Efficiency Impact of Convergence Bidding on the California Electricity Market

Ruoyang Li · Alva J. Svoboda · Shmuel S. Oren

September 1, 2014

Abstract The California Independent System Operator (CAISO) has implemented Convergence Bidding (CB) on February 1, 2011 under Federal Energy Regulatory Commission's (FERC) September 21, 2006 Market Redesign and Technology Upgrade (MRTU) Order. CB is a financial mechanism that allows market participants, including electricity suppliers, consumers and virtual traders, to arbitrage price differences between the day-ahead (DA) market and the real-time (RT) market without physically consuming or producing energy. In this paper, we analyze market data in the CAISO electric power markets, and empirically test for market efficiency by assessing the performance of trading strategies from the perspective of virtual traders. By viewing DA-RT spreads as payoffs from a basket of correlated assets, we can formulate a chance constrained portfolio selection problem, where the chance constraint takes two different forms as a value-at-risk (VaR) constraint and a conditional value-at-risk (CVaR) constraint, to find the optimal trading strategy. A hidden Markov model (HMM) is further proposed to capture the presence of the time-varying forward premium. Our backtesting results cast doubt on the efficiency of the CAISO electric power markets, as the trading strategy generates consistent profits after the introduction of CB, even in the presence of transaction costs. Nevertheless, by comparing with the performance before the introduction of CB, we find that the profitability decreases significantly, which enables us to identify the efficiency gain brought about by CB.

R. Li · S. S. Oren
Department of Industrial Engineering and Operations Research,
University of California Berkeley,
4141 Etcheverry Hall, Berkeley, CA 94720, USA
E-mail: oren@ieor.berkeley.edu

R. Li
E-mail: ruoyang@berkeley.edu

A. J. Svoboda
Pacific Gas and Electric Company,
77 Beale Street, San Francisco, CA 94105, USA
E-mail: ajsh@pge.com

Keywords Convergence bidding · Market efficiency · Trading strategy · Value-at-risk · Conditional value-at-risk · Hidden Markov model

1 Introduction

Since 1992, the electricity sector in the United States began the process of deregulation in the pursuit of competitiveness and efficiency. The Independent System Operator (ISO) was formed to administer regional wholesale electricity markets, and ensure reliability for grid operations. Several regional wholesale electricity markets were established under the management of the ISOs: ISO New England (ISO-NE), New York ISO (NYISO), Pennsylvania - New Jersey - Maryland Interconnection (PJM), Midwest ISO (MISO), Electric Reliability Council of Texas (ERCOT), and CAISO. To provide hedging instruments against volatile wholesale spot prices, forward contracts and other financial derivatives have been introduced into these deregulated electricity markets. Financial incentives attract virtual traders to play their critical role in price discovery and market efficiency through exploiting arbitrage opportunities. CB is a financial mechanism that allows market participants, including electricity providers, retailers and virtual traders, to arbitrage price discrepancies between the forward and spot electricity markets. After the introduction of CB in the other five regional wholesale electricity markets, the CAISO has implemented CB on February 1, 2011 under FERC's September 21, 2006 MRTU Order. The central question of this study is to address whether the CAISO's forward and spot electricity markets are efficient, and if not, to what extent CB improves market efficiency.

Recently, Jha and Wolak [2013] test the efficiency of the CAISO electric power markets with hypothesis testing. They assume that in the presence of risk neutral traders CB reduces the differences between the forward prices and the expected spot prices to the extent that trading profitability no longer exists accounting for trading costs. By estimating the implied trading costs derived from several heuristic trading strategies, they present statistical evidence that demonstrates the improvement of market efficiency after the introduction of CB.

In this study, we examine the theoretical and empirical tools intended for other financial markets to help us understand the efficacy of CB in the forward and spot electricity markets. The efficient market hypothesis first formalized by Samuelson [1965] and Fama [1970], asserts that at any given time asset prices should always reflect all available information, and change quickly to incorporate new information. Jensen [1978] defines market efficiency in terms of trading profitability – "a market is efficient with respect to [an] information set, if it is impossible to make economic profits by trading on the basis of [this] information set." In particular, if anomalous returns are not high enough for a sophisticated trader to generate consistent profits after allowing for transactions costs, they are not economically significant. The definition of market efficiency by Jensen [1978] directly converts the test of market efficiency into the assessment of return behavior. Following this methodology, we test the efficiency of the forward and spot electricity markets by developing robust forecasting models and exploring profitable trading strategies. The trading strategy implemented is backtested using market data in the CAISO electric power markets. Market efficiency is then evaluated in the context of trading performance.

This paper is organized as follows. Section 2 introduces the CAISO's two-settlement electricity markets and the current market design for CB. Section 3 presents the formulation of the virtual trader's optimization problem. Section 4 presents the regime switching model to capture the time-varying forward premium in electricity markets. Section 5 describes the data used in the study. Section 6 examines market efficiency and presents some empirical evidence. Section 7 discusses the implication of market efficiency. Section 8 summarizes the results.

2 CAISO Electric Power Markets

2.1 Pricing Mechanism

Locational marginal prices (LMPs) are the prices used for the settlement of power purchases and sales in the wholesale electricity markets. LMPs are determined by the ISO to maximize social welfare with respect to the physical constraints of the transmission system, and expose producers and consumers to the marginal cost of electricity delivery at different locations. Unlike traditional commodity markets, the wholesale electricity market cannot be cleared with a single clearing-price auction, where the aggregate supply and demand curves are formed and the single clearing price is set to balance the supply and demand. The physical laws governing power flow and the capacity of the transmission lines prevent electricity from flowing freely between producers and customers on the electric power network. When the transmission lines are congested and the import of electricity from cheap producers are constrained, the ISO is forced to use some local but expensive producers for power generation in order to satisfy the demand. As a result, LMPs are high in the downstream areas of the congested transmission lines, and low in the upstream areas. The differences between LMPs in the downstream areas and the upstream areas are congestion rents that reflect the marginal values of the scarce transmission resources. LMPs are calculated for a number of locations on the electric power network. These locations are called nodes, and each node represents the geographic region where physical resources are aggregated.

2.2 Two-Settlement Electricity Markets

The two-settlement electricity markets consist of two interrelated markets: DA market, and RT market. The DA market is a forward market, where energy can be purchased at forward prices, also called day-ahead LMPs (DA LMPs). The RT market is a spot market, where energy can be purchased at spot prices, also called real-time LMPs (RT LMPs). DA LMPs are generally considered more stable than RT LMPs. In the RT market, price spikes are often triggered by unplanned outages of generation plants and transmission facilities, and unpredictable weather, while the DA market is less affected due to a longer planning horizon.

The DA market includes three sequential processes: market power mitigation and reliability requirement determination (MPM-RRD), integrated forward market (IFM), and residual unit commitment (RUC). The MPM-RRD starts the day before delivery. Market participants are allowed to submit supply and demand bids for both physical and virtual trades until the start of the MPM-RRD. In the MPM-RRD, the ISO mitigates bids from physical resources that exercise locational market power, and ensures the availability of physical resources whose outputs are required to maintain local reliability. The results of the MPM-RRD are a pool of bids that is ready for the IFM. In the IFM, the ISO economically clears the supply bids against demand bids with the transmission constraints enforced, determines DA schedules and DA LMPs, and procures ancillary services. When the CAISO forecast of demand exceeds the total physical supply cleared in the IFM, the additional capacity is procured by the ISO in the RUC to satisfy reliability requirements. Note that the additional resources procured in the RUC are not directly used for production, and hence do not receive DA LMPs. However, there are still costs to keep these resources staying online, namely start-up costs and minimum load costs, as discussed later.

In the RT market, the ISO runs the economic dispatch process every 5 minutes to rebalance the residual demand, which is the deviation between the instantaneous demand and the scheduled demand in the DA market. RT LMPs are determined to settle the residual demand and the supply used to balance the residual demand.

The uplift costs are the costs that are incurred in the DA and RT market but are not covered by DA and RT LMPs. The uplift costs include but are not limited to start-up costs and minimum load costs, and are allocated among market participants based on a two-tier cost allocation scheme that considers both causation and socialization. The tier 1 uplift costs account for cost causation, and the tier 2 uplift costs account for cost socialization. Start-up costs are the costs that are incurred when generation plants are turned on, and minimum load costs are the costs that maintain generation plants to operate at the minimum load level.

2.3 CB in Two-Settlement Electricity Markets

CB allows market participants to arbitrage between the DA and RT markets through a financial mechanism, exempting them from physically consuming or producing energy. A virtual demand bid is to make financial purchases of energy in the DA market, with the explicit requirement to sell back that energy in the RT market at the same location. Conversely, a virtual supply bid is to make financial sales of energy in the DA market, with the explicit requirement to buy back that energy in the RT market at the same location. On the physical side, the positions taken in the DA market are offset by the opposite positions in the RT market, which leaves the market participants with no physical obligation. In anticipation of DA LMPs being less than RT LMPs, market participants can make profits by using virtual demand bids to effectively buy energy in the DA market and sell it back in the RT market. These virtual demand bids result in the additional demand in the DA market that increases DA LMPs, and the additional supply in the RT market that decreases RT LMPs. This yields the desired outcome of CB – price convergence.

Price convergence is regarded as a benefit to the DA and RT markets. It reduces the incentive for market participants to defer their physical resources to the RT market in expectation of favorable RT LMPs. The improved stability of the DA market is beneficial from reliability perspectives. To ensure the reliability of the power grid, the ISO is required to procure sufficient capacity in the RUC, when the total physical supply cleared in the IFM is not enough to meet the CAISO forecast of demand. With physical resources withheld by market participants, the ISO tends to over-procure capacity in the RUC. This raises the RUC uplift costs, and increases the risk of decommiting scheduled resources in the RT market when deferred physical resources show up.

