Energy Policy 39 (2011) 2711–2721

Contents lists available at ScienceDirect

Energy Policy

journal homepage: www.elsevier.com/locate/enpol

A novel approach for modeling deregulated electricity markets

Ofir D. Rubin [a,*], Bruce A. Babcock [b,c,1]

[a] Department of Agricultural Economics and Management, The Hebrew University of Jerusalem, P.O. Box 12, Rehovot 76100, Israel
[b] Department of Economics, Iowa State University, 578F Heady Hall, Ames, IA 50011-1070, USA
[c] Center for Agricultural and Rural Development (CARD), Iowa State University, Ames, IA 50011-1070, USA

ARTICLE INFO

Article history:
Received 21 October 2010
Accepted 4 February 2011
Available online 1 March 2011

Keywords:
Deregulated electricity markets
Electricity forward premium
Oligopoly pricing

ABSTRACT

The theoretical framework developed in this study allows development of a model of deregulated electricity markets that explains two familiar empirical findings; the existence of forward premiums and price-cost markups in the spot market. This is a significant contribution because electricity forward premiums have been previously explained exclusively by the assumptions of perfect competition and risk-averse behavior while spot markups are generally the outcome of a body of literature assuming oligopolistic competition. Our theoretical framework indicates that a certain premium for forward contracting is required for efficient allocation of generation capacity. However, due to the uniqueness of electricity and the design of deregulated electricity markets this premium might be substantially higher than its optimal level.

1. Introduction

Although the electricity sectors in many countries have been deregulated over the last 20 years, there is still no satisfactory explanation for why deregulated electricity markets are characterized by forward prices that exceed spot prices (Benth et al., 2008; Bessembinder and Lemmon, 2002; Cartea and Villaplana, 2008; Douglas and Popova, 2008; Longstaff and Wang, 2004; Pirrong and Jermakyan, 2008) and why spot prices exceed marginal costs (Borenstein and Bushnell, 1999; Borenstein et al., 2002; Green, 1999; Mansur, 2008; Puller, 2007; Wolfram, 1999). One body of literature explains forward premiums by assuming risk aversion when firms bid for future supplies of electricity. Another body of literature explains marked up prices by the market power of suppliers. Our contribution is a new modeling approach that simultaneously generates forward premiums and price markups.

Bessembinder and Lemmon (2002) (henceforth BL, 2002) developed a general equilibrium model where both quantity and price of electricity forward contracts are determined endogenously in an electricity market governed by a two-settlement process (i.e. one market for forwards and one spot market for balancing power in real time). BL (2002) assume that the forward price is based mainly on the expected spot price. They show that the assumption of risk aversion coupled with positive skewness of the expected spot price will generate forward prices that exceed expected spot prices.

This explanation for forward premiums relies on risk-averse firms. But firms operating in the electric industry are typically large and well-capitalized. In addition, these firms are continuously in the market for electricity. It is simply not realistic that these firms would have value functions that place less weight on upside income than downside income generated from an hourly or daily market for electricity. It is much more plausible that these firms would operate with an objective of maximizing long-term income. Furthermore, this body of literature ignores the fact that buyers and sellers of electricity are large in the sense that the electric industry is often concentrated therefore raising the possibility that firms will exploit their market power.

Prices that exceed marginal costs have been explained using models of oligopolistic competition. The most suitable approach is a Cournot game in which a power supplier acts as a Cournot competitor choosing its own quantity taking the quantities of its rivals as given. Nash equilibrium in this game is reached where all suppliers simultaneously choose profit maximizing quantities. Standard Cournot models usually overestimate market power of power supplies because they do not consider entry and exit of firms. Since supernormal profits encourage the entry of new firms, incumbents may not be able to exercise market power up to the Cournot equilibrium. Second, being a single-shot game the standard Cournot framework is not suitable to consider the two-settlement process implemented by most deregulated electricity markets. This is limiting because forward contracts are the main pricing tool in these markets. Moreover, the literature suggests that the presence of contracts in a Cournot setting drives suppliers to act more competitively in the spot market and move away from the Cournot equilibrium (Allaz and Vila, 1993).

Supply function equilibrium (SFE) is another oligopolistic modeling approach that is frequently used to study electricity

* Corresponding author. Tel.: +972 8 9489227; fax: +972 8 9466267.
E-mail addresses: rubino@agri.huji.ac.il, ofirubin@gmail.com (O.D. Rubin), babcock@iastate.edu (B.A. Babcock).
[1] Tel.: +1 515 2946785; fax: +1 515 2946336.

0301-4215/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.enpol.2011.02.040

Nomenclature

Abbreviations

CFD	contract for differences
FOC	first-order condition
GF	generation firm
ISO	independent system operator
LSE	load serving entity
MAPE	mean absolute percentage error
SFE	supply function equilibrium
SOC	second-order condition

Symbols

X	realized load
$\hat{X}$	short-term load forecast
$\sigma_{\hat{X}}$	standard deviation of short-term load forecast
$f_X(X\|\hat{X})$	the conditional probability distribution function of load
q^i	generation output of GF i, $i=1,...,M$
$\bar{\bar{q}}^i$	maximum generation output of GF i
q_F^i	GF i's supply of short-term forwards
$\bar{q}_F^i$	GF i's supply of short-term forwards, taken by all other firms as given
x_F^j	LSE j's demands for short-term forwards, $j=1,\ldots,N$
$\bar{x}_F^j$	LSE j's demands for forwards, taken by all other firms as given
C	total variable costs of generating power
θ	operational costs related to ramping constraints
α_t	cost parameter of generating power
α_s^-	cost parameter of turning off generators
α_s^+	cost parameter of turning on generators
δ	α_s^-/α_s^+
P_s	spot price
P_F	forward price
P_R	retail price
π_{LSE}^j	LSE j's overall profits
π_{GF}^i	GF i's overall profits
*	denotes variables at the optimal level
**	denotes variables at the market equilibrium level

markets. This is a theoretical framework developed by Klemperer and Meyer (1989) and employed for modeling electricity market in Green and Newbery (1992) and others. In this single settlement model suppliers bid supply curves rather than price-quantity pairs. Since most deregulated markets are governed by uniform price auctions, the SFE setting describes actual suppliers' behavior more closely than the Cournot model. Unfortunately, the solution of a SFE model is characterized by multiple equilibria. The range of possible equilibrium may be narrowed down by capacity constraints, firm entry and market for contracts (Green, 1999; Green and Newbery, 1992; Newbery, 1998).

A shortcoming of the oligopolistic modeling approach is the lack of a realistic representation of consumers' behavior. Kian et al. (2005) proposed a model of double-sided auctions for spot power. They develop bidding strategies for suppliers and buyers in a dynamic system. On the downside, this study does not consider the market ramifications of trade governed by a two-settlement process.

To summarize, oligopoly modeling approaches are capable of explaining the existence of price-cost markups. This markup is caused by producers having an incentive to withhold generation capacity in a single-shot game. But this is a major drawback because it is only a partial representation of the demand side in electricity markets. Therefore, although oligopoly models explain the empirical evidence of market power it does not relate to the financial aspect of spot price volatility and the existence of forward premiums.

The fact that both spot markups and forward premiums are well documented in the empirical literature implies that the state of knowledge regarding modeling electricity markets is incomplete. Our novel approach relies on two additional features describing the reality of deregulated electricity markets. First, the supply curve of electricity is dynamic due to ramping constraints; frequent start-ups and shut-downs of generators increase the costs of generating power. Therefore, it is an important aspect of electricity pricing. This concept alone is not new. Mansur (2008) claims that by ignoring production constraints such as ramping costs, several studies overestimated the exercise of market power in electricity markets. Second, adequate regulation of the power system relies on scheduling power for future delivery from designated generators. This makes the market for short-term forwards illiquid since traders cannot make commitments to supply future power and secure the required amount only in the spot market. We show that by accounting for these fundamental elements of electricity markets jointly our theoretical framework is capable of explaining the coexistence of spot markups and forward premiums. First, unobserved startup costs give rise to the gap between the competitive spot price and the direct marginal costs of generating power. Our computational experiment illustrates how overlooking ramping constraints may generate this gap and lead to a biased measure of market power. Second, and perhaps the main contribution of this article is that forward premiums are modeled as the outcome of oligopolistic competition and do not require the imposition of risk-aversion. We show that although a certain premium is required for efficient allocation of generation capacity, in practice, this premium might be considerably higher than its optimal level. This deviation depends primarily on how concentrated the electric industry is and how flexible it is in adjusting to unexpected changes in load (i.e. real-time electricity demand).

The remainder of this paper is organized as follows. In Section 2 we develop the theoretical framework; we compute market equilibrium and compare it with the efficient allocation of generation capacity in Section 3. In Section 4, we put the theoretical framework developed in this study to work by providing detailed computational experiments examining electricity market outcomes under various settings. Section 5 concludes our findings and highlights the policy implications of this paper.

