PIVOTAL SUPPLIERS AND MARKET POWER IN EXPERIMENTAL SUPPLY FUNCTION COMPETITION*

Jordi Brandts, Stanley S. Reynolds and Arthur Schram

We use experiments to study market power with supply function competition, akin to the competition in electricity markets. Our treatments vary the distribution of demand levels as well as the amount and symmetry of the allocation of production capacity between different suppliers. We relate our results to a descriptive power index and to the predictions of two models: a supply function equilibrium (SFE) model and a multi-unit auction (MUA) model. Observed behaviour is consistent with the equilibria of the unrestricted SFE model and inconsistent with the unique equilibria of two refinements of the SFE model and of the MUA model.

In the worldwide process of regulatory reform in the electricity industry, the possible existence of market power is one of the basic problems analysts and policy makers have to deal with. Field data document the existence of reduced competition due to market power in some electric power markets (Wolfram, 1999; Borenstein *et al.*, 2002). The severe welfare losses this may cause are a major concern that needs to be addressed to fully assess the success of the reforms. If non-competitive prices can easily persist in these markets, this creates the need to find measures to mitigate market power.

Among the features of markets that need to be taken into account in relation to market power is the presence of one or more pivotal suppliers. In a general sense, a producer can be considered to be pivotal if, without his capacity, the supply cannot serve the whole demand. It is important that the issue is not simply one of insufficient total capacity to serve the market demand but one of particular producers controlling large enough parts of the capacity. We will refer to market power due to pivotal suppliers as pivotal power.

Concerns about pivotal power are the basis for some energy policy provisions. Several organisations that coordinate and regulate wholesale electricity markets in the US – PJM, Electric Reliability Council of Texas (ERCOT) and California Independent System Operator (ISO) – employ pivotal supplier screening tests as a trigger for market power mitigation measures (Reitzes *et al.*, 2007). In addition, the US Federal Energy Regulatory Commission (FERC) may block a generation company from charging market-based rates for energy if the company fails a pivotal supplier test for market power. Under the FERC test, a generation supplier is deemed pivotal and, therefore,

* Corresponding author: Jordi Brandts, Department of Business Economics, Universitat Autònoma de Barcelona, Institut d'Anàlisi Econòmica (CSIC) and Barcelona GSE, Campus UAB, 08193 Bellaterra (Barcelona), Spain. Email: jordi.brandts@uab.es.

Financial support by the Spanish *Ministerio de Economía y Competitividad* (Grant: ECO2011-29847-C02-01), the *Generalitat de Catalunya* (Grant: 2009 SGR 820), the Antoni Serra Ramoneda Research Chair (UAB-CatalunyaCaixa), Consolider-Ingenio 2010 (CSD2006-00016) and the Barcelona GSE Research Support Programme is gratefully acknowledged. David Rodríguez provided excellent research assistance. Three anonymous referees and the editor provided very helpful suggestions for improving a previous version of this article. Remaining errors are ours.

fails the test, if peak demand cannot be met in the relevant market without production from the supplier's capacity.¹

A major goal of this article is to provide evidence from laboratory experiments regarding the effects of pivotal power. Our design of these experiments and analyses of the data they generate is guided by a combination of theoretical analysis and tools used in the field. To our knowledge this is the first experimental study focusing on the specific issue of pivotal power. A potentially important distinction for policy makers is the presence of a pivotal supplier *versus* the supplier's incentive to exercise market power. We examine in the laboratory the extent to which pivotal suppliers actually exercise market power under varying market conditions. We study both the cases where pivotal power is evenly spread among producers and where it is concentrated in a subset of producers. The first case corresponds more to a situation of tight market capacity and the second more to a case in which particular firms may have strong influence on market outcomes, even though market capacity is large. To study more closely the circumstances under which pivotal power may matter, we also analyse the impact on market power of a variation in the extent of demand uncertainty.

The use of laboratory experiments makes it possible to implement variations in capacity distributions with a high degree of control, to isolate their effects under conditions that are strongly *ceteris paribus*. This control makes the experimental method a useful tool for studying electric power markets (Rassenti *et al.*, 2002; Staropoli and Jullien, 2006). In fact, experimental research has been very influential in studying and designing mechanisms in a variety of real world markets, including matching markets (Kagel and Roth, 2000), auctions (Brunner *et al.*, 2010) and railway competition (Cox *et al.*, 2002). Although pivotal power as such has not been studied experimentally, previous laboratory experiments do show that market power is easily exerted in environments that mirror the wholesale electricity market. Moreover, experiments have been useful for studying how certain market features can increase or limit market power; demand side bidding (Rassenti *et al.*, 2003) and forward trading (Brandts *et al.*, 2008) have been shown to enhance competition.

A second important goal of this article is to provide evidence from laboratory experiments on the predictive power of two types of theoretical models in settings with and without pivotal suppliers. One of the advantages of the use of laboratory experiments is that it facilitates the interplay between theory and data in a way that is not easy to accomplish with field data. The variations of the conditions in the different treatments of our experiments are guided by theory and yield specific theoretical predictions for the different cases that can be directly tested with the experimental data. Our experimental data can, hence, be used to evaluate theories pertaining to the effect of pivotal suppliers in electric power markets. (Falk and Heckman, 2009 for a recent methodological discussion of laboratory experiments.)

¹ The following quote is from FERC Order No. 697 (Federal Energy Regulatory Commission, 2007, pp. 18–19): 'The second screen is the pivotal supplier screen, which evaluates the potential of a seller to exercise market power based on uncommitted capacity at the time of the balancing authority area's annual peak demand. This screen focuses on the seller's ability to exercise market power unilaterally. It examines whether the market demand can be met absent the seller during peak times. A seller is pivotal if demand cannot be met without some contribution of supply by the seller or its affiliates'.

Theoretical analyses of strategic behaviour and market power in wholesale electricity markets have been based on either a multi-unit auction model (Anwar, 1999; Fabra *et al.*, 2006), hereafter MUA, or the supply function equilibrium model (Klemperer and Meyer, 1989), hereafter SFE. Both are models of one-shot strategic interaction. The MUA is a discrete unit model in which each supplier submits price offers for units of capacity under their control. The SFE assumes a completely divisible good and has been used to study a variety of issues related to electric power markets (Newbery, 1998; Green, 1999; Bolle, 2001; Baldick *et al.*, 2004; Genc and Reynolds, 2011). We explain below that both the MUA and SFE models can be used to generate equilibrium predictions for subjects' behaviour in our experiments; moreover, there are differences in equilibrium predictions for these two models.

A third goal of this article is to link our analysis of experimental results to the Residual Supply Index, a measure of pivotal power that has been used in empirical studies with field data from electric power markets. The Residual Supply Index (hereafter, RSI) measures the aggregate capacity of all suppliers except the largest as a fraction of total demand. The largest supplier is pivotal when this index is less than one; lower values of RSI can be interpreted as yielding more market power for the largest supplier. Rahimi and Sheffrin (2003) find that higher values of the RSI yield significantly lower price-cost margins for California wholesale electricity market data for summer peak hours in the year 2000. Wolak (2009) develops a measure of a supplier's ability to exercise unilateral market power in each half-hour period in his study of the New Zealand wholesale electricity market. He notes that this ability to exercise market power is strengthened the greater the probability that the supplier is pivotal during the period.² Wolak finds a positive correlation between the average half-hourly firm-level ability to exercise unilateral market power and half-hourly market prices.

Our experimental design involves three pivotal supplier treatments; two treatments involve symmetrical reductions in sellers' capacities, and the third involves a reallocation of sellers' capacities to create two large, pivotal suppliers. The experimental results allow us to reject the hypotheses that pivotal power does not matter for these three treatments. Average market prices are higher in treatments with pivotal power; the way in which average prices change is intuitive and consistent with the qualitative predictions of the RSI. In contrast to these effects of capacities, variations in the distribution of demand quantities have smaller effects on observed market prices. In none of the cases we study does pivotal power lead to monopoly prices.

Both MUA and SFE models have multiple equilibria for one or more of our experimental treatment conditions; multiplicity is more pronounced for SFE. Overall, we find that observed behaviour is consistent with the range of equilibria of the unrestricted SFE model and inconsistent with the equilibria of the MUA model and two refinements of the SFE model. In treatments without pivotal power, observed prices are on average higher than the (near) marginal cost level of the MUA equilibrium, though there is a tendency for them to move towards the competitive

² Wolak discusses pivotal suppliers and their significance in the New Zealand wholesale market in subsection 3.4 of part 2 of his report. He defines a pivotal supplier as follows: 'A supplier that faces a residual demand curve that is positive for all possible positive prices is said to be a pivotal because some of its supply is necessary to serve the market demand regardless of the offer price' (Wolak 2009, p. 115).

MUA prediction. For treatments with pivotal power observed prices are consistent with the qualitative feature of the SFE that market power caused by a symmetrical reduction in capacity has a stronger effect on prices than market power caused by an asymmetrical distribution of overall high capacity (which is also predicted by the RSI). With caution one can say that our results suggest that the continued use of the SFE model to study electric power market as in the work of Niu *et al.* (2005), Hortacsu and Puller (2008), Sioshansi and Oren (2007) and Vives (2011) seems warranted.

We also study the best response behaviour of our subjects. We compare actual profits to optimal profits for subjects and compare these results to those from a similar analysis with field data from the Texas (ERCOT) wholesale power balancing market (Hortacsu and Puller, 2008). We find that the behaviour we observe in our experiments is closer to best responding than behaviour in the particular field environment we compare our data to.

Our experiments bear some similarity to multi-unit auction experiments reported in Sefton and Zhang (2010); these are sales auctions with three buyer subjects, in contrast to our procurement auctions with four seller subjects. Subjects in their experiments bid on multiple discrete units, and bids for individual units are from a discrete price grid. The bidding game corresponding to their experiments has multiple equilibria, with equilibrium prices ranging from the (common) valuation of bidders (competitive bidding) to a price of zero (tacitly collusive bidding). Experimental results in Sefton and Zhang are consistent with bids equal to or slightly below values; they find relatively little evidence of bid shading. This is in contrast with results in our high-capacity treatments, in which average offers remained well above marginal cost for some groups. We discuss possible reasons for the differences in experimental results below.

