UNIVERSITY OF
CAMBRIDGE

**Electricity Policy
Research Group**

The supply function equilibrium and its policy implications for wholesale electricity auctions

Pär Holmberg and David Newbery

Abstract

The supply function equilibrium provides a game-theoretic model of strategic bidding in oligopolistic wholesale electricity auctions. This paper presents an intuitive account of current understanding and shows how welfare losses depend on the number of firms in the market and their asymmetry. Previous results and general recommendations for divisible-good/multi-unit auctions provides guidance on the design of the auction format, setting the reservation price, the rationing rule, and restrictions on the offer curves in wholesale electricity auctions.

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

www.eprg.group.cam.ac.uk

The supply function equilibrium and its policy implications for wholesale electricity auctions[*]

Pär Holmberg [†] and David Newbery [‡]

April 14, 2010

Abstract

The supply function equilibrium provides a game-theoretic model of strategic bidding in oligopolistic wholesale electricity auctions. This paper presents an intuitive account of current understanding and shows how welfare losses depend on the number of firms in the market and their asymmetry. Previous results and general recommendations for divisible-good/multi-unit auctions provides guidance on the design of the auction format, setting the reservation price, the rationing rule, and restrictions on the offer curves in wholesale electricity auctions.

Key words: Wholesale electricity markets, supply function equilibria, competition policy, market regulation

JEL Classification D43, D44, C62, L94

1 INTRODUCTION

The wave of restructurings in the electricity supply industry prompted by experience in Britain and Chile, and in Europe under pressure from a succession of EU Directives, raises new issues for regulators. The former vertically integrated industry needed regulation of the final prices (explicitly, if privately owned, or implicitly, if state owned). Restructuring aimed to create competing generating companies selling into a wholesale market, with competing retailers buying to supply their customers. As electricity cannot readily be stored, a system operator is required to take charge of balancing instantaneous demand and supply and ensuring that the current flowing through the transmission links does not exceed safe limits by calling on generators in different locations to adjust their output. Restructuring may result in too few generating companies located within constrained market areas (which are unable to import alternative generation from outside the zone because of transmission constraints), raising issues of market power. Finally, demand and supply vary considerably over the course of a day and season, and both are subject to sudden shocks, caused by plant or line failures, weather changes and even the half-time break in a major sporting event.

As a result the wholesale market and the balancing market or mechanism need careful design to ensure efficient dispatch at acceptable prices. This paper addresses the question of what we have learned from the analysis of such markets, an active topic in the economics of Industrial Organization and auction theory, as the traditional models of imperfect competition have proven unsatisfactory for these very specific features of electricity markets. In contrast to most other markets, the way price is determined is very well defined in the standard model of a wholesale electricity market. Each producer submits an offer curve that specifies how much it is willing to produce at different prices. Similarly, consumers and retailers (suppliers), who represent small consumers, submit demand (or bid) curves specifying how much electricity they want to buy at different prices. The design of the market can influence price formation and how competitive the market will be by choice of the auction format, the level of any price cap, the rationing rule, and by making restrictions on the offer curves. When making their choice regulators should consider the impact on participants' contracting and investment incentives under various market designs and rules.

Competition authorities also need to predict electricity prices under various counterfactuals – what might happen if a merger or acquisition is accepted or an interconnector built? Often authorities are content with using concentration measures, such as the Herfindahl-Hirschman index (HHI), to assess the degree of competition in the market. However, these measures work poorly for electricity markets, where demand and supply must be instantly balanced and where the tightness of reserve margins and transmission capacity constraints can vary considerably

over short periods with significant impacts on prices (Borenstein et al., 1999, Ofgem, 2000). Thus given installed production capacities, it depends very much on the level and location of instantaneous demand whether the market will suffer from the exercise of market power.

Fortunately, we have made considerable progress in developing a more suitable model - the supply function equilibrium - to address these questions. This paper qualitatively assesses the two leading auction formats, the uniform-price and pay-as-bid formats, and other rules of electricity auctions using supply function equilibria under uncertain demand. We provide new results highlighting how short-run welfare losses depend on the number of firms in the market and their asymmetry.

The paper is organized as follows. The remainder of this section sets out the structure of the electricity wholesale market to motivate the justification for the supply function equilibrium (SFE) model, Section 2 characterises the SFE and surveys the literature. Section 3 draws out the implications for the analysis of market power, derives expressions for the deadweight cost of that market power and the effect of forward contracting on both. The section concludes with a brief summary of the empirical support for the SFE model. Section 4 examines possible market design remedies and section 5 concludes.

1.1 The wholesale electricity market

Electricity is produced by many different technologies that often have very different marginal costs. The production cost of a plant is primarily determined by fuel costs and its efficiency that are well-known and common knowledge. The plants of a producer are used in merit order, starting with the lowest marginal cost, such as nuclear power or hydro-power. Last in the merit-order are peaking plants, such as open-cycle gas turbines burning natural gas or oil with high variable and low capital costs. The merit-order implies that producers' marginal costs increase with output. There are some local deviations from this trend, as start-up costs introduce local non-convexities, but these are normally neglected in market analyses (though not in optimal scheduling programmes). Although electricity may be produced by various technologies, it is still a completely homogeneous good suitable for trading on commodity exchanges and auctions.

In wholesale electricity markets, producers sell electricity to retailers. In their turn, retailers sell electricity to consumers in the retail market. Electricity consumption is to a large extent determined exogenously, e.g. by the weather and work-days or holidays, and is very inelastic, especially close to the time of consumption.[1] This limited flexibility means that retailers' market power is small compared with that of generators, which can be significant.

---

[1]Demand can bid usually into the market, but in Britain the amount in the past was small - 2,000MW compared to peak demand of over 50,000 MW. Smart metering may change this in future.

Due to restrictions on the rate that fossil generation can ramp-up output, particularly from a cold start, production plans are scheduled the day before delivery, and the day-ahead (or prompt or spot) market is an important component in this planning process. A well-designed liquid market can provide the strike prices for financial contracts. The day-ahead market is typically organized as a double auction to which retailers and producers submit non-increasing bid curves and non-decreasing offer curves, respectively, as shown in figure 1. The market clearing price (MCP in figure 1) is determined by the intersection of the bids and offers. There is normally a separate price and auction for each delivery period, typically a half-hour or hour but which can be as short as 5-15 minutes, even if, as in many markets, the generators' offer curves must be valid for the whole of the next day.

Electricity is special in that supply must equal consumption at every instant, because it is very expensive to store electrical energy on a large scale. The system operator uses a real-time or balancing market to make necessary adjustments to production (and consumption to the extent that it bids into the market) during the delivery period by accepting additional power production from, or by selling back electricity to, producers. Offers to the balancing market are submitted before the delivery period starts but demand is uncertain when offers are submitted. System imbalances arise because of unexpected changes in wind and temperature, unexpected production outages or unexpected transmission-line failures. These are normally handled by the available incremental and decremental production capacity in the real-time market.

In very extreme cases, this available incremental production capacity may be insufficient to meet the system imbalance caused by multiple unexpected events, and an outage or loss of load occurs. The loss-of-load probability (LOLP) is typically very small, but always positive. No matter how large the reserve margin (available incremental capacity), sufficiently many simultaneous unexpected events that decrease the production or reserve capacity, or increase demand, will lead to a power shortage. LOLP during any delivery period can be estimated *ex* ante from the reserve margin, the probability distribution of demand and from the probabilities of having production failures in individual plants. Newbery (1998b) shows that the LOLP estimated ex ante by the system operator in Britain decreases exponentially with the reserve margin. Using data provided in Newbery's paper, and adjusting for the system operator's consistent overestimation of LOLP, we roughly estimate LOLP during an half-hour to 0.1% for a reserve margin of 10% for the British market from 1990-95, and that it roughly decreased by a factor 100 for every additional 10% of reserve capacity.[2] However, these estimates are very uncertain, and only intended to give

---

[2]The values are estimated from Fig. 5 in Newbery (1998b) and that capacity payments in the old pool were proportional to LOLP. It is assumed that the system operator has overestimated the LOLP by a factor of 50. This very uncertain factor is based on the reasoning in the appendix of Newbery (1998b).

some feeling for the magnitude of LOLP and its relation to the reserve margin. Normally, the variance of the demand distribution is small and predictable, with most uncertainty lying on the supply side.

In the rare situations when electricity demand exceeds market supply, demand has to be rationed (first by reducing voltage that cuts demand automatically, then by load shedding) to avoid a system collapse, and this occurs when the price reaches the price cap. One reason for price caps is that consumers who do not switch off their equipment when the electricity price becomes very high do not necessarily have a high marginal benefit of power. It may be that the residential consumer is not at home or not aware of the high price. Moreover, residential consumers typically do not face the real-time price but buy whatever they want at a contracted price, and distribution companies do not observe who consumes what in real-time. Such market or information imperfections make it welfare improving to ration demand at some very high reservation price, often set equal to the value of lost load (VOLL) (Stoft, 2002). In the English Electricity Pool, VOLL was set at £2,500/MWh in 1990 (5,000 euros/MWh at 2009 prices and exchange rates). These market/information imperfections may be reduced by installing real-time ("smart") meters combined with time-of-use pricing, and could be completely removed if the meters could automatically control desired household consumption.

Most electricity wholesale and real-time (or balancing) markets are organized as uniform-price auctions in which all accepted bids from retailers pay the market clearing price and all accepted offers from producers are paid the same price. The British balancing market (strictly, the Balancing Mechanism) is an exception. In this market, all accepted bids and offers pay or are paid their bid or offer. This discriminatory format has also been seriously considered in other electricity markets, e.g. California (Kahn et al., 2001) and recently also in Italy.[3] In several, mainly European, markets, production is adjusted after market-clearing to ensure that transmission constraints are not violated within any price zone. This is called counter-trading and offers accepted in this post-clearing process are paid their offer price, not the market clearing price (MCP) of that price zone, according to a producer's individual supply function to the balancing market.[4] Thus many balancing markets are actually a blend of the uniform-price and

---

[3]The Italian Law Decree n. 185/2008, better known as the "Anti-Crisis Decree" was converted into law by January 27, 2009. As a part of the decree, the Ministero per lo Sviluppo Economico in consultation with the Italian Authority for electricity and gas ("AEEG"), have been instructed to change the trading mechanism on the Day-Ahead Market (or "MGP") from the current system based on a marginal price rule to a new system based on a pay as bid rule. The change is intended to decrease the average price of electricity traded on the "MGP".

[4]Plant that is constrained-off, that is requested not to supply because of excessive local production, is typically paid its lost profit, equal to the MCP less its offer price. Effectively, the producer sells accepted power production at the MCP when the market is cleared but if it turns out that his production is infeasible, he has to buy this power back at its offer price in the post-clearing adjustment.

the pay-as-bid format. Treasury auctions are similar to electricity auctions in that both are divisible-good/multi-unit auctions and bidders commit to a bid/offer curve in both auctions. Unlike electricity markets, the pay-as-bid format is used in most treasury auctions. Bartolini and Cottarelli (1997) find that 39 out of 42 countries surveyed by them use discriminatory auctions. But it should be noted that the U.S. treasury gradually switched from a pay-as-bid to a uniform-price format during 1992-1998 (Ausubel and Cramton, 2002).

Electricity prices are volatile because electricity is not suitable for large-scale storage, short-run demand is very inelastic and short-run supply can also be inelastic (and in any case *the marginal cost* varies substantially depending on the fuel and efficiency of the marginal plant). To hedge their risks, market participants can buy and sell various derivative contracts, e.g. futures and forward contracts. These contracts commit the parties to buy and sell the contracted quantity in the spot or real-time market at the agreed delivery price. In most cases the contracts are financial, so no physical transaction takes place. Forward contracts are normally traded over the counter, and futures contracts on power exchanges.

2 The supply function equilibrium

The supply function equilibrium (SFE) model was originally developed by Klemperer and Meyer (1989) and first applied to the electricity market by Green and Newbery (1992) and Bolle (1992). It is a game-theoretic model of competition in wholesale electricity markets. It assumes that each producer chooses its offer curve in order to maximize its profit, given demand and offer curves chosen by competitors. In concentrated markets, producers may be able to tacitly collude to higher electricity prices, as Sweeting (2007) suggests happened in Britain, but this effect is only briefly considered below. The SFE is a static equilibrium, i.e. it is assumed that all producers will play their profit-maximizing strategies repeatedly, irrespective of what happened in previous periods. In practice it may take a while before each producer has figured out its best offer strategy given the residual demand it faces. This learning is facilitated in markets that disclose individual or aggregated offer curves to the auction, such as the Amsterdam Power Exchange (APX), illustrated in figure 1. This figure shows the determination of the market clearing price (MCP) for hour 12 on 26 June 2007, illustrating a part of the ladders of offers and bids and showing that 1,942.4 MW was traded at a MCP of 58.83 Euros/MWh.[5]

The setting of the SFE assumes that production costs are common knowledge and that

---

[5] Typical capacity connected to the Dutch system would be over 15,000 MW and the APX covers a wider area than just the Netherlands, so a relatively small fraction of capacity is traded on the APX. Note that price responsiveness on the demand side is mainly provided by producers who bid to buy back electricity that they have sold in the forward market.
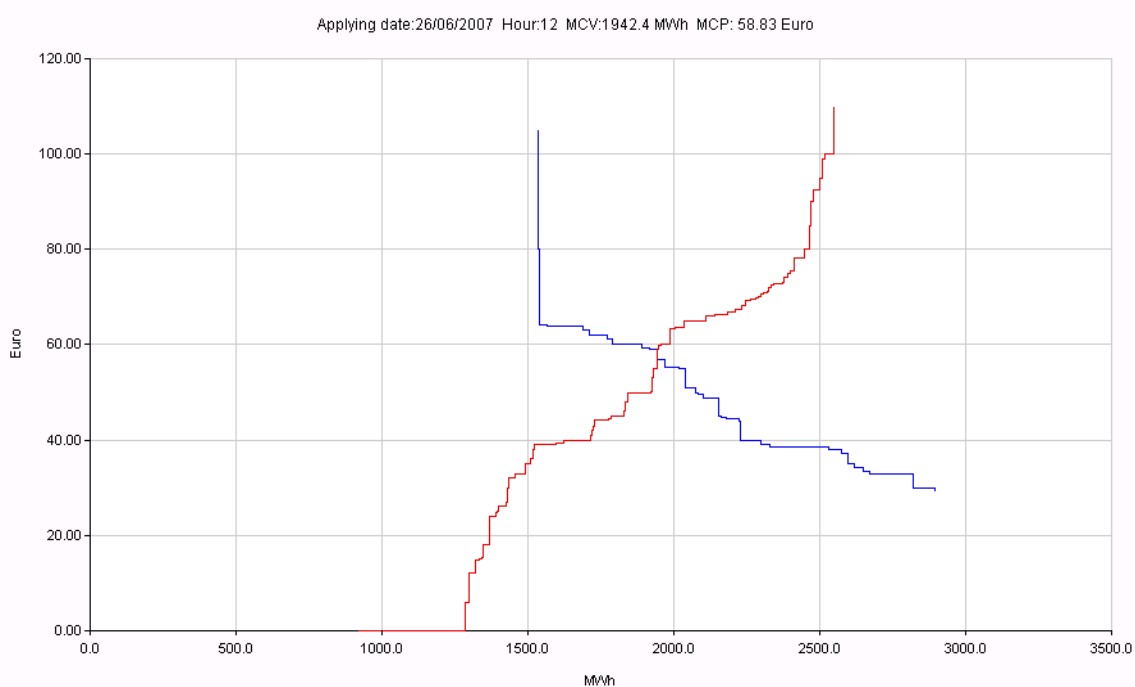
Figure 1: *Market clearing price detail from APX Hour 12, 26 June 2007*

demand is uncertain or time-varying. This is a good description of electricity markets, where technology characteristics and fuel prices are transparent and producers make offers before demand has been realised. They further assume that the shock ε is additive to the demand schedule, so that $D(p,\varepsilon) = D(p) + \varepsilon$, and production uncertainties are neglected for strategic producers. This considerably simplifies the determination of the residual demand schedule of a producer i, for which shocks will again be additively separable, $R_i = R_i(p) + \varepsilon$. These assumptions make the SFE model tractable in comparison with multi-unit/divisible good auction models with common or affiliated uncertain values/costs by Wilson (1979) and Ausubel and Cramton (2002), which are often used to analyze treasury auctions. For example, SFE can be determined analytically for cases with constant marginal costs (Newbery, 1998; Holmberg, 2007) and linear marginal costs with linear demand (Klemperer and Meyer, 1989; Green, 1996; Baldick et al., 2004; Newbery, 2008b).[6] Closed form solutions are also available for symmetric firms and perfectly inelastic demand (Rudkevich et al, 1998; Anderson and Philpott, 2002a). Numerical algorithms make it possible to calculate SFE of markets with asymmetric firms and general cost functions (Anderson and Hu, 2008a; Aromí, 2007; Edin, 2007; Holmberg, 2009a).