2. The model

The theoretical framework we present is based on an oligopolistic competition in electricity markets administrated by a two-settlement process. The first trading period takes place in the market for short-term forward contracts (e.g. day-ahead or hour-ahead). The second trading period is in the spot market. When the spot market is settled, real-time supply equals demand and power is generated, transmitted and consumed. To be clear, we model a two-settlement process with respect to the residual demand (residual of the quantity that is forward contracted) which is traded in wholesale markets. We do not model longer-term contracts which are traded over the counter.

The model consists of an independent system operator (ISO) and firms on each side of the market; load serving entities (LSEs) and generation firms (GFs), which are buyers and sellers, respectively.

The ISO manages the power system and administrates wholesale electricity markets. In addition, the ISO makes predictions regarding real-time electricity demand at the beginning of the trading period. These predictions are accessible to all market participants. LSE firms are the natural buyers of power in the model. They are committed to deliver any realized load to their end-users for a fixed short run retail price. Since LSEs face inelastic demand in real-time, they have an incentive to trade power via forward contracts and by that reduce their exposure to upward spikes in the spot price. GFs generate and supply power in real-time; they are strictly sellers in the market for forward contracts but may buy back contracts in the event of excess supply in the spot market. GFs have a time-sensitive convex cost function which characterizes the various types of power generators and a range of fuel inputs in use. The model is based on a double-sided auction where both LSEs and GFs engage in a Cournot competition. We assume that all players have perfect information about the distribution of the spot price and they are risk neutral.

The model does not attempt to examine electricity commodities other than spot power. The outcomes on subsequent markets (e.g. ancillary services, reserve capacity and others) are not considered.

2.1. Power generation

The portfolio of power generators owned by a GF may be ordered in terms of their marginal cost to obtain a cost curve, which is increasing and convex (see for example BL, 2002). Another fundamental feature of this curve is that it is non-stationary. In fact it is time-sensitive because turning on generators is constrained by ramp-up time and the associated start up and shut down costs.

This convex curve accounts for the generation capacity which has been turned-on in advance to be able to produce power in a particular delivery period in the future. The set of generators that are turned-on in advance characterizes the GF's supply curve in real-time. Clearly, this formation of the cost function has significant financial implications. For example, when producers turn on some of their generators because they are able to sell their output in advance power is less costly to generate because the capacity which has relatively longer ramp-up time is characterized by lower heat rate. More specifically, the day-ahead supply curve (corresponding to a day-ahead forward market) includes more generators that are able to respond to changes in load at lower cost than the applicable generators in real-time. Therefore load forecasting error a day before the delivery period has different financial consequences than the same error made just a few hours before the delivery.

We assume that GFs have an identical set of generating technologies and that capacity of peaking power plants is large enough to accommodate any possible load realization. Essentially, this is equivalent to the supposition that the system operator manages capacity and ancillary services adequately. Start by denoting q_F as the amount of electricity that is pre-scheduled by a particular GF for delivery in real-time. Then, the costs of power generation are governed by two possible states of the world. If production (denoted by q) is lower than q_F then there is a need to turn off some generators before the delivery period. In contrast, if realized production (load) is higher than q_F, generators have to be turned on. In this state of the world the cost of generating power is higher for two reasons. First, the cost function of any applicable set of generation capacity is convex; therefore higher production level necessarily means higher marginal cost of production. Second, startup costs during ramp-up time drive marginal costs up. The startup costs account for the time which generators are turned on and operating but their output level is still low.

Quadratic cost functions are being used commonly for modeling the cost of generating power. For example, the supply function equilibrium (SFE) model introduced by Klemperer and Meyer (1989) has been applied in numerous studies of electricity markets. Other examples can be found in Bjorgan et al. (1999), Sun and Tesfatsion (2007), Tseng and Barz (2002) and Twomeya and Neuhoff (2010). We extend the typical static quadratic form to model GF's total variable cost in the following way:

$$C(q,q_F) = 0.5\alpha_t q^2 + \theta(q,q_F), \tag{1}$$

where

$$\theta(q,q_F) = \begin{cases} 0.5\alpha_s^-(q_F - q)^2 & \text{if } q_F \geq q > 0 \\ 0.5\alpha_s^+(q - q_F)^2 & \text{if } \overline{\overline{q}} > q \geq q_F \end{cases},$$

α_t, $\alpha_s^-,\alpha_s^+ \geq 0$ are parameters and $\overline{\overline{q}}$ is the upper bound for the output of generation capacity owned by each GF.

The parameter α_t describes the cost related to the volume of electricity production. $\theta(q,q_F)$ is a function that accounts for the cost components which are related to the state of the generators in real-time. If GF's production level happens to be higher than the pre-scheduled capacity, an incremental cost α_s^+ is involved in turning on additional generators toward the delivery period. Likewise, α_s^- is the incremental cost involved in shutting down generators. The t subscript is used to differentiate between the direct cost of generating power and the ramping costs taking place toward the delivery period (spot), denoted by subscript s.

Incorporating a two-state of the world cost function has a significant advantage because it allows for the modeling of spikes in production costs which are not entirely explained by high realizations of load. For example, sudden increase in generation costs can also be caused by scheduling insufficient generating capacity in advance. This may be the outcome of profit-maximizing behavior or errors in load forecasting. In any case, in real-time the economy cannot avoid startup costs which may drive total variable cost up rapidly.

2.2. Spot power bids and firm entry

Markets for spot power (day-ahead, real time and others) are generally administrated by uniform price auctions. GF and LSE firms submit their bids to the market administrator (usually the ISO itself) and generators are dispatched by their lowest bids until system demand is met. The bid of the marginal unit clears the market and determines the market price. This method is commonly adopted on the ground of the efficiencies associated with the competitive behavior of market participants. If sellers and buyers bid their marginal costs and maximum willingness to pay, respectively; economic theory tells us that the allocation of resources will be efficient. However, the empirical evidence cited above suggests that the assumption of competitive behavior may not be suitable for electricity markets, since typically each LSE firm represents large numbers of consumers and the number of GFs in a region is small.

If GFs exercise market power, the degree to which they are able to manipulate market prices depends on the timing of market operation. Although the same homogeneous commodity is traded in both forwards and spot markets, the cost structure is very different. Due to ramp-up time and fuel costs, peaking plants are turned on mainly for balancing power in real-time, whereas base and intermediate load plants are the core supply of power in forward contracting. Moreover, peaking plants are relatively small and do not require high construction costs. Hence supernormal profits in the spot market may encourage the entry of new peaking generators (e.g. Newbery, 1998). Base and intermediate load plants on the other hand are more expensive and require more time to build. Therefore construction of these plants may be considered only in the long run. Consequently, unlike peaking plants, the strategy for trading the energy output of these plants is less threatened by entry. This environment gives rise to our claim that the degree of competitiveness of electricity markets governed by uniform price auctions increases the closer the trade occurs to the time of delivery. Based on the motivation of preventing entry, GFs behavior in spot and forward markets may diverge greatly. First, we focus on the spot market. The assumption of zero construction costs of peaking plants motivates perfectly competitive behavior in real time.[2] On the other hand, market power may be exercised in markets for forward contracts. We will analyze the spot market outcome taking forward positions as given and then work back to study separately LSEs' and GFs' strategies in the market for forward contracts to characterize a symmetric Cournot–Nash equilibrium in both markets.

2.3. The spot market

If the threat of entry motivates GFs' to bid competitively in the spot market then we know that the realized spot price reflects true real-time marginal costs of electricity production. That is

$$P_s = C'(q^1, q_F^1) = \cdots = C'(q^M, q_F^M), \tag{2}$$

where P_s is the spot price and M is the number of GFs in the electric industry.

Notice that the competitive price accounts only for real-time production costs. While turning on generators in real-time is considered to be marginal costs, shutting down generators are seen as sunk costs. That is because firms cannot price their output above the competitive level in the spot market to compensate for poor decisions which they had made in an earlier stage (i.e. turning on excess generation capacity in the market for forward contracts).

Explicitly, the competitive price can be expressed as

$$P_s = \begin{cases} \alpha_t q & \text{if } q_F \geq q > 0 \\ \alpha_t q + \alpha_s^+ (q - q_F) & \text{if } \overline{\overline{q}} > q \geq q_F \end{cases}. \tag{3}$$

In words, in the event that generators need to be started, marginal costs are higher which makes the spot power supply curve steeper.