The remainder of this article is organised as follows. In the following section, we present our experimental design and procedures. Specific theoretical predictions are provided in Section 2. The results follow in Section 3. Section 4 concludes and discusses the implications of our findings.

1. Design and Procedures

In the experiment, there are 25 rounds each consisting of five periods. The demand is simulated using a simple box-design (Davis and Holt, 1993). In each of the five periods t in round r , a perfectly inelastic demand d_t^r is randomly chosen from the set $d_t^r \in \{d^{\min}, \dots, d^{\max}\}$ with equal probabilities for each element in the set. In all of our treatments, $d^{\max} = 35$. We define $l \equiv d^{\min}/d^{\max}$ as the load ratio, which is our first treatment variable. All sessions have either $l = 4/7$ (i.e. $d^{\min} = 20$) or $l = 6/7$ ($d^{\min} = 30$).³ There is a price cap given by $p^{\max} = 25$, i.e. no units can be traded above this price. Prices and quantities are restricted to integers.

On the supply side, there are four firms in each market. Each subject represents one firm. Each firm j offers a discrete number of units in round r , which will apply to each period t in r . Any units sold are produced at constant marginal costs $c = 5$.⁴ Individual supply is limited by an exogenously enforced maximum capacity $s_j^{\max}$. These determine industry capacity, which is given by $S^{\max} \equiv \sum_{j=1}^4 s_j^{\max}$.

³ The load ratio affects the theoretically predicted outcomes (Section 3).

⁴ Given our focus on pivotal power, the assumption of constant marginal costs is not restrictive.

Our second treatment variable is this industry capacity. This is given by either $S^{\max} = 48$ or $S^{\max} = 36$. Note that in both cases $d^{\max} < S^{\max}$, i.e. in all cases industry capacity suffices to satisfy the maximum demand. Our third treatment variable pertains to the $S^{\max} = 48$ case. We distinguish between the case where this capacity is distributed evenly across firms ($s_j^{\max} = 12, j = 1, \dots, 4$) and the case where there is asymmetrical capacity ($s_j^{\max} = 5, j = 1, 2, s_j^{\max} = 19, j = 3, 4$). For the $S^{\max} = 36$ treatment we only consider the symmetrical case ($s_j^{\max} = 9, j = 1, \dots, 4$).

Firm j offers units for sale in round r by bidding a discrete ‘supply function’, s_j^r . This is a vector of up to $s_j^{\max}$ supply prices, p_{jk}^{sr} , ordered from low to high at which firm j is willing to sell units: $s_j^r = (p_{j1}^{sr}, p_{j2}^{sr}, \dots, p_{js_j^{\max}}^{sr})$, with $p_{jk}^{sr} \geq p_{j,k-1}^{sr}, k \geq 2$.⁵ Subjects can offer fewer than $s_j^{\max}$ units by not entering prices for them. Equivalently, they can offer $m < s_j^{\max}$ units by setting $p_{jk}^{sr} = 26, k = m + 1, \dots, s_j^{\max}$.⁶ The individual supply functions are combined and supply prices are ordered from low to high to obtain the market supply function for round r , denoted by $s^r = [s^r(1), \dots, s^r(S^{\max})]$, which is a vector of the $S^{\max}$ submitted supply prices ordered from low to high (if necessary, supplemented with infinite prices for units not supplied). Finally, a uniform transaction price p_t^r , is determined in each period t of r by comparing d_t^r to s^r : $p_t^r \equiv \min\{s^r(d_t^r), p^{\max}\}$. Note that if $s^r(d_t^r) > p^{\max}$, then supply cannot satisfy demand at a price below $p^{\max}$ and $k < d_t^r$ units are sold, where k is uniquely determined by $s^r(k) \leq 25$ and $s^r(k + 1) > 25$.

Finally, the payoffs of firm j in round r , π_j^r , are determined by the uniform prices in each of the five periods of a round and the marginal costs:

$$\pi_j^r = \sum_{t=1}^5 (p_t^r - c) q_{tj}^r, \quad j = 1, \dots, 4,$$

where q_{tj}^r denotes the number of units sold by j in period t of round r .

In the experiment, subjects submit supply functions by entering a price for each possible unit in a table. To ease the task, the software fills gaps between units priced. For example, if a subject enters a price of 5 for unit 1 and then 7 for unit 5, then units 2–4 are automatically priced at 7, though the subject can subsequently change them. In addition, the software does not allow decreasing prices across units. The subject is free to withhold units from the market by leaving them unpriced, as long as all subsequent units remain unpriced as well. No supply price is submitted until the subject finalises and confirms the complete set. There is no time limit for submission of the supply functions.

After all four subjects have submitted a supply function, they are aggregated and the result is confronted with five subsequent demand realisations – the five periods of a round – yielding five prices. Each realisation appears on the subjects’ monitors for 5 s. After the five periods, the subject can page back and forth between the periods until satisfied. After everyone has indicated that they are ready the next round commences.

The results of a period appear on the screen graphically and in numbers. Figure 1 shows an example of the graph a subject could see. This was a treatment where each

⁵ The ‘s’ superscript on the price variable indicates that the variable is a price offer made by a seller.

⁶ Units offered at a price above the price cap, $p^{\max} = 25$ will not be sold.

Fig. 1. Screenshot of Period Results

subject had a maximum of nine units to supply. The graph shows the results for period 3 (shown at the bottom on the right), where $d_3^l = 30$ and $p_t^r = 13$ were realised (bottom left) and this subject sold seven units (bottom centre). The graph shows the realised demand function (including the price cap), the price (light coloured line) and the nine supply prices submitted by this subject for this round (rising from 6 to 22). The seven units sold by the subject are shown in dark grey; the two that were not sold are light grey.⁷ Notice that this subject was willing to sell her 8th unit at price 13 as well but other subjects had also submitted units at this price. In this case, the computer randomly appoints subjects to slots on the aggregate supply function.⁸

The three treatment variables presented above (load ratio, industry capacity and symmetry of capacity) are varied between subjects. Table 1 presents an overview of

Table 1
Treatment Overview

		Low capacity: $s^{\max} = 36$		High capacity: $s^{\max} = 48$	
		Symmetric: $s_j^{\max} = 9, j = 1, \dots, 4$	Symmetric: $s_j^{\max} = 12, j = 1, \dots, 4$	Asymmetric: $s_j^{\max} = 5, j = 1, 2$ $s_j^{\max} = 19, j = 3, 4$	
Load ratio	Low: $l = 4/7$	$lsk; n = 6^*; \text{€}45.14$	$hsk; n = 6; \text{€}15.89$	–	
	High: $l = 6/7$	$lsh; n = 5; \text{€}67.13$	$hsh; n = 5; \text{€}27.28$	$hah; n = 6; \text{€}43.60$	

Notes. The entries in the cell show the acronym, xyz (x = capacity, y = symmetry and z = load ratio); the number of markets we have data for, n ; and the average earnings in euro (€). *For ls , we organised two sessions with three markets each. All data for the other treatments were obtained in one session each.

⁷ Subjects did not receive information about the offer price of units sold by other suppliers. A reader pointed out that, although generators typically do not get to see the previous day's bidding data immediately, this information is often released quite quickly and a generator may have plenty of ways to find out its competitors' supply. We agree that different information feedback structures may affect the offers generators make. However, we think that the feedback rule we use is a good place to start.

⁸ We use pro-rating on the margin rationing in the experiments. This scheme for rationing supply is commonly used in wholesale electricity auctions. The way in which excess supply is rationed may have an impact on bidding; see Kremer and Nyborg (2004) for a theoretical analysis.

these treatments and the number of markets we ran for each of them. In addition, it gives the average subject earnings for each treatment.

Note that this simple design allows us to investigate the role of pivotal suppliers in a straightforward manner. First, in treatments *hsl* and *hsh* no single firm is pivotal because any three firms can cover maximum demand ($d^{\max} = 35$). Second, *lsl* and *lsh* deal with the case where every firm is pivotal for at least some possible demand quantities: together the four firms have sufficient capacity for maximum demand but three firms can only supply 27 units. Third, *hah* covers the situation where two of the four firms (i.e. firms 3 and 4) are pivotal. Any combination of three firms including these two, suffices for maximum demand but a combination of firms 1 and 2 with either 3 or 4 can only supply 29 units. Finally, note the large variation in earnings. This is a first indication that market power matters.

The experiment was conducted in six sessions at the CREED laboratory for experimental economics of the University of Amsterdam. Overall, 112 subjects were recruited by public advertisement on campus and were mostly undergraduate students in economics, business and law. They were allowed to participate in only one experimental session. Each session lasted for about 2–3 hours. Earnings in the experiment were denoted in experimental francs. We used an exchange rate of 250 francs to 1 euro. All subjects received a starting capital of 1,250 francs, which was part of their earnings. There was no show-up fee. Subjects earned between €12.40 and €112.20 with an average of €42.71.

At the outset of each session, subjects were randomly allocated to the laboratory terminals and were asked to read the instructions displayed on their screens.⁹ Then they were introduced to the computer software and given five trial rounds to practise with the software's features. Subjects were told that during these trial rounds other subjects' decisions would be simulated by the computer, which was programmed to make random decisions, and that gains or losses made during those rounds would not count for their final earnings from participation. Once the five trial rounds were over, the pool of subjects was divided into independent groups (markets) of four subjects.

Each session then consisted of 25 repetitions (rounds), each round taking approximately 3–4 minutes to be completed. In each session only one treatment was run with five or six groups per session (for *lsl*, two sessions were needed to obtain data from six groups). As mentioned at the beginning of the section, each round consisted of five periods, meant to correspond to different times of the day, with possibly different demand realisations. Subjects made supply decisions for each round; decisions were held fixed across the five periods within the round.

Subjects stayed in the same market for the whole session and did not know which of the other subjects were in the same market as themselves. The interaction in fixed groups approximates best actual circumstances in the kind of electric power markets that we are interested in. Fixed interaction is used in all previously cited experiments on electric power markets. The procedure has also the advantage that the observations from the different groups are statistically independent from each other.