The equations originally derived by Klemperer and Meyer (1989) can be explained in the

---

[6]There are linear SFE for the case of linear marginal costs, and analytic non-linear solutions have been derived by Newbery (2008b).

following intuitive way. Even if the demand shock is unknown *ex ante* a producer can still construct an offer curve that leads to an ex post optimal profit for each shock outcome. In a pure-strategy Nash Equilibrium (NE) of the game, competitors' offer strategies are known, so that for each shock outcome a producer knows its residual demand. Thus it can calculate the optimal mark-up for each outcome by applying the monopoly mark-up rule or Ramsey pricing to the elasticity of its residual demand (defined to be a positive number) γ_i^{res} (Tirole, 1988):

$$\frac{p - C'(q_i(p))}{p} = \frac{1}{\gamma_i^{res}}. \tag{1}$$

The elasticity of residual demand is straightforward to calculate as long as there are no binding transmission constraints in the power system. With binding constraints, changing an offer in one node will influence the power flows in the system in a non-trivial way, but it is still possible with suitable software to derive the transmission-constrained elasticity of residual demand (Xu and Baldick, 2007). Figure 2 shows that the optimal output for each demand outcome is chosen so that the marginal revenue (MR) equals the marginal cost (MC). This gives a point in the producer's optimal offer curve. By repeating the procedure for each shock outcome ε, the optimal offer curve can be derived. The same approach works where demand is certain but time varying and offers must be valid for each time period (48 half-hours in Britain), provided that the time dependence is separable so that $D(p, t) = D(p) + f(t)$. For the tractable case of linear demand this would have $D(p, t) = a - \mu t - bp$, $0 \le t \le 1$.

Each producer calculates its offer curve in a similar way and the system is in equilibrium when each producer is satisfied that its offer schedule is optimal given the behaviour of all other suppliers. Equilibrium is then calculated from a system of equations as in (1). The equations are differential equations as the optimal mark-up of one producer depends on the slope of competitor's offer curves. The second-order condition for profit maximizing offer curves in uniform-price auctions is that the marginal cost should increase faster than the marginal revenue at each extremum point, which is the case in figure 2. Holmberg et al. (2008) show that the second-order condition is satisfied for all sets of increasing offer curves satisfying (1) if demand is weakly concave (which includes linear demand).

Without an initial condition or end-condition there is generally a continuum of equilibria bounded by the Bertrand and Cournot equilibria. Figure 3 shows the range of possible equilibria estimated by Green and Newbery (1992) for two possible market structures in the English Electricity Pool. The Cournot line is the optimal offer if all other producers offer a fixed supply - as would be the case if competitors' capacity constraints are binding in the high demand state, that is, his supply is required to balance demand and supply then. The wide range of equilibria can be explained in the following way. Assume for simplicity that marginal costs are constant. Now if competitors' choose to play Bertrand strategies, i.e. offers are perfectly elastic, then the residual
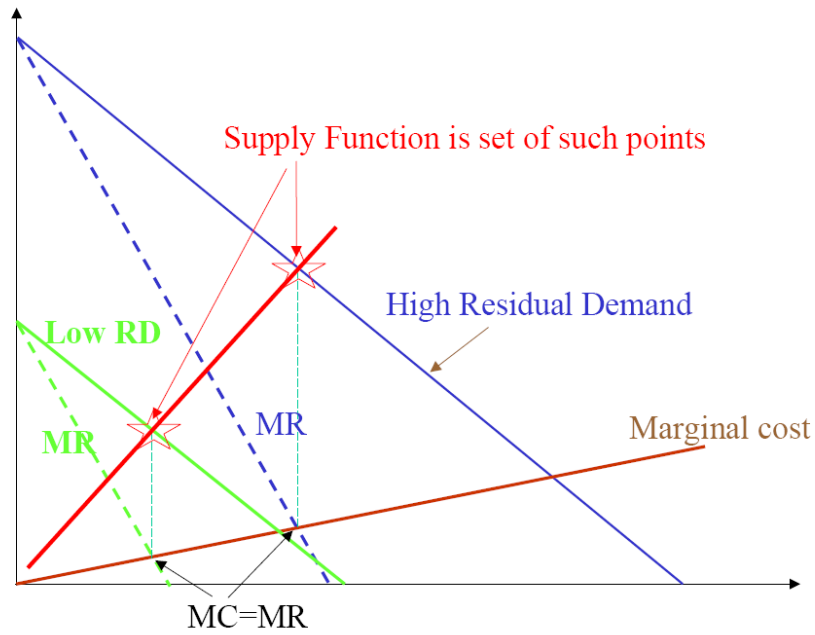
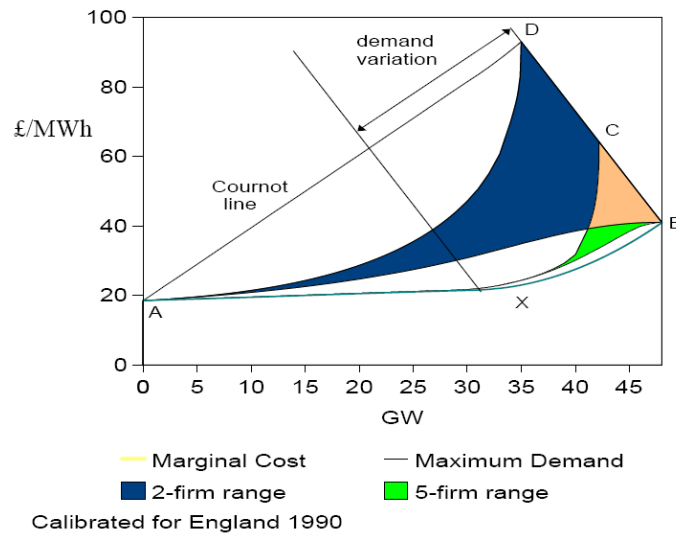Figure 2: *The SF is the best response to all residual demands*

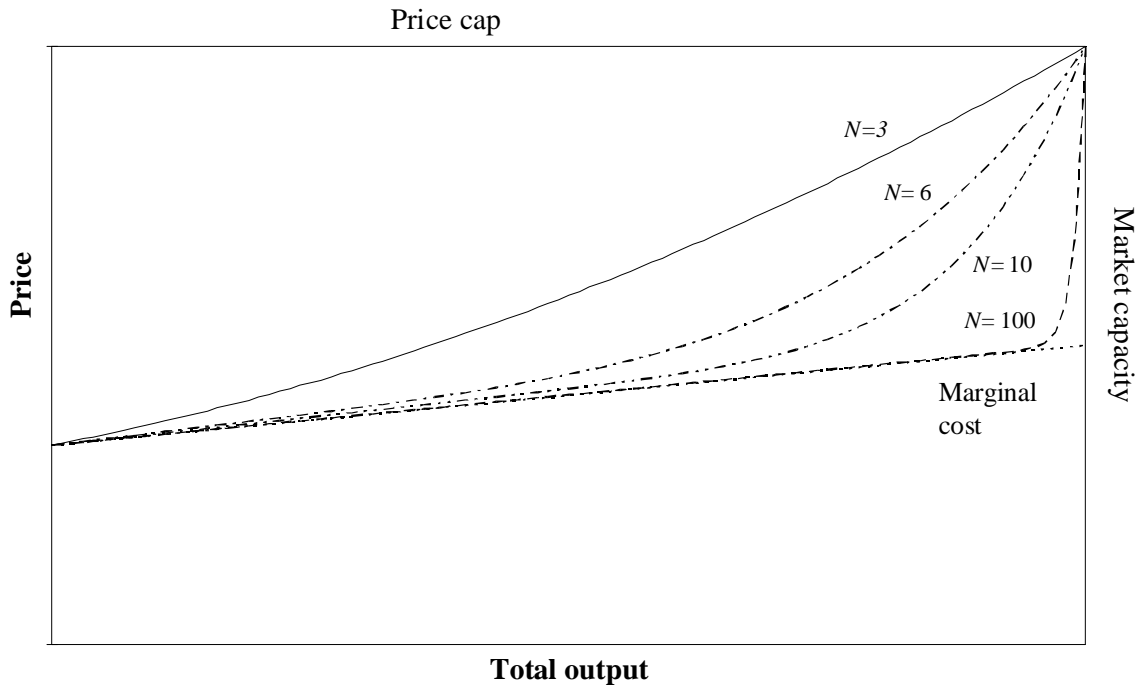Figure 3: *Estimated SFE for England and Wales, 1990.*

Figure 4: *The effect of capacity and number of competitors on the SF*

demand of producer becomes perfectly elastic and the best response is, according to (1), to have zero mark-ups, so that the best offer is a Bertrand offer. If competitors' offers are less elastic then the best response will have increasing mark-ups, so the optimal offer is also less elastic. But the range of equilibria shrinks as the maximum demand shock increases and with possible infinite demand shocks a unique equilibrium can be found (Klemperer and Meyer, 1989). For bounded demand shocks, the range of equilibria is constrained by capacity (Green and Newbery, 1992; Baldick and Hogan, 2002; Anderson and Hu, 2008a). Genc and Reynolds (2004) analyse in detail how pivotal producers reduce the range of equilibria. In particular, a unique equilibrium will be singled out if maximum demand is high enough to make the capacity constraints of all (but possibly one) firms bind with a positive probability, which could be arbitrarily small (Aromí, 2007; Holmberg, 2007; Holmberg, 2008a). If demand is sufficiently inelastic, the unique equilibrium is selected by the price cap and the capacity constraint. Figure 4, taken from Holmberg (2008a), illustrates this for the case of perfectly inelastic demand. We see that the market price is near marginal cost for low demand outcomes and near the price cap for high demand outcomes.

To avoid discontinuities in its optimal offer curve, each producer needs to face a smooth residual demand curve. If demand is sufficiently elastic so that the price cap never binds for the optimal offer curves, then the unique symmetric equilibrium for a positive loss of load probability
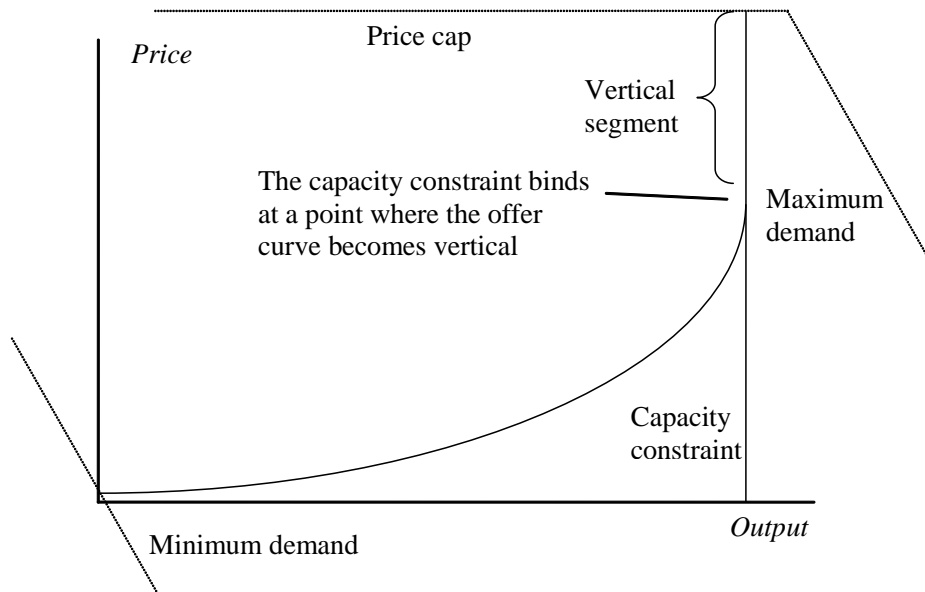
Figure 5: *Uniqueness determined by the SF becoming vertical at full capacity*

(LOLP) is given by the symmetric solution where the slope of the supply curve becomes vertical at the capacity constraint, i.e. supply curves touch the Cournot line at that point. This ensures that all producers have a smooth residual demand at the capacity constraint and corresponds to the upper supply schedule in Green and Newbery (1992). Note that maximum demand schedule need not cross the smooth offer curve implied by the first–order condition in (1), because at the capacity constraint this curve can be extended vertically up to the price cap without any kinks, as shown in figure 5, with the price then on the demand curve (in this case at the cap).

2.1   Generalisations of the supply function equilibrium

The SFE model developed by Klemperer and Meyer (1989) relied on *ex post* optimality, i.e. a producer could choose its optimal supply function (SF) such that its profit was maximized for each shock outcome. This may not always be possible, if producers choose the best expected outcome before knowing the shock, but would like to have chosen a different offer curve once they know the actual shock. Thus Wilson (2008) analyses SFE in a transmission network with multiple nodes, where each node has an individual demand shock. When the shocks are multi-dimensional, it is generally not possible to find an *ex post* optimal SF. Wilson was nevertheless able to derive first-order conditions for the *ex ante* optimal choices and hence characterise an equilibrium for a general transmission network. The equations are quite complicated, as they depend on the probability distributions of the demand shocks.

Anderson and Philpott (2002b) characterize the residual demand by its market distribution function, which implicitly determines the contour of the residual demand curve for each probability level. This general formulation allows for equilibria that are *ex ante*, but not *ex post*, optimal and residual demands that are not restricted to the linear shifts of the SFE model. Analogously, one can introduce offer distribution functions, which implicitly determine the contour of each producer's offer curve for each probability level, to analyse mixed strategy equilibria in electricity auctions and other multi-unit auctions (Anderson et al., 2009).

3 The determinants of market power

One of the key questions in restructuring the industry is whether the resulting structure will be workably competitive, and to that extent not to need additional regulation. Similar questions arise if a merger is proposed, or whether a proposed demerger will be sufficient to satisfy a regulator concerned with market power (as with the 1994 British Pool Inquiry that resulted in the divestiture of 6,000 MW of plant from National Power and Powergen - see Offer, 1994, 1995). For that we need models that relate the price-cost mark-up to the industry structure and demand characteristics. Newbery (1998) analytically solves for the SFE for two symmetric duopolies with constant marginal costs c, linear demand $D(p) = a - bp$, and additive shocks or time varying demand (a varies with t). In the appendix, this solution is generalised to $N > 2$ symmetric firms with market capacity $\overline{Q}$ to given an implicit equation for the price-cost mark-up $\frac{p-c}{c}$:

$$ x = \frac{Q}{\overline{Q}} = \frac{N\gamma}{(N-2)} \frac{p-c}{c} \left((N-1) \left(\frac{1}{N\gamma} \frac{c}{p-c} \right)^{(N-2)/(N-1)} - 1 \right), \tag{2} $$

where $x = Q/\overline{Q}$ is the load factor (output as a fraction of total capacity) and $\gamma = cb/\overline{Q}$ is the elasticity of demand at the efficient price c (i.e. the marginal cost) when the demand shock ε (or the time varying value of a) is such that the linear demand at this price equals the market capacity. The relation is based on the upper supply schedule, so it is a worst case (highest price) scenario. If the price cap is sufficiently low, so that it binds, the market will have a more competitive unique equilibrium.

We can use (2) to prove that it is beneficial for competition to connect two symmetric markets. This will double the capacity, demand, including its slope, and the number of producers in the integrated market. Hence the dimension-less constant $\gamma = cb/\overline{Q}$ is the same in the integrated market as for the separate markets, but N has doubled. Thus mark-ups, $\frac{p-c}{c}$, will be reduced for every load factor x.[7] But if instead the same firms operate on the two separate markets, market integration neither influences N nor mark-ups as a function of the load factor.

