Forthcoming, *Rand Journal of Economics*

Retail Electricity Competition[*]

Paul Joskow [†] Jean Tirole[‡]

November 22, 2005

Abstract

We analyze a number of unstudied aspects of retail electricity competition. We first explore the implications of load profiling of consumers whose traditional meters do not allow for measurement of their real time consumption. In general, the combination of retail competition and load profiling does not yield the second best prices given the non price responsiveness of consumers. By contrast, when consumers have real time meters and are billed based on real time prices and consumption, retail competition yields the Ramsey prices even when consumers can only partially respond to variations in real time prices. We then examine the incentives competitive retailers have to install one of two types of advanced metering equipment. Competing retailers overinvest in real time meters compared to the Ramsey optimum, but the investment incentives are constrained optimal given load-profiling and retail competition. Finally, we consider the effects of physical limitations on the ability of system operators to cut off individual customers. Competing retailers have no incentive to determine the aggregate value of non-interruption of consumers in the zones they serve, preferring instead to free ride on other retailers serving consumers in the same zones.

1 Introduction

Competition in the supply of electricity to retail consumers is being introduced in many countries. In the U.S. more than a dozen states have introduced retail electricity competition and the European Union has issued directives requiring that retail electricity markets be opened to competitive supply in all member countries. Yet there has been little study of the properties of retail electricity competition in the context of some of the unusual attributes of electricity supply, demand, pricing and metering. The paper analyzes a number of hitherto unstudied aspects of retail competition in electricity markets. Its starting point is that final consumers may not react to the real time prices that emerge in wholesale electricity markets for (at least) three reasons: First, they do not have incentives to properly adjust their consumption to real-time prices (RTPs) if only their total consumption over a given period is recorded, i.e., they

[*] We are grateful to Severin Borenstein, Claude Crampes, Raymond Deneckere, Alvin Klevorick, Bruno Jullien, Stephen Holland, Patrick Rey, an anonymous referee, and the participants at the IDEI-CEPR conference on "Competition and Coordination in the Electricity Industry," January 16–17, 2004, Toulouse and at the ninth annual POWER conference, UC Berkeley, March 19, 2004, for helpful discussions and comments.

[†] Department of Economics, and Center for Energy and Environmental Policy Research, MIT.

[‡] IDEI and GREMAQ (UMR 5604 CNRS), Toulouse, and MIT.

are on a traditional meter. Second, even if their consumption is recorded on a real-time (RT) basis, transaction costs associated with monitoring the evolution of hourly prices and constantly optimizing the use of equipment may be large for small consumers, although they may be reduced by Internet-based technologies in the future. Third, consumers, even if they want to, may not be able to adjust their consumption freely. They may be constrained by the physical attributes of distribution networks as they are presently configured; in particular, rationing usually occurs at the level of zones rather than individual consumers (what we refer to as the joint interruptibility problem).

In order to analyze competition among electricity retailers or Load Serving Entities (LSEs) for the final (retail) consumers, it is convenient to refer to the taxonomy of Table 1.

INSERT TABLE 1 ABOUT HERE

Price-sensitive consumers, who are studied in standard theory, are endowed with real-time (RT) meters and either autonomously or through communication with the LSE, adjust their demand efficiently to the evolution of the wholesale spot market price.

Price-insensitive or partially price sensitive consumers with real-time meters are endowed with RT meters, but are unaware or only partially aware of RT prices and therefore do not adjust their consumption perfectly as real time prices vary from minute to minute and hour to hour. At the extreme, they are fully (RT) price-insensitive. Such consumers are not irrational; rather they trade off the transaction costs outlined above and the savings in their electricity bill. While these consumers do not react to real time prices their actual real time consumption can be measured and assigned to their LSE for wholesale market settlement purposes.

Consumers on traditional meters are metered only once a month or every few months (in some countries meters are read even less frequently), and pay a per-kWh electricity charge that is independent of the actual timing of their overall consumption. The case of consumers on traditional meters can be decomposed into two subcases, depending on the way the consumers' LSE is charged for its energy purchases.

In the case of a *monopoly* local distribution company, this company pays the real-time wholesale market price of the consumer's consumption: Even though the LSE is then unable to measure the realized profile of any given consumer with a traditional meter in its distribution area, it observes and pays for the costs it incurs in the wholesale markets associated with the realized total consumption profile of all such customers in the area in the aggregate.