⁹ A transcript of the instructions (translated from Dutch) is included in online Appendix A.

2. Theoretical Predictions and Hypotheses

As mentioned in the introduction, we centre our theoretical analysis on a descriptive index, the RSI (Residual Supply Index) and two theoretical models, the MUA (multi-unit auction model) and SFE (supply function equilibrium model). In this Section, we formulate specific hypotheses derived from these benchmarks. Our hypotheses will pertain to prices. In particular, we will use the volume weighted average price (VWAP, hereafter), which is defined as the monetary value of trades divided by the number of units traded per round. The VWAP provides a useful way to compare observed prices in experiments to theoretical predictions. Formally, let $p(d)$ be the expected price when d units of output are demanded. The expected VWAP is defined as

$$\bar{P}^e = \sum_{d=20}^{35} \frac{\delta dp(d)}{E(d)},$$

where superscript e denotes expectation and δ is the probability of each possible demand level.

The RSI is an indicator of market power given by the following expression:

$$RSI(d) = \frac{\text{total capacity} - \text{largest seller's capacity}}{\text{demand quantity}} = \frac{s^{\max} - s_3^{\max}}{d},$$

where we have used the fact that in our notation seller $j = 3$ has the highest capacity in all treatments (as does seller $j = 4$). Table 2 shows the range of RSI for each of our treatments, as well as the midpoint of the interval.

If no firms are pivotal for any demand quantity, as in high-capacity treatments *hsh* and *hsl*, then the RSI exceeds one in all periods of all rounds; in this case, any group of three firms has enough capacity to meet demand. If the largest firm is pivotal for all demand quantities, as in treatments *lsh* and *hah*, then the RSI is less than 1 for all periods of all rounds. For treatment *lsl* firms are pivotal for low demand quantities but not for high quantities. RSI is less than 1 for some periods and greater than 1 for other periods in treatment *lsl*. We expect treatments with higher average values of RSI to have lower market prices.

The descriptive index is useful but has clear limitations, as it does not propose specific price levels for the different parameter configurations we analyse. The MUA and the SFE take us a bit further in this regard. Both models analyse the interaction

Table 2
RSI

		Low capacity:		High capacity:	
		Symmetric:		Symmetric:	Asymmetric:
Load ratio	Low	[0.71,1.25], 0.98, <i>lsl</i>		[1.03,1.80], 1.415, <i>hsl</i>	
	High	[0.71,0.83], 0.77, <i>lsh</i>		[1.03,1.20], 1.115, <i>hsh</i>	
					[0.83,0.97], 0.90, <i>hah</i>

Notes. The first entry in the cell shows the range within which the RSI falls for the various possible demand realisations in the treatment concerned. The second entry shows the midpoint of that interval. The third entry is the treatment acronym defined in Table 1.

between firms as a one-shot strategic game, where the strategies consist of supply functions. Of course, the subjects in our experiment are engaged in a 25-round repeated game, so that the equilibrium prescriptions do not exactly pertain to the environment we study. However, as in many other studies, the equilibria of the one-shot game are relevant benchmarks, particularly given the known and finite time horizon we use. The central difference between the two models is that the MUA pertains to discrete production units while the SFE specifies continuously divisible output.

2.1. MUA Predictions

The MUA considers the game as an auction in which each firm j submits a vector of offer prices, selected from non-negative real numbers, for discrete units of output, s_j^r . This game is analysed in Anwar (1999) and Fabra *et al.* (2006). For our parameters, this formulation yields pure strategy equilibria for treatments *hsh*, *hsl*, *hah* and *lsh*; the equilibrium is in mixed strategies for treatment *lsl*. Details are given in online Appendix B.

Equilibria for high-capacity treatments *hsh* and *hsl* involve market-clearing prices equal to or slightly above marginal cost ($c = 5$); the intervals of predicted equilibrium VWAP are provided in Table 3. The excess capacity present in treatments *hsh* and *hsl* provides incentives for price cutting when offers are above marginal cost. Price cutting incentives are mitigated to some degree because price offers are restricted to discrete units in the experiments.¹⁰

Pure strategy equilibria of MUA for *hah* and *lsh* involve asymmetrical strategies, in which three firms offer all units at a low price, and the 4th firm (one of the two high-capacity firms for *hah*) offers all its units at 25 (the price cap). The equilibrium price is equal to 25 for all demand realisations. The firm submitting high-price offers earns lower expected profit than its rivals; the low-price offers of rivals leave the high-price firm with no incentive to reduce its offers. These equilibria embody maximal exercise of market power; firms extract the maximum possible surplus in equilibrium.^{11,12}

¹⁰ The MUA model with continuous price offers provides a simple prediction for these treatments; any pure strategy Nash equilibrium (PSNE) yields market-clearing prices equal to marginal cost ($c = 5$). This prediction is driven by strong incentives to undercut rivals price offers in the symmetric, high-capacity treatments. Our experimental design departs from the continuous price interval assumption by specifying that price offers must be chosen from a set of discrete prices. When price offers are restricted to a set of discrete price offers, the incentive to undercut is weakened, as a discrete price cut may involve a significant loss of revenue for infra-marginal units. Indeed, Sefton and Zhang (2010) have an experimental design for a multi-unit sales auction with discrete units and $n = 3$ bidders for which a PSNE of the one-shot game can sustain collusive bidding, even in a setting in which $n - 1$ bidders value units above the cost of all of the units available from the seller. Derivation of equilibrium strategies for our procurement auction setting is more complex than for the Sefton and Zhang design because we allow $n = 4$ subject bidders, and have a larger number of quantity units per subject to bid on and a larger number of discrete prices to choose from.

¹¹ This effect of the asymmetrical distribution of capacities is reminiscent of the price competition environment with capacity constraints studied experimentally by Davis and Holt (1994). With a symmetrical distribution of capacity, the (pure strategy) equilibrium price is equal to marginal cost, while for a certain asymmetrical distribution of the same capacity the equilibrium prices are above marginal cost.

¹² As the pure strategy equilibria for *lsh* involve strong asymmetries in strategies and payoffs across players, we also compute a lower bound for expected VWAP (reported in Table 3) that is based on necessary conditions for a symmetrical mixed strategy equilibrium.

In treatment *lsl*, firms are pivotal for high demand quantities but not for low demand quantities. There are no pure strategy equilibria of the MUA for *lsl*. Fabra *et al.* (2006) derive a mixed strategy equilibrium for the case in which firms are restricted to make a single price offer for their capacity. However, Anwar (1999) shows that mixing over a single offer is not an equilibrium when each firm can make distinct offers for multiple units. We are not aware of analytical results for mixed strategy equilibria of MUA in which each firm can submit multiple offers. However, we can provide bounds on mixed strategy prices. Our experiment requires firms to submit offers in discrete price units from the set $\{5, 6, \dots, 25, 26\}$; a unit offered at 26 will not be accepted and is equivalent to withholding the unit. In *lsl*, each firm submits offers for nine units. A firm's strategy is a non-decreasing offer schedule for nine units; each firm has a finite set of strategies to choose from.¹³ It is well known that any finite n -person non-cooperative game has at least one mixed strategy Nash equilibrium. In online Appendix B we show that expected equilibrium profit for a seller has a positive lower bound. This profit bound permits us to bound the equilibrium VWAP for treatment *lsl* $\bar{P}^e \geq 11.55$.

2.2. SFE Predictions

The second theoretical approach permits firms to submit continuous supply functions to an auctioneer. Klemperer and Meyer (1989) formulate and analyse game-theoretic models in which demand is uncertain and strategies are continuous, non-decreasing supply functions for infinitely divisible output. A Nash equilibrium for such a game is termed a supply function equilibrium (SFE). Genc and Reynolds (2011) extend the SFE analysis of Klemperer and Meyer to permit capacity constraints and supply functions with discontinuities (e.g. step functions).

The SFE formulation has been used in a number of studies to predict behaviour in naturally occurring wholesale electricity markets (Newbery, 1998; Green, 1999; Bolle, 2001; Baldick *et al.*, 2004) in which suppliers submit offers for discrete units of output.¹⁴ Output is not infinitely divisible in our experiments either. Each firm (subject) submits offers for between 5 and 19 discrete units of output, depending on the treatment. This permits us to explore whether or not the SFE model provides useful predictions of behaviour in an environment with discrete units. In addition, our experiments permit us to compare the predictive power of the SFE model to that of the MUA model in a particular setting.¹⁵

¹³ The strategy set is finite but very large. There are 14,307,150 strategies to choose from.

¹⁴ The assumption of completely divisible goods yields equilibria where prices are continuous functions of quantity. However, in real world electricity markets (as in our experiments), prices are discrete step functions. Holmberg *et al.* (2008) show that under certain conditions such step functions converge to continuous supply functions as the number of steps increases. This provides a justification for approximating step functions with smooth supply functions.

¹⁵ Under some market rules, one theory may be much more suitable than the other. If market rules limit firms to submitting offers with one or two steps, then the MUA model seems more appropriate than SFE. Some market rules allow firms to submit upward sloping supply functions. For example, the Southwest Power Pool RTO runs an energy balancing market in which each firm submits multiple price-quantity pairs. This RTO interpolates linearly between pairs to yield a piece-wise linear, upward sloping supply curve for the firm. See: <http://www.spp.org/section.asp?group=328&pageID=27>. An SFE model would seem more appropriate than a MUA model for such market rules.

Details of the SFE method applied to our parameters are presented in online Appendix C. Here, we present the main results derived from this theory. The first point to make is that Nash equilibrium pure strategies of the MUA model are also equilibrium strategies of the SFE model. Second, the SFE model admits additional pure strategy equilibria compared to the MUA model.