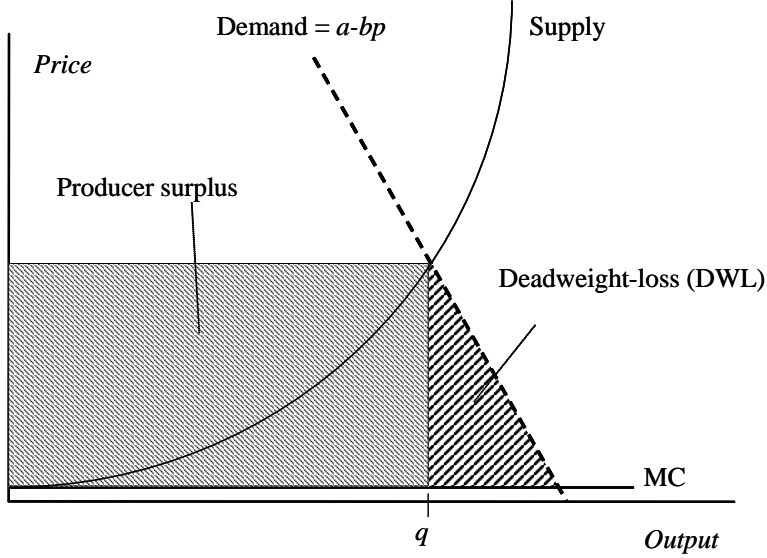[7] This is not immediate but is established in the appendix.

Figure 6: *Mark-ups and resulting deadweight loss (DWL).*

Hence, large cross-elasticities for consumption in adjacent delivery periods (e.g. if load can switch from high price to a following low price period) will not make bidding more competitive if the demand curves (and hence realised prices) in the two periods are identical.

As illustrated in figure 6, the mark-up results in under-production, which gives a deadweight loss, DWL. In the appendix we relate the mark-up $\frac{p-c}{c}$ to the relative DWL (i.e. expressed as a fraction of the short-run industry profits) ω, so that (2) can be written as follows:

$$x = 2N(N-1)^{N-1}\frac{\omega}{(N(1+2\omega)-2)^{N-1}}. \tag{3}$$

We note that the relative DWL ω is independent of the demand elasticity, γ, for constant marginal costs, although of course profits will depend on γ. It should also be noted that the expression is not valid when the capacity constraint binds for the welfare maximizing output, so that the welfare loss triangle is truncated by the capacity constraint. When output equals market capacity there are no welfare losses even if there are mark-ups in the market. Because of the truncation, the relation between welfare losses and mark-ups will be more complicated near the capacity constraint. We use numerically calculated symmetric SFE to consider this effect. In the numerical simulations we also consider increasing and more realistic marginal costs, when the relative deadweight losses depend on the demand elasticity. In these calculations we use production costs of the English Pool in 1988/89 (Green and Newbery, 1992). In figure 7 the relation in (3) is plotted as a function of the load factor for 5 and 10 firms comparing the constant cost case with those estimated for the English Pool. The short-run demand elasticity of
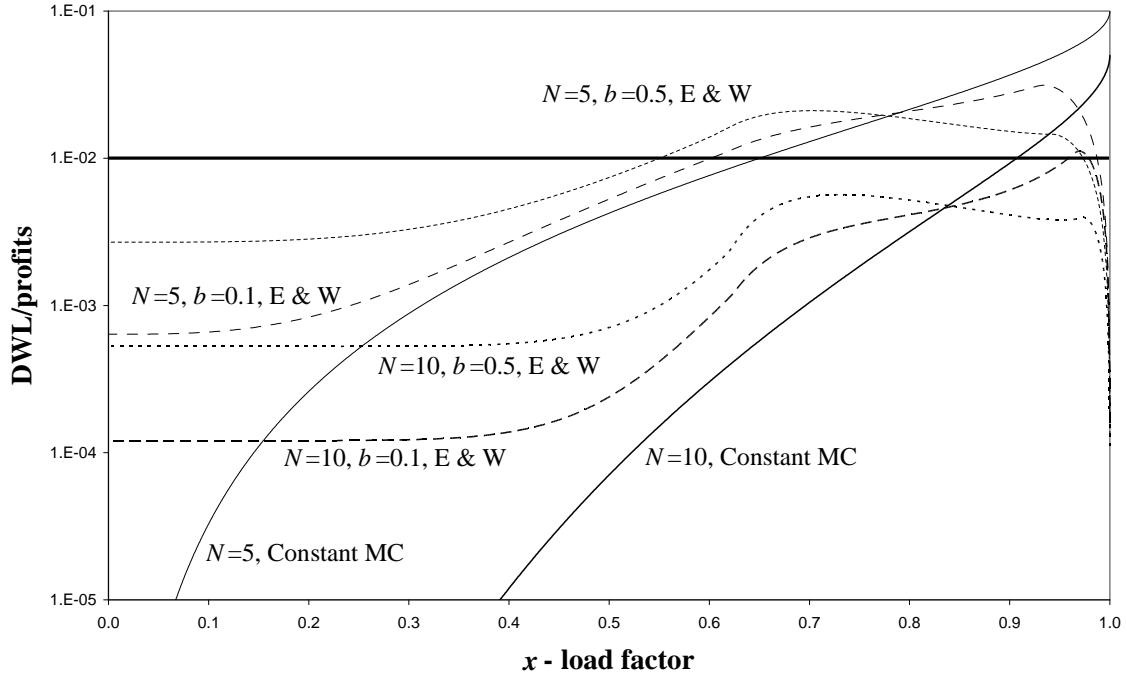
Figure 7: *Dependence of deadweight loss (DWL) on load factor and number of firms (N).*

electricity markets have been estimated to be in the range 0.05 to 0.4.[8] Thus the results for the English Pool have been calculated for a wide interval: $b = 0.1$ and $b = 0.5$ correspond to $\gamma = 0.08$ and $\gamma = 0.4$. We see that typically the relative DWL is increasing with the load factor. The reason is that the mark-up is normally convex in output. This convexity is less pronounced for increasing marginal costs (which makes supply functions steeper) and more elastic demand, when the slope of competitors' supply functions has relatively less influence on mark-ups in comparison to the slope of the exogenous demand, and the relative DWL curve is flatter in that case. In the cases studied here, five symmetric firms are enough to keep relative DWL below 1% if the load factor is below 50%. This corresponds to a Herfindahl-Hirschman index (HHI) of 2000.[9] Ten symmetric firms (HHI = 1000) are enough to keep relative DWL below 1%. Note that the graphs for constant marginal costs are calculated from (3), which disregards that welfare losses are truncated near the capacity constraint.

---

[8] von der Fehr et al (2005) estimate the consumer demand elasticity at 0.3 from the high price episode in Norway following hydro shortages, while Patrick and Wolak (1997) cite Borenstein and Bushnell's (1997) estimates from California of 0.1-0.4, and estimate the short-run demand elasticity by the water industry in England as 0.05-0.27 (depending on the time of day and hence price level). Other individual industries had lower elasticities.

[9] The HHI is defined as the sum of the squared percentages of market shares, so 10 firms with 10% each would have an HHI of 10x100=1000. See Tirole (1988) for more details.

In practice it is the average DWL that matters. Considering the load duration curve for the market in England and Wales, Green and Newbery (1992) estimate the average DWL relative to average short-run profits to 1.3% for five symmetric producers and $\gamma = 0.2$. With two producers, and otherwise unchanged market conditions, the average DWL relative to average profits is significantly higher, $8.5\% - 9.8\%$ (depending on the slope of the demand curve). Green and Newbery (1992) also calibrate their model to consider different entry scenarios in 1994. All entry is with high-efficient combined cycle gas turbines (CCGT), while the oligopoly is mostly unchanged for the more costly and price-setting coal-fired stations, although incumbents replace some of these stations by CCGT. With five symmetric producers, the market is sufficiently competitive to discourage entry. However, they estimate that there will be significant entry for a duopoly market with $\gamma = 0.2$. Even if the duopoly of the price-setting units is by and large unaffected by the entry, the output of these units is reduced, so that market shares for the two incumbents reduce to 26% each. This also reduces the strategic mark-ups and the average DWL relative to average profits is reduced to 4.9%. Green and Newbery (1992) show that duopoly with entry results in lower concentration and lower average DWL relative to average profits when producers offer below the upper supply schedule.

With asymmetric firms the equilibrium is calculated from a system of equations as in (1). As in the symmetric case a unique equilibrium is singled out if maximum demand is high enough to make the capacity constraints of all (but possibly one) firms bind. To model strategic bidding of 153 hydro-producers in Norway, Holmberg (2007) assumed constant marginal costs and asymmetric capacities. The result is that small firms with less capacity and less market-power offer their capacity with lower mark-ups compared to large firms. This implies that the capacity of small firms will bind at lower prices compared to large firms. Hence, many firms in the market compete for the marginal offer at low prices, but the number of competing firms decreases as the price increases. Thus the asymmetric market is more competitive for low demand shocks compared to a symmetric market with a similar market concentration index (HHI index), but less competitive for high demand shocks (see figure 8).

In a symmetric pure-strategy equilibrium of a market with symmetric producers, production is always efficient for a given total output, so that the total production cost of the market is minimized. This is also true for an asymmetric market with identical constant marginal costs. This is not true if mark-ups are asymmetric and marginal costs are increasing, because then there will be circumstances when costly production from a low mark-up producer will be accepted instead of less costly production from a high mark-up producer. Figure 9 shows an asymmetric SFE for five uncontracted strategic producers in the England and Wales market during 1999. The costs of the strategic producers were approximated by quadratic cost functions as described
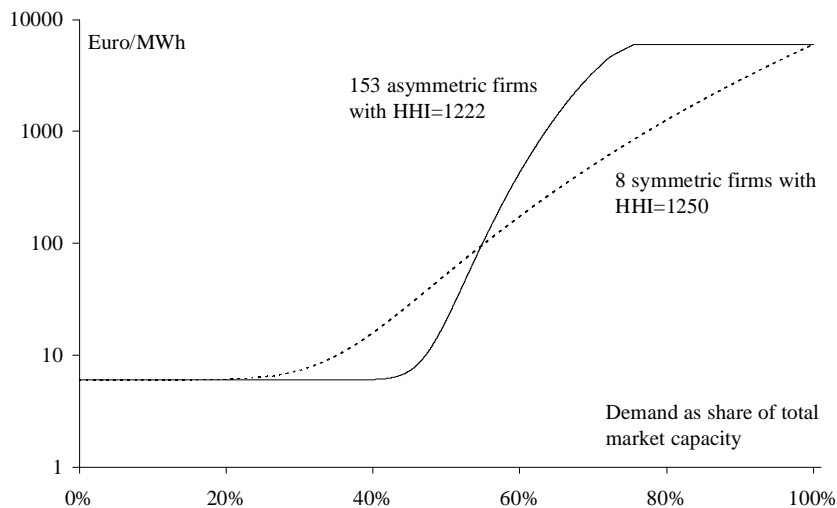
Figure 8: *Unique SFE of Norwegian real-time market compared with 8-firm symmetric SFE (Logarithmic price scale)*

in Green (1996) and Baldick et al. (2004). The nuclear power producers, BNFL and British Energy, were assumed to be non-strategic and the offers of their must-run plants were assumed to be inframarginal. The equilibrium was originally calculated by Anderson and Hu (2008a) and is replicated using an alternative numerical method by Holmberg (2009a). The market shares and market concentration index depend on the price; HHI varies in the range 2000-5000. To illustrate the production inefficiency we have highlighted the marginal costs of each producer at the price £30/MWh in Fig. 9.

We use the asymmetric SFE to calculate welfare loss relative to profits. The results are presented in Figure 10. We see that the calculated welfare loss relative to profits due to inefficient production is largest around the output 4 GW (4,000 MW) of the strategic producers. At this point National Power and PowerGen have significantly larger output and mark-ups compared to AES, Eastern and Edison Mission Energy that start production at a higher marginal cost. Apart from the peak at 4% the production inefficiency is fairly stable around 2% in the mid-range of the output. It is smaller for low and high outputs. Figure 10 also shows that the dead-weight loss caused by under-production is U-shaped. The curve is very steep for high outputs when the capacity constraints of AES and Edison Mission Energy starts to bind. The high relative dead-weight loss for low outputs can be explained by only National Power and PowerGen competing in the market for low outputs, which gives relatively high mark-ups. Moreover, as competition is increasing for higher outputs, the offer curves of National Power and Power Gen are locally concave for small outputs, see figure 9, which gives the relative welfare loss a decreasing shape.
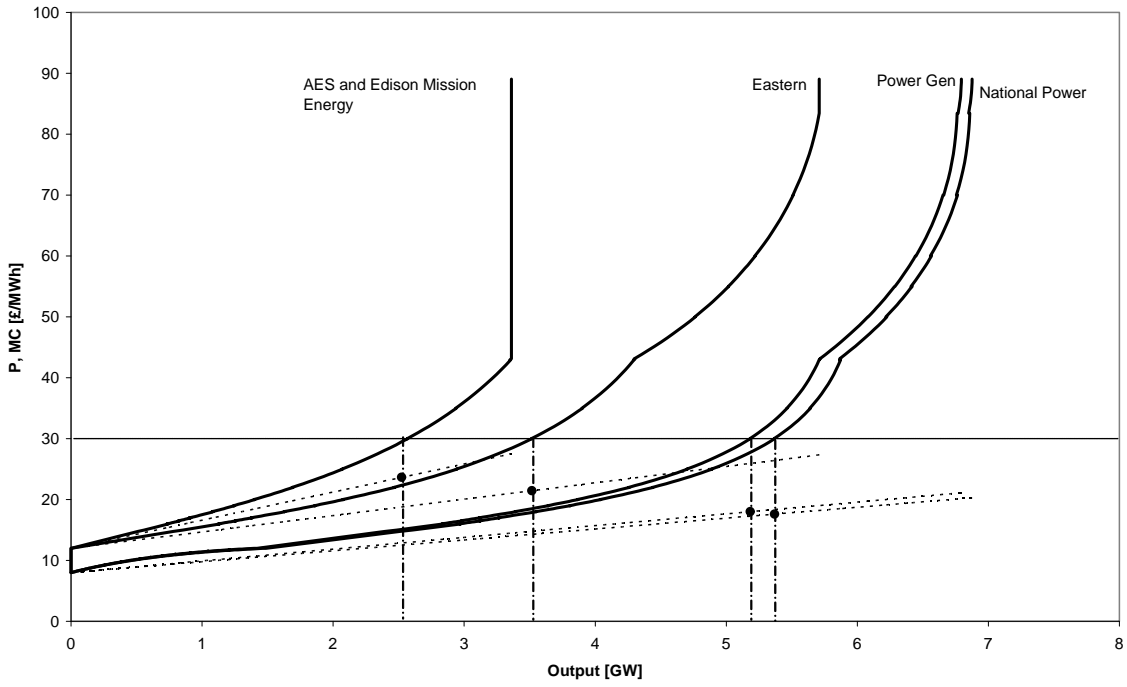
16

Figure 9: *Calculated SFs of strategic producers in England and Wales (1999) compared with their approximately linear marginal costs.*
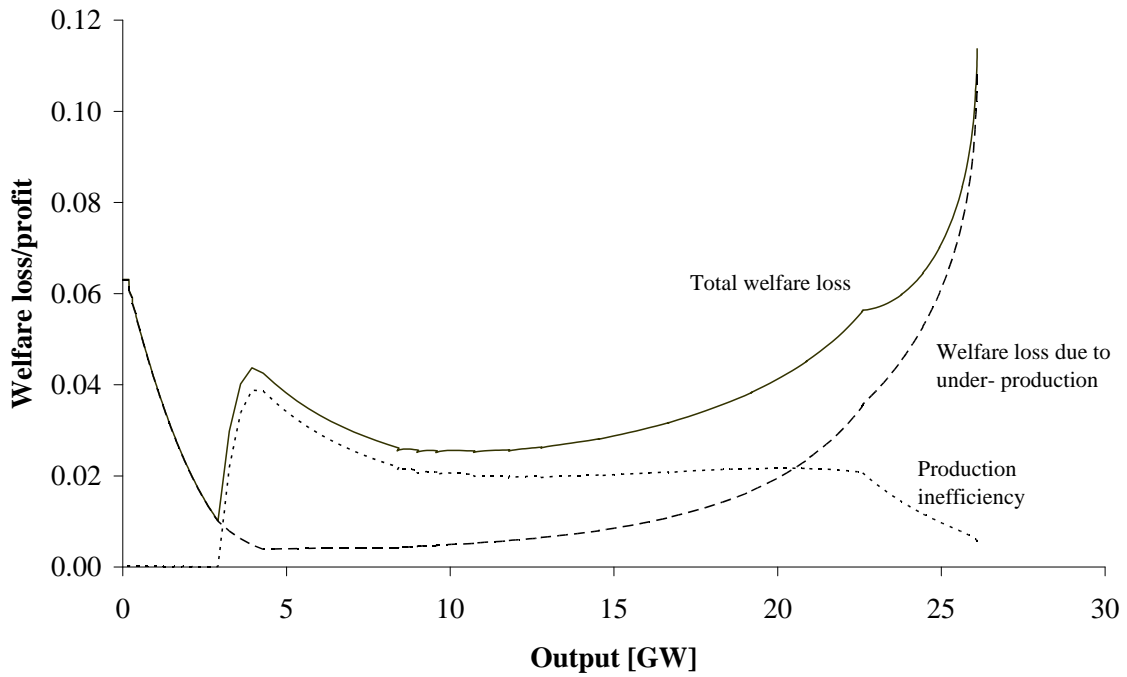

Figure 10: *Welfare loss in asymmetric model of the England and Wales market in 1999.*

In the asymmetric market with five strategic firms, all producers have similar production costs and capacities. Even so, the DWL due to inefficient production is of the same order of magnitude as the DWL due to underproduction. Thus one would expect the production inefficiencies to dominate if the asymmetry is larger, especially if the smallest firms, who have the lowest mark-ups, have a relatively large fraction of the stations with high production costs.