Denote load by X. Given the number of forward contracts offered by each firm and the fact that power must be balanced in all times (i.e. $\sum_1^M q^m = X$) one can solve for the spot price that clears the market and the quantities to be produced by each GF. Note that condition (2) provides M equations and balancing real-time power is an additional equation. Therefore, we have altogether a system of M+1 equations which can be solved for M firms production levels and one spot market price. For example, for M=2 and assume without loss of generality that $q_F^2 > q_F^1$ the spot price is

$$P_s(X, q_F^1, q_F^2) = \begin{cases} \frac{\alpha_t}{2} X & \text{if } 2q_F^1 \geq X \geq 0 \\ \alpha_t \frac{(\alpha_t + \alpha_s^+)X - \alpha_s^+ q_F^1}{2\alpha_t + \alpha_s^+} & \text{if } \frac{\alpha_s^+ (q_F^1 + q_F^2) + 2\alpha_t q_F^2}{\alpha_t + \alpha_s^+} \geq X > 2q_F^1 \\ \frac{\alpha_t}{2} X + \frac{\alpha_s^+}{2}(X - q_F^1 - q_F^2) & \text{if } X > \frac{\alpha_s^+ (q_F^1 + q_F^2) + 2\alpha_t q_F^2}{\alpha_t + \alpha_s^+} \end{cases}, \tag{4}$$

and firms' production levels are

$$\{q^1, q^2\} = \begin{cases} \left\{\frac{X}{2}, \frac{X}{2}\right\} & \text{if } 2q_F^1 \geq X \geq 0 \\ \left\{\frac{\alpha_s^+ q_F^1 + \alpha_t X}{2\alpha_t + \alpha_s^+}, \frac{(\alpha_t + \alpha_s^+)X - \alpha_s^+ q_F^1}{2\alpha_t + \alpha_s^+}\right\} & \text{if } \frac{\alpha_s^+ (q_F^1 + q_F^2) + 2\alpha_t q_F^2}{\alpha_t + \alpha_s^+} \geq X > 2q_F^1 \\ \left\{\frac{X}{2} + \frac{\alpha_s^+ (q_F^1 - q_F^2)}{2(\alpha_t + \alpha_s^+)}, \frac{X}{2} + \frac{\alpha_s^+ (q_F^2 - q_F^1)}{2(\alpha_t + \alpha_s^+)}\right\} & \text{if } X > \frac{\alpha_s^+ (q_F^1 + q_F^2) + 2\alpha_t q_F^2}{\alpha_t + \alpha_s^+} \end{cases}. \tag{5}$$

The first parts in Eqs. (4) and (5) describe the case that capacity traded via forwards is sufficient to meet realized load. The second is where only firm 1 adds capacity toward the delivery period and the third part is where both firms start up generators in real-time.

Next, and for the rest of this study, we focus on the existence and the characteristics of a symmetric forward position case in which we have M identical GFs where the Cournot–Nash equilibrium is $q_F^1 = \cdots = q_F^M$. Focusing on the symmetric case simplifies the analysis since we need to examine only two states of the world. One is where no generators are being turned on in real-time and the other is where all firms turn on generators in real-time. For these two states it can be verified that the spot price is

$$P_s\left(X, \vec{q}_F^{\,m}\right) = \begin{cases} \frac{\alpha_t}{M} X & \text{if } \sum_1^M q_F^m \geq X \geq 0 \\ \frac{\alpha_t}{M} X + \frac{\alpha_s^+}{M}\left(X - \sum_1^M q_F^m\right) & \text{if } X > \sum_1^M q_F^m \end{cases}, \tag{6}$$

where $\vec{q}_F^{\,m} = \{q_F^1, \ldots, q_F^M\}$, and production level of firm i is

$$q^i\left(X, \vec{q}_F^{\,m}\right) = \begin{cases} \frac{X}{M} & \text{if } \sum_1^M q_F^m \geq X \geq 0 \\ \frac{X}{M} + \frac{\alpha_s^+}{M(\alpha_t + \alpha_s^+)}\left((M-1)q_F^i - \sum_{1, m \neq i}^M q_F^m\right) & \text{if } X > \sum_1^M q_F^m \end{cases}. \tag{7}$$

Assuming symmetric forward positions, we can examine the changes in production levels and price caused by deviation of one producer (i.e. offering the marginal unit for forwards contracting). These changes become useful when we analyze the GF's maximization problem. In the event that power traded via forwards is larger than realized load, deviation in the forward position has no significance on the spot market. That is because no additional generators are needed in real-time and producers cannot recover shut-down costs. On the other hand, in the event that all firms generate additional power in real-time a deviation has an impact on generation cost and thereby market outcome. Suppose firm i chooses to deviate, the change in the level of output with respect to own forward position is[3]

$$\frac{\partial q^i(X, \vec{q}_F^{\,m})}{\partial q_F^i} = \frac{M-1}{M} \times \frac{\alpha_s^+}{\alpha_t + \alpha_s^+}, \tag{8}$$

and with respect to m's position it is

$$\frac{\partial q^i(X, \vec{q}_F^{\,m})}{\partial q_F^m} = -\frac{1}{M} \times \frac{\alpha_s^+}{\alpha_t + \alpha_s^+}, \quad m = 1, \ldots, M, \quad m \neq i. \tag{9}$$

Notice that since load must be met at all times we get

$$\sum_{m=1}^M \frac{dq^m(X, \vec{q}_F^{\,m})}{dq_F^i} = 0.$$

Finally, a Cournot firm that chooses to deviate from the symmetric position expects (in the case of turning on generators

[2] The zero construction cost assumption is made for convenient presentation of the concept. In the case of positive construction costs of small generators, GFs may price their output up to their average costs to prevent entry.

[3] Employing a Cournot approach implies that

$$\frac{\partial q_F^m(\cdot)}{\partial q_F^i} = 0, \quad m = 1, \ldots, M, \quad m \neq i.$$

in real-time) a spot price change of

$$\frac{\partial P_s(X, \overrightarrow{q}_F^{m})}{\partial q_F^i} = -\frac{\alpha_s^+}{M}. \tag{10}$$

The Cournot players (both GFs and LSEs) observe both load forecasts and forwards bids made by other firms.[4] Therefore we may treat the distribution of load and thereby the conditional distribution of spot price and expected production levels as common knowledge in our model. In the following, we analyze the market for forward contracts.

2.4. The demand for electricity forwards

The objective of this section is to evaluate the LSE's willingness-to-pay for electricity forwards. Assume N identical LSE Cournot firms where each is committed to deliver $1/N$ portion of the realized load. In the short run (a year or more in the context of building a new large generator) LSEs are compensated by a fixed electricity retail price P_R. In every period the ISO announces a load forecast which is superior to any private prediction. Since this forecast is adopted by all market participants, information about forecast is symmetric. When overall load forecast is $\hat{X}$, each LSE's expected real-time demand is $\hat{X}/N$. Armed with this information and taking its rivals' bids as given, LSE j maximizes profits by choosing a forward position x_F^j. That is

$$\begin{aligned} \mathbb{N} &\equiv \max_{x_F^j} E\left(\pi_{LSE}^j | \hat{X}, \sum_{n=1,n\neq j}^{N} \overline{x}_F^n\right) \\ &= [P_R - P_F]x_F^j + \int_0^{\infty} \left[P_R - P_s\left(X, \sum_{n=1,n\neq j}^{N} \overline{x}_F^n + x_F^j\right)\right] \\ &\quad \times \left(\frac{X}{N} - x_F^j\right) f_x(X|\hat{X})dX \end{aligned} \tag{11}$$

where $\overline{x}_F^n$ is the quantity bid of player n, P_F is the market price of a forward contract and $f_x(X|\hat{X})$ is the conditional probability distribution function of load at the time of trading forwards.

LSE's expected profit has two payoff components; the first component in Eq. (11) stands for the payoff in trading forward contracts while the second component is the expected payoff associated with balancing power in the spot market. Substitute for the expected spot price in the symmetric case (6), profit may be written as

$$\begin{aligned} \mathbb{N} &= [P_R - P_F]x_F^j + \int_0^{\sum_{1,n\neq j}^{N} \overline{x}_F^n + x_F^j} \left[P_R - \frac{\alpha_t}{M}X\right] \times \left(\frac{X}{N} - x_F^j\right) f_X(X|\hat{X})dX \\ &\quad + \int_{\sum_{1,n\neq j}^{N} \overline{x}_F^n + x_F^j}^{\infty} \left[P_R - \frac{\alpha_t}{M}X - \frac{\alpha_s^+}{M}\left(X - \sum_{n=1,n\neq j}^{N} \overline{x}_F^n - x_F^j\right)\right] \\ &\quad \times \left(\frac{X}{N} - x_F^j\right) f_X(X|\hat{X})dX. \end{aligned} \tag{12}$$

The two integrals in Eq. (12) account for the two cases of need to balance power in real-time. The first integral is the expected spot payoff when the LSE has over-purchased power via forward contracts while the second stands for under-purchase of power. In the former, Contract for Differences (CFD) is put into effect. A typical CFD states that any deviation between forward power and spot power may be traded for the realized spot price. Essentially, CFD is a financial settlement that helps LSEs to hedge against volumetric risk on one hand and on the other helps GFs avoid the cost and transmission problems associated with spot power surplus.[5]