The benefit of CB also comes from the fact that it relieves market participants from using physical resources to arbitrage price differences between the DA and RT markets, also called implicit virtual bidding in some literature. Implicit virtual bidding is the bidding strategy where market participants intentionally defer their physical resources to the RT market to take advantage of favorable RT LMPs, by bidding at prices that are unlikely to be cleared in the DA market rather than their economic costs and benefits. Although implicit virtual bidding can achieve price convergence in the absence of CB, it can also lead to disastrous effects that jeopardize the efficiency of the DA and RT markets. Without the revelation of the true economic costs and benefits of physical resources, it is difficult for the ISO to allocate resources efficiently and optimally. In addition, the prices at which market participants bid their physical resources largely depend on their own anticipation of DA and RT LMPs, and this introduces uncertainty into the DA market. In some cases, the ISO can either over-schedule physical supply in the IFM that has to be sold back in the RT market, or under-schedule physical supply in the IFM that relies on the procurement in the RUC to balance. These variations decrease the stability of the DA market, and significantly undermine the reliability of the power grid.

CB can be conducted at both nodes and trading hubs. In comparison to nodes, trading hubs provide more liquidity to trade large volumes of virtual bids. There are three trading hubs in the CAISO electric power markets, that corresponds to three congestion management zones: NP15, SP15 and ZP26. DA and RT LMPs at the trading hub represent the weighted average of prices at generation nodes within the corresponding congestion management zone. The weights are determined annually based on the seasonal generation in the previous year, and are differentiated by peak and off-peak hours. The virtual bids submitted at the trading hub are distributed to generation nodes in proportion to their weights, and are bound together so that they are cleared as a whole in the DA market.

The credit policy for CB requires that the current exposure of virtual bids submitted by a market participant may not exceed the collateral established with the ISO. The current exposure of virtual bids is calculated by the sum of the product of the quantity and the corresponding reference price of each virtual bid. For one node, the reference price is the 95th percentile value of the historical price differences between DA and RT LMPs. After the settlement of virtual bids, the collateral is adjusted based on the realized profits and losses of virtual bids.

There is no transaction fee imposed on submitted virtual bids, but cleared virtual bids are required to pay uplift costs. The costs allocated to cleared virtual bids include the IFM tier 1 uplift costs, and the RUC tier 1 uplift costs. In particular, cleared virtual demand bids are obligated to pay a proportion of the IFM tier 1 uplift costs, as virtual demand bids tend to increase physical supply procured in the IFM. Cleared virtual supply bids are subject to a proportion of the RUC tier 1 uplift costs, as the ISO tends to under-schedule physical supply in the IFM due to virtual supply bids and increase additional capacity procured in the RUC. The costs allocated to 1 MWh of cleared virtual position are estimated to be between \$0.065 and \$0.085 by the CAISO.

3 Portfolio Optimization

3.1 Formulation

In the DA and RT markets, the ISO determines DA LMPs $P_t^{DA} \in \mathbb{R}^{24}$ and RT LMPs $P_t^{RT} \in \mathbb{R}^{24}$ for one node on day t, for $t = 1, \dots, T$. Both P_t^{DA} and P_t^{RT} contain 24 hourly market-clearing prices for 1 MWh of electricity. DA-RT spreads can be expressed as $R_t = P_t^{DA} - P_t^{RT}$. The virtual trader's objective is to maximize the expected payoff of his virtual bids (1) with respect to a budget constraint (2), by entering virtual positions $x_t \in \mathbb{R}^{24}$ in the DA market and closing those positions in the RT market,

$$\text{(P0)} \max_{x_t} \; E\big[R_t^T x_t\big] - \tau \|x_t\|_1 \tag{1}$$

$$s.t. \quad C\|x_t\|_1 \leq W_0 \tag{2}$$

where τ is the costs allocated to 1 MWh of virtual position, C is the reference price for 1 MWh of virtual position, and W_0 is the initial collateral. In this formulation, we implicitly assume that virtual traders behave as price-takers, and that contract can be fractional. $x_t^{(j)} \geq 0$ denotes a virtual supply bid, and we can equivalently view it as taking a long position in the corresponding DA-RT spread, while $x_t^{(j)} < 0$ denotes a virtual demand bid, and we can equivalently view it as taking a short position in the corresponding DA-RT spread.[1] In the budget constraint (2), both supply and demand bids must provide collateral separately, as they are not allowed to offset each other under the current credit policy for CB. This formulation can be easily extended to multiple-node networks.

Without loss of generality, we assume $W_0 = 1$. The collateral used to establish virtual positions in DA-RT spreads is $y_t = Cx_t$ and the costs associated with 1 dollar of collateral are $\tau^c = \frac{1}{C}\tau$. By viewing DA-RT spreads as payoffs from a basket of correlated assets, the returns on DA-RT spreads are then defined as $R_t^c = \frac{1}{C} R_t = \frac{1}{C}\big(P_t^{DA} - P_t^{RT}\big)$. With these substitutions, (P0) is equivalent to (P1),

$$\text{(P1)} \max_{y_t} \; E[R_t^{cT} y_t] - \tau^c \|y_t\|_1 \tag{3}$$

$$s.t. \quad \|y_t\|_1 \leq 1, \tag{4}$$

which is a portfolio optimization problem in the presence of linear transaction costs. The budget constraint (4) requires that the absolute value of weights must sum up to one. This is different from the standard portfolio optimization problem where long and short positions can be netted out.

3.2 Portfolio Optimization under a VaR Constraint

VaR is a modern way of measuring the risk of a portfolio, based on computing probabilities of large losses of the portfolio. Mathematically, $\text{VaR}(z; \eta) = \inf\{\gamma | \text{P}(z \leq \gamma) \geq \eta\}$ is the level η-quantile of the random variable z denoting the losses. To put it another way, the confidence level η is the probability that losses do not exceed or

[1] $x_t^{(j)}$ is the j-th entry of x_t.

equal to $\mathrm{VaR}(z;\eta)$. (P1) can be reformulated as a portfolio optimization problem ($\mathrm{VAR0}(\gamma,\eta)$) under a VaR constraint (6),

$$
\begin{aligned}
(\mathrm{VAR0}(\gamma,\eta))\ \max_{y_t} \quad & E[R_t^{cT} y_t] - \tau^c \|y_t\|_1 && (5)\\
s.t. \quad & \mathrm{VaR}(-R_t^{cT} y_t;\eta) \le \gamma && (6)\\
& \|y_t\|_1 \le 1 && (7)
\end{aligned}
$$

where γ is the predetermined upper bound for the VaR of the portfolio.

As shown in Table 1, DA-RT spreads are negatively skewed in most of the hours, which cannot be modeled properly by a normal distribution. Without assuming normality, VaR cannot be written in a closed form, and there is no guarantee that VaR is convex. Nemirovski and Shapiro [2006] propose a computationally tractable approximation of the non-convex VaR constraint. Therefore, we can replace the VaR constraint (6) with the Chebyshev bound (43) yielding ($\mathrm{VAR1}(\gamma,\eta)$),[2]

$$
\begin{aligned}
(\mathrm{VAR1}(\gamma,\eta))\ \max_{y_t} \quad & \mu_t^T y_t - \tau^c \|y_t\|_1 && (8)\\
s.t. \quad & -E[(R_t^{cT} y_t + \gamma)] + (\eta E[(R_t^{cT} y_t + \gamma)^2])^{\frac{1}{2}} \le 0 && (9)\\
& \|y\|_1 \le 1. && (10)
\end{aligned}
$$

Note that the Chebyshev bound (43) is a conservative approximation of the VaR constraint (6), which implies that the confidence level realized is higher than the confidence level intended η.

3.3 Portfolio Optimization under a CVaR Constraint

Since VaR is incapable of addressing the distribution of losses beyond $\mathrm{VaR}(z;\eta)$, CVaR is introduced by Rockafellar and Uryasev [2000] as an alternative risk assessment technique to account for losses in the tail of the distribution. For continuous distributions, CVaR is defined as the conditional tail expectation exceeding $\mathrm{VaR}(z;\eta)$, $\mathrm{CVaR}(\mathrm{z},\eta) = \mathrm{E}[\mathrm{z}|\mathrm{z} \ge \mathrm{VaR}(\mathrm{z},\eta)]$. In this case, the optimization problem can be stated as follows,

$$
\begin{aligned}
(\mathrm{CVAR0}(\gamma,\eta))\ \max_{y_t} \quad & E[R_t^{cT} y_t] - \tau^c \|y_t\|_1 && (11)\\
s.t. \quad & \mathrm{CVaR}(-R_t^{cT} y_t;\eta) \le \gamma && (12)\\
& \|y_t\|_1 \le 1. && (13)
\end{aligned}
$$

VaR and CVaR can be characterized by function $g_\eta(z,\rho) = \rho + \frac{1}{1-\eta} E[(z-\rho)_+]$ in the following forms,

$$
\mathrm{CVaR}(\mathrm{z},\eta) = \min_\rho g_\eta(z,\rho), \quad (14)
$$

$$
\mathrm{VaR}(\mathrm{z},\eta) = \arg\min_\rho g_\eta(z,\rho). \quad (15)
$$

[2] See Appendix for details.

Thus, by substituting the CVaR constraint (12) with (14), (CVAR0(γ, η)) becomes

$$(\mathrm{CVAR1}(\gamma,\eta)) \ \max_{y_t} \ E[R_t^{cT} y_t] - \tau^c \|y_t\|_1 \tag{16}$$

$$s.t. \quad g_\eta(-R_t^{cT} y_t, \rho) \le \gamma \tag{17}$$

$$\|y_t\|_1 \le 1. \tag{18}$$

4 Regime Switching Model

4.1 Spot Price and Forward Price

In deregulated electricity markets, the prominent features of electricity spot prices include, mean-reversion, seasonality, and spikes. The causes of these features can be traced to the inherent characteristics of electricity. As the supply function of power generation becomes much steeper above a certain capacity level, the marginal production cost increases substantially with the aggregate demand. The consumer demand is highly inelastic and varies widely from season to season, resulting in seasonal variations in the levels of electricity spot prices. The difficulty of storing electricity further limits the feasibility of holding inventories to arbitrage and smooth price discrepancies across time periods. In some extreme cases, price spikes occur when the power system is not flexible enough in response to forced outages of power plants and unexpected contingencies in the transmission networks within a short time frame. Price spikes are frequently seen during the summer, when the demand is high.