Under *retail competition* by contrast, an LSE other than the local monopoly distribution grid owner and serving such a consumer pays a unit electricity charge based on the "load profile" of the consumer. That is, it pays *the average wholesale price for the load profile* that is representative of the consumer's class regardless of the actual time pattern of the individual customer's consumption and the relationship between this actual physical consumption and the contemporaneous RT wholesale prices.

Note that Table 1 does not consider cases in which the customer is endowed with a RT meter, yet the retailer is subject to load profiling; presumably, the presence of a RT meter at the customer level also allows the LSE to be charged for and pay the actual wholesale market costs associated with the customer's RT profile.[1]

The paper focuses on three possible failures of price signals to adequately reflect the scarcity conveyed by real time wholesale market price signals. The first failure arises at the consumer level when only her aggregate consumption is measured because she is not on a RT meter. Because the consumer then does not pay more when consuming mainly at peak when wholesale prices are high than when spreading consumption more equally across peak and off-peak hours, the consumer consumes too much at peak and too little off peak. By contrast with the non-responsiveness created by the absence of real time consumption measurement, non-responsiveness generated by the consumer's transaction costs is shown not to result in a market failure. The second failure occurs at the retailer's level, when the latter's individual consumers' real time intake of power purchased in the wholesale market again is not measured by the system, so that the retailer is charged for power purchased in the wholesale market on the basis of some estimated consumption load profile rather than the LSE's consumers' actual load profile. The third failure may arise from the joint interruptibility problem.

Section 2 analyzes the behavior of load serving entities (LSEs) in a world in which consumers are homogeneous (possibly up to a scale parameter) and on traditional meters. Section 2.1 characterizes the second-best optimum and shows that it can be implemented in the absence of retail competition provided that the monopoly retailer is permitted to use two-part tariffs and, in particular is not constrained to use linear prices.

By contrast, Section 2.2 shows that (with or without the incumbent distributor) retail competition, which implies load profiling of consumers with traditional meters, leads to a retail price equal to the average wholesale power cost and differing from the socially optimal retail price. In a sense, retail competition plus load profiling transforms a second-best situation (with consumer

[1] Our grouping is an oversimplification. There are a number of partially price sensitive categories, such as those subject to time-of-use pricing (retail prices are preset for certain blocks of time) or critical peak pricing (that combines time-of-use pricing with high retail prices for a number of critical hours per year to be declared by the utility). See Borenstein et al (2002) for a review of recent innovations.

free riding on the RT structure) into a third best one (in which both consumers and LSEs free ride on the RT structure).

Section 3.1 takes on the case in which consumers are on real-time meters, but, in order to economize on transaction costs, do not react or only partially react to the real-time prices. We find that with homogeneous and rational consumers retail competition leads to the second-best optimum; this is no longer true when a uniform pricing constraint is imposed (section 3.2). Section 3.3 compares and contrasts our results to those in Borenstein and Holland (2003), hereinafter BH.

Section 4 extends the analysis to situations in which consumers differ in other aspects than just scale, i.e., they have different load profiles, and investigates the possibility of adverse selection and competitive screening.

Section 5 shows that given the price inefficiencies associated with load profiling, LSEs face the right incentives when offering their customers enhanced metering equipment, and so subsidies for such equipments are not warranted.[2]

Last, Section 6 analyzes the implications of limitations in the controllability of the distribution circuits. These limitations imply that price sensitive consumers may be rationed along with everyone else, and that LSEs cannot generally demand any specific level of rationing that they desire to reflect their consumers' valuations. At best one can then elicit only the aggregate willingness to pay for reliability in any given joint interruptibility zone. The section discusses both market mechanisms that are needed to reach a "third best" and the difficulties that make the phasing out of non-market mechanisms unlikely when there is retail competition.

We re-examine and extend results in BH. BH analyze a model with identical retail consumers served by competing LSEs (retail competition). A fraction of these retail consumers have real-time meters and are billed based on the real time price. The rest of the retail consumers served by LSEs pay a uniform price for all of their consumption. BH also assume that competing LSEs are all billed based on the real time consumption and associated real time wholesale prices for all of the retail customers they have signed up to serve.[3] Finally, BH assume that retailers must charge linear prices – they are not permitted to use two-part tariffs.

Two criticisms can be levelled at this set of assumptions. The first is that non-linear prices are both common and more efficient than linear ones. Constraining retail suppliers to offer only linear prices, rather than allowing them potentially to emerge as a consequence of retail competition, is a rather restrictive assumption. The second concerns the absence of real-time pricing for a fraction of consumers. The natural rationale for this assumption is that these

[2] We do not consider metering technologies where there are economies of scale or density.