3.1 Forward contracting

It is well-known that forward sales mitigate the market power of electricity producers. This has been shown both empirically (Wolak, 2000; Bushnell et al., 2008) and theoretically (von der Fehr and Harbord, 1992; Newbery, 1998a; Green, 1999a). Chao and Wilson (2005) show that the mitigation can be even more successful with option contracts. Given the large potential market power that a low demand elasticity and concentrated market structures confer on incumbents in real-time, contracts are potentially very important. As generators in the world's restructured electricity markets tend to sell a large fraction of their output in the contract market (Green, 1999a; Bushnell, 2007) the effect in reducing deadweight loss is very welcome. But high contract cover is not true for all markets. For example, one of the alleged reasons for the failure of the California Power Exchange was the very limited use of forward contracting by market participants and a high reliance on spot market trading (Bushnell, 2004).

Mark-ups in the real-time market only influence the revenue from sales net of forward contracting. Hence, it is the residual demand net of forward contracts that are relevant for a profit maximizing producer. It is immediate to show (e.g. Anderson and Hu, 2008c; Holmberg, 2008b) that a producer's optimal supply curve is given by the natural generalisation of equation (1):

$$\frac{p - C'\left(q_i\left(p\right)\right)}{p} = \frac{1}{\gamma_i^{net-res}}.$$

where q_i is the producer's total output including forward sales and $\gamma_i^{net-res}$ is the elasticity of its residual net-demand. As before, a producer offers positive net-supply with positive mark-ups in the real-time market. If a producer has negative net-supply, i.e. it has to buy back electricity in the real-time market, then it will use its market-power to push down the price. Hence, mark-ups are negative for negative net-supply. Mark-ups are zero at the contracting point where net-supply is zero. These statements are formally proven by Newbery (1995, 1998a) and Anderson & Hu (2008c) and used to identify contract positions by Hortacsu and Puller (2008). Figure 11 shows qualitatively how the equilibrium in figure 4 is changed by forward contracting.

We note that adjusting contract cover closer to the expected output reduces mark-ups at that point, as mark-ups at the contracting point are zero. Hence, mark-ups would be small if all producers were almost fully contracted. To analyse the effect of the forward contracts, we revisit
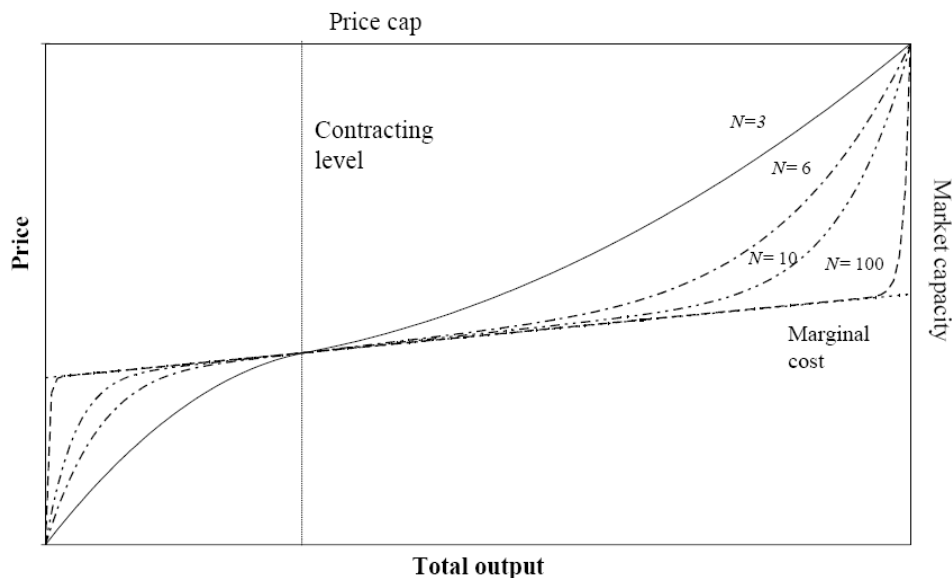
Figure 11: *Mark-ups are positive for outputs greater that contracts and negative for outputs less than contracts.*

the calculations in figure 7, but in figure 12 we now assume that half of the expected output is sold in the forward market.[10] Again it should be noted that the calculations for constant marginal costs are based on (3), which disregards that welfare losses are truncated near the capacity constraint.

Comparing 7 and figure 12 we see that the welfare losses are reduced by a factor 4-10 when half the expected output is sold in the forward market. The reduction of the welfare losses is likely to be even larger for most real electricity markets, where the contract cover is more like 80-100% (Green, 1999a; Sweeting, 2007). Hence, with significant forward contracting, sufficient competition in electricity markets can be maintained with only a few number of firms.

Obviously high contract cover reduces producers' market-power, so why do they sell in the forward market and how can market regulators stimulate such sales? One way is for regulators to mandate a sufficient volume of forward sales. According to Bushnell (2007) this is an important factor when explaining forward sales in many U.S. markets. This is becoming more common in Europe as well. For example, the European Commission has recently conditioned proposed mergers on that the merged electricity producers have to sell parts of their capacity in virtual

---

[10]Note that the amount of contracting is assumed to change for each delivery period, so that it is proportional to the expected output for every period. Thus a new SFE is calculated for each expected output level. The standard deviation in the demand is assumed to be small for each expected output level so that the forward price approximately equals the spot price for the expected output level.
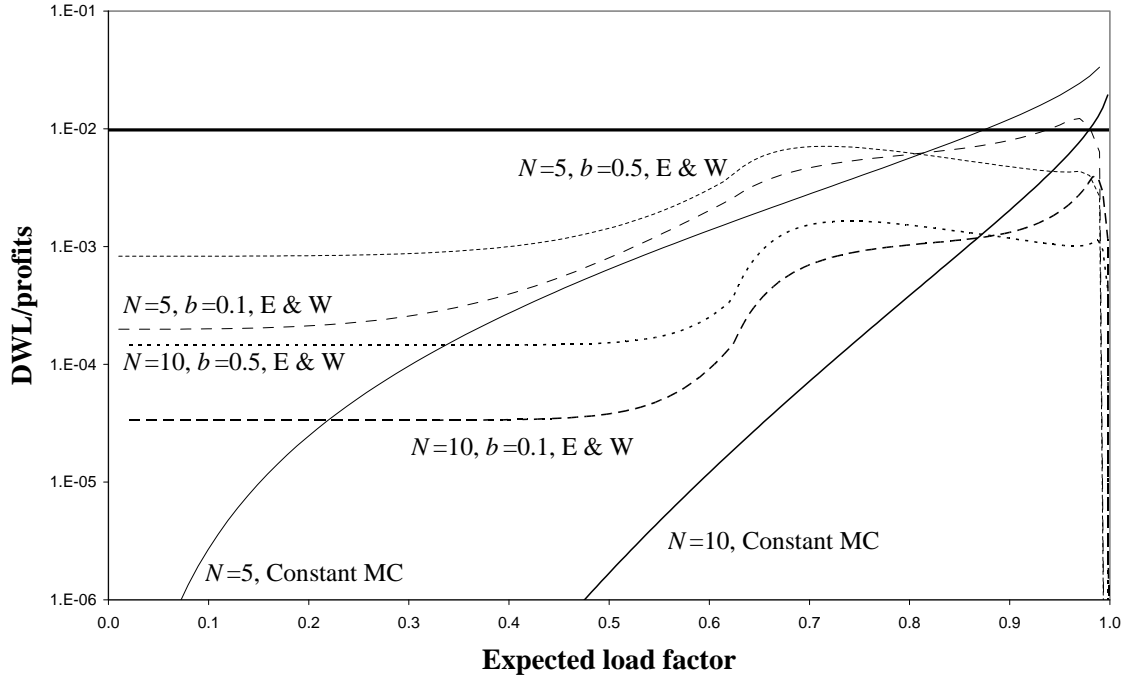
Figure 12: *Dependence of deadweight loss (DWL) on load factor and number of firms (N).*

power plant auctions. This temporary divestiture is equivalent to the sale of a package of forward contracts (Ofgem, 2009). It can also be in producers' own interest to sell in the forward market if they are risk-averse.

Forward contracting can also be driven by arbitrage opportunities. The risk in the electricity market is mainly non-systematic, i.e. the correlation between electricity prices and the stock market as a whole is weak. Hence, the Capital Asset Pricing Model (CAPM) would predict that the risk premium for risks carried in the electricity market would be small, because speculators can, at least in theory, completely eliminate non-systematic risks by holding a well-diversified portfolio (Hull, 1997). However, many empirical studies suggest that risk premia can be significant in electricity markets as in Longstaff's and Wang's (2004) study of the PJM market (U.S.). Kristiansen (2007) reaches a similar conclusion for Nord Pool, the electricity market of the Nordic countries. Anderson and Hu (2008b) use an SFE model to show that such a risk premium can arise when strategic retailers prefer to buy in the forward market in order to increase producer's forward sales, which lowers their mark-ups in the real-time market. Similarly, such a risk-premium can occur if consumers have to buy in the forward market, e.g. due to penalties on real-time imbalances. Green (1999b) motivates the risk premium by noting that retailer's have thinner margins and so are more risk-averse than producers. Given that much trading is

20

bilateral and not marked to market as in futures markets, the risks of bankruptcy are very real for retailers (and have certainly occurred in Britain, Texas and California) when there are sudden very large wholesale price movements, so retailer risk aversion is understandable.

But forward contracting in markets with imperfect competition is also driven by strategic mechanisms. Newbery (1998a) shows that producers may use contract sales to keep output high and spot prices low to deter entry. The risk of keeping prices high is that entrants will be attracted by the observation that the incumbents are able and willing to maintain profitability, but once they enter their capacity will remain to suppress prices for the next 20-40 years. Indeed, high prices encouraged a "dash for gas" and excess entry into the English market in the 1990s, and the incumbents appeared to have tacitly colluded to maintain prices while selling off plant, only for those prices to fall dramatically as the new owners, with smaller market shares, offered more aggressively and caused prices to collapse in 2000-01. Equilibrium in this oligopoly with free entry will be one in which the average price is equal to the long-run marginal cost (which includes the cost of investment) but the price is above the short-run marginal cost (SRMC), thus producing the short-run profits to cover the fixed costs.

Financial trading is anonymous in most power markets, and a firm's forward and futures positions are not revealed to competitors. Still, producers may be able to make rough estimates of changes in competitors' total futures positions by analysing changes in the turn-over in the forward market and the forward price. That competitors can deduce this information gives producers another strategic incentive, which was first illustrated by Allaz and Vila (1993). In their two-stage Cournot model, producers use forward contracts as a first-mover device. By selling to consumers in the forward market, a producer can shift consumer's real-time demand (net of forward contracts) inwards. If competitors observe or infer what is happening they would respond by shifting their optimal output inwards. For fixed contracting levels of competitors, this makes it possible for a producer to increase its market share and profit at competitors expense by selling in the forward market. However, all producers have the opportunity to sell in the forward market and they all have incentives to do so. Hence, instead of increased market-shares, strategic contracting just leads to more competitive bidding and lower profits for producers, which is good for welfare but bad for producers. In the limit as firms recontract (assuming they can observe earlier contract positions) the market becomes perfectly competitive. Actual markets have something between 3-8 rounds of contracting (as measured by the churn rate, Ofgem, 2009) but most of these subsequent rounds are between traders, retailers or generators adjusting their positions at the margin and do not amount to additional contracts modelled by Allaz and Vila.

Allaz and Vila's result appears attractive for the designers of electricity markets. Unfortunately the results are not robust to alternative specifications of competition. Murphy and

Smeers (2005) show in a multi-stage Cournot model that if capacity is chosen knowing that in the subsequent periods contracting will make competition more intense and profits lower, firms will invest less in order to keep capacity tight and prices higher in compensation, just as Kreps and Scheinkman (1983) showed that intense price (Bertrand) competition in the post investment period would lead to Cournot choices of capacity that would constrain output and support Cournot equilibria. Mahenc and Salanié (2004) show that strategic forward trading can have an anti-competitive effect if contracts are observable and producers compete in prices rather than quantities. The reason is that with Bertrand competition, competitors respond by reducing mark-ups if a producer commits to low mark-ups by making large forward sales. To avoid the aggressive response producers have incentives to buy in the forward market (negative contracting), which will push up mark-ups. While such squeezes are theoretically possible they are unusual (perhaps because they attract the attention of regulators).

Unlike the Bertrand and Cournot equilibrium, the SFE is *ex post* optimal to a range of additive demand shocks. A linear shift of the net-demand curve in the real-time market, due to increased forward sales, has no influence on competitors' optimal offers. To influence competitors' bidding in this model, the slopes of competitors' residual net-demand curves have to change. Moreover, they have to be able to predict these changes from observed changes in forward sales. Holmberg (2008b) shows that a strategic producer will use the forward market in order to commit itself to less elastic net-offer curves in the real-time market. This makes competitors' residual demand less elastic and their mark-ups will increase in accordance with (1). Thus competitors' output curves shift inwards, and the strategic producer can increase its market-share on their expense. Whether such a commitment of a strategic producer results from more contracting, which is pro-competitive, or less contracting, which is anti-competitive, depends on the curvature of the marginal cost and residual demand curves (Holmberg, 2008b). When the producer increases its forward sales, the marginal cost curve as a function of the net-supply (net of forward contracts) will shift inwards, and so will the net-supply curve. The inward shift of the net-supply curve is largest at points where the marginal cost curve is steep (as illustrated in figure 13) and at points where the mark-up is small. Hence, if marginal costs are convex and the residual demand is concave, selling in the forward market will shift the offer curve more inwards at high prices compared to low prices. This will make the producer's net-supply curve less elastic for all prices and competitors' optimal net-supply curves will become less competitive as well. However, the result is the opposite for concave marginal costs and convex residual demand, provided the producer's capacity is non-binding (pro-competitive effects cannot be ruled out for all prices if production capacities are binding). If both marginal costs and residual demand are linear (which requires that firms offer linear SFs as in the linear SFE model by Green, 1999a)
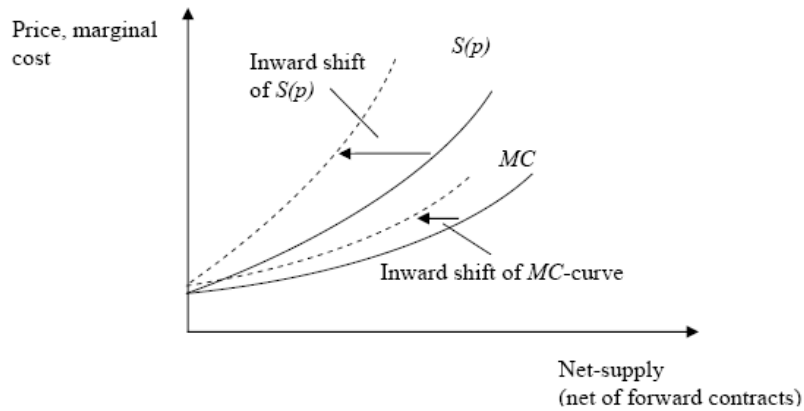
Figure 13: *Effect of an increase in forward sales for strictly convex marginal costs*

and the market capacity is non-binding, then strategies are neutral; competitors' observations of changes in a producer's forward sales do not influence competitors' bidding in the real-time market. Empirically, marginal costs and offer curves are typically convex. Convex offer curves result in convex residual demand if demand is convex or sufficiently inelastic. Hence, the Allaz and Vila effect is non-robust in electricity markets. Moreover, it is uncertain how strong this effect is as competitors' forward positions are not directly observable.