Consider our high-capacity *hsh* and *hsl* treatments. The equilibria for the MUA model have price equal to marginal cost (or slightly above marginal cost, for discrete prices); a firm has an incentive to undercut any rival offers that are above marginal cost. However, with infinitely divisible output, if a firm's rivals submit smooth upward sloping supply curves then the firm's best response is to offer its supply at prices above marginal cost. Klemperer and Meyer (1989) show that in general there are multiple supply function equilibria and these equilibria involve non-negative price-cost markups. Supply function equilibria for some of the treatments are illustrated in Figure 2. For *hsh* and *hsl* any aggregate supply function between (and including) the two bold curves indicated by A and B is consistent with a SFE.¹⁶

One way to characterise supply function equilibria is by the equilibrium price they generate when $d^{\max}$ is realised, i.e. $p_t^r|_{d^{\max}} = s^r(35)$. As illustrated in Figure 2, the set of SFE for *hsh* and *hsl* is characterised by $s^r(35) \in [5, 25]$, i.e. the price at maximum demand can lie anywhere between the competitive price and the highest possible price, which in this case is the same as the monopoly price.

Consider now treatments *lsh* and *lsl* for which each firm is pivotal; other firms cannot fully compensate if one firm withholds units. The market power induced in these treatments has as a consequence that supply functions at or near marginal cost for all

Fig. 2. *Aggregate Supply Functions for Supply Function Equilibrium (SFE)*

Notes. The curves show possible aggregate supply curves for smooth supply function equilibria. The vertical dashed line at $Q = 35$ indicates maximum demand $d^{\max}$. For *hsh* and *hsl*, any curve between A and B constitutes a SFE. For *lsh* and *lsl* the set of aggregate supply curves for smooth SFE is reduced to the curves between A and some curve C, above curve B.

¹⁶ We refer to a SFE in which the aggregate supply function is differentiable over the range of possible demand quantities as a smooth SFE. For example, supply function equilibria associated with aggregate supply functions labelled A and C in Figure 2 are smooth.

units for all firms are not equilibrium strategies. In fact only the aggregate supply functions between some function C (above B) and A are SFE for these low-capacity treatments (Figure 2). More specifically, for lsh , functions characterised by $s^r(35) \in [21, 25]$ are SFE and for sl this holds for functions with $s^r(35) \in [18, 25]$.

For lsh , we have to consider non-smooth SFE as well. If one allows firms to submit non-smooth step-function supply functions (formally, right-continuous functions of price) then the asymmetrical equilibria of the MUA with price equal to 25, in which three firms offer all units at a low price and the 4th firm offers all of its units at 25 are also SFE for treatment lsh (but not for any of the other treatments with symmetrical capacity distribution).

For treatment hah the asymmetrical equilibria of the MUA, with one of the two high-capacity firms bidding in all units at a price of 25 (yielding price equal to 25 for all demand realisations), are also supply function equilibria. There are additional supply function equilibria for treatment hah with prices below the price cap. In these equilibria, low-capacity firms offer all units at a low price and high-capacity firms have upward sloping (in fact, linear) supply functions, which are between supply curves A and B in Figure 3; $s^r(35) \in [11.9, 25]$ for these equilibria.

Table 3 summarises the theoretical predictions from the two approaches in terms of volume weighted average prices $\bar{P}^c$. In summary, the MUA yields relatively precise predictions for three of five cases. SFE predictions include the pure strategy equilibrium predictions of MUA, as well as additional predictions of intervals of average prices that are based on smooth supply function equilibria.

We will also apply two refinements to SFE and investigate how well they organise the data we observe. First, the ‘midpoint equilibrium’ selects the mean SFE in the range of symmetrical SFE. This is an easy heuristic that attempts to predict average behaviour across markets, assuming that distinct SFE occur with more or less equal probability.

Fig. 3. *Supply Function Equilibrium for Asymmetrical Capacity Treatment*

Notes. The small firms submit low price offers for their entire capacity, jointly 10 units (in the graph, these low offers are set equal to 5 but they may be larger than the marginal costs). Each large firm submits a linear, increasing supply function. The aggregate supply function is horizontal for units 1–10 and increases after unit 10. The most competitive of these equilibria reaches a price of 11.9 at $d^{\max} = 35$; the aggregate supply is labelled B . The least competitive of these equilibria reaches the price cap of 25 at $d^{\max} = 35$; the aggregate supply is labelled A .

Table 3
Theoretical Predictions of Volume Weighted Average Price

Capacity	Low load ratio				High load ratio			
	MUA	SFE	SFE: midpoint	SFE: payoff dominant	MUA	SFE	SFE: midpoint	SFE: payoff dominant
High	[5, 5.75]	[5, 16.4]	10.7	16.4	[5, 7.5]	[5, 21.3]	13.2	21.3
Low	≥11.6	[12.4, 16.4]	14.4	16.4	≥18.5	[18, 21.3] & {25}	19.7	25
High	n.a.	n.a.	n.a	n.a.	{25}	[11.2, 23.1] & {25}	17.2	25
asymmetrical								

Notes. The SFE refinements (midpoint and payoff dominant) are described in the main text. For the low capacity, high load ratio treatment Table 3 provides a range of MUA predictions based on mixed strategy equilibrium conditions. The MUA pure strategy prediction for this treatment is at the price cap of 25.

Second, the ‘payoff-dominant’ equilibrium is the SFE with the highest VWAP. This has the intuitive appeal of being best for the players concerned. Both refinements are given in Table 3.¹⁷

We will use these theoretical predictions to organise our data on volume weighted-average prices in several ways. We will study whether observed prices remain within the interval prescribed by the SFE and, if so, whether they are well approximated by the more extreme predictions of the SFE refinements or the MUA model, all shown in Table 3.

In addition, we take a more qualitative look at the data and test a set of formal hypotheses about the comparative static effects that the results in Table 3 predict for our treatment variables. The null hypothesis we use as a benchmark stems from the naive view that prices should not be expected to differ across treatments, as in all our treatments total capacity is sufficient to serve the maximum demand. The distinct alternative hypotheses are based on both the midpoint values of the RSI (Table 2) and the predictions that have been derived using MUA and the two SFE refinements (Table 3). The comparisons we perform pertain to the distinguished treatment variations of total capacity, capacity distribution and demand load ratio and to the direct comparison of the two ways in which pivotal power is present in our design.

In the following hypotheses, $\bar{P}_x$ stands for the volume weighted average price in treatment x . The first two hypotheses refer to the symmetrical reduction in capacity, with a high and low demand load ratio respectively:

- (i) With a high load ratio, the presence of pivotal firms, due to symmetrically distributed low total capacity, increases average prices (predicted by RSI, MUA and both SFE refinements). Formally:

$$H1_0 : \bar{P}_{hsh} = \bar{P}_{lsh} \quad \text{vs.} \quad H1_1 : \bar{P}_{hsh} < \bar{P}_{lsh}.$$

- (ii) With a low load ratio, market power caused by a symmetrical reduction in capacity causes an increase in average prices (predicted by RSI, MUA and the midpoint-SFE refinement). Formally:

$$H2_0 : \bar{P}_{hsl} = \bar{P}_{lsl} \quad \text{vs.} \quad H2_1 : \bar{P}_{hsl} < \bar{P}_{lsl}.$$

The next hypothesis refers to the change in distribution of the high total capacity level for the high demand load ratio.

- (iii) With a high load ratio, the presence of pivotal firms, due to asymmetrically distributed high total capacity, increases average prices (predicted by RSI, MUA and both SFE refinements). Formally:

$$H3_0 : \bar{P}_{hsh} = \bar{P}_{hah} \quad \text{vs.} \quad H3_1 : \bar{P}_{hsh} < \bar{P}_{hah}.$$

The next hypothesis refers to the two ways in which pivotal power can appear.

¹⁷ Alternatively, one may think that the set of SFE equilibria could be refined by a restriction to linear supply functions (as demonstrated by Klemperer and Meyer, 1989). However, this refinement only works with linear, downward sloping demand and linear marginal cost. In our environment, such an approach does not refine the set of SFE equilibria.

- (iv) With a high load ratio, market power caused by a symmetrical reduction in capacity has a stronger effect on average prices than market power caused by asymmetry (predicted by RSI and midpoint-refined SFE). Formally:¹⁸

$$H4_0 : \bar{P}_{lsh} = \bar{P}_{hah} \quad vs. \quad H4_1 : \bar{P}_{lsh} > \bar{P}_{hah}.$$

The two remaining pair-wise comparisons pertain to the impact of changing the load ratio.

- (v) With a high symmetrical capacity, the change from a low to a high load ratio yields higher prices (predicted by RSI and both SFE refinements). Formally:

$$H5_0 : \bar{P}_{hsh} = \bar{P}_{hsl} \quad vs. \quad H5_1 : \bar{P}_{hsh} > \bar{P}_{hsl}.$$

- (vi) With a low symmetrical capacity, the change from a low to a high load ratio leads to higher prices (predicted by RSI, MUA and both SFE refinements). Formally:

$$H6_0 : \bar{P}_{lsl} = \bar{P}_{lsh} \quad vs. \quad H6_1 : \bar{P}_{lsh} > \bar{P}_{lsl}.$$

Observe that both the RSI and the midpoint-SFE refinement prescribe a directional shift for all the cases we consider. The prescriptions of the MUA and of the payoff-dominant SFE do not change for two of the parameter changes.

3. Results

We start with a broad overview of the prices we observe, followed by a general qualitative overview of the supply functions submitted by our subjects. Subsequently, we provide an analysis of the aggregate supply functions. We then present data on average volume weighted average prices, compare them with the equilibrium predictions and formally test our hypotheses H₁–H₆. In the latter part of the Section, we analyse individual best responses and assess the theoretical predictions of the two models.

To obtain a first impression of prices observed in our experiments and how these relate to the theoretical predictions, Figure 4 shows average prices across rounds separately for each market.¹⁹

To highlight the effects of learning, we distinguish between the average across all rounds and the average across the final five rounds. This figure shows that the average prices per market lie largely within the bounds predicted by SFE. In the absence of market power, the observations for *hsh* and *hsl* appear to be drawn towards the competitive prices predicted by MUA. For each of these two treatments, all groups but one had average prices at the MUA prediction for the last five rounds. For the

¹⁸ The RSI predicts a shift simply because the aggregate capacity that is left after a pivotal supplier withdraws his capacity from the total is smaller under *lsh* than under *hah*. The midpoint-SFE picks this up; some of the lower prices that are equilibrium for the *hah* treatment are not part of the equilibria for *lsh*. In contrast, the MUA model and the payoff-dominant SFE do not suggest a difference between these two cases; both ways of introducing pivotal power lead to the same (asymmetric) equilibrium with the highest possible price.