Taking the derivative of N with respect to the decision variable and employing the *Leibniz integral rule*, we get

$$\begin{aligned} \frac{\partial \mathbb{N}}{\partial x_F^j} &= [P_R - P_F] - \int_0^{\sum_{1,n\neq j}^{N} \overline{x}_F^n + x_F^j} \left[P_R - \frac{\alpha_t}{M}X\right] f_X(X|\hat{X})dX \\ &\quad - \int_{\sum_{1,n\neq j}^{N} \overline{x}_F^n + x_F^j}^{\infty} \left[P_R - \frac{\alpha_t}{M}X - \frac{\alpha_s^+}{M}\left(X - \sum_{n=1,n\neq j}^{N} \overline{x}_F^n + x_F^j\right)\right. \\ &\quad \left. - \frac{\alpha_s^+}{M}\left(\frac{X}{N} - x_F^j\right)\right] f_X(X|\hat{X})dX. \end{aligned} \tag{13}$$

Therefore, the first-order condition (FOC) for interior profit maximization is

$$\begin{aligned} P_F &= \int_0^{\sum_{1,n\neq j}^{N} \overline{x}_F^n + x_F^j} \frac{\alpha_t}{M} X f_X(X|\hat{X})dX \\ &\quad + \int_{\sum_{1,n\neq j}^{N} \overline{x}_F^n + x_F^j}^{\infty} \left[\frac{\alpha_t}{M}X + \frac{\alpha_s^+}{M}\left(X - \sum_{n=1,n\neq j}^{N} \overline{x}_F^n - x_F^j\right)\right. \\ &\quad \left. + \frac{\alpha_s^+}{M}\left(\frac{X}{N} - x_F^j\right)\right] f_X(X|\hat{X})dX \end{aligned} \tag{14}$$

which can be expressed as

$$P_F = E[P_S(\cdot)] + \frac{\alpha_s^+}{M} \int_{\sum_{1,n\neq j}^{N} \overline{x}_F^n + x_F^j}^{\infty} \left(\frac{X}{N} - x_F^j\right) f_X(X|\hat{X})dX. \tag{15}$$

Notice that the second-order condition (SOC) is clearly satisfied here as

$$\frac{\partial^2 \mathbb{N}}{\partial x_F^{j2}} = -\frac{2\alpha_s^+}{M} \int_{\sum_{1,n\neq j}^{N} \overline{x}_F^n + x_F^j}^{\infty} f_X(X|\hat{X})dX < 0. \tag{16}$$

Condition (15) describes the firm's inverse demand function for forward contracts. It is interesting to see that an LSE's willingness to pay for a forward contract exceeds the expected spot power price. Assuming risk neutrality generally drives the price of forward contracts to the commodity's expected spot price. However, this result need not hold for the case of electricity. Since electricity has to be consumed at the time of production there is an economic value for pre-scheduling power for production (e.g. forward contracting). While for most commodities the time of production does not impact production cost, it does affect electricity generation cost. For that reason, the LSE in our model maximizes profits by choosing a forward position such that the marginal contract bought for price P_F is higher than the expected spot price. The wedge can be explained simply by the financial consequences of not contracting the marginal unit. In this case the marginal unit is not scheduled in

[4] Conceivably, participating and observing the outcome of 24 day-ahead and 24 real-time electricity markets being cleared on a daily basis may be considered as having complete information about spot price distribution. In addition, electricity markets, unlike any other commodity markets, are unique due to the presence of an ISO. As system operator and in most times the market administer, the ISO reports the conditions of the power system continuously and make forecasts accessible to all. The ISOs' reports also include supply and demand bids; volume of forwards traded and market prices. Doing so, the ISOs act as coordinators and diminish the value of private information. In addition to these transparencies, electric industries are typically more concentrated than other industries thereby making strategic modeling approach most relevant.

[5] The stochastic nature of load gives rise to realizations where actual electricity demand is lower than the amount settled for delivery via forward contracts. In this case the excess amount cannot and will not be produced for physical and economical reasons. Physically, the excess amount causes transmission and reliability problems in the power system. These and the associated costs are not treated in this study. Economically, electricity surplus cannot characterize equilibrium in electricity markets. Therefore, the existence of a settlement which enables the buy-out of surplus is an important financial instrument of electricity market operation.

advance; therefore its price also includes the expected cost of starting up additional generators. The expected additional cost of not scheduling the marginal unit is expressed by the RHS term in Eq. (15). Next, we focus on the supply side of the market.

2.5. The supply of electricity forwards

The optimization problem of each GF is similar to that of a monopoly which faces an inverse demand function (Eq. (15)) and takes its rivals output as given. Formally, GF i chooses forward position q_F^i to maximize its expected profits

$$\mathcal{M} \equiv \max_{q_F^i} E\left(\pi_{GF}^i \,|\hat{X}, \sum_{m=1,m\neq i}^{M} \overline{q}_F^m\right) = P_F\left(\hat{X}, q_F^i + \sum_{m=1,m\neq i}^{M} \overline{q}_F^m\right) \times q_F^i + \int_0^\infty \left[P_s\left(X, q_F^i + \sum_{m=1,m\neq i}^{M} \overline{q}_F^m\right) \times (q^i - q_F^i) - C(q^i, q_F^i)\right] f_X(X|\hat{X})dX, \tag{17}$$

where $\overline{q}_F^m$ is the quantity offered by rival m taken by i as given.

Expressing $\mathcal{M}$ with respect to the two states of the world, we get

$$\mathcal{M} = P_F\left(\hat{X}, q_F^i + \sum_{m=1,m\neq i}^{M} \overline{q}_F^m\right) q_F^i + \int_0^{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i} \left[P_s(X) \times (q^i - q_F^i) - C(q^i, q_F^i)\right] f_X(X|\hat{X})dX + \int_{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i}^{\infty} \left[P_s\left(X, q_F^i + \sum_{m=1,m\neq i}^{M} \overline{q}_F^m\right) \times (q^i - q_F^i) - C(q^i, q_F^i)\right] f_X(X|\hat{X})dX. \tag{18}$$

Recall that when excess capacity has been contracted, it has no effect on the magnitude of real-time production and the spot price. Therefore the first derivative is

$$\frac{\partial \mathcal{M}}{\partial q_F^i} = \frac{\partial P_F(\cdot)}{\partial q_F^i} q_F^i + P_F(\cdot) - \int_0^{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i} P_s(\cdot) + C'\left(q^i, q_F^i\right) f_X(X|\hat{X})dX + \int_{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i}^{\infty} \left[\frac{\partial P_s(\cdot)}{\partial q_F^i} \times (q^i - q_F^i) + P_s(\cdot) \times \left(\frac{\partial q^i}{\partial q_F^i} - 1\right) - C'(q^i, q_F^i)\right] f_X(X|\hat{X})dX. \tag{19}$$

Writing explicitly the spot price and the derivative of the cost function

$$\frac{\partial \mathcal{M}}{\partial q_F^i} = \frac{\partial P_F(\cdot)}{\partial q_F^i} q_F^i + P_F(\cdot) - \int_0^{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i} P_s(\cdot) f_X(X|\hat{X})dX + \alpha_s^-(q_F^i - q^i) + \int_{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i}^{\infty} \left[\frac{\partial P_s(\cdot)}{\partial q_F^i} \times (q^i - q_F^i) + P_s(\cdot) \times \left(\frac{\partial q^i}{\partial q_F^i} - 1\right) - \alpha_t q^i \frac{\partial q^i}{\partial q_F^i} - \alpha_s^+ (q^i - q_F^i) \times \left(\frac{\partial q^i}{\partial q_F^i} - 1\right)\right] f_X(X|\hat{X})dX. \tag{20}$$

Collecting terms

$$\frac{\partial \mathcal{M}}{\partial q_F^i} = \frac{\partial P_F(\cdot)}{\partial q_F^i} q_F^i + P_F(\cdot) - E[P_s(\cdot)] - \alpha_s^- \int_0^{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i} (q_F^i - q^i) f_X(X|\hat{X})dX + \alpha_s^+ \int_{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i}^{\infty} (q^i - q_F^i) + \int_{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i}^{\infty} \frac{\partial P_s(\cdot)}{\partial q_F^i} (q^i - q_F^i) f_X(X|\hat{X})dX. \tag{21}$$

Then, the FOC can be written as the sum of the components affecting GF's expected revenues in the following way:

$P_F(\cdot)=$	GF's willingness to offer a marginal contract is
$E[P_s(\cdot)]$	the expected spot price, plus
$+\alpha_s^- \int_0^{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i} (q_F^i - q^i) f_X(X\|\hat{X})dX$	expected loss caused by shutting down excess generation capacity in real-time, minus
$-\alpha_s^+ \int_{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i}^{\infty} (q^i - q_F^i) f_X(X\|\hat{X})dX$	expected revenue in the event that additional capacity needs to be started in real-time, plus
$+\int_{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i}^{\infty} \frac{\partial P_s(\cdot)}{\partial q_F^i} \times (q^i - q_F^i) f_X(X\|\hat{X})dX$	overall change in revenues in the spot market, and
$+\frac{\partial P_F(\cdot)}{\partial q_F^i} \times q_F^i$	the overall change in revenues in the market for forward contracts

(22)

We show in Appendix A that the GF's maximization problem is strictly concave in q_F^i. This confirms that if a symmetric Cournot–Nash equilibrium exists it is a unique symmetric solution for the GF's problem.