Regime switching models seem to be natural candidates to study the dramatic alternations in the behavior of electricity spot prices. Deng [2000] proposes several mean-reversion jump-diffusion models with parameters varying in different regimes to capture the systematic alternations of electricity spot prices among different equilibrium states of supply and demand. Mount, Ning, and Cai [2006] investigate the predictability of price spikes in electricity markets using daily on-peak average spot prices and loads. They adopts a probabilistic model with two regimes, where the state variables are the load and the reserve margin. However, the prediction accuracy decreases substantially when forecasts of the state variables are used.

In electricity forward markets, there is a wide range of tradable instruments with maturities varying from a day, a week, a month, to a year. Here we mainly present studies that focus on modeling forward prices that are settled one day ahead of delivery by regime switching. De Jong [2006] provides statistical evidence that the regime switching model outperforms the generalized autoregressive conditional heteroskedasticity (GARCH) model and the stochastic Poisson jump model. The consistent test results from various day-ahead spot markets in Europe and the United States make a convincing case for the use of regime switching models to capture price dynamics in electricity markets.[3] Haldrup and Nielsen [2006] analyze market data in Nord Pool with a regime switching model that features long memory. They find that the regime switching model is superior to the non-switching

[3] The day-ahead spot market or the spot market in Europe is similar to the DA market in the United States, where the delivery of electricity for each of the 24 hours is settled one day in advance.

model in terms of out-of-sample forecasting performance. Some other successful applications of regime switching models to electricity forward prices are presented in Huisman and Mahieu [2003], Weron [2009], and Janczura and Weron [2010].

4.2 Time-Varying Forward Premium

The forward premium is defined as the difference between the forward price and the expected spot price. In electricity markets, the 24 hourly forward premia FP_t on day t take the form,

$$FP_t = E_{t-1}[P_t^{DA} - P_t^{RT}] = E_{t-1}[R_t]. \tag{19}$$

There exists extensive literature on the time-varying property of the forward premium – a situation where the forward premium varies through time to reflect economic risk. The time-varying forward premium is observed and well documented in exchange rates and traditional commodity markets. In one of the seminal papers, Fama [1984] first attributes the behavior of forward exchange rates to a time-varying forward premium, and finds that the variation in the forward premium accounts for a substantial proportion of the variation in forward exchange rates. In addition to Fama [1984], other papers focusing on explaining the determination of the time-varying forward premium include Fama and French [1987], Bekaert and Hodrick [1993], Backus, Foresi, and Telmer [2001], and Baillie and Kilic [2006].

Recently, there is a growing literature investigating the time-varying forward premium in electricity markets. These studies present empirical evidence that supports the risk-factor-related time variation in the electricity forward premium. Bessembinder and Lemmon [2002] develop a general equilibrium model for forward prices, where the difference between the equilibrium forward price and the expected wholesale price can be explained by risk-related factors that reflect the net hedging pressure of producers and consumers. The risk-related factors are approximated in terms of the central moments of the distribution of wholesale spot prices. To be specific, the electricity forward premium is negatively correlated to spot price volatility, but positively correlated to spot price skewness. The model is empirically verified by using data from the PJM power market and the California Power Exchange (CALPX) at a monthly level.[4] The one-month forward price is estimated by the average of one-month forward prices prior to the delivery month. They also point out that in a frictionless market with risk-neutral outside speculators, the forward prices would converge to the expected spot prices. Based on a data set of hourly spot and forward prices in the PJM power market, Longstaff and Wang [2004] find evidence that supports the structural model presented in Bessembinder and Lemmon [2002] at an hourly level. They also conclude that the forward premium is fundamentally related to the risk premium required by market participants to compensate for uncertainty.

Shawky, Marathe, and Barrett [2003] conduct studies on the spot and future price relationship, based on the contracts traded on the New York Mercantile

[4] The CalPX was founded in 1998. It declared bankruptcy and permanently ceased market operations during 2000-2001 California energy crisis. During its existence, the CALPX administered market transactions, while the CAISO ensured the reliable management of transmission network.

Exchange and delivered at the California-Oregon Border. They find the forward premium of electricity is larger than those of other commodities. An exponential GARCH specification is employed to model the time-varying volatility clustering in the forward premium time series. Cartea and Villaplana [2008] propose a model to forecast wholesale electricity prices in different states identified by two observable state variables – demand and capacity. By testing their model in the PJM, England and Wales, and Nord Pool markets, they present empirical results that the forward premium exhibits a seasonal pattern. The forward premium is high during the months of high demand volatility. Benth, Cartea, and Kiesel [2008] provide a framework to explain the forward premium with two market factors – the levels of risk aversion of buyers and sellers, and the market power of producers relative to that of consumers.

As mentioned, the existing literature extensively studies the time-varying forward premium by statistical models with observable state variables, namely the volatility and skewness of spot prices, the level of risk aversion, market structure, and demand and supply capacity. The choice of state variables is largely predetermined and varies across different electricity markets, which limits the possibility to arrive at a generalization. From a different perspective, the time-varying forward premium can be subject to regime shifts, where the behavior of the forward premium exhibits dramatic changes. Lucia and Schwartz [2002] propose a factor model with unobservable state variables, for the purposes of derivative pricing. These unobservable state variables can be further interpreted as latent market regimes. However, their model is primarily aimed to forecast the forward curve – forward prices with different maturities, rather than the forward premium. To the authors' best knowledge, there is no paper on modeling the electricity forward premium with unobservable states. Our study therefore is intended to fill this gap by introducing a HMM framework to model the regime shifts in the electricity forward premium.

4.3 Model Description

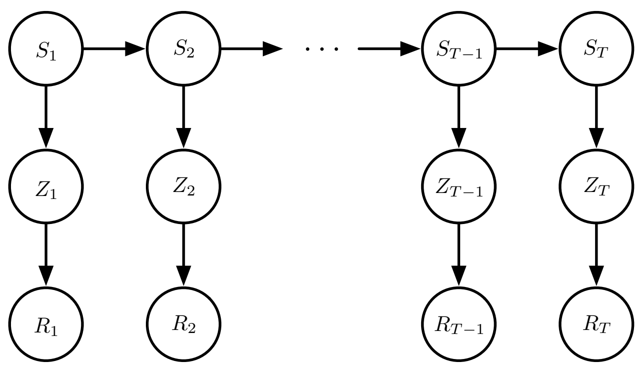

Fig. 1 GMHMM

A HMM can be presented as a dynamic bayesian network model in which the underlying state transition follows a Markov process. Each state has a probability distribution over the possible observations. The state is assumed to be invisible to the observer, but the observation is visible. Therefore some information about the sequence of states can be inferred from the sequence of observations. In the context of CB, $\{S_t, R_t\}_{t=1}^{T}$ is a discrete-time stochastic process, where the sequence of states $\{S_t\}_{t=1}^{T}$ is an unobserved Markov chain. Given $\{S_t\}_{t=1}^{T}$, the observed sequence of DA-RT spreads $\{R_t\}_{t=1}^{T}$ is a sequence of conditionally independent random variables with the conditional distribution depending on $\{S_t\}_{t=1}^{T}$ only through the current state of the chain S_t. In this study, we assume the conditional probability density function of R_t, given the occurrence of S_t, follows a Gaussian mixture distribution. This HMM variant is also called Gaussian mixture hidden Markov model (GMHMM). The GMHMM is illustrated in Figure 1.

We assume there exist M different states in the GMHMM and N different clusters in the Gaussian mixture distribution. The equation for DA-RT spreads R_t given the cluster z_t, for $z_t = 1, \ldots, N$, can be expressed as,

$$R_t = \mu_{z_t} + \Sigma_{z_t}^{\frac{1}{2}} \epsilon_t, \tag{20}$$

where μ_{z_t} denotes the conditional mean given the cluster z_t, Σ_{z_t} denotes the conditional covariance given the cluster z_t, and ϵ_t denotes the noise. Both μ_{z_t} and Σ_{z_t} can take different values depending on the realization of the cluster z_t. The noise term ϵ_t follows a standard multivariate Gaussian distribution $\epsilon_t \sim \mathbf{N}(0, I_{24})$. The cluster z_t follows a multinomial distribution, and occurs with probability $P(z_t|s_t) = c_{s_t,z_t}$, conditioned on the state s_t, for $s_t = 1, \ldots, M$ and $z_t = 1, \ldots, N$. The transition from the present state s_t to the future state s_{t+1} is governed by a transition probability matrix, and the transition probability is $P(s_{t+1}|s_t) = a_{s_t,s_{t+1}}$, for $s_t, s_{t+1} = 1, \ldots, M$.

The GMHMM offers a flexible framework where both the inferences of unobservable states and the estimations of forward premium statistics can be obtained from market data. We denote the historical DA and RT LMPs by p_t^{RT} and p_t^{DA}. Let $r_t = p_t^{DA} - p_t^{RT}$ denote the historical DA-RT spreads. The forward-backward algorithm and the expectation-maximization algorithm are adopted to compute the posterior marginals of state variables and update maximum likelihood estimators respectively, given a sequence of DA-RT spreads r_t. A detailed discussion of the forward-backward algorithm and the expectation-maximization algorithm can be found in Bilmes et al. [1998]. The maximum likelihood estimators are denoted as $\Theta = \{\pi_k, \mu_{k,h}, \Sigma_{k,h}, a_{k,l}, c_{k,h} : k, l = 1, \ldots, M, h = 1, \ldots, N\}$.

4.4 In-Sample and Out-of-Sample Test

We implement the in-sample and out-of-sample test to measure and evaluate the performance of the trading strategy using the historical data. In both tests, the two chance constrained portfolio selection problems (VAR1(γ, η)) and (CVAR1(γ, η)) can be approximated and solved with sampling for a given GMHMM. To illustrate the sampling procedure, we calculate the expected value of a function in general form $f(R_t)$.