¹⁹ A further disaggregation, at the level of each round per group is available from the authors upon request.

Fig. 4. *Average Price per Market*

Notes. Markers denote the volume weighted average price per market across all rounds (triangles) or last five rounds (crosses). Rectangles and lines connecting rectangles denote supply function equilibrium (SFE) predictions. Ovals and the dashed line between ovals denote multi-unit auction predictions. The SFE midpoint refinement is found at the middle point of each SFE line. The payoff-dominant SFE is the 'highest' rectangle for each treatment.

treatments with market power, the observations appear to be more or less uniformly spread over the predicted interval of smooth SFE prices.²⁰

Our symmetrical, high-capacity treatments (*hsh* and *hsl*) are similar in some respects to the multi-unit sales auction experiments reported in Sefton and Zhang (2010). In their no-communication treatment, they find that subjects' bids converge to their values, which is consistent with pure strategy Nash equilibrium.²¹ By contrast, some offers remained above the discrete units Nash prediction of offers equal to marginal cost for some groups in our *hsh* and *hsl* treatments. Factors that might account for the differences in results include relatively greater excess capacity in Sefton and Zhang and a random demand quantity in our experiments compared to the fixed sales quantity in Sefton and Zhang.

Turning now to the supply functions, we first note that when submitting their individual supply functions, subjects typically submitted all units that they had available. In the low-capacity treatments (where each subject had a capacity of nine units) on average 8.9 units were offered at a price lower than or equal to $p^{\max}$ (=25). In the symmetrical high-capacity cases (12 units each) on average 11.7 units were offered. In the asymmetrical treatment *hah* (two firms with five units and two with 19) the low-capacity firms always offered all units whereas the firms with high capacity on average offered 18.6 out of 19 units at a price lower than or equal to 25. This is an indication

²⁰ Note that for *lsh* there is one group that has VWAP at the pure strategy MUA prediction of 25 in the last five rounds. In the MUA equilibrium, three firms offer all units at a low price and one firm offers all units at a price of 25. The data for this group reveal that there are two firms offering all units at a low price and two firms offering at the price cap.

²¹ There are multiple pure strategy equilibria in the Sefton and Zhang design, including some with collusive bidding (footnote 10).

that attempts to exert market power were done by offering units at high prices, not by withholding them altogether.²²

Figure 5 gives the average aggregate supply function per treatment, distinguishing between the low load ratio and high load ratio cases. In both panels, the ranking of the functions, in relation to the load ratio, is the same. The highest prices are asked for the low-capacity treatments *lsh* and *lsl* and the lowest for the symmetrical high-capacity cases *hsh* and *hsl*. The supply function for the asymmetrical capacity case lies somewhere in between, in the top panel.

Note that the average supply function for *hah* combines units supplied by small and large firms. Figure 6 separates the two.²³ The figure shows that the 10 units of the small firms are offered at higher prices than the first units of large firms. In equilibrium, all units offered by the small firms are (on average) sold for any demand realisation (which is between 30 and 35 in this treatment), however, because all are within the first 29 units of the aggregate supply function.

This feature holds for a large majority of the markets; considering each of the supply functions (one for each round of each session), we observe that 15.8% of the units offered by small firms end up at the 30th unit or beyond in the aggregate supply function. If small firms' units were distributed randomly across the aggregate supply function, 39.5% (19/48) of their units would end up beyond the 30th position in the aggregate function. We therefore conclude that the small firms tend to offer their units at relatively low prices, as predicted by the equilibrium for *hah*.

Table 4 shows volume weighted average prices both for all rounds and for the last five rounds, averaged over all groups of each treatment, together with the equilibrium predictions for the two models we consider.

We calculated the realised VWAP by a group in a round by using the aggregate submitted supply function and determining the expected VWAP across the possible demand realisations.²⁴ In cases where fewer units are offered at a price lower than the maximum than are demanded for some realisations, total supply is offered at the maximum price and we use the quantity traded to determine the weights.²⁵

Focus first on the two cases with high symmetrical capacity, *hsh* and *hsl*. For both load ratios, prices are within the range of the SFE predictions but below the two refinements and above the prediction of the MUA. Note also that prices are lower in the last five rounds than in earlier rounds. This means that they are moving away from the SFE refinements and in the direction of the MUA predictions. For low symmetrical capacity with the high load ratio, *lsh*, prices are again within the interval of the SFE. They are

²² Across all treatments, in 82.2% of the rounds the aggregate supply function offered the maximum total capacity at prices lower than or equal to 25.

²³ We are grateful to an anonymous referee for suggesting this analysis.

²⁴ Alternatively, one could use the realised demand in the five periods of the round. The outcome then depends on the realised draws of demand, however. Because we aim at testing hypotheses derived from strategies used and because the information about the strategies is contained in the aggregate supply function, we prefer to weight the supply prices with the expected demand.

²⁵ For example, assume for a high load case that only 34 units are offered. Units 30, 31, ..., 34 are offered at prices 10, 11, ..., 14 respectively. For demand realisation $d < 35$, these are also the (uniform) trading prices and all units are sold. If $d = 35$, then 34 units are sold at $p = 25$. The average quantity traded is then $1/6 \times (30 + 31 + \dots + 34 + 34) = 32.33$. The VWAP is $1/6 \times \{(30/32.33) \times 10 + (31/32.33) \times 11 + \dots + (34/32.33) \times 14 + (34/32.33) \times 25\} = 14.33$.

Fig. 5. Aggregate Supply Functions

Notes. The lines show the average supply function across all rounds for a high load ratio (top panel) and low load ratio (bottom panel). The window with dashed contour shows the area $[d^{\min}, d^{\max}]$ within which demand may vary. Quantities are indicated on the horizontal axes, prices on the vertical axes.

Fig. 6. Aggregate Supply Functions Small and Large Firms

Notes. The lines show the average supply function across all rounds for small (short line) and large (long line) firms. Quantities are indicated on the horizontal axis, prices on the vertical axis.

Table 4
Predicted and Actual Volume Weighted Average Prices

	Equilibrium predictions				Actual	
	MUA	SFE	SFE: midpoint	SFE: payoff dominant	All rounds	Last five rounds
High load ratio						
High Sym Cap (<i>hsh</i>)	[5, 7.5]	[5, 21.3]	13.2	21.3	10.53	9.30
Low Sym Cap (<i>lsh</i>)	≥18.5	[18, 21.3] and {25}	19.7	25	20.48	20.02
High Asym Cap (<i>hah</i>)	{25}	[11.2, 23.1] and {25}	17.2	25	14.47	13.47
Low load ratio						
High Sym Cap (<i>hsl</i>)	[5, 5.75]	[5, 16.4]	10.7	16.4	8.07	7.00
Low Sym Cap (<i>lsl</i>)	≥11.6	[12.4, 16.4]	14.4	16.4	16.56	18.02

close to the SFE midpoint refinement and below the SFE payoff dominant. They are also above the lower limit of the mixed strategy equilibrium support for the MUA. For *lsl*, observe that prices are somewhat above the upper limit of the SFE interval and therefore above both refinements. They are also above the lower limit of the mixed strategy equilibrium support. Finally, for the high asymmetrical case, *hah*, prices are within the SFE interval and below all point predictions.²⁶ Formal tests of differences in average VWAP are presented below, when we discuss the results for our hypotheses testing.

Given that average prices are (slightly) different in the final five rounds than across all rounds, the dynamics of the VWAP may be important. Therefore, we now examine how these prices changed across rounds. Figure 7 presents their development across rounds, separately for each treatment. Note the difference with Figure 4, which shows average prices across rounds separately for each market. Figure 7 aggregates observations across markets but disaggregates across rounds. Starting with the symmetrical treatments, this shows that prices for both low-capacity treatments *lsl* and *lsh* are substantially and consistently above those for the high-capacity treatments *hsl* and *hsh*. The differences increase over rounds: the primary reason is that average prices for high symmetrical capacity treatments decrease steadily. This decrease across rounds is statistically significant. Linearly regressing the VWAP on the round gives coefficient -0.13 for *hsl* and -0.14 for *hsh*, both with p -values <0.001 .

Comparing *hsh* to *hsl* one can see that prices for the former are above those for *hsl* in all rounds. In addition, average prices in these high symmetrical capacity experiments are above the (highest) MUA predictions of 7.5 (*hsh*) and 5.75 (*hsl*) in all rounds. Thus, aggregate behaviour in these high symmetrical capacity experiments appears to be inconsistent with MUA predictions, although prices are slowly moving towards the MUA prediction over time. We explore this issue further when we examine data from individual markets. The ordering of average prices in *hsh* and *hsl* would be consistent with a single upward sloping aggregate supply function in a SFE. This is true because

²⁶ The fact that prices stay away from the extreme predictions of the MUA could be attributed to a behavioural tendency not to choose prices at the edges of the choice space. However, it is worth pointing out here that in experiments with the double auction and the box demand design prices often do go all the way to the extremes (Davis and Holt, 1993).

Fig. 7. Development of Volume Weighted Average Prices

Note. For each treatment the graph shows the volume weighted average price at each round.

the low-demand realisations in *hsl* would cut the aggregate supply curve at prices below the clearing prices for *hsh*.