3. Market equilibrium

At this point and for future reference it will be useful to identify the optimal allocation in the model. This can be addressed easily as a central planner problem. The optimization problem is to minimize the cost of supplying electricity by determining how much generation capacity (denoted also as X_F) should be scheduled for delivery (i.e. brought online) ahead of time. Formally

$$\min_{X_F} \mathcal{L} = \int_0^\infty M \times C\left(\frac{X}{M}, \frac{X_F}{M}\right) f_X(X|\hat{X})dX. \tag{23}$$

Then, it is straightforward to show that the FOC for an interior solution is

$$\frac{\int_{X_F^*}^{\infty} (X - X_F^*) f_X(X|\hat{X})dX}{\int_0^{X_F^*} (X_F^* - X) f_X(X|\hat{X})dX} = \frac{\alpha_s^-}{\alpha_s^+} \equiv \delta. \tag{24}$$

This says that the optimal scheduled generation capacity (X_F^*) should be such that the ratio between the expected values of under and over estimation of load equals the ratio of the incremental costs of shutting down generators to the incremental costs of starting them up.

Next, we examine market equilibrium by the intersections of demand and supply curves. Since LSEs are identical they all have the same FOC. Considering the symmetric Cournot–Nash equilibrium where $x_F^1 = x_F^2 = \cdots = x_F^N \equiv x_F$, the inverse aggregate demand is[6]

$$P_F = E[P_S(\cdot)] + \frac{\alpha_s^+}{MN} \int_{X_F}^{\infty} (X - X_F) f_X(X|\hat{X})dX \tag{25}$$

where here we denote $X_F \equiv N x_F$.

[6] Corner solutions for the LSEs problem may arise where at optimum (1) $X_F = 0$; the price of a forward contract is too high to enhance LSEs' expected profits or (2) $X_F \rightarrow \infty$; which is the case of fully hedged positions. That is the forward price is lower than the expected spot price for any amount of forward bought.

Notice that the first derivative of the forward price with respect to GF i's quantity is

$$\frac{\partial P_F(\cdot)}{\partial q_F^i} = -\frac{\alpha_s^+}{M}\left(1+\frac{1}{N}\right)\int_{\sum_{m=1,m\neq i}^{M}\overline{q}_F^m+q_F^i}^{\infty} f_X(X|\hat{X})dX. \tag{26}$$

Then, solving for the symmetric Cournot–Nash equilibrium on the supply side (where $q_F^1 = q_F^2 = \cdots = q_F^M \equiv q_F$) gives the following optimality condition for aggregate supply:

$$\begin{aligned} P_F(\hat{X}, Mq_F) = E[P_s(\cdot)] + \alpha_s^- \int_0^{Mq_F}(q_F - q)f_X(X|\hat{X})dX \\ + \frac{\alpha_s^+}{M}\left(1+\frac{1}{N}\right)q_F \int_{Mq_F}^{\infty} f_X(X|\hat{X})dX \\ -\alpha_s^+\left(1-\frac{1}{M}\right)\int_{Mq_F}^{\infty}(q-q_F)f_X(X|\hat{X})dX. \end{aligned} \tag{27}$$

Finally, if there is a forward price P_F^{**} at which the market clearing condition $q_F^{**}M = X_F^{**}$ holds, then we say that the market for forward contracts has an interior symmetric solution. Equating aggregate demand with aggregate supply (Eqs. (25) and (27)), one can solve numerically for the aggregate number of forward contracts traded in equilibrium. That is

$$X_F^{**} = \frac{\left(\frac{MN+M-N}{MN}\right)\times\int_{X_F^{**}}^{\infty} Xf_X(X|\hat{X})dX + \delta\int_0^{X_F^{**}} Xf_X(X|\hat{X})dX}{\left(\frac{MN+M+1}{MN}\right)\times\int_{X_F^{**}}^{\infty} f_X(X|\hat{X})dX + \delta\int_0^{X_F^{**}} f_X(X|\hat{X})dX}. \tag{28}$$

To evaluate how efficient this market outcome is, we can compare it with the solution obtained by the central planner problem. For this purpose, simply rewrite condition (24) to express the optimal amount of power to be scheduled in advance. That is

$$X_F^{*} = \frac{\int_{X_F^{*}}^{\infty} Xf_X(X|\hat{X})dX + \delta\int_0^{X_F^{*}} Xf_X(X|\hat{X})dX}{\int_{X_F^{*}}^{\infty} f_X(X|\hat{X})dX + \delta\int_0^{X_F^{*}} f_X(X|\hat{X})dX}. \tag{29}$$

Comparing the results it is obvious that there is no reason to believe that the equilibrium number of forward contracts is equal to the optimal level. In particular, when there is same number or fewer GFs than LSEs operating in the market, the amount of electricity scheduled in advance will always be less than optimal (i.e., $X_F^{**} < X_F^{*}$). The intuition is clear: a relatively concentrated supply side will exercise market power by withholding generation capacity.[7] Moreover, one may employ the implicit function theorem to show that X_F^{**} increases in M. As the number of GFs increases, it is less effective for these firms to maximize profits by withholding generation capacity from forward contracting. It is not a surprising result seeing that power producers engage in an oligopoly competition and real-time demand is completely inelastic.

Lastly, the equilibrium forward price and the consequent expected forward premium are computed numerically by substituting (28) back into the demand (or the supply) curve.

4. Computational experiments

We conducted many simulations to evaluate the effectiveness of the model in capturing the economic determinants in electricity markets. We present and discuss in this section model predictions and the sensitivity of the results to the parameters employed in the analysis. Although the magnitude of our results depends on the particular parameters chosen, general conclusions regarding the significance of economic determinants in this market are clearly identifiable. In some parts of the simulations we examine electricity market outcomes while varying M, the number of GFs. To enable a direct comparison, we need to make sure that regardless of the size of M, the industry supply curve remains the same. For this purpose, we parameterize the cost function as proposed in Twomeya and Neuhoff (2010)

$$C(q,q_F) = 0.5M\alpha_t q^2 + M\theta(q,q_F) \tag{30}$$

We consider the following figures for the base-case scenario in the numerical analysis. Assume that there are 5 GF and 5 LSE firms, the cost parameters of generating power are $\alpha_t = 1$, $\alpha_s^- = 0.5$ and $\alpha_s^+ = 2$. This assumption implies that overestimating is less costly than underestimating load (i.e. $\delta < 1$). This is sensible because generators that have relatively shorter response time are characterized by relatively higher heat rate. Consequently, it is expected that the optimal allocation in our numerical experiments is to schedule more electricity than the expected load. In addition, we assume that load is normally distributed (truncated at zero) with mean 100. Various sources indicate a mean absolute percentage error (MAPE) of overall load prediction in the range of 3–5% when forecasting demand a day ahead. The size of the error depends mainly on the season, size of the region and day of the week (e.g. Holttinen, 2005; Soares and Medeiros, 2008; Soares and Souza, 2006; Taylor et al., 2006). Applying a standard deviation of 5 for load forecast, we generate a computed MAPE of 4% for our base-case scenario. All the numerical results consist of 5000 draws.

Given these assumptions we can compute the costs of electricity production when both the pre-scheduled amount of electricity and realized load are exactly 100. This is useful because it represents the most efficient production scheme of one hour power supply in the simulated market. Therefore it can be used as a benchmark. Employing the cost function and symmetry, the marginal cost is \$100 and total variable costs of each GF are \$1000.

The breakdown of direct generation costs and operational costs which are linked to ramping constraints are of main interest in our experiment. We would like to illustrate our assertion that the price markups reported in the literature (cited above and elsewhere) are explained partly by ramping constraints rather than uncompetitive behavior. For this purpose, let us assume that the ISO adopts the forecasted level of load in the sense that the pre-scheduled amount of electricity is exactly at that level. Then if we allow realized load to fluctuate we can identify and examine the associated operational costs. In our illustration real-time variable generation costs are on average 4.81% above the direct marginal cost of generating power. The distribution of this gap (depicted in Fig. 1) is the operational costs caused by load uncertainty. Because generation output is determined by the ISO and not by the market, the simulated real-time operational costs do not stand for market power. This makes obvious that one should infer the reason for empirical price-cost markups with caution.

