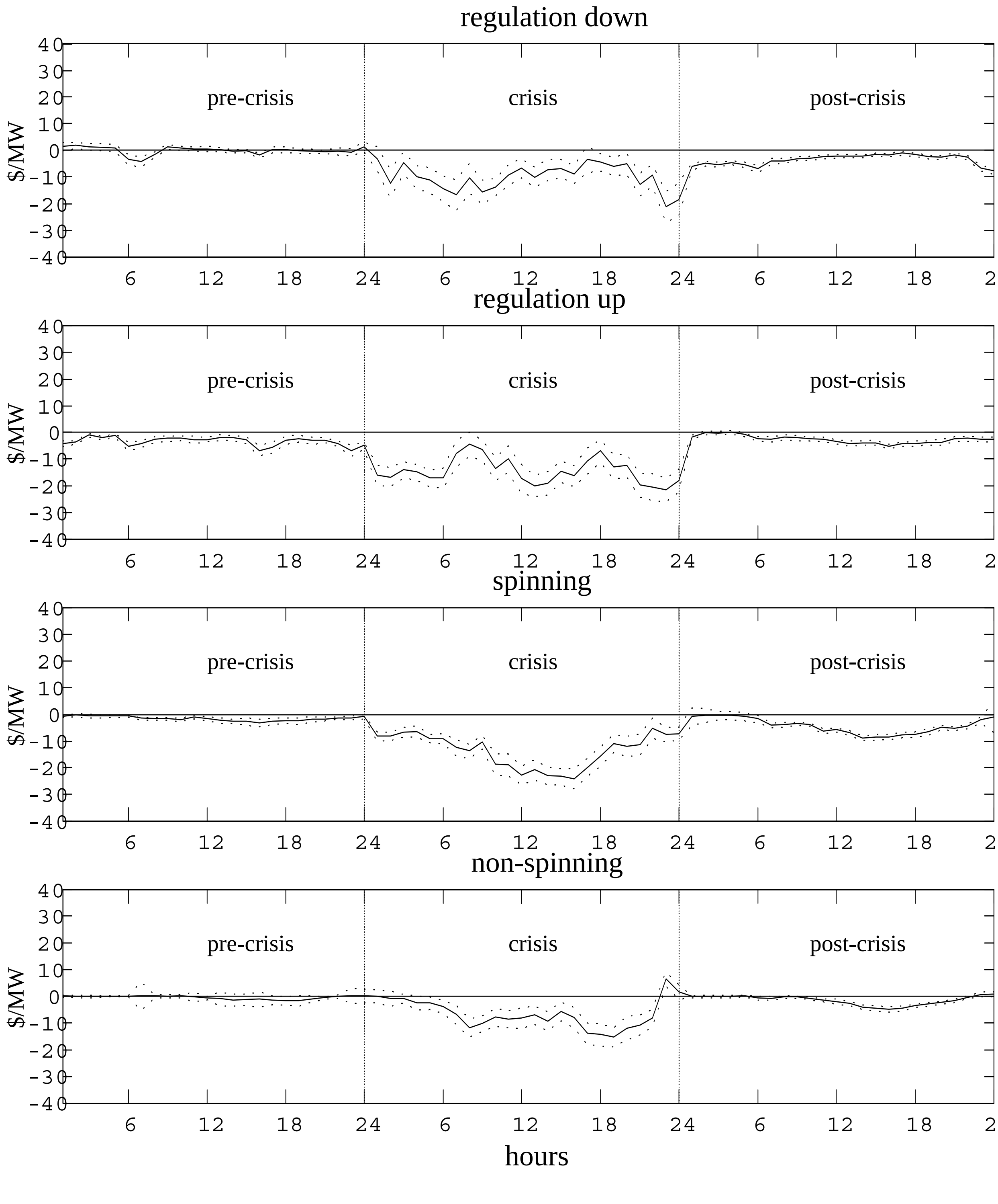

Figure 4: Moving Average Coefficient Estimates; regulation down; hour 18:00, dashed lines: 95% confidence intervals; pre-crisis: August 1999-April 2000, crisis: May 2000-June 2001, post-crisis: July 2001-August 2002