[3] This assumption is implicit in the profit function on page 4 of Borenstein and Holland (2003).

consumers have traditional meters that do not measure real time consumption. This rationale however seems inconsistent with the other assumptions that there is retail competition and that competing LSEs can be billed for the wholesale market costs associated with the real time consumption of their retail customers. If there are competing LSEs it is not possible to measure their aggregate real time wholesale market supply responsibilities unless all of their customers are on real time meters. When some consumers have traditional meters, the LSEs' responsibilities for the cost of wholesale power must be based in part on load profiling. Accordingly, we allow retail suppliers to offer two part tariffs and examine whether competitive equilibria support two-part tariffs under a variety of assumptions. These assumptions lead to a number of results that differ from those in BH: see Section 3.3 for a detailed comparison. We also consider a number of additional cases that are not covered by Borenstein and Holland, including cases where consumers have real time meters but respond only partially to real time prices, where consumers are heterogeneous, and where non-price rationing must be accomplished on a zonal basis rather than for individual consumers.

2 Consumers on traditional meters

2.1 Model and social optimum

States of nature (or, equivalently, periods) are indexed by $i \in [0,1]$. Let f_i denote the frequency of state i and $E[.]$ denote expectations with respect to this distribution. Because we focus on competition on the demand side, we take the wholesale prices as exogenous, and we identify states of nature by the wholesale price p_i, with p_i increasing in i. Taking the wholesale prices as exogenous does not reflect any implicit assumption on the supply elasticities. However, the normative Ramsey analysis (but not the retail competition equilibrium) will assume that the wholesale prices are competitive. If generators had market power, the choice of the demand-side institution could affect this market power and so a complete comparison of the properties of alternative demand-side institutions would reflect their properties in terms of market-power alleviation (needless to say, supply-side market power could alternatively be addressed directly through supply-side policies such as forward sales requirements).

For the sake of simplicity, let us ignore rationing for the moment. We consider a representative retail consumer with gross surplus $S_i(q_i)$ from consuming q_i in state i . The consumer is on a traditional meter and has demand $D_i(p)$ in state i for (uniform) price p, with $S_i'(D_i(p)) = p$ and $D_i' < 0$. Note that consumers are assumed to be homogeneous. They may differ in the size of their demand, though: That is, they can be indexed by $\sigma > 0$, such that a consumer of type σ has demand $q_i = \sigma D_i(p)$ and gross surplus $\sigma S_i(q_i/\sigma)$. We normalize σ to be equal to 1, but nothing is changed if consumers differ in their scale σ.[4] [More general forms of heterogeneity are discussed in Section 4.]

Assumption 1. The function $E\left[\left(p-p_i\right)D_i'(p)\right]$ is decreasing in p.[5]

The retail consumers are physically served by a local monopoly regulated grid owner (usually also called the incumbent distributor, or transmission and distribution service provider). Because we are not interested here in the price of access to the grid, we normalize to zero any delivery, metering and customer service costs that continue to reflect responsibilities of the distribution grid owner. Thus, when we later introduce LSEs, their only cost will be either the purchase of energy from the wholesale market in real time or, in the case of load-profiled consumers, the load profiled variable charge for power supplied from the wholesale market to be paid for power delivered by the local grid owner. We consider two-part tariffs, consisting of a monthly subscriber charge and a per-kWh variable charge.[6] We will later note that focusing on two-part tariffs involves no loss of generality.

A consumer on a traditional meter cannot obtain the first-best utility, U^{FB}, that she would obtain if her demand were controlled to perfectly adjust to the RTP:

$$U^{FB} \equiv E\left[S_i\left(D_i\left(p_i\right)\right)-p_iD_i\left(p_i\right)\right]. \tag{1}$$

A Ramsey social planner for consumers with traditional meters chooses prices, namely single per unit retail price p^* and fixed fee A^*, so as to maximize the consumer's expected net surplus subject to the budget balance constraint:

$$\begin{aligned} U^* \quad \equiv \quad & \max_{\{p,A\}} E\left[S_i\left(D_i\left(p\right)\right)-pD_i\left(p\right)\right]-A \\ & \text{subject to} \\ & \qquad E\left[\left(p-p_i\right)D_i\left(p\right)\right]+A\geq 0. \end{aligned} \tag{2}$$

At the optimum, the budget constraint is binding, and the Ramsey planner maximizes the joint surplus:

$$U^* = \max_{p} E\left[S_i\left(D_i\left(p\right)\right)-p_iD_i\left(p\right)\right], \tag{3}$$

[4] Neither the social planner nor the LSEs need to observe the consumer's scale σ in advance: They can infer it ex post from the consumer's total consumption.