In the in-sample test, the whole sequence of DA-RT spreads, $r_1, \ldots, r_T$, is used to train the parameters of GMHMM Θ on day t. The expected function value of

DA-RT spreads $f(R_t)$, conditioned on the whole sequence of DA-RT spreads, can be derived as,

$$E[f(R_t)|r_1,\ldots,r_T] = \sum_{s_t=1}^{M} E[f(R_t)|s_t,r_1,\ldots,r_T]P(s_t|r_1,\ldots,r_T) \quad (21)$$
$$= \sum_{s_t=1}^{M} E[f(R_t)|s_t]P(s_t|r_1,\ldots,r_T) \quad (22)$$
$$= \sum_{s_t=1}^{M}\sum_{z_t=1}^{N} E[f(R_t)|z_t,s_t]P(z_t|s_t)P(s_t|r_1,\ldots,r_T) \quad (23)$$
$$= \sum_{s_t=1}^{M}\sum_{z_t=1}^{N} E[f(R_t)|z_t]c_{s_t,z_t}P(s_t|r_1,\ldots,r_T), \quad (24)$$

where $E[f(R_t)|z_t]$ can be simulated since R_t follows a multivariate Gaussian distribution given the cluster z_t, c_{s_t,z_t} is the maximum likelihood estimator of the cluster probability obtained by the expectation-maximization algorithm, and $P(s_t|r_1,\ldots,r_T)$ is the posterior state probability computed by the forward-backward algorithm.

In the out-of-sample test, only the sequence of available DA-RT spreads up to day t, $r_1,\ldots,r_{t-2}$, is used to train the parameters of GMHMM Θ on day t. We exclude r_{t-1}, because virtual positions for day t must be taken in the morning of day $t-1$, when RT LMPs for the rest of the day are still unavailable for the calculation of r_{t-1} .

The probability of being in the state s_t, conditioned on the sequence of available DA-RT spreads up to day t, can be derived as,

$$P(s_t|r_1,\ldots,r_{t-2}) = \sum_{s_{t-2}=1}^{M} P(s_t,s_{t-2}|r_1,\ldots,r_{t-2}) \quad (25)$$
$$= \sum_{s_{t-2}=1}^{M} P(s_{t-2}|r_1,\ldots,r_{t-2})P(s_t|s_{t-2},r_1,\ldots,r_{t-2}) \quad (26)$$
$$= \sum_{s_{t-2}=1}^{M} P(s_{t-2}|r_1,\ldots,r_{t-2})P(s_t|s_{t-2}), \quad (27)$$

where $P(s_t|s_{t-2})$ is the probability of going from the state s_{t-2} to the state s_t in 2 time steps. The n-step transition probability satisfies the Chapman - Kolmogorov equation, and thus (27) can be rewritten as,

$$P(s_t|r_1,\ldots,r_{t-2}) = \sum_{s_{t-2}=1}^{M} P(s_{t-2}|r_1,\ldots,r_{t-2}) \sum_{s_{t-1}=1}^{M} P(s_t|s_{t-1})P(s_{t-1}|s_{t-2}) \quad (28)$$

The expected function value of DA-RT spreads $f(R_t)$, conditioned on the sequence of available DA-RT spreads up to day t, can be derived as,

$$E[f(R_t)|r_1,\ldots,r_{t-2}] = \sum_{s_t=1}^{M} E[f(R_t)|s_t,r_1,\ldots,r_{t-2}]P(s_t|r_1,\ldots,r_{t-2}) \quad (29)$$

$$= \sum_{s_t=1}^{M} E[f(R_t)|s_t]P(s_t|r_1,\dots,r_{t-2}) \tag{30}$$

$$= \sum_{s_t=1}^{M}\sum_{z_t=1}^{N} E[f(R_t)|z_t,s_t]P(z_t|s_t)P(s_t|r_1,\dots,r_{t-2}) \tag{31}$$

$$= \sum_{s_t=1}^{M}\sum_{z_t=1}^{N} E[f(R_t)|z_t]c_{s_t,z_t}P(s_t|r_1,\dots,r_{t-2}), \tag{32}$$

where $E[f(R_t)|z_t]$, c_{s_t,z_t} and $P(s_t|r_1,\dots,r_{t-2})$ can be computed in the same way as mentioned in the in-sample test.

One distinction between the two tests lies in the fact that virtual positions constructed in the out-of-sample test only relies on the distribution of past DA-RT spreads, while the distribution of both past and future DA-RT spreads are used to determine virtual positions in the in-sample test. By using the predicted distribution of DA-RT spreads, the out-of-sample test produces a robust and credible assessment of the trading strategy. The in-sample test contributes to the evaluation of the trading strategy by allowing us to obtain the most efficient portfolio of virtual positions and achieve the best attainable performance, under the true distribution of DA-RT spreads.

5 Data

The data for this study consists of the historical DA and RT LMPs at the CAISO NP15 EZ Gen Hub before and after the implementation of CB. The data in the pre-CB period includes the historical DA and RT LMPs from January 1st, 2010 to December 31st, 2010, and the data in the post-CB period includes the historical DA and RT LMPs from January 1st, 2012 to December 31st, 2012. For each day, the data contains DA and RT LMPs for each of the 24 hours during that day. The CAISO NP15 EZ Gen Hub is one of the trading hubs in the CAISO electric power markets, and covers the current CAISO congestion management zone NP15.[5]

6 Numerical Results

Here we only present results in the post-CB period. Similar features are observed in the pre-CB period, which we do not report to save space.

6.1 Summary Statistics for the DA and RT Markets

Table 1 presents summary statistics for post-CB DA-RT spreads in dollars per megawatt hour. Post-CB DA-RT spreads can also be viewed as realized or ex post forward premia. The mean of post-CB DA-RT spreads varies throughout the day. Large negative spreads are observed during peak hours. The volatility of post-CB DA-RT spreads is higher during peak hours than during off-peak hours.

[5] The majority of Pacific Gas and Electric Company's load is located in NP15.

Table 1 Summary Statistics for Post-CB DA-RT Spreads

Hour	Mean	Standard Deviation	Skewness
1	1.81	7.01	2.93
2	1.98	9.31	2.61
3	2.34	11.24	2.03
4	3.08	13.44	2.62
5	1.32	11.05	1.90
6	-0.10	14.23	-7.33
7	1.45	19.88	-4.94
8	-0.56	38.08	-12.09
9	-0.54	30.92	-12.10
10	-2.31	43.33	-10.24
11	-2.68	56.07	-12.03
12	0.22	35.92	-12.32
13	0.14	36.41	-9.19
14	2.19	32.08	-13.23
15	-0.29	39.99	-7.67
16	-2.43	77.89	-10.70
17	-5.68	78.68	-6.23
18	-5.39	73.41	-6.24
19	-3.35	50.58	-4.61
20	-1.23	46.79	-6.91
21	1.57	26.62	-6.90
22	-1.74	42.67	-7.49
23	0.07	17.32	-4.38
24	1.26	24.67	-10.32
Overall	-0.37	40.67	-11.71

Post-CB DA-RT spreads are negatively skewed in most of the hours, because price spikes occur frequently in the RT market during the summer. The overall mean of post-CB DA-RT spreads (-\$0.37) is closer to zero than the overall mean of pre-CB DA-RT spreads (-\$2.36), which indicates better price convergence after the introduction of CB.[6]

Table 2 and Table 3 show the seasonal means and standard deviations of post-CB DA-RT spreads in dollars per megawatt hour. Both exhibit strong seasonal patterns, especially for peak hours. In particular, the means of post-CB DA-RT spreads for 5 p.m. range from a low of -\$23.82 during the period from May to July to a high of \$3.55 during the period from November to January. The large negative mean values of post-CB DA-RT spreads are observed during the period from May to July, as a result of the price spikes that occur regularly throughout the summer in the RT market. The lowest overall mean of post-CB DA-RT spreads is observed during the period from May to July, and the highest overall standard deviation of post-CB DA-RT spreads is also observed during the same period. This seasonal variation is consistent with the Bessembinder and Lemmon [2002] model in that downward hedging pressure is imposed on the forward premium by the variance. The strong seasonal patterns raise the need to incorporate a time-varying property in the forward premium model, and support the use of the GMHMM characterized by the time-varying conditional mean and variance. It is also worth noting that off-peak hours do not display significant seasonal effects as peak hours do. The means and standard deviations of post-CB DA-RT spreads in off-peak hours show

[6] Summary statistics for pre-CB DA-RT spreads are not reported in this paper.

Table 2 Seasonal Means of Post-CB DA-RT Spreads

Hour	November - January	February - April	May - July	August - October
1	1.84	1.18	2.25	1.95
2	1.98	2.71	1.63	1.61
3	1.47	4.25	2.02	1.68
4	2.93	3.46	3.19	2.78
5	1.51	1.44	1.79	0.55
6	-0.74	-1.73	1.85	0.18
7	-0.16	0.34	5.82	-0.29
8	2.32	-3.53	-3.70	2.58
9	-3.99	-0.60	2.44	-0.14
10	-1.44	-10.37	2.03	0.27
11	0.15	-11.31	-0.72	0.90
12	-0.70	-5.47	1.85	4.99
13	-2.25	-1.93	0.92	3.70
14	2.76	-0.13	-2.22	8.27
15	1.09	-2.11	-1.71	1.43
16	0.58	1.15	-8.49	-3.04
17	2.68	-0.98	-14.45	-9.92
18	3.55	2.02	-23.82	-3.18
19	0.35	2.46	-7.51	-8.79
20	3.32	-0.39	-7.79	-0.17
21	0.93	3.39	1.38	0.44
22	-1.85	1.27	-10.84	4.32
23	-0.75	0.43	-1.19	1.66
24	1.88	2.02	-2.47	3.61
Overall	0.73	-0.52	-2.41	0.64

Table 3 Seasonal Standard Deviations of Post-CB DA-RT Spreads

Hour	November - January	February - April	May - July	August - October
1	7.40	7.31	8.20	4.77
2	10.37	10.48	9.36	6.72
3	8.54	13.02	13.68	8.78
4	11.23	15.42	15.22	11.65
5	9.53	10.54	14.43	9.05
6	17.46	20.04	9.15	5.22
7	19.30	22.21	14.97	22.00
8	9.44	55.29	51.24	8.76
9	56.30	17.93	7.34	17.61
10	31.29	77.13	15.91	20.01
11	30.15	98.84	27.36	36.17
12	29.90	60.60	23.17	10.13
13	53.25	20.47	36.14	27.79
14	9.72	10.40	61.68	7.89
15	16.01	21.82	56.80	49.59
16	14.09	6.75	99.58	118.79
17	14.82	20.07	111.17	108.03
18	22.96	10.59	117.88	81.25
19	40.83	12.25	47.47	78.20
20	13.03	29.47	60.89	63.16
21	20.30	7.40	21.73	43.60
22	40.20	7.09	70.50	23.48
23	18.00	10.51	22.21	16.58
24	14.12	14.65	44.51	5.70
Overall	25.50	33.81	51.67	46.29