At the margin, increased forward sales of a producer can also result in a lower forward price and less forward sales for competitors (rather than increased consumer contracting). Green (1999a) and Holmberg (2008b) show that this makes competitors' bidding less competitive. This mechanism gives a producer incentives to increase its own forward sales, so that its output can be increased at competitors expense. But for given strategies of competitors, any producer has similar incentives. Hence, in equilibrium all producers will increase their forward sales, which makes the market more competitive. The pro-competitive effect increases with liquidity provided by competitors in the forward market. As competitors can always observe their own contracting level, observability is not an issue in this case.

3.2 Empirical results

The SFE model can be tested qualitatively and quantitatively against observed bidding behaviour. With many firms in the market, the SFE model predicts that electricity prices are near the marginal cost until the capacity constraint binds, where there is a steep increase in the price (see figure 4). This phenomenon is called "hockey-stick bidding", which has been observed in U.S. markets (Hurlbut et al., 2004). Figure 14 shows the same effect observed on a European power exchange, where the price-cost mark-up only becomes appreciable as the available supply falls
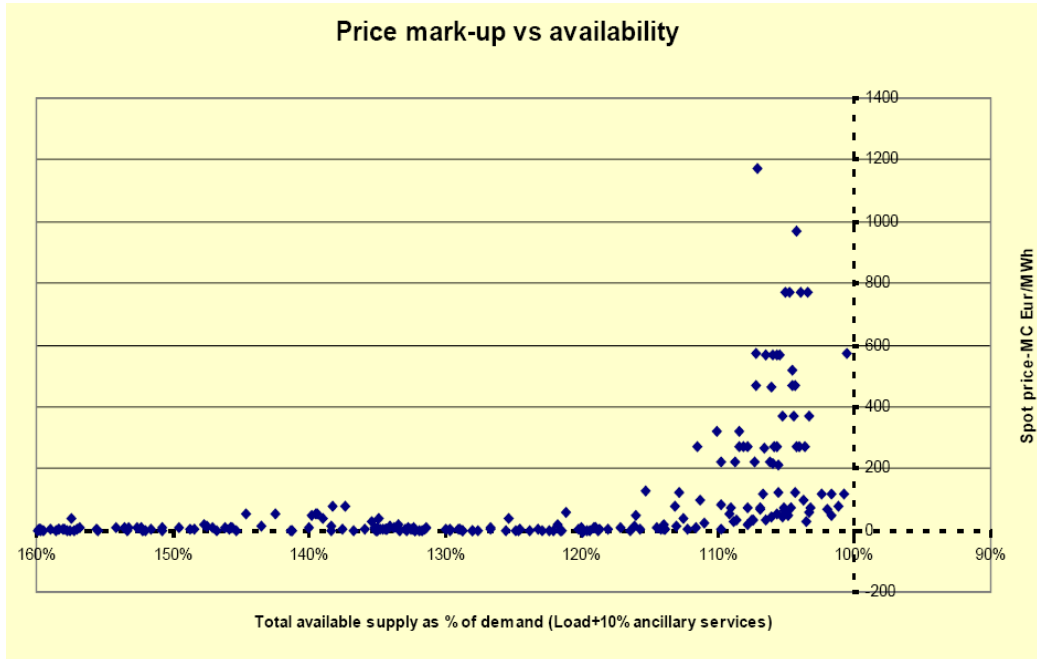
Figure 14: *Price-cost mark-up observed on a European power exchange*

to less than 110% of the estimated demand (so that the reserve margin falls below 10%).[11]

The SFE model also has empirical support in that two empirical studies of ERCOT (a balancing market in Texas) suggest that the offers of the two to three largest firms do indeed match the first-order condition of the continuous SFE (Hortacsu and Puller, 2008; Sioshansi and Oren, 2007). Niu et al. (2005) are also able to replicate observed market prices in this market using a model with linear SFE and fixed forward contracting. Willems et al. (2009) are able to replicate mark-ups in the German electricity market by using a SFE model with calibrated contracting cover. They also find a similar fit for a Cournot model if they allow for a higher contracting level.

There are other empirical approaches that test the underlying assumptions of the SFE model. Wolak (2003a) backs out the unobserved underlying cost and contract positions of generators bidding into the Australian market. He cannot reject the hypothesis that producers use pure and static strategies to maximize their expected profit with respect to an uncertain smoothed residual demand. Hence, his results indirectly support the SFE model.

The SFE is a static model and may not be a suitable model for predicting behaviour in markets where prices are driven by tacit collusion, although it can be used to test whether tacit collusion is occurring. Sweeting (2007) makes the assumption that producers' contract cover

---

[11] The EU Sector Inquiry gives many excellent examples of similar scatter diagrams, see London Economics (2007).

is 80% (or larger) and is able to characterize the various phases of market evolution and the exercise of market power in the English Electricity Pool. He concludes that the behaviour of the two largest generators was consistent with either tacit collusion or an attempt to raise the Pool prices, so that they could negotiate higher forward prices and sale prices of the plants they were divesting.

The SFE is primarily used to predict bidding behaviour, but the first-order conditions can also be used to back-out mark-ups from observed offer curves.Wolak (2003b) uses observed individual bid and offer curves to calculate the elasticity of the residual demand for five large producers in the California market. From these results he can estimate potential mark-ups using (1). On average the potential mark-ups were 15% in 2000, which was 3-5 times higher than in 1998 and 1999.

4 Regulation of electricity markets

4.1 Should offer curves be disclosed?

Some markets, e.g. in Britain,[12] New Zealand and Australia, disclose individual offer curves with some time delay. In this case, it is easy to calculate the residual demand elasticity of each producer at the clearing point as in Wolak (2003b). This greatly simplifies market surveillance for market regulators and competition authorities, as no information about production costs is needed to calculate potential mark-ups. The advantages are especially large for markets where production costs or opportunity costs are particularly difficult to estimate for outsiders, which is the case in hydro-dominated markets where opportunity costs are determined by the producers' prognoses of future inflows to the reservoirs and future electricity prices. A disadvantage with disclosing offer curves is that producers can use the disclosed individual offer curves to monitor competitors' signals and how well they follow an implicit or explicit collusive strategy. But this risk is reduced if only parts of the offer curves around the clearing point are disclosed, which provides sufficient information for the approach by Wolak (2003b), and if disclosure is significantly delayed, perhaps up to one year. The obvious solution is for individual offers/bids to be immediately made available to the regulator, but not the market, at least without a suitable delay.

Some markets (e.g. APX in figure 1) disclose aggregated supply and demand curves. This makes it possible for producers, who know their own offer curves, to exactly calculate their residual demand, so that they can monitor their competitors' aggregate offers. But this type

---

[12]In the Pool until it was abolished by the New Electricity Trading Arrangements in 2001, and thereafter in the Balancing Mechanism.

of disclosure does not give regulators the same possibility for market surveillance. Unless the disclosure of aggregated offers/bids are delayed by a long time, it seems that it facilitates tacit collusion without improving regulators' monitoring possibilities.

4.2 The price cap and investment

Most of the analysis so far has focused on short-run effects for given production capacities. But in the long-run investors need to recover their fixed costs in order to invest. The variation of demand over the day and annually means that systems with an optimal mix of production technologies will have low variable cost base-load plant, higher cost mid-merit plant, and peaking plant that has low capital costs but high variable costs. In a perfectly competitive market with zero mark-ups, base-load and mid-merit will still recover some of their fixed costs during peak demand periods where the price is set by peaking plants. But peaking plants always require some mark-ups to recover their marginal costs. Section 3 demonstrates that these mark-ups should only occur at maximum capacity if short-run social welfare losses are to be avoided. This is the case in perfectly competitive markets when peaking power plants cover their fixed costs during the rare events when there is an electricity shortage, so that the market is cleared at the price cap. For similar reasons the transmission charges in Britain are collected from consumers in the three peak half-hours (separated by ten days), efficiently collecting those fixed costs at the peak. Stoft (2002) demonstrates that risk-neutral producers will have the right investment incentives in perfectly competitive markets when the price cap is set equal to the Value of Lost Load (VOLL). But in an oligopoly with market power the fixed costs are not collected only at the peak, but over the whole range of outputs, and as demonstrated in Section 3 such mark-ups results in welfare losses.

An advantage with the regulation where peaking plants can recover their fixed costs because of a high price cap is that producers always want to offer all of their available production capacity to the market, because it is always better to offer capacity at the price cap than to withhold it, at least in a static equilibrium (Holmberg, 2008a). To perfectly hedge against uncertainties in their residual demand, producers can issue call options (or equivalently one-way contracts for differences) for each production unit with a quantity and strike price corresponding to the unit's capacity and marginal cost (Chao and Wilson, 2005). As suggested by Oren (2005), a market for call options is a natural way of implicitly introducing capacity payments (discussed below).

Some markets in Latin America have replaced the high price cap with explicit capacity payments and a lower price cap, p_{cap}. To get optimal investments in a competitive market, the administratively determined capacity payment should equal LOLP$*$(VOLL $- p_{cap}$), which ensures that producers in competitive markets are paid the same amount on average as in markets
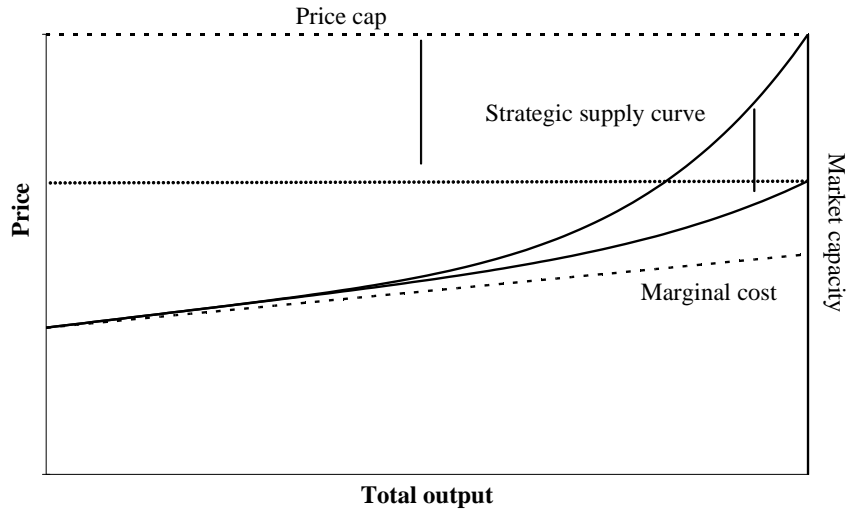
Figure 15: *Reducing the price cap pushes down the equilibrium price for every demand outcome.*

without capacity payments and the price cap equal to VOLL. (The former British pool had an implicitly determined capacity payment in which generators were paid LOLP∗(VOLL-SMP) where SMP was the system marginal price, with much the same effect). Capacity payments are determined on the basis of LOLP for every period, and although they will normally be very low, for some hours they can be very high when reserve margins are below a critical level. Capacity payments are independent of the bids (but increase exponentially as the reserve margin falls), so the SFE can still be used to calculate equilibria in such markets. Hence, an advantage with capacity payments is that producers will offer more competitively if the price cap is sufficiently low so that it binds for the highest offers (see figure 15), and the capacity payment will not compensate strategic producers for the reduction in their mark-ups. This result in lower average prices and reduces over-investment due to excessive entry.

However, a problem is that it might be profitable for producers to withhold production from the auction in order to increase LOLP and the capacity payments. This was a problem in England and Wales (Newbery, 1995) and led to a change in the Pool rules so that plant would not influence the calculated LOLP until 8 days after withdrawal. Another potential problem with capacity payments is that as with mark-ups they will result in welfare losses if they are collected from consumers for every delivery period. An advantage with capacity payments, and also to the issue of call options, is that it stabilizes risk-averse producers' incomes. A major problem with capacity payments is that even with detailed information, it is likely that there will be large errors in the calculated capacity payments, because it is very difficult for a system operator to correctly estimate LOLP (Newbery, 1998b). Partly for this reason PJM and other

American markets have instead imposed reserve requirements on distributors which results in market based capacity payments to producers. However, it is not necessarily easier for regulators to estimate welfare maximizing reserve requirements than correct values of LOLP.

A potential problem for investments is that producers may not trust market regulators to keep a high price cap and high capacity payments during a period of extended electricity shortages, as it would lead to very high electricity prices. Although home consumers can contract to reduce the adverse consequences of the resulting price spikes, there might be political pressure on regulators to lower the price cap or capacity payments, in order to push down electricity prices as illustrated in figure 15, even if such an irrational measure would exacerbate the shortage, especially in interconnected markets, where neighbouring markets with electricity shortage need to compete for reserve capacity (Stoft, 2002).[13] Hence, even if the price cap is set to VOLL, producers might under-invest due to the regulatory risk that the price cap might be lowered in the future. As with granting independence to central banks, it should be possible to reduce this potential time-inconsistency problem by making regulators of electricity markets more independent.

Entry and investments will be stimulated by high mark-ups. Mark-ups for production late in the merit-order are to a large extent determined by the price cap and LOLP. Hence, for a given price cap and LOLP, marginal investments in peaking power plants should be less dependent on the number of firms in the market compared to investments in off-peak production. Hence, compared to a competitive market with the same market capacity, one would expect more entry with off-peak power plants into a non-competitive market, which may be true for incumbents' investments as well.

4.3   Restrictions on the offer curves

The SFE outlined in Section 2 is an equilibrium for any demand uncertainty. The equilibrium is unique when no offer along the upper supply schedule is accepted or rejected with certainty. But if the demand uncertainty is bounded or if producers neglect very unlikely outcomes then the equilibrium offer curves may have sections that are either accepted or rejected with certainty. These sections are only price-setting for out-of-equilibrium events. Hence, producers can choose the shape of these sections without regard to the expected equilibrium profit and a wide-range of SFE can be supported under these circumstances (Klemperer and Meyer, 1989; Green and Newbery, 1992). Figure 16 shows how the lower section of an offer curve, where offers are always inframarginal (accepted with certainty), can be chosen to discourage competitors from

---

[13]The highest price spike in the English Pool occurred when the French market needed to import power instead of its normal export behaviour, and at the same time a British nuclear power station tripped.
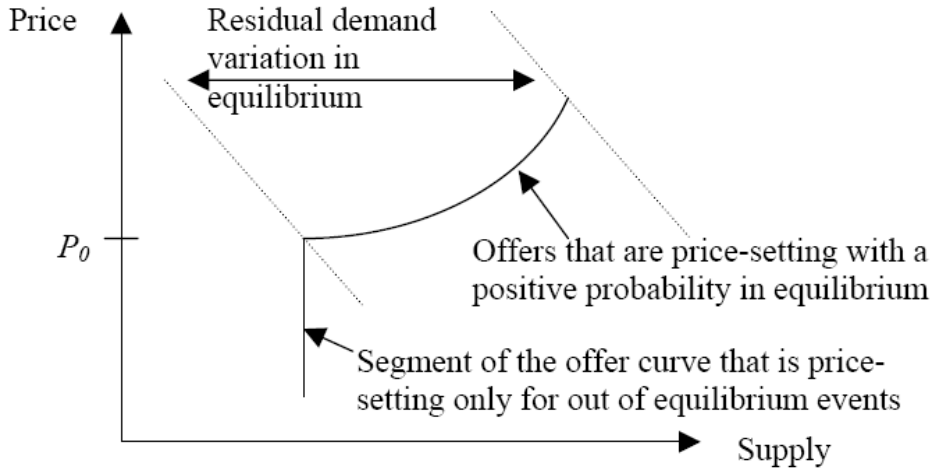
28

Figure 16: *Infra-marginal parts of an offer curve to discourage deviations*

undercutting the lowest realized equilibrium price, because in this case the output of the producer will remain fixed for such an out-of-equilibrium event. This is a costless threat that would support a less-competitive static equilibrium than would have been possible with a wider support of the demand uncertainty. In addition, sections that are only price-setting for out-of-equilibrium events can be used as costless signals to coordinate collusion (Klemperer, 2004).

A larger demand variation for a given offer curve widens the range of bids that are price-setting with a positive probability. This measure reduces the range of SFE that can be supported and the opportunities for costless signaling. In the pre-2001 pool of England and Wales, each producer had to stick to the same offer curve during the whole day. In the Australian market producers are restricted to choose ten price levels per production unit, which they must maintain during the whole day, although the quantities offered at the chosen prices are allowed to differ in each delivery period. It is clear that such restrictions on the offer curves can make bidding more competitive in one-shot games. But welfare is reduced if offers are too constrained, as this may lead to inefficient production if a unit's costs vary during the day.