Average prices for the low symmetrical capacity treatments (solid lines in Figure 7) vary across rounds but tend to stay within to slightly above the intervals predicted for smooth SFE (Table 4). Note that average prices for *lsh* are inconsistent with the MUA prediction of 25. Their dynamics show no tendency towards this prediction (a linear regression of VWAP on round gives coefficient 0.05, with $p = 0.35$). The pure strategy equilibria from MUA for treatment *lsh* involve asymmetrical strategies in which one player earns a much lower payoff than the other three players. We emphasise these equilibria mainly because these are the only equilibria considered in the electricity market design paper by Fabra *et al.* (2006) and are the basis for several of their policy conclusions. However, results from experiments that share the feature of a symmetrical design with only asymmetrical pure strategy equilibria (e.g. the symmetric volunteer's dilemma, Goeree *et al.*, 2005) are often more consistent with a symmetrical mixed strategy equilibrium than with pure strategy equilibria. If a mixed strategy equilibrium exists for treatment *lsh* then expected VWAP for this equilibrium is bounded below by 18.5 (cf. Tables 3 and 4).²⁷ Average prices for *lsh* are consistent with this bound for expected VWAP. Average prices for *lsl* are consistent with the MUA prediction, in the sense that they are also above the lower bound prediction for the mixed strategy MUA equilibrium. There is a marginally significant upward trend (coefficient 0.06, with $p = 0.08$).

Finally, average prices for *hah* vary over rounds but tend to lie within the interval of equilibrium prices for smooth SFE. In later rounds, average prices are in the lower portion of this predicted interval. The decreasing trend is marginally significant (coefficient -0.09 , $p = 0.06$) Average prices for *hah* are clearly inconsistent with the

²⁷ The procedure for computing this lower bound for expected VWAP follows the procedure used in online Appendix B for treatment *lsl*.

pure strategy equilibrium prediction of 25 for MUA. If anything, they are converging away from this predicted level.

We conclude that the time trends in our data could be interpreted as convergence in the direction of the MUA prediction only in the symmetrical high-capacity cases. As for the two SFE refinements, a comparison between Figure 7 and the predictions in Table 3 reveals that the data do not appear to be converging towards either prediction in any of the treatments.

We now move to the tests of the hypotheses 1–6 about differences in prices across treatments presented in Section 3.²⁸ Table 5 presents the results of Mann–Whitney tests for all pair-wise differences in means across the five treatments. It takes the (volume weighted) average price per market (across all rounds) as the unit of observation. The p-values pertaining to our six hypotheses are shown in italics. For the hypotheses, we only need to consider these results. Observe that five of six of the differences in italics are statistically significant at the 10% level or better (four are significant at the 1% level) and therefore support the alternative hypotheses against the null of no differences in average prices.

We summarise the results of our hypotheses testing in the following way:

- (i) for H1–H3, the alternative hypotheses are supported as follows: symmetrically decreasing capacity (with either load ratio) and asymmetrically redistributing a given total capacity (with the high load ratio) all have the positive effects on prices predicted by the RSI and both theoretical models. In other words, when the RSI, MUA and at least one SFE refinement redistribute, all yield the same comparative static prediction, this is confirmed by our data.
- (ii) for H4, the alternative hypothesis is supported as follows: with a high load ratio, market power caused by a symmetrical reduction in capacity has a stronger effect than market power caused by asymmetry; this is in accordance with the hypothesis based on the RSI and the SFE midpoint refinement, while MUA and the payoff-dominant SFE are mute on this particular comparison.

Table 5
Pairwise Mann–Whitney Tests for Volume Weighted Average Prices

	<i>hsl</i>	<i>hah</i>	<i>lsh</i>	<i>lsl</i>
<i>hsh</i>	<i>0.165</i>	<i>0.089</i>	<i>0.004</i>	<i>0.009</i>
<i>hsl</i>	–	<i>0.002</i>	<i>0.002</i>	<i>0.001</i>
<i>hah</i>	–	–	<i>0.002</i>	<i>0.047</i>
<i>lsh</i>	–	–	–	<i>0.002</i>

Notes. Cell entries give the p-value for the Mann–Whitney test for the null hypothesis that the difference in means between the treatments in the row and column concerned are equal to zero. Group averages across rounds are taken as the units of observation. Results in italics are relevant for the hypotheses developed in Section 3, as explained in the text.

²⁸ Given the directional nature of our hypotheses, the p-values reported in Table 5 are based on one-tailed tests. All other p-values in this Section are based on two-tailed tests.

- (iii) for H5, the null hypothesis cannot be rejected: with high symmetrical capacity the change from a low to a high load ratio does not significantly affect prices, an effect predicted both by RSI and the two SFE refinements. Note that the VWAP for a low load ratio the possible demand realisations include the demand set for a high load ratio. In addition, lower levels of demand (i.e. between 20 and 29) are possible. With a non-decreasing aggregate supply function this may load the deck in favour of a higher VWAP when the load ratio is high. Although a formal test of H4 requires comparison across the two demand domains, for completeness' sake we also present the comparison for the case where demand is restricted to {30,...,35}. We then find that the average VWAP is still higher with a high load ratio (10.53 *versus* 8.59) but the difference is not statistically significant (Mann–Whitney, $N = 11$, $p = 0.165$).²⁹
- (iv) for H6, the alternative hypothesis is supported as follows: for low symmetrical capacity, the change in load ratio does lead to higher prices, an effect predicted by RSI, MUA and both SFE refinements. We again present the comparison for the case where demand is restricted to {30,...,35}. The average VWAP is again higher with a high load ratio (20.48 *versus* 19.57) but the difference is not statistically significant (Mann–Whitney, $N = 11$, $p = 0.396$).

Next, we consider heterogeneity in bidding. We focus on the firm that determines the supply price of the 35th unit, as this firm's supply schedule is likely to determine the market clearing price. If the supply price of the 35th unit is consistently chosen by one or two firms in an experiment, this suggests differences in bidding behaviour across firms. Alternatively, if this supply price is chosen roughly equally proportions by firms in an experiment, this suggests more similar bidding across subjects. Table 6 shows the extent to which the price of the 35th unit was determined by one or two firms in the market.

First, assume that each firm is equally likely to determine $s^*(35)$.³⁰ With such random positioning, one can derive 95% confidence intervals for the across-market averages given in Table 6.³¹ With $n = 5$ markets, and a 0.25 chance of any particular firm

Table 6
Firms Determining $s^(35)$*

	<i>hsh</i>	<i>hsl</i>	<i>hah</i>	<i>lsh</i>	<i>lsl</i>
One firm	0.48	0.56	0.70	0.56	0.66
Two firms	0.82	0.84	0.91	0.87	0.9

Notes. For each treatment denoted in the first row, numbers give the average (across markets) fraction of the 25 rounds that the price of unit 35 was determined by a single firm (2nd row) or two (of the four) firms (3rd row). In each market, the (two) firm(s) with the highest fraction is (are) used to determine this average.

²⁹ We thank an anonymous referee for suggesting this additional test.

³⁰ For the asymmetrical treatment, *hah*, one may expect the two large firms to alternate, yielding fractions 0.5 and 1 respectively.

³¹ The binomials needed to calculate these intervals are too involved to allow for analytical solution. Instead, we ran simulations to determine them. We thank Yang Yang for her help in doing so.

determining $s^r(35)$ in any round, the 95% confidence interval for the average fraction is $[0.312, 0.408]$; for $n = 6$ this is $[0.320, 0.400]$. Similarly, the 95% confidence interval for two firms determining the price with random positioning is $[0.592, 0.680]$ for $n = 5$ and $[0.593, 0.673]$ for $n = 6$. Given that all the observations in Table 6 lie outside the relevant intervals, we conclude that random positioning is not determining the price setter of $s^r(35)$.

On average there appear to be relatively small differences across the symmetrical treatments. 48%–66% of these prices are determined by a single firm in any market and two firms account for more than 80%. We conclude that there is a strong asymmetry in the bidding by distinct firms in a market. Even when all four firms have (equal) market power in *lsh* and *sl*, (almost) 90% of the prices at unit 35 are determined by only two of the four firms. While we clearly observe heterogeneous bidding by subjects, it is not consistent with the MUA predictions of high-price, asymmetrical equilibria for treatments *lsh* and *hah*. These equilibria involve low offers for all units by three subjects and all capacity of the fourth subject offered at the price cap. We did not observe this in *lsh* or *hah* markets. Note that the measure of asymmetry in bidding from Table 6 was higher for treatment *sl* than for *lsh*.

Now we take a closer look at the degree of rationality of the behaviour we observe. As we have detailed data on round-by-round offers submitted by subjects it is possible for us to assess the extent to which individual choices are best responses to choices made by rival subjects.³² We would not expect outcomes in the experiment to be consistent with Nash equilibrium predictions unless subjects are making best responses to rivals' choices. The game that subjects are playing is complex with a very large strategy set. Both MUA and SFE predict multiple equilibria. In addition, our subjects have only limited feedback regarding auction results. After each period (there are five periods per round) subjects observe the market clearing price, the quantity demanded and the position of their own offers in the aggregate offer queue; see Figure 1. Subjects do not directly observe the offers made by other subjects, although they may be able to infer approximate offers of rivals. Given the size of the strategy set, multiplicity of equilibria and limited information feedback, it is not at all obvious that subjects would play best responses in the experiments.

To assess individual choices we compare the actual profit of subjects to what we call *ex post*-optimal profit.³³ We calculate the *ex post*-optimal profit for a subject in a round of play by finding an offer that yields the highest possible profit given the actual offers submitted by other subjects for that round.³⁴ Note that a subject's *ex post*-optimal offer in a particular round need not be unique. For example, if rival subjects submit relatively high offers then a subject's best response would be any offer schedule that

³² Note that we consider here the round-by-round best response and neglect the possibility of strategic play across rounds (which would be very difficult to implement). Moreover, *ex post* (i.e. after realisation of demand), the optimal profit based on a best response may not be attainable.

³³ A similar approach of comparing actual profit to *ex post*-optimal profit was used in Hortacsu and Puller (2008) in their examination of behaviour of electricity generation suppliers in the Texas ERCOT wholesale power balancing market. We compare the results of our analysis to those of theirs, see below.