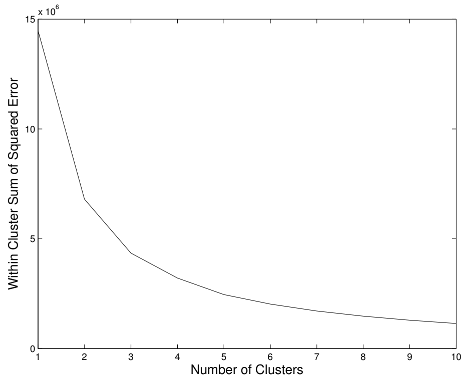

Fig. 2 Post-CB Within-Cluster Sum of Squared Error

Table 4 Transition Probabilities of the Post-CB GMHMM

	State 1	State 2
State1	95.23%	4.77%
State2	4.00%	96.00%

Table 5 Cluster Probabilities of the Post-CB GMHMM

	Cluster 1	Cluster 2	Cluster 3
State1	89.65%	10.35%	0.00%
State2	56.84%	42.10%	1.05%

relatively small variation across different periods, compared with those in peak hours.

6.2 Summary Statistics for the GMHMM

Several heuristic procedures for model selection are applied to determine the number of states and the number of clusters in the GMHMM. We choose the number of states $M = 2$ to avoid the overfitting problem commonly encountered in learning a large state-space HMM. A simple model also allows us to provide clear economic interpretations for different states, which are discussed later. One common method of choosing the appropriate number of clusters is to graph the within cluster sum of squared error against the number of clusters in Figure 2. The appropriate number of clusters can be defined as the number at which the reduction in the within cluster sum of squared error slows significantly. As demonstrated in Figure 2, to increase the number of clusters reduces the the within cluster sum of squared error, but at 3 clusters the marginal gain drops suggesting that additional clusters do not have a substantial impact on the within cluster sum of squared error. It produces an "elbow" in the graph at 3 clusters. Hence, we choose the number of clusters $N = 3$, according to this "elbow criterion". After model selection, statistical inference and estimation can then be conducted by applying the forward-backward

Table 6 Summary Statistics for DA-RT Spreads in the Clusters of the Post-CB GMHMM

	Mean			Standard Deviation		
Hour	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
1	2.20	0.75	1.73	7.21	6.38	2.68
2	1.95	2.08	0.83	8.98	10.22	2.25
3	2.50	1.93	0.59	11.13	11.57	1.43
4	3.54	1.91	1.21	13.55	13.21	1.88
5	2.13	-0.84	0.22	11.57	9.28	1.38
6	1.18	-3.53	-1.30	10.28	21.27	1.59
7	1.87	0.33	-2.27	20.95	16.81	3.36
8	3.18	-5.89	-236.75	7.11	53.68	240.32
9	2.67	-9.28	-1.39	5.92	57.98	3.11
10	3.70	-18.71	0.07	6.95	80.59	4.32
11	4.70	-22.88	3.69	7.67	104.85	3.55
12	4.84	-12.16	-7.55	6.37	66.95	0.87
13	4.73	-11.70	-29.70	6.62	67.73	27.41
14	5.17	-3.67	-107.51	7.91	56.28	97.00
15	6.04	-9.11	-412.31	8.12	44.78	77.20
16	7.35	-9.84	-944.17	10.75	59.77	119.64
17	9.99	-36.38	-583.82	13.61	113.61	244.14
18	9.55	-37.59	-412.65	12.06	109.88	355.19
19	6.17	-29.78	19.05	22.38	84.75	40.92
20	7.92	-26.74	28.00	9.14	83.73	32.35
21	5.95	-10.86	22.44	8.52	47.05	23.38
22	5.35	-21.47	16.41	5.77	78.43	13.56
23	1.18	-3.09	1.20	14.77	22.62	3.48
24	3.91	-6.01	4.91	5.89	45.84	2.16

algorithm and the expectation-maximization algorithm on the historical data. In particular, note that the maximum likelihood estimators presented in this section are estimated using the whole sequence of DA-RT spreads to obtain a complete picture of the property of the forward premium across seasons.

The transition probabilities of the post-CB GMHMM are shown in Table 4. The transition probability from one state to itself is over 90%, which implies that the alternations between states occur at a relatively low frequency in the underlying state transition process. It captures the fact that the forward premium time series exhibits seasonal patterns and evolves slowly from season to season. Table 6 shows summary statistics for DA-RT spreads of the clusters of the post-CB GMHMM in dollars per megawatt hour.[7] Each cluster is represented by a multivariate Gaussian distribution characterized by its mean vector and covariance matrix. For most of the hours, the means are positive in cluster 1, and negative in cluster 2 and cluster 3. The standard deviations in cluster 2 are uniformly larger than those in cluster 1, indicating a higher level of volatility. However, cluster 3 behaves very differently from the other two clusters, and can be interpreted as a cluster where DA-RT spreads are highly volatile, especially during several specific peak hours, including 7 a.m. and 2 p.m. to 5 p.m. During these peak hours, the means in cluster 3 are lower than -\$400, while the lowest mean value in cluster 1 and cluster 2 is -\$37.59 during the corresponding hours. The standard deviations in cluster 3 are also significantly larger than those in cluster 1 and cluster 2 for these hours. Table

[7] A full covariance matrix is estimated in this study, but only diagonal elements are presented in Table 6 to convey insights.

Table 7 Summary Statistics for DA-RT Spreads in the States of the Post-CB GMHMM

	Mean		Standard Deviation		Skewness	
Hour	State 1	State 2	State 1	State 2	State 1	State 2
1	2.05	1.59	7.16	6.89	0.00	0.05
2	1.98	1.96	9.15	9.47	0.01	0.01
3	2.49	2.18	11.25	11.24	0.01	-0.01
4	3.41	2.76	13.51	13.31	0.00	0.01
5	1.86	0.85	11.36	10.73	0.03	0.07
6	0.77	-0.82	11.93	15.90	-0.27	-0.27
7	1.74	1.17	20.54	19.22	0.02	0.03
8	2.17	-2.86	18.86	48.21	-1.06	-4.50
9	1.46	-2.41	19.75	38.74	-1.41	-0.49
10	1.61	-5.69	26.90	54.06	-1.74	-0.67
11	2.06	-6.80	34.95	69.87	-1.56	-0.64
12	3.13	-2.29	22.63	44.73	-1.57	-0.54
13	3.14	-2.76	23.28	45.07	-1.55	-0.58
14	4.38	0.43	19.55	39.98	-0.93	-0.88
15	4.36	-4.34	16.95	51.33	-1.92	-5.51
16	5.50	-9.14	22.20	102.52	-1.47	-7.92
17	5.05	-15.15	41.28	98.80	-2.57	-3.18
18	4.34	-14.22	40.55	91.68	-2.81	-3.02
19	2.33	-8.59	36.32	60.44	-1.53	-0.79
20	4.22	-6.54	30.29	57.39	-2.77	-0.98
21	4.18	-1.00	18.04	32.37	-1.91	-0.80
22	2.53	-5.76	26.68	52.36	-2.40	-0.87
23	0.74	-0.60	15.83	18.63	-0.09	-0.16
24	2.94	-0.25	15.84	30.34	-1.36	-0.53

5 reports the cluster probabilities of the post-CB GMHMM. As shown in Table 5, cluster 3 is not historically observed in state 1 and occurs with very low probability in state 2. This is consistent with the fact that DA-RT spreads in cluster 3 exhibit occasional extreme price movements of magnitudes that can only be observed during the summer, but rarely seen for the rest of the year.

Table 7 shows summary statistics for DA-RT Spreads of the states of the post-CB GMHMM in dollars per megawatt hour. The means in state 1 are higher than those in state 2, since more observations in state 1 are drawn from cluster 1 as shown in Table 5 and cluster 1 exhibits higher means. Similarly, the standard deviations in state 1 are smaller than those in state 2. Therefore, we can interpret state 1 as a low volatility state, and state 2 as a high volatility state. Similar implications can be seen from Figure 3. In Figure 3, the posterior probability of being in state 1 is low during the summer, and high during the rest of the year. After the inference, the posterior probability of being in state 1 is adjusted based on the empirical evidence that DA-RT spreads are most volatile during the summer to correctly reflect the updated belief that the occurrence of state 1 which exhibits low volatility is rather unlikely during this period. Finally, we note that the negative skewness shown in Table 7 is consistent with summary statistics in Table 1.

Figure 4 and Figure 5 plot the marginal distribution of the post-CB DA-RT spread for for 1 a.m. and 1 p.m., representing peak hours and off-peak hours respectively. During off-peak hours, the marginal distributions of pre-CB DA-RT spreads are almost identical in the two states. During peak hours, however, the marginal distribution of pre-CB DA-RT spreads in state 1 has more density con-

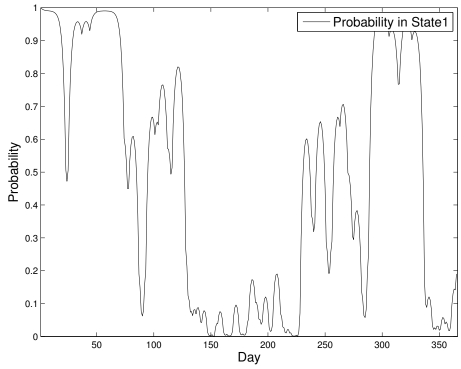

Fig. 3 Post-CB Posterior State Probability

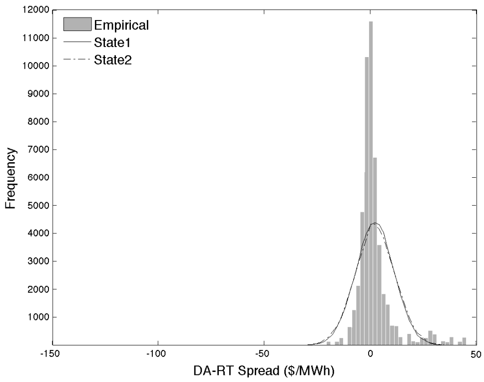

Fig. 4 Marginal Distribution of Post-CB DA-RT Spreads for 1 a.m.