[5] This assumption is made mainly for analytical convenience. It is satisfied in particular if the demand functions' curvature is small enough ($|D_i''/D_i'|$ small).

[6] Offers by retailers to residential customers in England and Texas that we have reviewed have a fixed monthly charge plus one or more tiers of kWh charges.

showing that the price p^* is a weighted average of the real-time prices:

$$E\left[\left(p^* - p_i\right) D_i'\left(p^*\right)\right] = 0. \tag{4}$$

Assumption A1 implies that (4) has a unique solution.

To get some feel for what the Ramsey price entails when consumers have traditional meters and pay a marginal price p^* (plus a fixed fee A), suppose for example that the elasticity of demand comes from the installation and operation of air conditioning units (used only during the summer months). Suppose further that there are only two periods: winter and summer, have respective wholesale prices p_1 and p_2 ($p_1 < p_2$). In this simple illustrative example demand is responsive to prices during the summer but is completely inelastic during the winter. In the U.S. the summer is the peak period and the winter is the off-peak period. Then, the Ramsey price is $p^* = p_2$ in the US. In France $p^* = p_1$, since summer is off peak in France.[7] Thus the Ramsey price would be greater than the average annual wholesale price of electricity in the US and below the average annual wholesale price in France. Retailer budget balance is achieved with the fixed component A of the two-part tariff.

Let us now consider the case of an LSE whose energy purchase cost corresponds to its customers' actual load profile. As we have argued, this is the case when customers in an area are served by a monopoly distribution company. The LSE then chooses the two-part tariff (p, A) so as to maximize its profit

$$E\left[\left(p - p_i\right) D_i(p)\right] + A$$

subject to the consumers being granted a certain utility level $\overline{U}$ (0 if the monopoly is unregulated, higher if regulated):

$$E\left[S_i\left(D_i(p)\right) - pD_i(p)\right] - A \geq \overline{U}.$$

This program is of course the dual of the Ramsey program above. We thus obtain:

Proposition 1 Traditional meters give rise to consumer moral hazard: Relative to their RTP-based demand, consumers consume relatively too much on peak and too little off peak.

(i) The Ramsey usage price is given by

$$E\left[\left(p^* - p_i\right) D_i'\left(p^*\right)\right] = 0.$$

(ii) The Ramsey (second-best) consumption prevails under a monopoly LSE (the Ramsey surplus allocation further holds if regulation extracts the LSE's profit).

[7] If "installation and operation" referred to electric heating, then $p^* = p_2$ in France since winter is the peak period.

Remark (optimality of two-part tariffs): We have assumed that the Ramsey planner offers two-part tariffs. With traditional meters, the social planner (or an LSE for that matter) actually cannot do better through more complex pricing structures than with a two-part tariff. At best he can hope to control total consumption through the marginal charge, while the load curve is chosen by the consumer without any concern for the actual cost of purchasing energy. More formally, the social planner is limited to total-consumption based tariffs $T(Q)$. Suppose that the planner selects the consumer's total consumption Q, and charges an amount T for this. The consumer then chooses her load curve so as to solve:

$$\max E\left[S_i\left(D_i\right)\right] \quad \text{subject to} \quad E\left[D_i\right] = Q.$$

Letting p denote the shadow price of the constraint, $S_i'\left(D_i\right) = p$, and so the allocation is the same as under a two-part tariff.

2.2 Retail competition for load-profiled consumers: independent retailers

We analyze the competitive outcome with load profiled customers in two environments. In the first, the local grid owner is subject to a line-of-business restriction. He provides access or delivery service to retailers, but is not allowed to compete for the final consumer. In the second, this line-of-business restriction is lifted and so the incumbent distributor is permitted to compete with independent retailers. We assume either that the distributor separates its retail supply business into a ring-fenced affiliate that is treated like any other retailer (as in the UK and in Texas), or that the retail arm maximizes the profit of the vertically integrated firm.