4.1. Spot price distribution

The model focuses on one particular delivery period at a time. Since there is a great variability in the seasonal, diurnal, hourly and other temporal characteristic of load we start by evaluating the flexibility of the model to accommodate analyses of different delivery periods. We consider periods that are characterized by expected loads in the range of 50–150. The standard deviation is computed as $\sigma_{\hat{X}} = 0.05{*}\hat{X}$. Doing so, we normalize the standard

---

[7] $M \le N$ is a sufficient condition for the suboptimal allocation $X_F^{**} < X_F^{*}$. Necessary condition may be expressed as

$$(\theta-1)\left[(N+1)\times\int_{X_F^{**}}^{\infty} Xf_X(X|\hat{X})dX + \delta(M+1)\int_0^{X_F^{**}} Xf_X(X|\hat{X})dX\right] + \theta\delta(M-N)\int_0^{X_F^{**}} Xf_X(X|\hat{X})dX > 0,$$

where $\theta = \int_{X_F^{**}}^{\infty} f_X(X|\hat{X})dX$.

deviation with respect to the expected load of the period in question.

Fig. 2 displays the densities of spot market prices by expected load. In particular, the middle density is the one that illustrates the base-case scenario. The densities are skewed as implied by the two-state cost function. For periods of higher than average expected load the spot price density is shifted right and its variation is higher. In reality, even hourly prices in a single day are drawn from very different distributions. Load (and thereby prices) at 2 am would correspond to a density in the left side of Fig. 2. Spot price at 5 pm of the same day might be characterized by a density on the right side of the same figure as this is usually the peak load hour of the day.

4.2. Uncertainty and market equilibrium

We let standard deviation of load to vary between 5 and 35 to explore the impact of load uncertainty on market equilibrium. Given spot price expectations, the demand for electricity forwards is illustrated in Fig. 3. As uncertainty regarding real-time demand increases, demand for forwards shifts upward and becomes smoother. This reflects the increase in risk exposure due to the higher probability of making errors in load forecasting. LSEs' willingness to pay for forward contracts increases with increase in the standard deviation of load forecasts because forecasting errors are translated to an asymmetric real-time payoff distribution (i.e. the penalty is higher in the case of scheduling insufficient generation capacity).

Given demand for forwards, each GF chooses how much electricity to offer in forward markets which then determines the aggregate supply curve in the spot market. Equilibrium in the forwards market is described in Fig. 4 for the considered levels of standard deviation of load. The intersections of the dashed vertical line and the demand curve for forwards in the graphs in Fig. 4 describe the equilibrium price and quantity of forward contracts in each case. Once the forward market is cleared, production cost of any possible realization of load and thereby spot price is determined. The equilibrium is characterized by a capacity withholding as one would expect in an oligopoly competition. The kink on the supply curve at the equilibrium quantity of forwards describes the shift into less efficient power production sources when realized load is higher than the pre-scheduled amount. Then, the equilibrium at the spot market may be depicted as the intersection of the illustrated supply curve and an inelastic demand curve created by any given realization of load.

The numerical experiments demonstrate that greater load uncertainty increases the equilibrium number of forward contracts. The GF response to the increased demand is to offer more electricity via forward contracts.

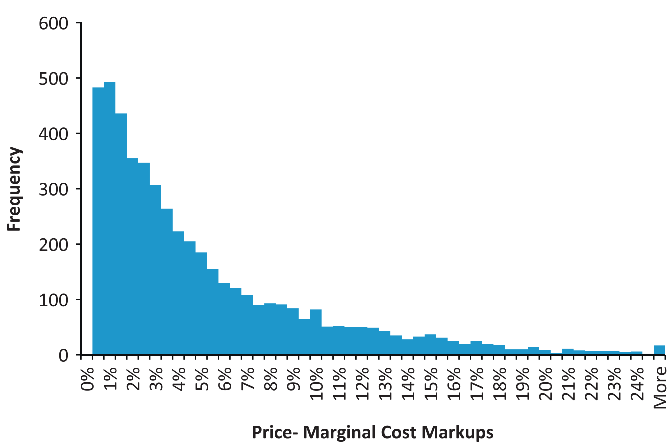

Fig. 1. Histogram of markups.

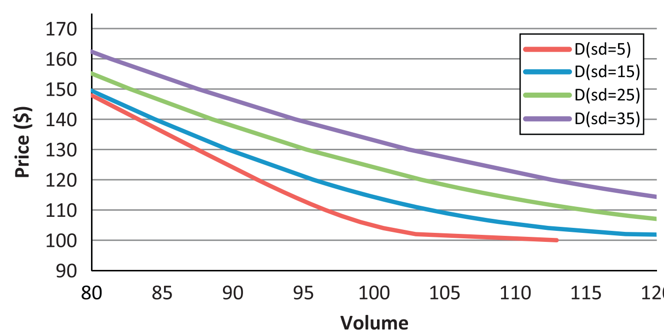

Fig. 3. Aggregate demand for electricity forwards as a function of standard deviation of load.

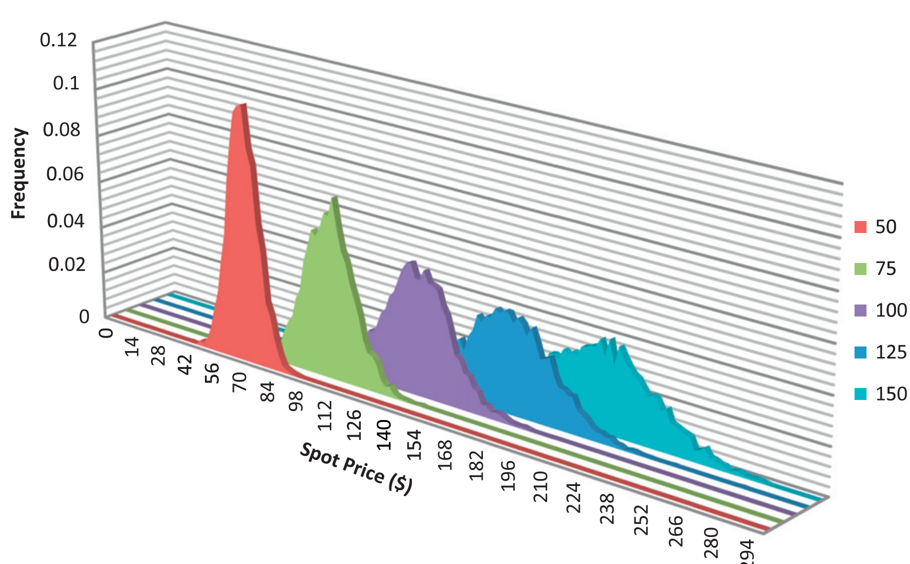

Fig. 2. Spot market price densities by expected electricity demand.

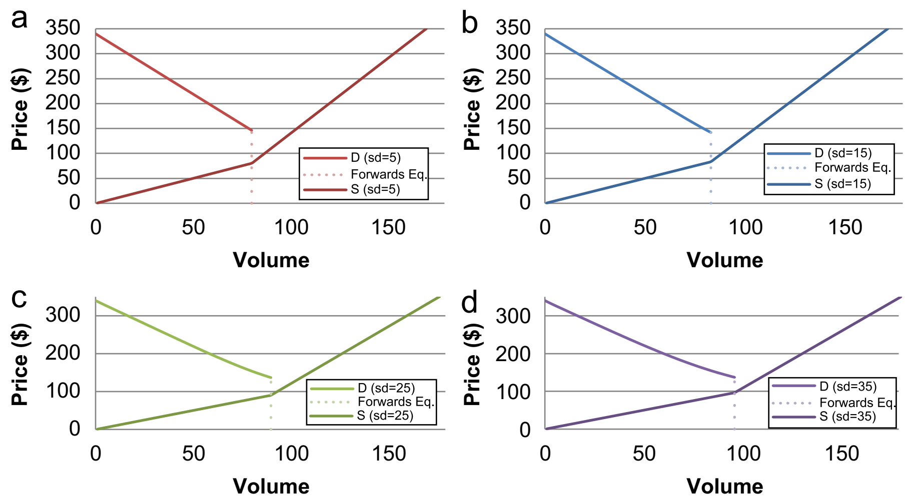

Fig. 4. Equilibrium in the forwards market and the consequence real-time supply curve. Results by standard deviation of load of sizes 5, 15, 25 and 35 are depicted in (a), (b), (c) and (d), respectively.