Auction rules can also constrain the shape of the offer curves. The SFE model assumes that offers consist of smooth supply curves. However, in practice electricity auctions require offers to be stepped or piece-wise linear curves. For administrative reasons there are also quantity multiples, price tick sizes, and restrictions in the number of steps per bidder (see Table 1). von der Fehr and Harbord (1993) and Kremer and Nyborg (2004b) show that a large quantity multiple and small price tick-sizes will encourage bidders to undercut each other, as in a Bertrand game, which will result in a competitive equilibrium with zero mark-ups for non-pivotal producers

with constant marginal costs. But this mechanism does not work for demand outcomes where producers are pivotal[14] because he will find it profitable to deviate from such a competitive outcome. Von der Fehr and Harbord (1993) show that the combination of large quantity multiples and small price tick-sizes kills any pure-strategy equilibrium if the demand variation is sufficiently large and at least one producer is pivotal at the highest demand outcome. Instead, as with the Bertrand-Edgeworth model (Edgeworth, 1925), there will be mixed-strategy equilibria with randomized bidding.

Table 1: *Constraints on the supply functions in various electricity markets.*

Market	Max steps	Price range	Price tick size	Quantity multiple
Nord Pool spot	64 per bidder	0-5,000 NOK/MWh	0.1 NOK/MWh	0.1 MWh
ERCOT balancing	40 per bidder	-$1,000/MWh-$1,000/MWh	$0.01/MWh	0.01 MWh
PJM	10 per genset	0-$1,000/MWh	$0.01/MWh	0.01 MWh
UK (NETA)	5 per genset	-£9,999/MWh-£9,999/MWh	£0.01/MWh	0.001 MWh
Spain Intra-day market	5 per genset	Yearly cap on revenues	e0.01/MWh	0.1 MWh

Analogously, Anderson and Xu (2004) show that strategic producers in the Australian market design should randomise their choice of stated price levels. It is unclear whether the large quantity multiples assumed in the von der Fehr and Harbord (1993) model might result in lower average mark-ups for mixed-strategy equilibria, but it is clear that the market participants would bear the cost of uncertainty caused by the inherent price instability. Moreover, randomized bidding will cause production inefficiencies. For example, symmetric producers with strictly increasing marginal costs do not cause any production inefficiencies for a given market output in a symmetric pure-strategy equilibrium, but they do in a symmetric mixed-strategy equilibrium. On the other hand an advantage with mixing is that it widens the range of bids that are price-setting with a positive probability. This decreases the risk that out-of-equilibrium bids are used as costless signals or costless threats. Newbery (1998a) conjectures that the mixing may not be too severe if each producer is allowed to choose many steps in their offer curves. Parisio and Bosco (2003) show that pure-strategy equilibria can occur also in markets with large

---

[14]A producer is pivotal if competitors do not have enough capacity to meet market demand in his absence.

quantity multiples relative to the price tick size if there is sufficient uncertainty in competitors' cost functions.

Holmberg et al. (2008) show that if the market design instead has small quantity multiples relative to the price tick size and if the market has lax or no restrictions on the number of steps, then undercutting down to the marginal cost is avoided and the market has a pure-strategy equilibrium that converges to the smooth SFE as the number of price levels increases, also for pivotal producers. Hence, if market designers want to avoid inherent price instabilities, they can choose large price tick sizes, small quantity multiples, and lax restrictions on the number of steps per bidder. Under beneficial circumstances, Anderson and Hu (2008a) show that pure-strategy equilibria exist also for piece-wise linear supply functions, and that they converge to smooth SFE. Electricity auctions with piece-wise linear offer curves are used in Nord Pool (Nordic countries) and PowerNext (France). This seems to be the easiest way to avoid a market design with the inherent price instability caused by mixed strategies.

Rationing of excess supply at the clearing price is often necessary in multi-unit auctions and so market designs must specify how rationing will take place. This is normally by pro-rata on-the-margin rationing (Kremer and Nyborg, 2004a) in which only the incremental supply at the clearing price is rationed and the accepted share of each producer's incremental supply at this price is proportional to the size of this increment offered. Holmberg et al. (2008) note that this rationing rule has the advantage that profits in a one-shot game are maximised when supra-marginal offers (offers that are never accepted in equilibrium) are offered with a perfectly elastic segment along the highest realised equilibrium price. This strategy maximizes the size of a producer's incremental supply and accepted output at the highest shock outcome. Hence, pro-rata on-the-margin rationing will make it costly to use supra-marginal offers for signalling and threats in multi-unit auctions. This form of supra-marginal offers is also beneficial in supporting the most competitive SFE consistent with producers' capacity constraints.

4.4   Pay-as-bid auctions

The pay-as-bid auction is used in most treasury auctions around the world, but it has so far been less popular in electricity markets. One exception is England & Wales, which switched to such a format in 2001 for the balancing mechanism (the only remaining auction "market" as the day-ahead market is effectively an OTC market). Italy seems to become another exception, a recent law dated January 2009 says that their day-ahead market should switch to a pay-as-bid format.

If producers would offer the same curves in pay-as-bid auctions as they use in uniform-price auctions, then average prices in electricity markets would be significantly lower with the pay-as-
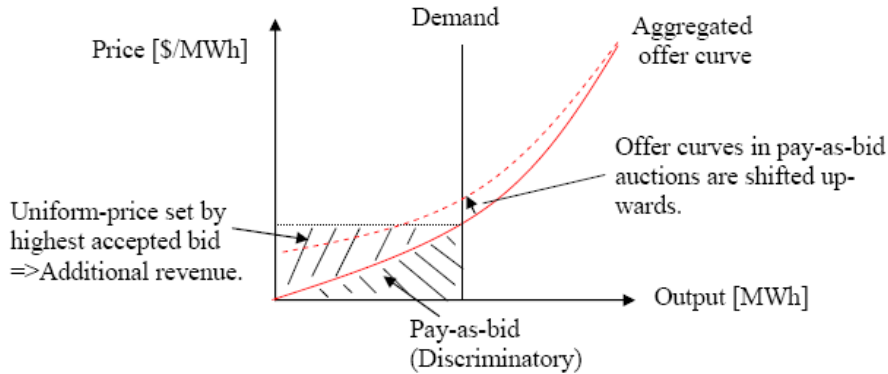
Figure 17: *Producer surplus in uniform-price and pay-as-bid auctions.*

bid format as illustrated in figure 17. However, as pointed out by Kahn et al. (2001), Wolfram (1999) and others, this naïve assumption is not reasonable, as strategic producers will change their optimal bidding strategies when the auction format changes.[15] Moreover, theoretical and empirical results applicable to treasury auctions show that rankings of the auctioneer's revenue are ambiguous for the two formats (Ausubel and Cramton, 2002). In this auction buyers have common/affiliated uncertain values given by the value of the security in the secondary market. Hence, these results should be applicable to day-ahead auctions in electricity markets, where producers' opportunity costs of selling in a later market are common/affiliated.

Electricity prices went down in Britain after their real-time market switched to a pay-as-bid format. But Evans and Green (2005) use a SFE model to control for investments and market concentration in Britain and they conclude that low electricity prices after the introduction of pay-as-bid pricing in 2001 can be explained by added capacity as well as forced and voluntary divestitures rather than the changed auction format. An experiment by Rassenti et al. (2003) even suggests that average prices might be higher in pay-as-bid auctions, at least if demand is certain.

Game-theoretic models of real-time markets with static strategies, however, have so far concluded that short-run average prices are lower with the pay-as-bid format, at least if contracting is neglected. Fabra et al. (2006) came to this conclusion using an auction model, where producers choose one price for all of their capacity, as did Son et al. (2004) with a similar model. Holmberg (2009b) reaches similar conclusions for an SFE model. His results suggest that switching from a uniform-price to a pay-as-bid format will be most beneficial for the auctioneer/consumers when the risk of power shortage is neither extremely small nor extremely large and when competition

---

[15]Indeed, in single object auctions the Revenue Equivalence theorem implies that payments to producers will be the same under either format. See e.g. Klemperer (2004).

is imperfect without being extremely bad. Under perfect competition producers' expected profits are the same regardless of the auction format. Wang and Zender (2002) use a model of strategic bidding, which is essentially a SFE model, to compare auction formats in treasury auctions. They also conclude that the auctioneer would prefer the pay-as-bid format.

These theoretical results were all derived for oligopoly markets with inelastic demand and a fixed price cap. Elastic demand in pay-as-bid auctions creates some difficulties. As with Bertrand models (Maskin, 1986), the question is how much of the excess demand at lower offered prices will remain at higher prices. Federico and Rahman (2003) assume that demand depends on the highest accepted offer in the market though much of the demand is met at lower prices, because of the pay-as-bid mechanism. The corresponding assumption in Bertrand models is called parallel rationing (Maskin, 1986), which is equivalent to the assertion that demand does not depend on income (Levitan and Shubik, 1972). The result of Federico and Rahman's (2003) assumption is that the non-strategic demand side of their model is the same in the pay-as-bid and uniform-price format, which simplifies their analysis. Using an SFE model, Federico and Rahman (2003) show that consumer surplus is larger in pay-as-bid auctions compared to uniform-price auctions for both monopolistic and perfectly competitive producers. However, output and total welfare are sometimes lower in the pay-as-bid format and these models do not discuss long-run equilibrium in which capacity is endogenous.

Under suitable conditions there will be a pure strategy NE in pay-as-bid auctions, where each producer chooses a smooth optimal supply curve given the properties of its residual demand. But compared with uniform-price auctions, it is normally harder to derive an SFE for pay-as-bid auctions. *Ex post* it is always optimal to offer all accepted bids at the same price, the marginal bid. Hence, *ex post* optimality for a range of demand shocks in a pay-as-bid auction would require horizontal offers. As with the Bertrand NE, pure-strategy equilibria with horizontal offers are possible if capacity constraints are non-binding and marginal costs are constant (Wang and Zender, 2002). But this is an exception, typically equilibrium offers to the pay-as-bid auction are not *ex post* optimal. Hence, as with the generalisations of the SFE model discussed in Section 2.1, optimal offers will normally depend on the probability distribution of the demand shock. The pay-as-bid format also implies that the offer price of one unit does not influence the profit from the producer's other units, as long as the units are offered in merit-order. Hence, the expected profit from each unit can be optimized independently. Rather than choosing the offer curve such that the profit is maximized for each shock outcome, which is the case in the uniform-price format, producers in the pay-as-bid format choose an offer curve so that the expected profit from each production unit is maximized (given the properties of the residual demand).

Let $F(\varepsilon)$ be the probability distribution of the demand shock and $f(\varepsilon)$ the density function

of this shock. Assume that a unit q^* of the producer is offered at the price p^* and that this offer is marginal for the shock outcome ε^*. Then the optimal offer of a production unit is given by the following relation (Holmberg, 2009b, Anderson et al., 2009):

$$\left[p - C'\left(q^*\right)\right] f\left(\varepsilon^*\right) \Delta\varepsilon - \left[1 - F\left(\varepsilon^*\right)\right] \underbrace{\frac{\Delta\varepsilon}{\left|D'_{res}\left(p^*\right)\right|}}_{\Delta p} = 0. \tag{4}$$

The first term is the expected marginal gain from the unit q^* if its offer is lowered, so that the range of demand shocks for which the offer is accepted increases from ε^* to $\varepsilon^* + \Delta\varepsilon$. Hence, the probability that the bid of unit q^* is accepted increases by $f(\varepsilon^*)\Delta\varepsilon$, which considering the unit's mark-up, gives the expected marginal gain $(p - C\prime(q^*))f(\varepsilon^*)\Delta\varepsilon$. The second term is the loss associated with increased sales. To achieve this the offer price is decreased by the amount $\Delta p = \frac{\Delta\varepsilon}{\left|D'_{res}(p^*)\right|}$, which results in the marginal loss $\left[1 - F\left(\varepsilon^*\right)\right] \frac{\Delta\varepsilon}{\left|D'_{res}(p^*)\right|}$, where $1 - F\left(\varepsilon^*\right)$ is the probability that the offer of unit q^* is accepted. The optimal offer price p^* is determined by the requirement that the marginal gain and marginal loss exactly balance each other. Moreover, for a profit maximum, we have the necessary second-order condition that the marginal loss must increase faster with respect to ε^* compared to the marginal gain. In contrast to the second-order condition for uniform-price auctions, this is quite a restrictive condition. It is straightforward to show by means of (4) that this will require the hazard rate of the demand shock $\frac{f(\varepsilon)}{1-F(\varepsilon)}$ to be decreasing with respect to ε at points where the marginal cost is locally constant (Holmberg, 2009b, Anderson et al., 2009). This restriction on the probability distribution rules out many demand shock distributions that one would encounter in practice, including the normal distribution. An exception, where pure-strategy SFE exists also for locally constant marginal costs is the Pareto distribution of the second-kind, which Holmberg (2009b) uses in his calculations. Pure-strategy SFE exists for a broader class of demand distributions if marginal costs are increasing and mark-ups are sufficiently low.

In uniform-price auctions it is optimal to offer the first unit at marginal cost, also for strategic producers (see figure 4). In pay-as-bid auctions such bidding would result in zero profits from the first unit and they are therefore generally offered with a mark-up. Hence, for low demand outcomes prices are generally higher in pay-as-bid auctions. If demand is sufficiently inelastic so that the highest offer is determined by the price cap, then the highest offer in both auction formats will be equal to the price cap. In uniform-price auctions the highest offer is price-setting for the whole market, so all producers get this price. In pay-as-bid auctions on the other hand, the average of the price paid to accepted offers (the system buy price) is generally lower than the highest offer. Hence, for high demand outcomes, the system buy price is lower than in uniform-price auctions. In summary, the price variation (the difference between the lowest and highest

High mark-ups=>
Horizontal mixture

Medium mark-ups =>
Hockey-stick mixture

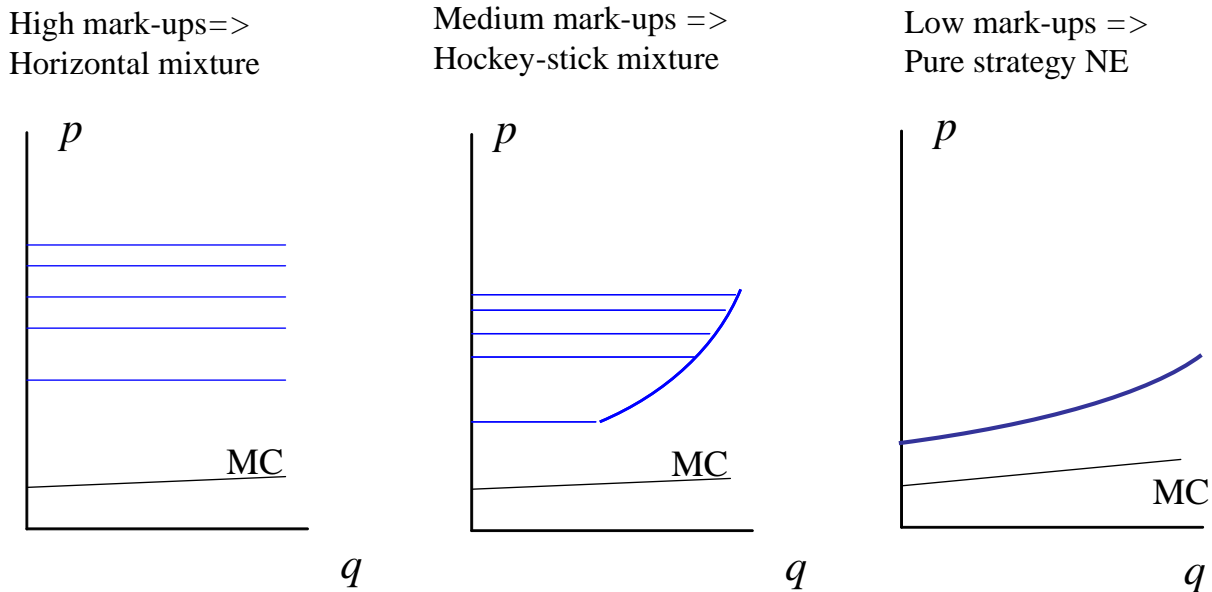Low mark-ups =>
Pure strategy NE

Figure 18: *Mixed-strategy NE for high and medium mark-ups in pay-as-bid auctions but a pure-strategy NE for low mark-ups.*

system buy price) is lower in pay-as-bid auctions when there is a pure-strategy equilibrium. This has also been confirmed in the experiments by Rassenti et al. (2003).