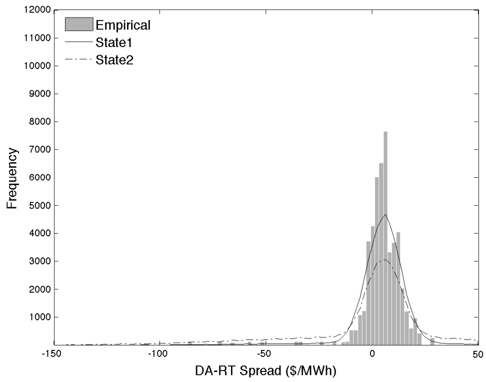

Fig. 5 Marginal Distribution of Post-CB DA-RT Spreads for 1 p.m.

centrated around the mean and less in both tails, compared to that in state 2. The difference of the marginal distributions between the two states is supported by the findings we report in Table 2 and Table 3 that seasonal patterns are stronger for off-peak hours.

All of these results demonstrate that many stylized facts of the time-varying forward premium can be well captured and accommodated in the GMHMM framework.

6.3 Test for the Bessembinder and Lemmon [2002] Model

Bessembinder and Lemmon [2002] model the forward market as a closed system, where the only participants are producers and consumers. In their general equilibrium model, the forward premium reflects the net hedging pressure of producers and consumers, and the sign of the forward premium is indeterminate. The forward premium can be expressed as,

$$P_t^{DA} - E[P_t^{RT}] = \theta_1 Var[P_t^{RT}] - \theta_2 Skew[P_t^{RT}], \qquad (33)$$

where $\theta_1 \leq 0$, and $\theta_2 \leq 0$, implying that the forward premium is negatively related to the variance of RT LMPs, and positively related to the skewness of RT LMPs. To express forward premia in terms of DA-RT Spreads $R_t = P_t^{DA} - P_t^{RT}$, we can rewrite (33) as,

$$E[R_t] = \theta_1 Var[R_t] + \theta_2 Skew[R_t]. \tag{34}$$

To further test the implications of Bessembinder and Lemmon [2002], we regress the means for each of the 24 hours in the 2 states of the post-CB GMHMM on the corresponding variance and skewness measures in Table 7. The regression specification can be written in the form of (35),

$$StateMean_i = \theta_0 + \theta_1 StateVar_i + \theta_2 StateSkew_i + \epsilon_i. \tag{35}$$

As shown in Table 8, θ_1 is negative with a t-statistic of -13.4369, and θ_2 is negative with a t-statistic of -4.2229. Both of θ_1 and θ_2 are significant at 1% level and the R-squared value for the regression is 0.8246, which strongly supports the empirical implications of the Bessembinder and Lemmon [2002] model.

These test results provide evidence against the efficiency of the current CAISO DA and RT markets. The theoretical framework of Bessembinder and Lemmon [2002] essentially describes the characteristics of the electricity forward premium in an inefficient market, where the risks are borne within the industry by a few producers and consumers due to a lack of risk-sharing mechanisms. As financial mechanisms are implemented to allow risk neutral outside traders to enter the market and share the risks, the risk premium should decline and these characteristics are expected to disappear for the market to be fully efficient. Nevertheless, in the post-CB period, we demonstrate that the electricity forward premium still displays characteristics similar to what Bessembinder and Lemmon [2002] describe for an inefficient market in the absence of risk neutral outside traders, which to some extent implies market inefficiency.

Table 8 Post-CB Regression Analysis

θ_0	θ_1	θ_2	t_{θ_0}	t_{θ_1}	t_{θ_2}	R Squared	DF
1.8041	-0.0021	-1.0617	4.6235	-13.4369	-4.2229	0.8246	45

6.4 Pre-CB and Post-CB Performance

To test for market efficiency, we backtest the trading strategy on the last 120 days in the pre-CB and post-CB period respectively, and all the rest of the data is used in training. We adopt several popular metrics for performance assessment. The annualized expected return and the annualized standard deviation directly measure the reward and risk of the trading strategy converted to an annual basis. The Sharpe ratio, also known as the reward-to-variability ratio, is a risk-adjusted measure used to evaluate the quality of the return. The ratio is calculated by using excess return and standard deviation to determine reward per unit of risk.[8]

[8] We assume the risk-free interest rate is 3% in the calculation of excess return.

These standard measures of risk, however, do not account for the risk exposure associated with skewness, kurtosis, and serial correlation of the return distribution. In such cases, we include the maximum drawdown, that measures the greatest loss from a historical peak in the cumulative return, as an additional measure of the worst-case risk.

In the backtest, the predetermined upper bound γ for both the VaR and CVaR constraints is set to 0.02. To investigate the robustness of the trading strategy, we vary the choice of confidence levels η. As the Chebyshev bound (9) in (VAR1(γ,η)) is a conservative approximation of the VaR constraint (6), we tend to use lower confidence levels for the VaR constraint than for the CVaR constraint to ensure the risks of the two portfolios obtained from (VAR1(γ,η)) and (CVAR1(γ,η)) comparable. Therefore we vary η from 0.90, 0.85 to 0.80 for the VaR constraint, and from 0.99, 0.98 to 0.95 for the CVaR constraint. The costs τ are assumed to be \$0.085 for 1 MWh of cleared virtual position and the reference price C for 1 MWh of cleared virtual positionand is calculate by the 95th percentile value of the historical price differences between DA and RT LMPs.[9]

Table 9 Performance under a VaR Constraint

Strategy Parameter	Expected Return	Standard Deviation	Sharpe	Maximum Drawdown
		Pre-CB In-Sample Performance		
$\gamma = 0.02$ $\eta = 0.90$	153.42%	14.60%	10.34	1.12%
$\gamma = 0.02$ $\eta = 0.85$	187.50%	19.54%	9.48	1.53%
$\gamma = 0.02$ $\eta = 0.80$	210.34%	24.71%	8.43	2.09%
		Pre-CB Out-of-Sample Performance		
$\gamma = 0.02$ $\eta = 0.90$	90.36%	24.50%	3.58	6.91%
$\gamma = 0.02$ $\eta = 0.85$	118.70%	29.42%	3.95	8.96%
$\gamma = 0.02$ $\eta = 0.80$	150.92%	35.15%	4.23	9.45%
		Post-CB In-Sample Performance		
$\gamma = 0.02$ $\eta = 0.90$	78.03%	11.10%	6.79	2.23%
$\gamma = 0.02$ $\eta = 0.85$	86.66%	13.08%	6.42	2.82%
$\gamma = 0.02$ $\eta = 0.80$	87.61%	15.21%	5.59	3.72%
		Post-CB Out-of-Sample Performance		
$\gamma = 0.02$ $\eta = 0.90$	33.47%	13.03%	2.35	2.73%
$\gamma = 0.02$ $\eta = 0.85$	35.11%	14.35%	2.25	3.19%
$\gamma = 0.02$ $\eta = 0.80$	38.03%	16.55%	2.13	5.08%

Table 9 and Figure 6 - Figure 9 reports the trading performance under a VaR constraint in the pre-CB and post-CB period. To begin with, we investigate the behavior of the trading strategy under different confidence levels η. In Table 9, both the annualized standard deviation and the maximum drawdown are negatively related to the confidence level, as a higher confidence level imposes a tighter bound for probability of tail events and hence lower risks are undertaken in the portfolio. The annualized expected return is positively related to the annualized standard deviation. This is what one might expect based on economic theory that high potential returns are associated with high levels of uncertainty. In terms of risk-adjusted performance measures, the low risk trading strategy exhibits a high

[9] The upper bound of the estimated costs allocated to 1 MWh of cleared virtual position is used to ensure the robustness of our results.

Sharpe ratio in all cases except for the out-of-sample performance in the pre-CB period, as the Sharpe ratio penalizes the trading strategy that generates high but volatile returns.

Table 10 Performance under a CVaR Constraint

Strategy Parameter	Expected Return	Standard Deviation	Sharpe	Maximum Drawdown
Pre-CB In-Sample Performance				
$\gamma = 0.02$ $\eta = 0.99$	267.77%	37.04%	7.18	1.19%
$\gamma = 0.02$ $\eta = 0.98$	296.75%	44.60%	6.61	1.14%
$\gamma = 0.02$ $\eta = 0.95$	341.83%	59.30%	5.74	1.42%
Pre-CB Out-of-Sample Performance				
$\gamma = 0.02$ $\eta = 0.99$	245.86%	47.78%	5.10	6.59%
$\gamma = 0.02$ $\eta = 0.98$	266.79%	55.65%	4.76	7.23%
$\gamma = 0.02$ $\eta = 0.95$	284.81%	66.01%	4.29	7.80%
Post-CB In-Sample Performance				
$\gamma = 0.02$ $\eta = 0.99$	47.35%	11.54%	3.86	2.75%
$\gamma = 0.02$ $\eta = 0.98$	52.93%	13.29%	3.77	3.16%
$\gamma = 0.02$ $\eta = 0.95$	60.82%	16.58%	3.50	4.60%
Post-CB Out-of-Sample Performance				
$\gamma = 0.02$ $\eta = 0.99$	22.58%	16.56%	1.19	4.02%
$\gamma = 0.02$ $\eta = 0.98$	25.55%	17.87%	1.27	5.29%
$\gamma = 0.02$ $\eta = 0.95$	22.18%	22.83%	0.84	9.01%

We further compare the performance in the pre-CB and post-CB periods. There is little disparity between the in-sample and out-of-sample tests in their relative performance before and after the implementation of CB. In particular, to compare the post-CB metrics against the pre-CB metrics, we see the annualized expected returns and the Sharpe ratios in both of the tests drop dramatically. It is plausible because in the post-CB period virtual traders who engage in arbitrage trades tend to use the trading strategy, which proves to be profitable in the pre-CB period. As these arbitrage trades have the effect of causing price convergence between the DA and RT markets, the profitability is significantly eroded in the post-CB period, which serves to be convincing evidence for the improvement of market efficiency brought about by CB.