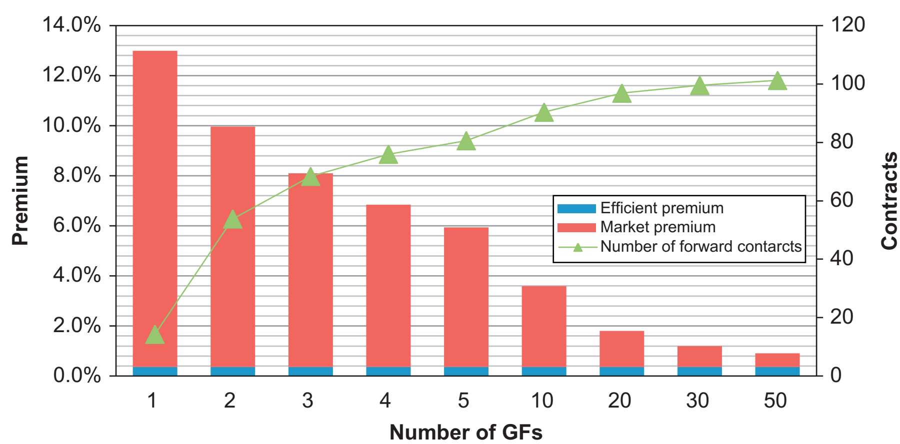

Fig. 5. Equilibrium electricity forwards and premiums by number of GFs in the market (base-case scenario assumptions).

4.3. The number of GFs

In a standard Cournot competition model market power is determined by the number of oligopoly players. With a single producer, the model outcome coincides with the solution for the monopoly's profit maximization problem. With two firms aggregate profits will be less than the situation of two producers acting as monopolies, and lastly, for a sufficient number of producers the profits and market outcomes are similar to those generated under perfect competition. The model developed here generates market power dynamics which is similar to the results described by a standard Cournot model.

In Fig. 5 we examine market allocation of forward contracts and expected premiums varying the number of power producers. As the number of GFs increases, more electricity is settled via forward contracts. This is due to the decrease in GFs ability to exercise market power by withholding capacity. For example, the particular parameters employed here shows that without any regulation a single GF would offer less than 15% of the expected load for forward contracting and enjoy an expected forward premium of \$34.30 (12.6%). The amount increases at a decreasing rate; about one half of the expected load is contracted in a duopoly electric industry and the expected premium is \$18.50 (9.6%) and so on. Interestingly, the illustration suggests that a certain premium is required even if the market allocation is at the optimal level (we denote this as *efficient premium*). This is because there is always a positive probability to be short of spot power. That being said, even with 50 power producers and an equilibrium number of forward contracts which is higher than the expected load; expected premium is still significantly higher than the efficient level (\$0.56 vs. \$0.36).

4.4. The cost of electricity production and deadweight loss

In this section we are interested in exploring the impact of cost parameters employed in our simulations and the subsequent deadweight loss associated with withholding capacity. Starting with α_t; the absence of the direct cost of generating power from the equations that describe the equilibrium number of forward contracts (28) and the central planner solution (29) is a noteworthy result. The reason for that is the fact that α_t is an essential cost which cannot be avoided or altered for any possible realization. On the other hand, δ is important as it characterizes the penalty for balancing generation capacity in real-time. Next, we look at the sensitivity of model predictions with respect to δ. We perform that by varying the incremental cost α_s^+.

Recall that δ expresses the ratio between the incremental costs of shutting-down and starting-up generators. As expected, for a relatively lower penalty for scheduling insufficient amount (i.e. higher δ) the optimal allocation of pre-scheduled capacity is smaller (Fig. 6).

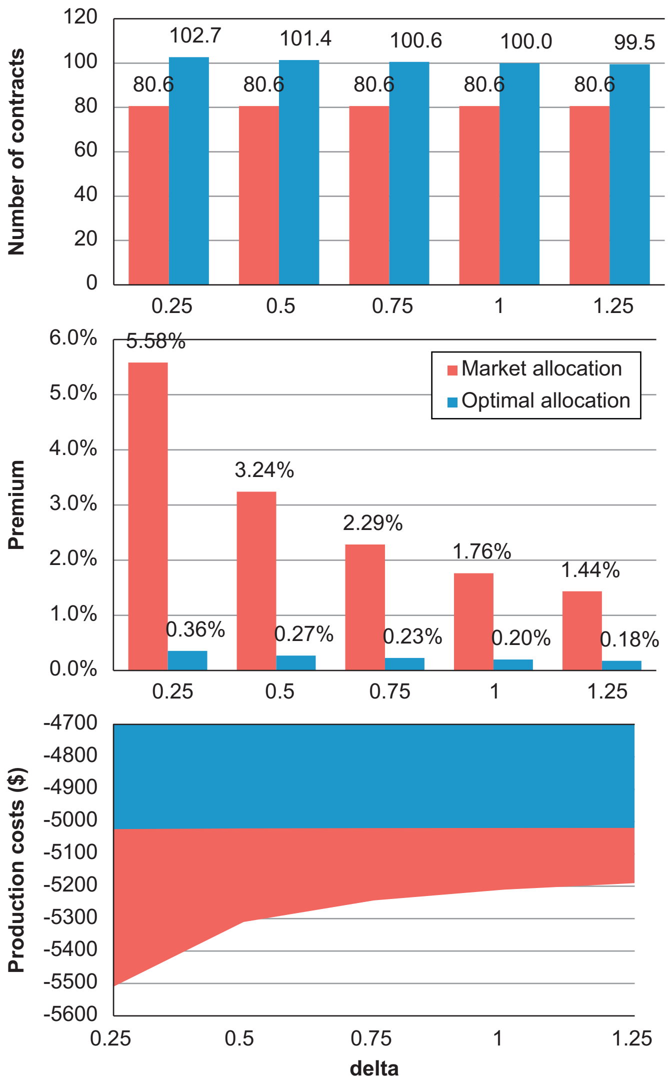

Fig. 6. The cost of balancing power and market outcomes.

For example, for $\delta=1$, the optimal allocation is to pre-schedule exactly the expected load (i.e. 100). This however still justifies a certain premium (of size 0.2). Although the penalty is symmetric the direct cost of generating power is convex which explains the rationale for the existence of a positive premium. As δ increases the forward premium decreases because the financial consequences of starting up generators in real-time are less severe.

In our illustration the unconstrained aggregate amount of electricity offered by five power producers via forward contracts is 80.6% of the expected load. This amount is invariant to the value of δ since in this range it is always the case that GFs choose to withhold generation capacity. Since this happens to the extent that the probability of shut downs is close to zero, market allocation does not vary. On the other hand, the premium is adjusted with respect to the change in starting-up costs.

Finally, we compute the deadweight loss caused by misallocating generation capacity. Deadweight loss is simply the difference between the cost of generating power with the market allocation rule and the cost of power following the central planner solution. This difference is deadweight loss because real-time demand for electricity is completely inelastic. The cost of generating power in our example exceeds the cost of the optimal allocation by 3.1–9.7%, depending on the values of cost parameters employed (Fig. 6). Although the magnitude of these results varies according to the portfolio of energy resources in a region in question, our predictions are unambiguous—higher cost of starting up generators translates into GFs exercising more market power which then results in larger deadweight loss.

5. Conclusions

In this paper we present a new theoretical framework to model firm behavior in deregulated electricity markets. The framework was used to perform computational experiments to demonstrate how the model can be employed in studying applied problems. The sensitivity analyses show that the model is flexible in accommodating diverse assumptions regarding the electric sector and the characteristics of any delivery period in question. Similar to real-world electricity markets, the distributions of the spot price at different delivery periods in the model diverge greatly. Also, the model generates a wedge between the forward price and the expected spot price. This result is in line with the extensive empirical evidence suggesting the existence of a forward premium. Yet, it is important to perceive a fundamental difference between our work and the related body of literature in this area. The results presented by our model are not driven by a risk preferences assumption but by the basic properties of power generation cost structure, number of firms, and the commonly adopted design of deregulated electricity markets.

Our results make it clear that GFs have an incentive to manipulate market prices by withholding generation capacity to maximize the joint profits from the spot and forward markets. Although deadweight loss is typically expected in a Cournot competition, the loss in electricity markets is more severe than in other markets in which market power is present. That is because real-time load is inelastic; the power is being generated eventually in spite of any capacity withholding behavior. Consequently, electricity production may be inefficient and cause misuse of energy resources, which in turn generates sizeable premiums for forward contracting. Policy makers and administers of deregulated electricity markets are aware of this problem. Yet, in practice it is not an easy task to measure when market power is exercised. Overall operating costs and production constraints are not transparent. Operating costs include the incremental cost involved in shifting to power plants of higher heat rates, start up and shut down costs as well as other costs. Add in the complications such as the physical withholding of power due to outages and periods of maintenance and it is clear that precise markups are complicated to estimate.