There is, however, a considerable risk that the second-order condition is not satisfied in balancing markets. Garcia and Kirschen (2006) find that system imbalances in Britain are approximately normally distributed. Given the central limit theorem, this is not surprising as the system imbalance results from a large number of actions by different market participants acting independently. Second, marginal electricity production costs are roughly stepped, i.e. approximately locally constant. Hence, one can expect that offer curves submitted to electricity auctions with pay-as-bid formats are at least partly randomized. Genc (2009) shows that mixtures over horizontal offers occur in discriminatory divisible-good auctions for cases with constant marginal costs and uniformly distributed demand (see left graph of figure 18). His equilibrium is unchanged if producers are restricted to choose one price for their whole capacity as in the auction model by Fabra et al. (2006). These strategic mixtures are essentially Bertrand-Edgeworth equilibria (Edgeworth, 1925), but with uncertain demand. Fabra et al.'s (2006) comparison with mixed-strategy equilibria in uniform-price auctions indicates that the auctioneer prefers the pay-as-bid format also when producers use mixed strategies. Anderson et al. (2009) generalize the results by Genc (2009) by considering mixtures for increasing marginal costs. In this case, mixed strategy equilibria will cause production inefficiencies. With increasing marginal costs and a sufficiently low price cap there will be new types of mixed strategy equilibria where some of the

offer curves have strictly increasing parts (see middle graph of figure 18).

As a final remark on the game-theoretic analysis we must caution that possible lower short-run prices in the pay-as-bid auction may not be true in a long-run equilibrium in which investors need to earn a normal return on their investment; they may under-invest in market designs with discriminatory pricing, leading to higher long-run average prices.

An advantage with pay-as-bid auctions compared to uniform-price auctions is that all infra-marginal bids are price-setting. Hence, the risk that out of equilibrium bids are used as costless signals or costless threats is lower in pay-as-bid auctions compared to uniform-price auctions (Klemperer, 2004). Fabra (2003) and Klemperer (2004) show that the risk for tacit collusion is lower in pay-as-bid auctions compared to uniform-price auctions.

The pay-as-bid format is disadvantageous for must-run plants that have to bid sufficiently low that their bids are surely accepted. But this is less of a problem if they can sell most of their production in forward markets or if the System Operator invites longer-term tenders for such plant. To make the design of real-time markets more favourable to must-run plants one can also allow for non-competitive offers (Klemperer, 2004). These offers are always accepted and they are paid the system buy price. Non-competitive bids are often allowed in treasury auctions. Similar solutions seem sensible for small electricity producers, because producers need good information of the properties of the residual demand to bid intelligently (Klemperer, 2004). Otherwise pay-as-bid auctions might discourage potential bidders who have only small amounts to trade and for whom the cost of obtaining market information might not be worth paying. For the same reason, a major concern is that the pay-as-bid format might deter potential entrants.

One issue with pay-as-bid auctions is that they have multiple prices, so there is no obvious strike price that can be used to clear financial contracts. Before 1991 the electricity Pool of England and Wales acted as both a day-ahead market and balancing market with a well-defined strike price on which contracts-for-differences could be written. With the abolition of this gross pool (into which all generators had to offer) under the New Electricity Trading Arrangements, the day-ahead market became a small (2-3%) residual market and the balancing mechanism was separated to become a real-time market, operated as a pay-as-bid auction with no single strike price. The extra risk created by the balancing mechanism arguably encouraged generators and supply businesses to vertically integrate to reduce wholesale and balancing risks, which considerably reduced liquidity in forward markets. According to Ofgem (2009) liquidity is only one-third that in Germany and the Nordic countries. Another concern with the multiple prices associated with pay-as-bid auctions is that they may cause problems for the system operator when calculating optimal dispatches in a transmission-constrained system with locational pricing. In such a system it will also become less obvious how a strike price should be calculated for each

node or price zone. Even if the real-time market has a pay-as-bid format, it is still possible to trade day-ahead in a uniform-price auction, provided there is adequate liquidity, and in that case it might still be possible for a uniform-price day-ahead market to provide strike prices for financial contracts (such as contracts-for-differences). Proponents of nodal pricing for Britain (which are seen as increasingly necessary to handle large intermittent wind generation) also argue that a return to central dispatch and a gross pool would be needed to create the necessary pricing environment for liquid hedging contracts to re-emerge.

5   Conclusions

The supply function model describes optimal bidding in wholesale electricity auctions where demand varies and/or is uncertain. It is a static model, in that producers choose their offers curves to maximize profits at each moment (hour or half-hour) given their residual demand at that moment, without regard to bidding in future periods. Empirical studies of bidding in electricity markets by and large support this model and its underlying assumptions, i.e. ex-post (after uncertainties have been realised) producers maximize their profits for each period given their smoothed residual demand curve. But there are exceptions. For example, Sweeting (2007) shows that producers coordinated to offer curves higher than the SFE model would have predicted during a period when producers were divesting plants, in effect tacitly colluding to keep prices higher and thus maintain the appearance of a more profitable market, leading to higher sales values for their divested plant. After divestiture, margins rapidly collapsed (Newbery, 2005).

This paper provides new results that highlight how short-run welfare losses depend on the number of firms in the market and their asymmetry. We show that market integration reduces mark-ups when demand shocks in the integrated markets are not perfectly correlated or when different producers operate on the integrated markets. Mark-ups depend very much on the elasticity of the demand, but we show that social welfare relative to profits is less sensitive to this elasticity. In this case it is mainly the load factor and the number of producers that matters. Without forward contracts, 5-10 symmetric producers are needed to keep the average welfare loss relative to the average industry profit below 1%. This corresponds to a market concentration index (HHI) of 1000-2000. With asymmetric firms, lower market concentration indices are needed to keep relative average welfare losses below 1% due to the production inefficiencies caused by asymmetric mark-ups. These inefficiencies increase with the asymmetry in the market. More research is needed to analyse long-run welfare losses resulting from inefficient investments.

Forward contracting reduces strategic producers' mark-up incentives. If producers sell half their expected output in the forward market, relative welfare losses decrease by a factor 4-10. This is of the same order of magnitude as if the number of producers would have been

doubled. The question is what incentives producers have to sell in the forward market when this decreases their market-power. Sometimes forward sales are forced upon large producers. For example, competition authorities can order them to sell parts of their capacity in virtual power plant auctions. Introducing penalty charges for deviations from contracted power give consumers incentives to buy their expected demand in the day-ahead and other forward markets. Producers' forward sales are also driven by their risk aversion and by arbitrage opportunities. As retailers have very thin margins they tend to be more risk-averse, which pushes up forward prices, encouraging producers to offer forward sales.

Forward sales can also be driven by strategic contracting, i.e. the producer wants to increase its market-share at competitors' expense. This mechanism has pro-competitive effects when competitors are marginal buyers in the forward market. Results are less robust when consumers are marginal buyers in the forward market. A market designer can keep this latter effect weak by introducing penalty charges for deviations from contracted power and by not disclosing producers' forward positions to the market.

Requiring individual offer curves to be disclosed to regulators makes it straightforward to monitor potential mark-ups and market manipulation. This is a great advantage especially for hydro-dominated markets, which are very complicated to monitor, as production decisions are determined by a prognosis of future inflows to the reservoirs and future electricity prices. Public disclosure should be delayed, perhaps by a year, so that it does not facilitate collusion. To reduce signalling opportunities, it may also be beneficial to only disclose parts of the offer curves around the clearing point, which is enough to monitor potential mark-ups.

Combining capacity payments with a low price cap reduces the risk for investors and makes bidding more competitive. However, one has to make sure that this does not provide incentives to withhold power from the auction, and it is difficult to calculate the loss of load probability accurately, which capacity payments are based on. Moreover, capacity payments result in welfare losses if they are collected from consumers every delivery period. If one wants to minimize welfare losses it makes more sense to collect the compensation for investors' fixed costs when the capacity constraint binds, which is the case in perfectly competitive markets without capacity payments and a high price cap.

Restrictions on the offer curves and long-lived offers decrease the possibilities for tacit collusion, but such restrictions may cause production inefficiencies. Large quantity multiples are useful to ensure competitive bidding when no producer is pivotal, but with pivotal producers, such a design will cause inherent price instabilities due to randomized bidding. These cause production inefficiencies and unnecessary uncertainty for market participants.

For given production capacities theoretical results indicate that average electricity prices

are lower if real-time markets use the pay-as-bid format instead of the uniform-price format. However, empirically the effect is not significant and for results applicable to day-ahead markets both empirical and theoretical comparisons of the two formats are ambiguous. One advantage with the pay-as-bid auction is that the risk for collusion is smaller, at least in theory. On the other hand, there is a considerable risk that pay-as-bid clearing results in inherent price instabilities. Moreover, detailed market prognoses are required to bid in pay-as-bid auctions, which disadvantages small producers, deters entry and in any case may have no beneficial long-run effect once investment incentives are considered.

References

Allaz, B. and J-L Vila (1993), "Cournot competition, forward markets and efficiency", *Journal of Economic Theory*, Vol. 59, pp. 1-16.

Anderson, E.J. and A. B. Philpott (2002a). 'Using supply functions for offering generation into an electricity market', *Operations Research* 50 (3), pp. 477-489.

Anderson, E. J., A. B. Philpott. (2002b). Optimal offer construction in electricity markets. *Mathematics of Operations Research* 27 82–100.

Anderson, E. J. and H. Xu (2004). 'Nash equilibria in electricity markets with discrete prices', *Mathematical Methods of Operations Research* 60, pp. 215–238.

Anderson, E.J. and H. Xu (2005), "Supply Function Equilibrium in Electricity Spot Markets with Contracts and Price Caps", *Journal of Optimization Theory and Applications*, Vol. 124, No. 2, pp. 257–283.

Anderson, E. J. and X. Hu (2008a). 'Finding Supply Function Equilibria with Asymmetric Firms', *Operations Research* 56(3), pp. 697-711.

Anderson, E.J., and X. Hu (2008b), "Forward contracts and market power in an electricity market", *International Journal of Industrial Organization*, Vol. 26, 3, May 2008, pp. 679-694.

Anderson, E.J. and X. Hu (2008c), Asymmetric Supply Function Equilibria with Forward Contracts, Presented at Auckland University, 2008.

Anderson, E. J., P. Holmberg, A. B. Philpott (2009), Mixed strategies in discriminatory divisible-good auctions, Mimeo, University of Sydney.

Aromí, J.D. (2007), Asymmetric Supply Function Equilibrium with Applications to Investment Decisions in the Electricity Industry, Mimeo, University of Maryland.

Ausubel, L. and P. Cramton (2002), Demand Reduction and Inefficiency in Multi-Unit Auctions, Working Paper No. 96-07, Department of Economics, University of Maryland. Revised 2002.

Baldick, R., and W. Hogan (2002), Capacity constrained supply function equilibrium mod-

els for electricity markets: Stability, non-decreasing constraints, and function space iterations, POWER Paper PWP-089, University of California Energy Institute.

Baldick, R., R. Grant, and E. Kahn (2004). 'Theory and Application of Linear Supply Function Equilibrium in Electricity Markets', *Journal of Regulatory Economics* 25 (2), pp. 143-67.

Bartolini, Leonardo and Carlo Cottarelli (1997), "Treasury Bill Auctions: Issues and Uses," in Mario I. Blejer and Teresa Ter-Minassian, eds., *Macroeconomic Dimensions of Public Finance: Essays in Honour of Vito Tanzi*, London: Routledge, 1997, pp. 267–336.

Bolle, F. (1992), "Supply function equilibria and the danger of tacit collusion. The case of spot markets for electricity", *Energy Economics*, Vol. 14, pp. 94-102.

Borenstein, S., J. Bushnell, and C. Knittel (1999), "Market Power in Electricity Markets, Beyond Concentration Measures", *The Energy Journal*, Vol. 20, No. 4, pp. 65-88.

Bushnell, J. (2004), "California's Electricity Crisis: a market apart?" *Energy Policy*, Vol. 32, pp. 1045-1052.

Bushnell, J. (2007), "Oligopoly Equilibria in Electricity Contract Markets", *Journal of Regulatory Economics*, Vol. 32, No. 3, pp. 225–245.

Chao, H. and Wilson, R. (2005), Implementation of Resource Adequacy Requirements via Option Contracts, EPRI, Palo Alto, CA and EdF: 2005 1010712.

Edgeworth, F. (1925), Papers Relating to Political Economy, London: Macmillan.

Edin, K-A. (2007), Hydro-Thermal Auction Markets, Mimeo, Tentum, Stockholm.

Evans, J. and R. Green (2005), Why did British electricity prices fall after 1998?, Working Paper, Birmingham University, UK.

Fabra, N. (2003), 'Tacit collusion in repeated auctions: uniform versus discriminatory', *The Journal of Industrial Economics*, Vol. LI, No. 3, pp. 271-293.

Fabra, Natalia, Nils-Henrik M. von der Fehr and David Harbord, (2006) 'Designing Electricity Auctions,' *RAND Journal of Economics* 37 (1), pp. 23-46.

Federico, G., and D. Rahman (2003), 'Bidding in an electricity pay-as-bid auction', *Journal of Regulatory Economics*, Vol. 24, No. 2, pp. 175-211.

von der Fehr, N-H. M. and D. Harbord (1992), Long-Term Contracts and Imperfectly Competitive Spot Markets: A Study of the UK Electricity Industry, University of Oslo Economics Memorandum No. 14/1992.

von der Fehr, N-H. M. and D. Harbord (1993). 'Spot Market Competition in the UK Electricity Industry', *Economic Journal* 103 (418), pp. 531-46.

von der Fehr, N-H, E. Amundsen and L. Bergman (2005) 'The Nordic market: signs of Stress', *The Energy Journal Special Issue on European Electricity Liberalisation*, 71-98.

Gans, J.S. and Wolak, F.A. (2007), 'A Comparison of Ex Ante versus Ex Post Vertical Market Power: Evidence from the Electricity Supply Industry', available at ftp://zia.stanford.edu/pub/papers/vertical_mkt_power.pdf

Genc, T. and Reynolds S. (2004), 'Supply Function Equilibria with Pivotal Electricity Suppliers', Eller College Working Paper No.1001-04, University of Arizona.

Genc, T. (2009), 'Discriminatory Versus Uniform-Price Electricity Auctions with Supply Function Equilibrium', *Journal of Optimization Theory and Applications*, Vol. 140, No. 1, pp. 9-31.

Green, R.J. and D.M. Newbery (1992). 'Competition in the British Electricity Spot Market', *Journal of Political Economy* 100 (5), October, pp 929-53.

Green, R.J. (1996), 'Increasing competition in the British Electricity Spot Market', *Journal of Industrial Economics,* 44 (2), pp. 205-216.

Green, R.J. (1999a), 'The Electricity Contract Market in England and Wales', *Journal of Industrial Economics*, Vol. XLVII, No. 1, pp.107-124.

Green, R.J. (1999b), Appendix to 'The Electricity Contract Market in England and Wales'. Available at editorial website of *Journal of Industrial Economics.*

Hästö, P. and P. Holmberg (2006), 'Some inequalities related to electricity auctions', *Applied Mathematics Letters*, vol. 19, no. 8, pp. 814-819.

Holmberg, P. (2007), 'Supply Function Equilibrium with Asymmetric Capacities and Constant Marginal Costs', *The Energy Journal,* 28 (2), pp. 55-82.

Holmberg, P. (2008a), 'Unique Supply Function Equilibrium with Capacity Constraints', *Energy Economics*, vol. 30, pp. 148–172.

Holmberg, P. (2008b), Game-theoretical, Strategic Forward Contracting in the Electricity Market, Working Paper 756, Research Institute of Industrial Economics (IFN), Stockholm.

Holmberg, P. (2009a). 'Numerical calculation of asymmetric supply function equilibrium with capacity constraints', *European Journal of Operational Research*, Vol.199, no. 1, pp. 285-295.

Holmberg, P. (2009b), 'Supply Function Equilibria of Pay-as-Bid Auctions', *Journal of Regulatory Economics*, vol. 36, no. 2, pp. 154-177.

Holmberg, P., D. Newbery, D. Ralph (2008), Supply Function Equilibria: Step Functions and Continuous Representations, EPRG Working Paper 0829, available at http://www.eprg.group.cam.ac.uk/wp-content/uploads/2009/01/main-body3.pdf and Working Paper 788, Research Institute of Industrial Economics (IFN), Stockholm.

Hortacsu, A. and S. Puller (2008), 'Understanding Strategic Bidding in Multi-Unit Auctions: A Case Study of the Texas Electricity Spot Market', *Rand Journal of Economics,* 39 (1), pp.

86-114.

Hull, J.C. (1997), *Options, futures and other derivatives*, Prentice-Hall International (UK): London.