In this subsection, we assume that (pure) retailers, but not the local regulated monopoly grid owner, compete for load-profiled consumers and can offer two-part tariffs if they choose to do so.[8] Retailers' settlement obligations for wholesale power costs are then based on their customers' load-profiled consumption. The aggregate demand of all consumers served through a particular distribution network is measured on a real time basis. Since the aggregate real time consumption obligations must add up to the aggregate real time supplies of power delivered over the distribution network, a set of "load profiles" must be applied to the monthly, bi-monthly or quarterly consumption measured for customers without real-time meters. For example, consider a customer with a standard meter read on a monthly basis with 1000 kWh of consumption recorder for the previous month. The 1000 kWh of monthly consumption then must be allocated to the 720 hours of the previous month for settlement purposes. This is accomplished by

[8] We are interested solely in the price effects of retail competition. We thereby ignore some benefits of competition (such as improved incentives to offer better metering, tariffs, total energy management services or hedging packages) as well as some potential costs of retail competition (such as consumer churn and poaching, duplicative or misleading advertising expenditures, and competitive screening for credit quality and high volume consumers).

assigning the customer to a group or class of customers thought to have similar consumption. A consumption or load profile is developed for each group based on real-time metered consumption patterns of a sample of customers in each class. To compute the "average wholesale power cost" a, suppose that, in equilibrium, retailers' variable (per-kWh) charge to consumers is p. Average consumption (load profiled) per consumer is $E\left[D_i\left(p\right)\right]$ and the wholesale price paid by the retailers for energy is

$$a\left(p\right)=\frac{E\left[p_iD_i\left(p\right)\right]}{E\left[D_i\left(p\right)\right]}. \tag{5}$$

We use the notation a for "access charge" by analogy with the economics literature on variable charges paid by entrants for access to regulated bottlenecks (local loop, etc.).[9] This access charge must be understood as the average wholesale power cost paid by retailers.

Let $\widehat{p}$ denote the price that balances the LSE's budget in the absence of fixed charge, when they pay the average wholesale power cost to the grid: $\widehat{p}=a\left(\widehat{p}\right)$, or:

$$E\left[\left(\hat{p}-p_i\right)D_i\left(\hat{p}\right)\right]=0.$$

Intuitively, $\hat{p}$ exceeds the Ramsey price p^* if the state of nature impacts average demand more than it does marginal demand. We will therefore be led to consider three cases:

Case 1: $$\frac{E\left[p_iD_i(p)\right]}{E\left[D_i(p)\right]}>\frac{E\left[p_iD_i'(p)\right]}{E\left[D_i'(p)\right]}. \qquad \text{for all } p.$$

In this case, $p^*<\hat{p}$.

Case 2: $$\frac{E\left[p_iD_i(p)\right]}{E\left[D_i(p)\right]}<\frac{E\left[p_iD_i'(p)\right]}{E\left[D_i'(p)\right]}. \qquad \text{for all } p.$$

In this case, $p^*>\hat{p}$.

Case 3: $$\frac{E\left[p_iD_i(p)\right]}{E\left[D_i(p)\right]}=\frac{E\left[p_iD_i'(p)\right]}{E\left[D_i'(p)\right]}. \qquad \text{for all } p.$$

In this case, $p^*=\hat{p}$.

Examples: For the additive linear with state-contingent intercept case $D_i(p)=d_i-h(p)$, we are in case 1. For the multiplicative case, $D_i(p)=d_ih(p)$, then $p^*=\hat{p}$ (case 3).

Lemma 1. (i) *Cases 1 through 3 can be characterized by how the average wholesale cost price varies with the marginal retail prices:*

[9] Note that our setup is equivalent to assuming that the distribution grid owner purchases the power in the wholesale market and then resells it to each LSE based on the real time metered or load profiled consumption of the customers they have signed up. The access charge a is then the price LSEs pay to compensate the distribution grid for the costs of the wholesale power it has purchased on their behalf.

$a' > 0$ *in case 1*

$a' < 0$ *in case 2*

$a' = 0$ *in case 3.*

(ii) *In all cases,* $a(p) > p$ *for* $p < \widehat{p}$

$a(p) < p$ *for* $p > \widehat{p}$*.*

Proof: Part (i) is obtained by differentiating (5). To demonstrate part (ii), it suffices to show that $a'(p) < 1$ whenever $a(p) = p$, or after a few computations:

$$H(p) = E\left[(p - p_i)\, D_i' + D_i\right] > 0.$$

We know that $a(p) > p$ for p small (since $a(p) \geq E[p_i]$) and $a(p) \leq p_1 < p$ for p going to infinity. Hence, if the equation $a(p) = p$ has multiple solutions (an odd number greater than one) the function $H(p)$ must be increasing over at least some range. But $H'(p) = E[2D_i' + (p - p_i)\, D_i''] < E[D_i' + (p - p_i)\, D_i''] < 0$, a contradiction. Q.E.D.