³⁴ The set of possible offer schedules for high-capacity subjects in *hah* is extremely large. For these subjects we approximate the best response in each round by sampling from the set of possible offers.

offers all units up to capacity at prices below rivals' offers, allowing the market price to be dictated by rivals' offers.³⁵

Table 7 summarises results for actual profit as a percentage of *ex post* optimal profit over all rounds for each subject. This figure ranged between 100% and 29%, with a median of 79% across all 112 subjects in the experiments. There are differences in actual/*ex post*-optimal profit across decision-making conditions. Differences across all symmetrical treatments are statistically significant ($KW, \chi^2 = 15.10, p = 0.00, N = 22$), as are differences between high and low-capacity subjects in the asymmetrical treatment ($MW, Z = 2.21, p = 0.03, N = 6$ paired observations). Subjects with a small amount of capacity (low-capacity subjects in *hah* and subjects in treatments *lsh* and *lsl*) have higher average actual/*ex post*-optimal profit than subjects with higher capacity (high-capacity subjects in *hah* and subjects in treatments *hsl* and *hsh*). More specifically, from high to low, the treatment-average actual/*ex post*-optimal profit is ordered as follows:

$$hah_{low} > 0.662 lsh > 0.792 lsl > 0.030 hah_{high} > 0.004 hsh > 0.004 hsl,$$

where hah_{low} (hah_{high}) refers to the low- (high-) capacity traders in *hah* and $>xx$ indicates the p-value of the Mann–Whitney test of the difference concerned (using market averages as units of observations). These tests confirm that the percentage of actual to *ex post*-optimal profit is significantly greater the lower the capacity per subject. It appears that the larger strategy sets associated with greater capacity contribute to a more complex decision-making environment and greater departures from optimality.³⁶ Finally, note that the numbers in euro show that the losses due to suboptimal responses are substantial.

Hortacsu and Puller (2008) conduct a similar analysis using market data and report on actual profit *versus ex post*-optimal profit for firms that offer electricity generation

Table 7
Actual Subject Profit as a Percentage of Ex Post-Optimal Profit

			No. of Units	Median	Max	Min
High load ratio	Symmetric	High capacity (<i>hsh</i>)	12	66 (€14.05)	79	42
		Low capacity (<i>lsh</i>)	9	84 (€12.79)	100	73
	Asymmetric	(<i>hah high-capacity trader</i>)	19	80 (€14.42)	96	61
		(<i>hah low-capacity trader</i>)	5	88 (€2.89)	100	61
Low load ratio	Symmetric	High capacity (<i>hsl</i>)	12	45 (€19.42)	66	29
		Low capacity (<i>lsl</i>)	9	85 (€7.97)	99	66
All subjects			–	79	100	29

Notes. Numbers represent average/*ex post*-optimal profit for the treatment concerned. *Ex post*-optimal profits were calculated in the way described in the main text. The column No. of units gives the maximum number of units each trader had available to offer. To see why a ratio larger than 100 may occur, see footnote 23. The amount in parentheses in euro denotes the median loss in euro's that follows from the median suboptimal choice.

³⁵ Similarly, for actual profit, we do not take the realised profit after demand realisation but the expected profit, given the subject's own supply curve, its rivals' actual aggregate supply curve and the distribution of demand quantities. This ensures that actual profits cannot exceed *ex post*-optimal profits. We thank an anonymous referee for suggesting this procedure.

³⁶ Note, however, that the percentage is significantly higher in the 19-unit *hah_{high}* case than in the 12 unit *hsh* and *hsl* cases. This may be due to the fact that subjects in *hah_{high}* have to deal with only one other large firm.

into the ERCOT wholesale power balancing market. It is possible to make these calculations because of the detailed information available about generation costs and about bids submitted by firms. They use data from a single trading period within each day (18:00–18:15) for days that did not experience transmission congestion across zones within ERCOT. The reported results for actual to *ex post*-optimal profit range from a high of 79% to a low of – 81%; the median figure for the sample of 35 firms is 15%. This contrasts with results from our experiments; approximately half of our subjects achieved higher actual to *ex post*-optimal profit than the firm with the highest percentage in the Hortacsu and Puller study.

Hortacsu and Puller (2008) also found differences in performance across firms but seemingly in the opposite direction of our experimental results. They find that large firms (those with a high volume of sales under *ex post*-optimal bidding) have significantly higher actual/*ex post*-optimal profit than smaller firms. Hortacsu and Puller attribute this result to the fixed costs associated with activities required to profit in the balancing market: acquiring information, analysing information and running a trading operation. The higher profit stakes available to larger firms made it worthwhile for them to invest in the fixed costs but the lower profit stakes for small firms left them with weak incentives to invest. By contrast, subjects in our experiments did not bear any costs of participating in the market except perhaps the opportunity cost of their attention.

This difference in results from the field and the laboratory are interesting.³⁷ They point to the particular advantages both methods have. On the one hand, the advantage of laboratory control is that it allows us to isolate causal effects when comparing realised-profit-to-optimal-profit ratios across distinct environments. They are less informative about the actual level of such ratios, however. For this, data from the field are more relevant. Moreover, our laboratory data allow us to isolate the effects of production capacity holding other characteristics (such as fixed costs) constant, while the field data allow for a comparison between large and small firms that take into account all differences between the two.

As a final exercise, we turn to an assessment of the theoretical predictions of market prices. The SFE and MUA models each provide equilibrium predictions of prices. We do not attempt to compare directly the predictive success of the MUA model to the complete set of equilibria predicted by the SFE model because the latter model predicts large intervals of equilibrium prices for every treatment and MUA equilibrium prices are a small subset of SFE prices. What we do instead is to simply report on test statistics for predictive success for MUA, for unrestricted SFE and for the two refinements of SFE distinguished above.

³⁷ Our laboratory market experiments differ from the ERCOT energy balancing market in several ways. Subjects in our experiments make repeated decisions in markets with stable costs and capacities, and a fixed group of participants. The ERCOT market environment is considerably more complex. It involves costs and capacities that change from day to day (e.g. due to generation outages, fuel cost changes and contractual commitments), several different types of generation, start-up and ramping costs for generation and changes in the set of market participants over time. Given these differences between our experiments and the ERCOT environment, we would not necessarily expect to see similar ratios of actual to optimal profit.

The following statistical model is the basis for the tests. Let,

p_{ij} = average VWAP for group i of treatment j , $j \in \{lsl, lsh, hsl, hsh, hah\}$

n_j = number of groups for treatment j ; $n = \sum_j n_j$.

The data generating process is assumed to be as follows:

$$p_{ij} = \tilde{p}_j + \varepsilon_{ij},$$

where $\tilde{p}_j$ is the latent, underlying price for treatment j and ε_{ij} is a zero mean random error term that reflects decision errors of subjects and/or aspects of subjects' payoffs that are not controlled in the experiment. Each theory makes predictions about values for the $\tilde{p}_j$ terms.³⁸ We assume that error terms are normally distributed with unknown variance, σ^2 ; let $f(\cdot)$ be the density function for error terms. Below, we develop likelihood ratio tests for the theories.

Each theoretical model makes predictions about prices for the five treatments. Let P_k be the set of prices predicted by model k for the five treatments; this set is a subset of the set of five-dimensional vectors of real numbers. Define the following likelihoods:

$$L_k = \left(\tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4, \tilde{p}_5^{\max} \right) \in P_k, \sigma \in \mathbb{R}_+ \prod_{j=1}^5 \prod_{i=1}^{n_j} f(p_{ij} - \tilde{p}_j)$$

$$L = \left(\tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4, \tilde{p}_5^{\max} \right) \in \mathbb{R}^5, \sigma \in \mathbb{R}_+ \prod_{j=1}^5 \prod_{i=1}^{n_j} f(p_{ij} - \tilde{p}_j). \quad (1)$$

L_k is calculated by choosing latent prices from the set of predicted prices for the model that are closest to average prices for treatments and by choosing an error variance equal to the average squared deviation from the best predicted prices. L gives the unrestricted maximum likelihood; latent prices are equal to treatment-average prices and error variance is equal to sample variance. The likelihood ratio statistic for model k is $\lambda_k = L_k/L$. For large n , the distribution of the test statistic $-2\ln(\lambda_k)$, approaches the chi-square distribution with 6 degrees of freedom. Table 8 reports results for tests of

Table 8
Likelihood Ratio Tests

Model		Test statistic $-2\ln(\lambda_k)$	$\chi^2_{0.05,6}$
MUA		52.58	12.59
SFE	Unrestricted	0.01	12.59
	Midpoint	19.98	12.59
	Pareto dominant	69.52	12.59

Notes. Test results are based on the likelihood functions described in (1). Prices used are group averages, therefore $n = 28$ observations.

³⁸ The data generating process could be modified so that actual prices are the minimum of the RHS and the price cap. This would be important if the data (group average prices) included observations equal to the price cap. Price is equal to the price cap for some individual rounds in experiments but none of the observed group average prices are at the price cap.

the hypothesis that latent prices for treatments are in the set of equilibrium prices predicted by theoretical models. We use a 95% confidence interval for the tests. Values of the test statistic that are larger than $\chi^2_{0.05,6}$ therefore reject the null hypothesis that our data are generated by the model concerned.

We observe that the only model that is not rejected by the test is the unrestricted SFE model, for which we calculate $\lambda_{\text{SFE}} = 0.993$, yielding a test statistic close to zero. This result reflects that there are SFE equilibrium price predictions that are either equal to or close to average observed prices for each treatment. For the restricted SFE models and the MUA model we find that the test statistics exceed the χ^2 statistic and therefore reject the hypothesis that observed group average prices are drawn from distributions with mean values in the equilibrium sets for these models.³⁹ As mentioned above, this higher predictive power in the unrestricted SFE model comes at a cost, however. The unrestricted SFE model is far less parsimonious than either of its refinements or the MUA alternative. The bottom line remains, however, that we have found no more parsimonious model capable of organising our data in a satisfactory way.

4. Concluding Discussion

We set out to study the effects of pivotal power experimentally, motivated both by the results of empirical field data studies and by the predictions of recent theoretical models. A first conclusion from our experiments is that the more fundamental intuitions about the impact of pivotal power are supported by our data. Prices are higher when (some) firms are pivotal. The existence of aggregate excess capacity is not enough to guarantee competitive prices. This general result is in accordance both with the predictions of the intuitive RSI and of the SFE model based on divisible output and MUA model based on discrete output units.