Hurlbut, D., K. Rogas, and S. Oren (2004), 'Hockey Stick Pricing: How the Public Utility Commission of Texas is Dealing with Potential Price Gouging', *The Electricity Journal*, Vol. 17, No. 3, pp. 26-33.

Kahn, A.E., P.C. Cramton, R.H. Porter, and R.D. Tabors, 'Uniform Pricing or Pay-as-Bid pricing: A dilemma for California and beyond', *The Electricity Journal*, Vol. 14, No. 6, pp. 70-79, 2001.

Klemperer, P. D. and M.A. Meyer, (1989). 'Supply Function Equilibria in Oligopoly under Uncertainty', *Econometrica*, 57 (6), pp. 1243-1277.

Klemperer, P. D. (2004), *Auctions: Theory and Practice*, Princeton:Princeton University Press.

Kremer, I and K.G. Nyborg (2004a). 'Divisible Good Auctions: The Role of Allocation Rules', *RAND Journal of Economics,* 35, pp. 147–159.

Kremer, I. and K.G. Nyborg (2004b). 'Underpricing and Market Power in Uniform Price Auctions', *The Review of Financial Studies,* 17 (3), pp. 849-877.

Kristiansen, T., (2007) 'Pricing of monthly forward contracts in the Nord Pool market', *Energy Policy*, Vol. 35, No. 1, pp. 307-316.

Levitan, R., M. Shubik. 1972. 'Price duopoly and capacity constraints', *International Economic Review,* 13, pp. 111-122.

London Economics (2007) *Structure and Performance of Six European Electricity Markets in 2003, 2004 and 2005*, Part II, DG Comp

Kreps, D. and J. Scheinkman, 'Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes', *Bell Journal of Economics*, 14 (Autumn 1983), 326-337

Longstaff, F. A., Wang, A. W., (2004) 'Electricity Forward Prices: A High-Frequency Empirical Analysis', *The Journal of Finance*, Vol. 59, No. 4.

Madlener, R. and M. Kaufmann (2002), Power exchange spot market trading in Europe: theoretical considerations and empirical evidence, OSCOGEN Report D 5.1b, Centre for Energy Policy and Economics, Switzerland.

Mahenc, P. and F. Salanié (2004), 'Softening competition through forward trading', *Journal of Economic Theory*, Vol. 116, pp. 282-293.

Maskin, E. (1986). 'The existence of equilibrium with price-setting firms', *The American Economic Review,* 76, pp. 382-386.

Murphy, F. and Y. Smeers (2004). Forward Markets may not decrease Market Power when

Capacities are endogenous, CORE-Working Paper.

Newbery, D.M. (1995) 'Power Markets and Market Power', *The Energy Journal,* 16(3), 41-66.

Newbery, D. M. (1998a). 'Competition, contracts, and entry in the electricity spot market', *RAND Journal of Economics,* 29 (4) pp. 726-749.

Newbery, D.M. (1998b), 'Pool Reform and Competition in Electricity', Ch. 5, pp.117-166 in M. Beesley (ed.) *Regulating Utilities: Understanding the Issues*, London: Institute of Economic Affairs.

Newbery, D.M. (2005) 'Electricity liberalisation in Britain: the quest for a satisfactory wholesale market design', *The Energy Journal, Special Issue on European Electricity Liberalisation,* ed. D. Newbery, 43-70.

Newbery, D.M. (2008a) Predicting market power in wholesale electricity markets, EPRG Working Paper 0821 available at http://www.eprg.group.cam.ac.uk/

wp-content/uploads/ 2009/02/eprg0821.pdf

Newbery, D.M. (2008b) Analytic Solutions for Supply Function Equilibria: Uniqueness and Stability, EPRG Working Paper 0824 available at http://www.eprg.group.cam.ac.uk/wp-content/uploads/2009/02/eprg0824.pdf

Niu, H., R. Baldick, and G. Zhu (2005), 'Supply Function Equilibrium Bidding Strategies With Fixed Forward Contracts', *IEEE Transactions on power systems,* 20 (4), pp. 1859-1867.

Offer (1994), Decision on a Monopolies and Mergers Commission Reference, Office of Electricity Regulation, Birmingham.

Offer (1995), Pool Prices and the Undertaking on Pricing Given by National Power and PowerGen Press Notice R3/95, Office of Electricity Regulation, Birmingham.

Ofgem (2000) Ofgem's Investigation Of Edison Under Market Abuse Licence Condition Initial Findings, Ofgem: July 2000 available at http://www.ofgem.gov.uk/Markets/WhlMkts/CustandIndustry/DemS 12jul00.pdf

Ofgem (2009) Liquidity in the GB wholesale energy markets, Discussion paper 62/09, Ofgem: 8 June 2009

Oren (2005), 'Generation Adequacy via Call Options Obligations: Safe Passage to the Promised Land', *The Electricity Journal*, Vol. 18, No. 9, pp. 28-42

Parisio, L. and B. Bosco (2003), 'Market power and the power market: Multi-unit bidding and (in)efficiency in electricity auctions', *International tax and public finance,* 10 (4), pp. 377-401.

Patrick, R.H. and F. Wolak (1997) Estimating the Customer-Level Demand for Electricity Under Real-Time Market Prices, draft available at ftp://zia.stanford.edu/pub/papers/rtppap.pdf

Rassenti, S., V. Smith and B. Wilson (2003), 'Discriminatory price auctions in electricity markets: low volatility at the expense of high price levels', *Journal of Regulatory Economics*, Vol. 23, pp. 109-123.

Roques, F.A., D. M. Newbery and W. J. Nuttall (2005), 'Investment Incentives and Electricity Market Design: the British Experience', *Review of Network Economics*, vol.4, no. 2.

Rudkevich, A., M. Duckworth and R. Rosen (1998). 'Modelling electricity pricing in a deregulated generation industry: The potential for oligopoly pricing in poolco', *The Energy Journal,* 19 (3), pp. 19-48, 1998.

Rudkevich, A. (2003), 'Supply function equilibrium: theory and applications', Proceedings of the 36th Annual Hawaii Conference on System Sciences, 6-9 Jan 2003.

Sioshansi, R. and S. Oren (2007), 'How Good are Supply Function Equilibrium Models: An Empirical Analysis of the ERCOT Balancing Market', *Journal of Regulatory Economics,* 31 (1), pp. 1-35.

Son, Y.S., R. Baldick, K.H. Lee, and S. Siddiqi, 'Short-Term Electricity Market Auction Game Analysis: Uniform and Pay-as-Bid Pricing', *IEEE Transactions on Power Systems*, Vol. 19, No. 4, pp. 1990-1998, 2004.

Stoft, S. (2002), *Power System Economics: Designing Markets for Electricity.* New York: Wiley-IEEE Press.

Sweeting, A. (2007) 'Market Power in the England and Wales Wholesale Electricity Market 1995-2000', *Economic Journal*, Vol. 117, Issue 520, 654-85.

Tirole, J. (1988). *The Theory of Industrial Organization*, Cambridge: MIT Press.

Wang, J. and J. Zender (2002), 'Auctioning divisible goods', *Economic Theory*, vol. 19, pp. 673-705, 2002.

Willems, B., I. Rumiantseva, H. Weigt. (2009). 'Cournot versus supply functions: What does the data tell us?' *Energy Economics*, Vol. 31, No. 1, pp. 38-47.

Wilson, R. (1979), 'Auctions of Shares', *Quarterly Journal of Economics*, Vol. 93, No. 4 , pp. 675-689.

Wilson, R. (2008), 'Supply Function Equilibrium in a constrained transmission system', *Operations Research*, Vol. 56, pp. 369-382.

Wolak, F. (2000), 'An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behaviour in a Competitive Electricity Market', *International Economic Journal*, Vol. 14, No. 2, pp. 1-43.

Wolak, F. A. (2003a), 'Identification and Estimation of Cost Functions Using Observed Bid Data: An Application to Electricity Markets' *Advances in Economics and Econometrics: Theory and Applications*, Eighth World Congress, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky,

vol. 2, pp. 115–149. Cambridge University Press, Cambridge.

Wolak, F.A. (2003b), 'Measuring Unilateral Market Power in Wholesale Electricity Markets: The California Market 1998-2000', *American Economic Review*, 93 (2), pp. 425-430.

Wolak, F.A. (2004), Quantifying the Supply-Side Benefits from Forward Contracting in Wholesale Electricity Markets, available from http://www.stanford.edu/∼wolak.

Wolfram, C. (1999), 'Electricity markets: Should the rest of the world adopt the UK reforms?', *Regulation*, Vol. 22, pp. 48-53, 1999.

Xu, L. and R. Baldick (2007), 'Transmission-constrained Residual Demand Derivative in Electricity Markets', IEEE Transactions on Power Systems, 22 (4), pp. 1563-1573.

Appendix

Consider N symmetric producers with constant marginal costs c and linear demand $D(p) = a - bp$, so that $D' = -b$. The relation in (1) can be written as follows:

$$q_i - (N-1) q_i' (p-c) = b (p-c).$$ (5)

Multiply both sides by the integrating factor $\frac{-(p-c)^{-1/(N-1)}}{(N-1)(p-c)}$:

$$\frac{-q_i (p-c)^{-1/(N-1)}}{(N-1)(p-c)} + q_i' (p-c)^{-1/(N-1)} = \frac{-b (p-c)^{-1/(N-1)}}{(N-1)}.$$

Noting that

$$\frac{d}{dp} \left[q_i (p-c)^{-1/(N-1)} \right] = \frac{-b (p-c)^{-1/(N-1)}}{(N-1)},$$

integrate both sides to give

$$q_i (p-c)^{-1/(N-1)} = A - \int \frac{b (p-c)^{-1/(N-1)}}{(N-1)} dp = A - \frac{b}{(N-2)} (p-c)^{(N-2)/(N-1)}.$$

Thus

$$q_i = A (p-c)^{1/(N-1)} - \frac{b}{(N-2)} (p-c),$$ (6)

and

$$q_i' = \frac{A (p-c)^{(2-N)/(N-1)}}{N-1} - \frac{b}{(N-2)}.$$

Let $\bar{p}$ be the price where the capacity constraint starts to bind. To ensure smooth residual demand for all producers, offers must be vertical at this point (see figure 5), i.e. $q_i'(\bar{p}) = 0$. Hence,

$$q_i'(\bar{p}) = \frac{A (\bar{p}-c)^{(2-N)/(N-1)}}{N-1} - \frac{b}{(N-2)} = 0,$$

which implies that

$$A = \frac{(N-1) b (\bar{p}-c)^{(N-2)/(N-1)}}{(N-2)}.$$

Accordingly, (6) can be written as follows:

$$q_i = \frac{b (p-c)}{(N-2)} \left((N-1) \left(\frac{\bar{p}-c}{p-c} \right)^{(N-2)/(N-1)} - 1 \right).$$

Let $\bar{q}$ be the symmetric capacity constraint. But from (5) $\bar{q} = b(\bar{p}-c)$. Hence, (6) can be written as:

$$q_i = \frac{b (p-c)}{(N-2)} \left((N-1) \left(\frac{\bar{q}}{(p-c) b} \right)^{(N-2)/(N-1)} - 1 \right).$$ (7)

46

Now define the load factor $x = q_i/\bar{q}$ and the elasticity at the marginal cost $\gamma = \frac{cb}{N\bar{q}}$, both of which are dimension-less numbers Thus (7) can be simplified to:

$$\frac{x}{\gamma} = \frac{N(p-c)}{(N-2)c}\left((N-1)\left(\frac{1}{N\gamma}\frac{c}{p-c}\right)^{(N-2)/(N-1)} - 1\right). \tag{8}$$

The next step is to derive a relation between the mark-up and the deadweight loss. In the case when demand is linear, this loss, illustrated in figure 6, can be calculated from

$$W = \frac{(p-c)^2 b}{2}.$$

The total (short-run) industry profit is $\Pi = (p-c)Nq_i$, so the ratio of deadweight loss to the industry profit ω is

$$\frac{W}{\Pi} \equiv \omega = \frac{b(p-c)^2}{2(p-c)Nq_i} = \frac{\gamma(p-c)}{2xc} \implies \frac{p-c}{c} = \frac{2x\omega}{\gamma}.$$

Then equations simplify as (8) becomes

$$\frac{x}{\gamma} = \frac{2Nx\omega}{(N-2)\gamma}\left((N-1)\left(\frac{1}{N\gamma}\frac{\gamma}{2x\omega}\right)^{(N-2)/(N-1)} - 1\right)$$

$$1 = \frac{2N\omega}{(N-2)}\left((N-1)\left(\frac{1}{2Nx\omega}\right)^{(N-2)/(N-1)} - 1\right),$$

$$2N\omega = x^{N-2}\left(\frac{N(1+2\omega)-2}{(N-1)}\right)^{N-1}, \tag{9}$$

$$x = 2N(N-1)^{N-1}\frac{\omega}{(N(1+2\omega)-2)^{N-1}}. \tag{10}$$

Proof that increasing N reduces the price-cost margin

Define two new variables, y and ϕ, where

$$y \equiv \frac{N\gamma(p-c)}{c} = \frac{b(p-c)}{\bar{q}} = \frac{p-c}{\bar{p}-c}, \text{ so } y \in [0,1]. \tag{11}$$

(The second equality follows from the definition of γ.) Define

$$\phi \equiv \frac{(N-1)}{(N-2)} \tag{12}$$

and insert (11) and (12) into (8) to give:

$$x = y\phi y^{-1/\phi} - \frac{y}{N-2} = \phi y^{1-1/\phi} - \frac{y}{N-2}.$$

This is an identity that is valid for any number of symmetric firms, N. Differentiate both sides of the equality with respect to N, holding the load factor x constant:

$$0 = \phi' y^{1-1/\phi} + (1 - 1/\phi)\phi y' y^{-1/\phi} + \frac{\phi'}{\phi^2}\ln(y)\phi y^{1-1/\phi} - \frac{y'}{N-2} + \frac{y}{(N-2)^2}.$$

We have from (11) and (12) that $y' = \frac{\gamma(p-c)}{c} + \frac{N\gamma p'}{c}$ and $\phi' = \frac{1}{(N-2)} - \frac{(N-1)}{(N-2)^2} = \frac{-1}{(N-2)^2}$. Hence,

$$0 = \frac{-y^{1-1/\phi}}{(N-2)^2} + (\phi-1)\left(\frac{\gamma(p-c)}{c} + \frac{N\gamma p'}{c}\right) y^{-1/\phi} \tag{13}$$

$$-\frac{\ln(y)\, y^{1-1/\phi}}{\phi(N-2)^2} - \frac{\gamma(p-c)}{c(N-2)} - \frac{N\gamma p'}{c(N-2)} + \frac{y}{(N-2)^2}.$$

From (12) $\phi - 1 = \frac{1}{N-2}$, and defining

$$Y \equiv \frac{y^{-1/\phi}}{N-2} - \frac{\gamma(p-c)}{cy} y^{-1/\phi} + \frac{\ln(y)\, y^{-1/\phi}}{\phi(N-2)} + \frac{\gamma(p-c)}{cy} - \frac{1}{N-2}$$

$$= \frac{y^{-1/\phi}}{N-2} - \frac{y^{-1/\phi}}{N} + \frac{\ln(y)\, y^{-1/\phi}}{N-1} + \frac{1}{N} - \frac{1}{N-2}. \tag{14}$$

equation (13) can be written

$$\frac{yY}{N-2} = \frac{N\gamma p'}{c(N-2)}\left(y^{-1/\phi} - 1\right). \tag{15}$$

It now remains to determine the sign of Y. Differentiate (14)

$$\frac{dY}{dy} = -\frac{y^{-1-1/\phi}}{\phi(N-2)} + \frac{y^{-1-1/\phi}}{\phi N} + \frac{y^{-1-1/\phi}}{N-1} - \frac{y^{-1-1/\phi}\ln y}{\phi(N-1)}$$

$$= y^{-1-1/\phi}\left(-\frac{1}{N-1} + \frac{1}{\phi N} + \frac{1}{N-1} - \frac{\ln y}{\phi(N-1)}\right)$$

$$= y^{-1-1/\phi}\left(\frac{1}{\phi N} - \frac{\ln y}{\phi(N-1)}\right) > 0,$$

because $y \in [0,1]$. Hence Y is largest when $y = 1$ and we can conclude from (14) that $Y \leq 0$, because $Y = 0$ if $y = 1$. We also realise that $\left(y^{-1/\phi} - 1\right) \geq 0$, because $y \in [0,1]$. Thus it follows from (15) that $p' = \frac{dp}{dN} \leq 0$ and we have established that mark-ups decrease when the number of firms increase (everything else equal).