Although there is an ongoing debate about what is the effective way to measure and monitor the degree of competitiveness of wholesale electricity markets, some provisions are commonly implemented. For example, the Hirschmann–Herfindahl index (HHI) is used as a first screening tool for market power by governmental agencies. Price caps are used frequently as an upper bound for spot price. The price cap is useful in stabilizing the volatility of spot markets and limiting LSEs' exposure to spot prices. Finally, and maybe the most effective and frequently used tool to deal with uncompetitive behavior is to impose a *must-offer* provision. Doing so, ISOs limit the ability of large producers to exercise market power by forcing them to participate and offer their capacity in forward and spot markets. In addition, the ISOs examine regularly whether the prices offered by GFs enable them to schedule considerable volume in advance, bilaterally and via forward markets. Our model suggests that sustainable forward premiums do not necessarily indicate the existence of a market failure. However, it is more likely that market power is present where the electric sector is still relatively concentrated and peaking power plants account for a relatively large share of electricity consumption, as in this case ramping costs play a central role in electricity pricing.

Appendix A. Second-order condition for the GFs maximization problem

The second derivative of $\mathcal{M}$ with respect to the number of forward contracts is

$$\frac{\partial^2 \mathcal{M}}{\partial q_F^{i2}} = \frac{\partial^2 P_F(\cdot)}{\partial q_F^{i2}} q_F^i + 2\frac{\partial P_F(\cdot)}{\partial q_F^i} - \int_0^{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i} \frac{\partial^2 C(q^i, q_F^i)}{\partial q_F^{i2}} f_X(X|\hat{X})dX$$
$$+ \int_{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i}^{\infty} \left\{ \frac{\partial^2 P_s(\cdot)}{\partial q_F^{i2}} \times (q^i - q_F^i) + \frac{\partial P_s(\cdot)}{\partial q_F^i}\left(\frac{\partial q^i}{\partial q_F^i} - 1\right) + \frac{\partial P_s(\cdot)}{\partial q_F^i}\left(\frac{\partial q^i}{\partial q_F^i} - 1\right) + P_s(\cdot)\left(\frac{\partial^2 q^i}{\partial q_F^{i2}}\right) - \frac{\partial^2 C(q^i, q_F^i)}{\partial q_F^{i2}} \right\} f_X(X|\hat{X})dX. \tag{31}$$

Notice that

1. $\dfrac{\partial^2 P_F(\cdot)}{\partial q_F^{i2}} = 0$ and $\dfrac{\partial^2 q^i}{\partial q_F^{i2}} = 0$.

2. $\displaystyle\int_0^{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i} \frac{\partial P_s(\cdot)}{\partial q_F^i} f_X(X|\hat{X})dX = 0,$

   which is to say that in the event that no additional generators are turned on after the forwards market is cleared the spot price is not affected by the marginal forward contract.

3. Recall that

   $\dfrac{\partial P_s(\cdot)}{\partial q_F^i} = -\dfrac{\alpha_s}{M}$, and $\dfrac{\partial q^i}{\partial q_F^i} = \dfrac{M-1}{M} \times \dfrac{\alpha_s}{\alpha_t + \alpha_s}$.

Therefore we can write

$$\frac{\partial^2 \mathcal{M}}{\partial q_F^{i2}} = 2\frac{\partial P_F(\cdot)}{\partial q_F^i} - \alpha_s^- \int_0^{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i} f_X(X|\hat{X})dX - 2\frac{\alpha_s^+}{M} \int_{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i}^{\infty} \times \left(\frac{M-1}{M} \times \frac{\alpha_s^+}{\alpha_t + \alpha_s^+} - 1\right) f_X(X|\hat{X})dX - \alpha_s^+ \int_{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i}^{\infty} f_X(X|\hat{X})dX. \tag{32}$$

Rearranging

$$\frac{\partial^2 \mathcal{M}}{\partial q_F^{i2}} = 2\frac{\partial P_F(\cdot)}{\partial q_F^i} - \alpha_s^- \int_0^{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i} f_X(X|\hat{X})dX - \alpha_s^+ \int_{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i}^{\infty} \times \left[\frac{2(M-1)\alpha_s^+ + M(M-2)(\alpha_t + \alpha_s^+)}{M^2(\alpha_t + \alpha_s^+)}\right] f_X(X|\hat{X})dX. \tag{33}$$

Substituting for the derivative of the forward price we get

$$\frac{\partial^2 \mathcal{M}}{\partial q_F^{i2}} = -2\frac{\alpha_s^+}{M}\left(1 + \frac{1}{N}\right) \int_{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i}^{\infty} f_X(X|\hat{X})dX - \alpha_s^- \times \int_0^{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i} f_X(X|\hat{X})dX - \alpha_s^+ \int_{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i}^{\infty} \times \left[\frac{2(M-1)\alpha_s^+ + M(M-2)(\alpha_t + \alpha_s^+)}{M^2(\alpha_t + \alpha_s^+)}\right] f_X(X|\hat{X})dX. \tag{34}$$

Finally

$$\frac{\partial^2 \mathcal{M}}{\partial q_F^{i2}} = -\alpha_s^- \int_0^{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i} f_X(X|\hat{X})dX - \alpha_s^+ \int_{\sum_{m=1,m\neq i}^{M} \overline{q}_F^m + q_F^i}^{\infty} \frac{2(M-1)\alpha_s^+ + M\left(M + \frac{2}{N}\right)(\alpha_t + \alpha_s^+)}{M^2(\alpha_t + \alpha_s^+)} f_X(X|\hat{X})dX \tag{35}$$

which is negative for any positive integers of N and M. Therefore, $\mathcal{M}$ is strictly concave in q_F^i, thus we know that if there is an interior symmetric solution it has to be unique.

References

Allaz, Blaise, Vila, Jean-Luc, 1993. Cournot competition, forward markets and efficiency. Journal of Economic Theory 59 (1), 1–16.

Benth, Fred, Espen, Álvaro, Cartea, Kiesel, Rüdiger, 2008. 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.

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

Bjorgan, Roger, Liu, Chen-Ching, Lawarree, Jacques, 1999. Financial risk management in a competitive electricity market. IEEE Transactions on Power Systems 14 (4), 1285–1291.

Borenstein, Severin, Bushnell, James, 1999. An empirical analysis of the potential for market power in California's electricity industry. The Journal of Industrial Economics 47 (3), 285–323.

Borenstein, Severin, Bushnell, James B., Wolak, Frank A., 2002. Measuring market inefficiencies in California's restructured wholesale electricity market. The American Economic Review 92 (5), 1376–1405.

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

Douglas, Stratford, Popova, Julia, 2008. Storage and the electricity forward premium. Energy Economics 30 (4), 1712–1727.

Green, Richard J., Newbery, David M., 1992. Competition in the British electricity spot market. Journal of Political Economy 100 (5), 929–953.

Green, Richard, 1999. The electricity contract market in England and Wales. The Journal of Industrial Economics 47 (1), 107–124.

Holttinen, Hannele, 2005. Optimal electricity market for wind power. Energy Policy 33 (16), 2052–2063.

Kian, Ashkan R., Jose Jr., B., Cruz, Thomas, Robert J., 2005. Bidding strategies in oligopolistic dynamic electricity double-sided auctions. IEEE Transactions on Power Systems 20 (1), 50–58.

Klemperer, Paul D., Meyer, Margaret A., 1989. Supply function equilibria in oligopoly under uncertainty. Econometrica 57 (6), 1243–1277.

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

Mansur, Erin T., 2008. Measuring welfare in restructured electricity markets. Review of Economics and Statistics 90 (2), 369–386.

Newbery, David M., 1998. Competition, Contracts, and entry in the electricity spot market. The RAND Journal of Economics 29 (4), 726–749.

Pirrong, Craig, Jermakyan, Martin, 2008. The price of power: the valuation of power and weather derivatives. Journal of Banking & Finance 32 (12), 2520–2529.

Puller, Steven L., 2007. Pricing and firm conduct in California's electricity market. The Review of Economics and Statistics 89 (1), 75–87.

Soares, Lacir J., Souza, Leonardo R., 2006. Forecasting electricity demand using generalized long memory. International Journal of Forecasting 22 (1), 17–88.

Soares, Lacir J., Medeiros, Marcelo C., 2008. Modeling and forecasting short-term electricity load: a comparison of methods with an application to Brazilian data. International Journal of Forecasting 24 (4), 630–644.

Sun, Junjie, Tesfatsion, Leigh S., 2007. Dynamic testing of wholesale power market designs: an open-source agent-based framework. Computational Economics 30, 291–327.

Taylor, James W., deMenezes, Lilian M., McSharry, Patrick E., 2006. A comparison of univariate methods for forecasting electricity demand up to a day ahead. International Journal of Forecasting 22 (1), 1–16.

Tseng, Chung-Li, Barz, Graydon, 2002. Short-term generation asset valuation: a real options approach. Operations Research 50 (2), 297–310.

Twomeya, Paul, Neuhoff, Karsten, 2010. Wind power and market power in competitive markets. Energy Policy 38 (7), 3198–3210.

Wolfram, Catherine D., 1999. Measuring duopoly power in the British electricity spot market. The American Economic Review 89 (4), 805–826.