Our experiments also permit us to assess the predictive power of these theoretical models in more detail, as well as of two refinements of the SFE model. The MUA model provides sharp predictions for three of our five treatments; for these treatments pure strategy equilibria of the MUA predict either near competitive pricing or monopoly pricing. The SFE model predicts a larger set of equilibria that includes the sharp MUA equilibria as well as additional equilibria based on upward sloping supply functions. One can see the SFE model prescriptions as a more modest proposal for organising observed behaviour, in contrast to the more stringent prescriptions of the MUA model and of the two refinements of SFE that we considered. On the other hand, as we argued above, this more general SFE model is not very parsimonious. It appears that none of the three ways that we tried to achieve more parsimony was able to maintain the SFE's ability to capture our data. Naturally, one could consider alternative refinements. We can think of no obvious one, however. As discussed above, the MUA and the two we consider come easily to mind.

We find that the additional equilibria of the SFE model proved necessary for explaining observed behaviour in two principle ways. First, for treatments with no

³⁹ The likelihood for MUA is computed using the lower bound for VWAP from a mixed strategy equilibrium condition for treatment *lsh*. If we restrict MUA predictions to those from pure strategy equilibria for *lsh* then the test statistic is higher and MUA is rejected more decisively.

market power, observed supply functions are upward sloping and prices tend to remain above the (near) competitive price prediction of the MUA. The multiple equilibria of the SFE capture this behaviour. Neither of the two (other) SFE refinements we considered captures these patterns, however. The movement of prices towards marginal cost for most markets in two of our treatments does appear to show dynamics in the general direction of the MUA prediction. Only in this weak way does the basic insight obtain support in our data, that the 'high-price' equilibria that arise with infinitely divisible offer prices and quantities are eliminated once some discreteness is introduced.

Second, pure strategy equilibria of the MUA model predict monopoly pricing (at the price cap) for two of our market power treatments. Observed behaviour in these two treatments is inconsistent with this sharp MUA prediction; behaviour is more consistent with additional equilibria from the SFE that involve upward sloping supply functions and lower market prices. Once again, it is inconsistent with the two SFE refinements, however. As mentioned above, we do not think that the fact that prices stay away from the extreme predictions of the MUA can be simply explained by a behavioural tendency not to choose prices at the edges of the choice space. In double-auction experiments with the box demand design, prices often do go all the way to the extremes. An additional consideration here is that our design involves a repeated game, albeit a finite one. Given that repeated interaction can facilitate tacit collusion in experimental oligopoly settings (Abbink and Brandts, 2008), it is noteworthy that in our case it does not lead to the attainment of a monopoly price one-shot equilibrium.

As noted in the introduction, market power due to the presence of pivotal suppliers has been documented to contribute to high prices and inefficiency in wholesale electricity markets and is a significant concern for public policy towards the electric industry. Our experimental results are consistent with evidence from naturally occurring electricity markets that pivotal power contributes to higher market prices. An important finding, however, is that the exercise of market power by pivotal suppliers in our experiments was not as severe as equilibrium predictions of the MUA model. These predictions require that agents adopt strategies that support an asymmetrical equilibrium with payoffs that differ substantially across agents (even for agents with identical costs and capacities). Experimental results for treatments with pivotal suppliers were more consistent with SFE predictions involving lower prices.

A final result that we wish to highlight here pertains to the effects of increasing the load ratios. We find that when both models suggest that it will affect prices it does have this effect. We interpret this as indicating that the models do identify a potential influence factor but that it only shows up in the data when it is a strong force. Stronger variations in demand reduce market power, in a situation where this power is otherwise strongest.

Universitat Autònoma de Barcelona and Barcelona GSE

University of Arizona

University of Amsterdam

Submitted: 22 October 2011

Accepted: 17 May 2013

Additional Supporting Information may be found in the online version of this article:

Appendix A: A Transcript of the Instructions (Translated from the Dutch).

Appendix B: Equilibria of the MUA.

Appendix C: Equilibria of the SFE Model.

References

- Abbink, K. and Brandts, J. (2008). '24. Pricing in Bertrand competition with increasing marginal costs', *Games and Economic Behavior*, vol. 63(1), pp. 1–31.
- Anwar, A.W. (1999). 'The case for a discriminatory pricing rule in competitive electricity pools', Working Paper, Department of Economics, University of Edinburgh.
- Baldick, R., Grant, R. and Kahn, E. (2004). 'Theory and applications of linear supply function equilibrium in electricity markets', *Journal of Regulatory Economics*, vol. 25(2), pp. 143–67.
- Bolle, F. (2001). 'Competition with supply and demand functions', *Energy Economics*, vol. 23(3), pp. 253–77.
- Borenstein, S., Bushnell, J. and Wolak, F. (2002). 'Measuring market inefficiencies in California's restructured wholesale electricity market', *American Economic Review*, vol. 92(5), pp. 1376–405.
- Brandts, J., Pezanis-Christou, P. and Schram, A. (2008). 'Competition with forward contracts: a laboratory analysis motivated by electricity market design', *Economic Journal*, vol. 118(525), pp. 192–214.
- Brunner, C., Goeree, J., Holt, C. and Ledyard, J. (2010). 'An experimental test of flexible combinatorial spectrum auction formats', *American Economic Journal: Microeconomics*, vol. 2(1), pp. 39–57.
- Cox, J., Olferman, T., Olson, M. and Schram, A. (2002). 'Competition for vs. on the rails: a laboratory experiment', *International Economic Review*, vol. 43(3), pp. 709–35.
- Davis, D. and Holt, C. (1993). *Experimental Economics*, Princeton, NJ: Princeton University Press.
- Davis, D. and Holt, C. (1994). 'Market power and mergers in laboratory experiments with posted prices', *RAND Journal of Economics*, vol. 25(3), pp. 467–87.
- Fabra, N., von der Fehr, N.-H. and Harbord, D. (2006). 'Designing electricity auctions', *Rand Journal of Economics*, vol. 37(1), pp. 23–46.
- Falk, A. and Heckman, J. (2009). 'Lab experiments are a major source of knowledge in the social sciences', *Science*, vol. 326(5692), pp. 535–8.
- Federal Energy Regulatory Commission. (2007), 18 CFR Part 35 (Docket No. RM04-7-000; Order No. 697), June 21.
- Genc, T. and Reynolds, S. (2011). 'Supply function equilibria with capacity constraints and pivotal suppliers', *International Journal of Industrial Organization*, vol. 29(4), pp. 432–42.
- Goeree, J., Holt, C. and Moore, A. (2005). 'An experimental examination of the volunteer's dilemma', Working Paper, University of Virginia.
- Green, R. (1999). 'The electricity contract market in England and Wales', *Journal of Industrial Economics*, vol. 47(1), pp. 107–24.
- Holmberg, P., Newbery, D. and Ralph, D. (2008). 'Supply function equilibria: step functions and continuous representations', Electricity Policy Research Group WP#EPRG0829, Cambridge: University of Cambridge.
- Hortacsu, A. and Puller, S. (2008). 'Understanding strategic bidding in multi-unit auctions: a case study of the Texas electricity spot market', *RAND Journal of Economics*, vol. 39(1), pp. 86–114.
- Kagel, J. and Roth, A. (2000). 'The dynamics of reorganization in matching markets: a laboratory experiment motivated by a natural experiment', *Quarterly Journal of Economics*, vol. 115(1), pp. 201–35.
- Klemperer, P. and Meyer, M. (1989). 'Supply function equilibria in oligopoly under uncertainty', *Econometrica*, vol. 57(6), pp. 1243–77.
- Kremer, I. and Nyborg, K. (2004). 'Divisible good auctions: the role of allocation rules', *RAND Journal of Economics*, vol. 35, pp. 147–59.
- Newbery, D. (1998). 'Competition, contracts, and entry in the electricity spot market', *RAND Journal of Economics*, vol. 29(4), pp. 726–49.
- Niu, H., Baldick, R. and Zhu, G. (2005). 'Supply function equilibrium bidding strategies with fixed forward contracts', *IESE Transactions on Power Systems*, vol. 20(4), pp. 1859–67.
- Rahimi, A. and Sheffrin, A. (2003). 'Effective market monitoring in deregulated electricity markets', *IEEE Transactions on Power Systems*, vol. 18(2), pp. 486–93.
- Rassenti, S., Smith, V. and Wilson, B. (2002). 'Using experiments to inform the privatization/deregulation movement in electricity', *Cato Journal*, vol. 21(3), pp. 515–21.
- Rassenti, S., Smith, V. and Wilson, B. (2003). 'Controlling market power and price spikes in electricity networks: demand-side bidding', *Proceedings of the National Academy of Science*, vol. 100(5), pp. 2998–3003.

- Reitzes, J., Pfeifenberger, J., Fox-Penner, P., Basheda, G., Garcia, J., Newell, S. and Schumacher, A. (2007), 'Review of PJM's market power mitigation practices in comparison to other organized electricity markets', Report by the Brattle Group for PJM Interconnection, LLC.
- Sefton, M. and Zhang, P. (2010), 'Divisible-good uniform price auctions: the role of allocation rules and communication among bidders', Working Paper, University of Nottingham.
- Sioshansi, R. and Oren, S. (2007), 'How good are supply function equilibrium models: an empirical analysis of the ERCOT balancing market', *Journal of Regulatory Economics*, vol. 31(1), pp. 1–35.
- Staropoli, C. and Jullien, C. (2006), 'Using laboratory experiments to design efficient market institutions: the case of wholesale electricity markets', *Annals of Public and Cooperative Economics*, vol. 77(4), pp. 555–77.
- Vives, X. (2011), 'Strategic supply function competition with private information', *Econometrica*, vol. 79(6), pp. 1919–66.
- Wolak, F. (2009), 'An assessment of the performance of the New Zealand wholesale electricity market', Investigation report, May 2009, available at <http://www.comcom.govt.nz/BusinessCompetition/Publications/Electricityreport/DecisionsList.aspx> (last accessed: 3 July, 2013).
- Wolfram, C. (1999), 'Measuring duopoly power in the British electricity spot market', *American Economic Review*, vol. 89(4), pp. 805–26.