We now proceed to explore whether the current CAISO DA and RT markets are efficient, and if not, the extent to which the implementation of CB enhances market efficiency in the post-CB period. For this purpose, we primarily focus on the out-of-sample test. Examining the out-of-sample performance metrics in Table 9 reveals that the trading strategy generates profits in the presence of transaction costs in the post-CB period. The out-of-sample Sharpe ratios under different parameters range from 2.13 to 2.35. For medium frequency strategies, the Sharpe ratio of the S&P 500 is commonly used as a benchmark, which Modigliani and Modigliani [1997] estimate to be about 0.30 based on quarterly returns ten years ending the second quarter of 1996.[10] All the out-of-sample Sharpe ratios exceed this benchmark by substantial margins in the post-CB period. In addition, the maximum drawdowns are small, if not negligible, in comparison with the corre-

[10] Generally, low, medium, and high trading frequencies are defined as position holding periods of months, days, and hours, respectively.

sponding annualized expected returns, which indicates that daily losses tend to be small and the occurrence of consecutive losses is uncommon. Taken together, these results provide compelling evidence that profitable trading opportunities still exist and consistent returns can be generated by exploiting these trading opportunities. Hence, market efficiency is not fully restored in the current CAISO DA and RT markets in the sense of Jensen [1978].

Table 10 and Figure 10 - Figure 13 reports the trading performance under a CVaR constraint in the pre-CB and post-CB period. Similar features are observed as Table 9 and Figure 6 - Figure 9 .

Combining the evidence against the efficiency of the current CAISO DA and RT markets with the evidence for the improvement of market efficiency brought about by CB, we are inclined to conclude that the current CAISO DA and RT markets lie somewhere between the two extremes, the inefficient markets as they were and the fully efficient markets that we pursue.[11] During this transition phase of CB, trading experience and market knowledge are accumulated among market participants within the industry, which serves as a necessary condition for the development of fully efficient DA and RT markets.

7 Market Power, Risk Averse Speculation and Implications

The efficient market hypothesis, which implicitly assumes that economic agents are risk neutral and have no market power, is called the simple efficiency hypothesis in Hansen and Hodrick [1980]. In the forward market, the simple efficiency hypothesis implies that forward prices are unbiased predictors of expected spot prices. These risk neutral and competitive market conditions are also implicitly assumed in the statement by Jensen [1978] that no trading strategy can consistently profit from an efficient market.

However, if the market is not competitive, economic agents can exercise market power by withholding their bidding quantities below the competitive levels to maximize their profits. Thus, as a result, price discrepancies are not fully arbitraged away and the zero-profit competitive equilibrium as described by Jensen [1978] is no longer attained. We argue this is an unlikely case in the current CAISO DA and RT markets. There are currently over 70 market participants registered with the CAISO to participate in CB, including electricity producers and consumers, their trading subsidiaries, investment banks, and energy trading firms.[12] The latter two are sophisticated virtual traders, who do not own physical resources and engage in arbitrage activities for pure financial incentives. They are generally large corporations, and have sufficient access to capital to perform intended trading strategies. Therefore, in the presence of low transaction costs and full nodal granularity, it is reasonable to assume a high degree of competition among these well-informed and

[11] This is consistent with Jha and Wolak [2013] that after the implementation of CB the implied trading costs decrease but the difference between the implied trading costs before and after the implementation of CB is relatively small at the NP15 trading hub for all three hypothesis tests.

[12] CAISO List of Scheduling Coordinators (SCs), Congestion Revenue Rights (CRR) Holders, CB Entities as updated on July 8th, 2014.
http://www.caiso.com/Documents/ISOListofSCsCRRsCBEs_July_2014pdf.pdf. Accessed July 23rd, 2014.

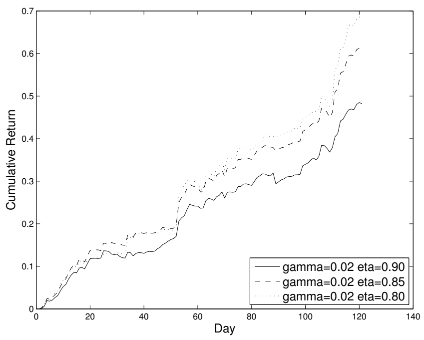

Fig. 6 Pre-CB In-Sample Performance under a VaR constraint

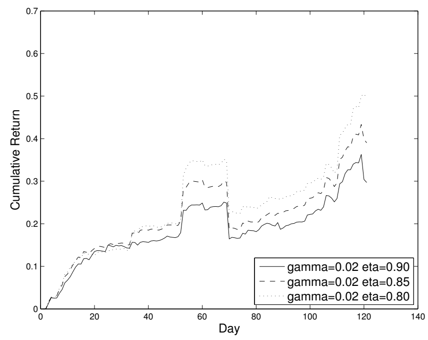

Fig. 7 Pre-CB Out-of-Sample Performance under a VaR constraint

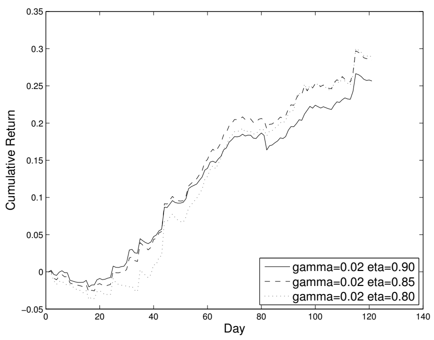

Fig. 8 Post-CB In-Sample Performance under a VaR constraint

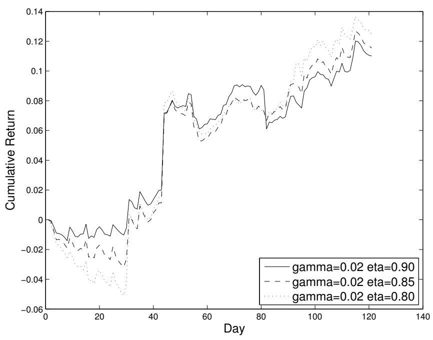

Fig. 9 Post-CB Out-of-Sample Performance under a VaR constraint

well-financed virtual traders that reduces the possibility of market manipulation and prevents the possession of excessive market power.

If, on the other hand, we extend the assumption to allow for risk averse economic agents, forward prices can systematically deviate from expected spot prices, which compromises the statement of Jensen [1978]. Some intuitions are provided by Keynes [1923]. He explains that, in a commodity market where only hedgers and speculators can take positions, the forward premium is determined by the net hedging demand.[13] As consumers traditionally show no interest in participation possibly due to informational setup costs, hedgers in this forward market are typically producers who are endowed with initial long positions in goods.[14] When facing price uncertainty, hedgers are net short to offset the exposures to their initial long positions, and therefore speculators have to be net long. Speculators are only willing to bear the risks if expected returns to speculation are positive, that is, forward prices are downward biased relative to expected spot prices. In

[13] To follow the convention in the finance literature, we use the term “speculators” referring to informed traders who explore price deviations and stabilize prices. The term “speculators” and “traders” are interchangeable in our paper.

[14] Informational setup costs is the implicit costs of collecting and analyzing market information.

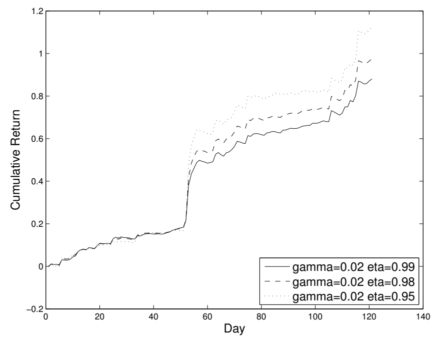

Fig. 10 Pre-CB In-Sample Performance under a CVaR constraint

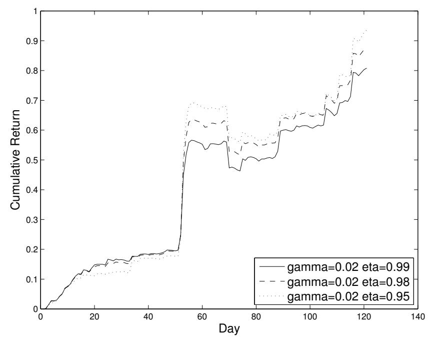

Fig. 11 Pre-CB Out-of-Sample Performance under a CVaR constraint

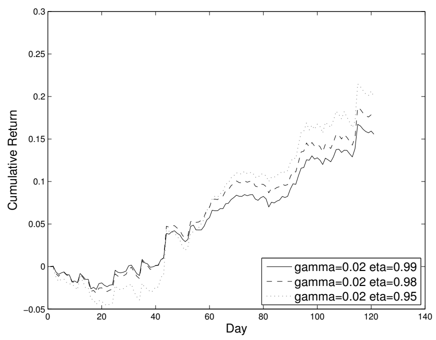

Fig. 12 Post-CB In-Sample Performance under a CVaR constraint

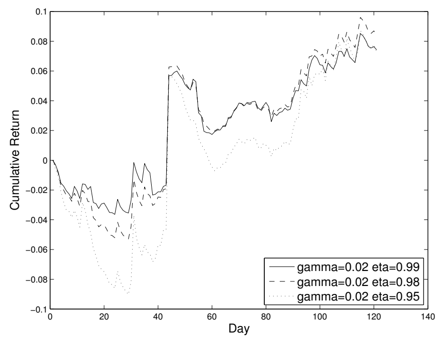

Fig. 13 Post-CB Out-of-Sample Performance under a CVaR constraint

contrast, Rolfo [1980], Anderson and Danthine [1983], and Hirshleifer [1990] argue that forward prices can be upward-biased predictors of expected spot prices, when hedgers are subject to both price and quantity uncertainties. If demand is elastic, quantity uncertainty has a more pronounced effect than price uncertainty from the perspective of hedgers, since small price fluctuations lead to large quantity fluctuations. Hedgers are then net long to reduce their quantity risks. Speculators are only willing to be net short, if forward prices constantly lie above expected spot prices. Hirshleifer [1990] shows that, apart from quantity uncertainty, the participation of consumers can also result in upward-biased forward prices, because the hedging incentives of consumers are opposite to those of producers. This way, we can think of both the upward and downward deviations of forward prices from expected spot prices as the costs of hedging, or the returns to speculation.

In this sense, rationally, risk averse virtual traders only employ trading strategies that are capable of generating returns high enough to compensate for the risks undertaken, to arbitrage cautiously between the DA and RT markets. For virtual traders with relatively low risk-tolerance, they might require a higher risk-adjusted return than that of the trading strategy we develop to enter the forward market. If that is the case, the DA and RT markets can still be efficient in the presence of profitable trading strategies, and the profitability of these trading strategies only reflects the competitive returns to induce the participation of those risk averse virtual traders.

We tend to believe this is not the case for three reasons. Firstly, virtual traders in the current CAISO DA and RT markets are investment banks and trading firms, who are by nature less risk averse, if not risk neutral. Secondly, the Sharpe ratios of our trading strategy in Table 9 are significantly higher than that of the S&P 500, we can reasonably assume that the trading strategy indeed provides a decent risk-adjusted return. Thirdly, when there exists sufficient competition among virtual traders, the price deviation between the DA and RT markets is expected to be kept small, even if virtual traders are risk averse. The small DA-RT spreads should certainly lead to the the demise of the profitability in the post-CB period, which is contrary to what we observe in Table 9.

Although our profitable trading strategy indicates market inefficiency, it is not in itself sufficient evidence for us to fully reject the efficiency of the current CAISO DA and RT markets, with these unjustified concerns on market power and risk aversion in mind. A thorough investigation of these alternative hypotheses requires the estimation of risk aversion parameters of speculators and the level of competition, which is beyond the scope of this study, and is left for future research.

8 Conclusion

In this study, we investigate whether the current CAISO DA and RT markets are efficient, and whether markets efficiency is improved by CB, based on the zero-profit condition of Jensen [1978]. In the backtest, our results show that our trading strategy continues to be profitable in the post-CB period, but the profitability is at a lower magnitude, compared to the profitability in the pre-CB period. Clearly, the deteriorated profitability in the post-CB period provides compelling evidence for the improved market efficiency, which demonstrates the benefit of CB. The profitability in the post-CB period, however, conveys empirical implications that can be interpreted differently, depending on the level of competition and the level of risk aversion of virtual traders. If virtual traders are risk-neutral and the competition among virtual traders is intense, the profitability in the post-CB period is convincing evidence against the fully efficient DA and RT markets. Otherwise, the profitability in the post-CB period might only rationally reflect the economic profit to incentivize the participation of risk averse virtual traders, which has nothing to do with market inefficiency and the mispricing of financial instruments.

Last but not least, market efficiency does not come for free. A market becomes or remains efficient as a result of the persistent efforts of market participants, who conduct research, identify inefficiencies, and trade until these inefficiencies disappear. In this sense, we encourage market participants to search for profitable trading strategies, make profits, and ultimately improve the efficiency of electricity forward markets.

References

Ronald W Anderson and Jean-Pierre Danthine. Hedger diversity in futures markets. *The Economic Journal*, pages 370–389, 1983.

David K Backus, Silverio Foresi, and Chris I Telmer. Affine term structure models and the forward premium anomaly. *The Journal of Finance*, 56(1):279–304, 2001.

Richard T Baillie and Rehim Kilic. Do asymmetric and nonlinear adjustments explain the forward premium anomaly? *Journal of International Money and Finance*, 25(1):22–47, 2006.

Geert Bekaert and Robert J Hodrick. On biases in the measurement of foreign exchange risk premiums. *Journal of International Money and Finance*, 12(2): 115–138, 1993.

Fred Espen Benth, Álvaro Cartea, and Rüdiger Kiesel. Pricing forward contracts in power markets by the certainty equivalence principle: explaining the sign of the market risk premium. *Journal of Banking & Finance*, 32(10):2006–2021, 2008.

Hendrik Bessembinder and Michael L Lemmon. Equilibrium pricing and optimal hedging in electricity forward markets. *the Journal of Finance*, 57(3):1347–1382, 2002.

Jeff A Bilmes et al. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *International Computer Science Institute*, 4(510):126, 1998.

Álvaro Cartea and Pablo Villaplana. Spot price modeling and the valuation of electricity forward contracts: The role of demand and capacity. *Journal of Banking & Finance*, 32(12):2502–2519, 2008.

Cyriel De Jong. The nature of power spikes: A regime-switch approach. *Studies in Nonlinear Dynamics & Econometrics*, 10(3), 2006.

Shijie Deng. *Stochastic models of energy commodity prices and their applications: Mean-reversion with jumps and spikes.* University of California Energy Institute, 2000.

Eugene F Fama. Efficient capital markets: A review of theory and empirical work*. *The journal of Finance*, 25(2):383–417, 1970.

Eugene F Fama. Forward and spot exchange rates. *Journal of Monetary Economics*, 14(3):319–338, 1984.

Eugene F Fama and Kenneth R French. Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. *Journal of Business*, pages 55–73, 1987.

Niels Haldrup and Morten Ørregaard Nielsen. A regime switching long memory model for electricity prices. *Journal of Econometrics*, 135(1):349–376, 2006.

Lars Peter Hansen and Robert J Hodrick. Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *The Journal of Political Economy*, pages 829–853, 1980.

David Hirshleifer. Hedging pressure and futures price movements in a general equilibrium model. *Econometrica: Journal of the Econometric Society*, pages 411–428, 1990.

Ronald Huisman and Ronald Mahieu. Regime jumps in electricity prices. *Energy economics*, 25(5):425–434, 2003.

Joanna Janczura and Rafal Weron. An empirical comparison of alternate regime-switching models for electricity spot prices. *Energy Economics*, 32(5):1059–1073, 2010.

Michael C Jensen. Some anomalous evidence regarding market efficiency. *Journal of financial economics*, 6(2):95–101, 1978.

Akshaya Jha and Frank A Wolak. Testing for market efficiency with transactions costs: An application to convergence bidding in wholesale electricity markets. Technical report, Working Paper, 2013.

John Maynard Keynes. Some aspects of commodity markets. *Manchester Guardian Commercial: European Reconstruction Series*, 13:784–786, 1923.

Francis A Longstaff and Ashley W Wang. Electricity forward prices: a high-frequency empirical analysis. *The Journal of Finance*, 59(4):1877–1900, 2004.

Julio J Lucia and Eduardo S Schwartz. Electricity prices and power derivatives: Evidence from the nordic power exchange. *Review of Derivatives Research*, 5 (1):5–50, 2002.

Franco Modigliani and Leah Modigliani. Risk-adjusted performance. *The Journal of Portfolio Management*, 23(2):45–54, 1997.

Timothy D Mount, Yumei Ning, and Xiaobin Cai. Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters. *Energy Economics*, 28(1):62–80, 2006.

Arkadi Nemirovski and Alexander Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

Jacques Rolfo. Optimal hedging under price and quantity uncertainty: The case of a cocoa producer. *The Journal of Political Economy*, pages 100–116, 1980.

Paul A Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial management review*, 6(2):41–49, 1965.

Hany A Shawky, Achla Marathe, and Christopher L Barrett. A first look at the empirical relation between spot and futures electricity prices in the united states. *Journal of Futures markets*, 23(10):931–955, 2003.

Rafał Weron. Heavy-tails and regime-switching in electricity prices. *Mathematical Methods of Operations Research*, 69(3):457–473, 2009.

Appendix

With $\phi(u) = (u+1)_+^2$, we can show $\phi(\frac{u}{\alpha}) \geq I(u \geq 0)$ for any $\alpha > 0$. By substituting $u = -R_t^T y_t - \gamma$, we have $\phi(\frac{1}{\alpha}(-R_t^T y_t - \gamma)) \geq I(-R_t^T y_t - \gamma \geq 0)$. Taking expectation on both sides yields,

$$E[\phi(\frac{1}{\alpha}(-R_t^T y_t - \gamma))] \geq P(-R_t^T y_t - \gamma \geq 0). \tag{36}$$

By multiplying α on both sides and replacing the positive part function, we can further show a conservative approximation of $P(-R_t^T y_t - \gamma \geq 0) \leq 1 - \eta$ in the following form,

$$\begin{aligned}
\alpha P(-R_t^T y_t - \gamma \geq 0) &\leq \alpha E[\phi(\frac{1}{\alpha}(-R_t^T y_t - \gamma))] && (37)\\
&= \alpha E[(\frac{1}{\alpha}(-R_t^T y_t - \gamma) + 1)_+^2] && (38)\\
&\leq \alpha E[(\frac{1}{\alpha}(-R_t^T y_t - \gamma) + 1)^2] && (39)\\
&\leq \alpha(1 - \eta). && (40)
\end{aligned}$$

Rearranging $\alpha E[(\frac{1}{\alpha}(-R_t^T y_t - \gamma) + 1)^2] \leq \alpha(1-\eta)$ yields,

$$\alpha E[(\frac{1}{\alpha}(-R_t^T y_t - \gamma) + 1)^2] - \alpha(1-\eta) \tag{41}$$

$$= \frac{1}{\alpha} E[(R_t^T y_t + \gamma)^2] - 2E[(R_t^T y_t + \gamma)] + \eta\alpha \leq 0. \tag{42}$$

Noticing that (42) is a quadratic function, we can minimize the function by setting $\alpha = (\frac{1}{\eta} E[(R_t^T y_t + \gamma)^2])^{\frac{1}{2}}$.[15] By substituting $\alpha = (\frac{1}{\eta} E[(R_t^T y_t + \gamma)^2])^{\frac{1}{2}}$ into (42), we derived the Chebyshev bound,

$$-E[(R_t^T y_t + \gamma)] + (\eta E[(R_t^T y_t + \gamma)^2])^{\frac{1}{2}} \leq 0. \tag{43}$$

[15] α is nonnegative, since $\alpha = (\frac{1}{\eta} E[(R_t^T y_t + \gamma)^2])^{\frac{1}{2}} \geq 0$.