A retailer designs his offer so as to solve:

$$\max_{\{p,A\}} \; E[(p - a)\, D_i(p)] + A$$

subject to

$$E[S_i(D_i(p)) - pD_i(p)] - A \geq \overline{U},$$

where $\overline{U}$ is the net surplus obtained by the consumer from subscribing with a rival retailer.

The retailer therefore selects p so as to maximize the joint surplus:

$$\max_p E[S_i(D_i(p)) - aD_i(p)],$$

or

$$(p - a)\, E\left[D_i'(p)\right] = 0,$$

yielding

$$p = a.$$

In equilibrium, a is given by (5). Hence

$$p = \widehat{p}.$$

Furthermore, $A = 0$: Retailers charge no monthly fee and just pass their variable cost of wholesale power through to the consumer.[10] Except in case 3, retail competition is, under load profiling, inconsistent with a Ramsey outcome.[11]

For future reference, let U^{RC} ("RC" for "retail competition") denote the consumers' equilibrium utility:

$$U^{RC} \equiv E\left[S_i\left(D_i\left(\widehat{p}\right)\right) - \widehat{p}D_i\left(\widehat{p}\right)\right]. \tag{6}$$

Proposition 2 Pure retail competition under load profiling delivers linear pricing at the average wholesale power cost $\widehat{p}$ despite the fact that LSEs have the possibility of offering two-part tariffs. The marginal price of electricity for the retail customer is therefore higher than the Ramsey price in case 1, and smaller in case 2; it is equal to the Ramsey price only in case 3.

Consider next the situation in which the distributor is also permitted to compete for load-profiled customers. We can assume either that the LSE behaves so as to maximize profits for the parent company as a whole, or that a ring-fencing rule requires the affiliate to maximize its own profits rather than those of the parent company. It can be shown that nothing is altered relative to pure retail competition:[12]

Proposition 3 Under load profiling and retail competition with the presence of an incumbent constrained or unconstrained by a ring-fencing requirement, the Ramsey optimum is generically not attainable. The consumers again face a linear price $\widehat{p}$ equal to the average wholesale power cost.

[10] Borenstein and Holland (2003) obtain the same average-wholesale-cost pricing result, although they rule out load profiling and assume that LSEs must offer linear tariffs: see sections 3.2 and 3.3.

[11] The Ramsey optimum can be achieved through a per customer subsidy or tax levied on retailers. Thus, let a retailer pay $\mathcal{A} + aQ$ when his customer consumes Q. The fixed charge $\mathcal{A}$ is over (or under) and beyond any delivery, metering and customer service costs that continue to reflect responsibilities of the distribution grid owner (these costs have been normalized at zero). Faced with an access tariff $(\mathcal{A}, a)$, retailers optimally pass this tariff through to their customers ($A = \mathcal{A}$ and $p = a$). The break-even constraint of the distribution grid owner is then:

$$\mathcal{A} + E\left[\left(a - p_i\right)D_i(a)\right] = 0.$$

The Ramsey outcome can be obtained by setting $a = p^*$, and then $\mathcal{A}$ so as to achieve budget balance, but (except in the non-generic case 3) this requires a departure from relying on load profiled consumption to calculate the wholesale price charged to retailers, in that the variable access charge differs (except in case 3) from the consumption-weighted average wholesale market price corresponding to the consumption induced by marginal price $p = a$.

[12] The proof of this result can be found at http://idei.fr/member.php?i=3.

3 Partially price responsive consumers with real-time meters

3.1 Rational consumers facing transaction costs

As noted in the introduction, even consumers on a RT meter may not be responsive to the RT price, due to transaction costs. We first investigate rational consumers. Rational consumers react imperfectly to the price profile that is offered to them by the LSE, but they make efficient use of the (endogenously imperfect) knowledge of this price profile and they trade off optimally the transaction cost involved in improving their monitoring of the price profile and in optimizing the usage of equipment, and the corresponding savings in their electricity bill.

Note first, that a consumer who does not bother to follow the evolution of the RT price may still be imperfectly responsive to it. For, a rational consumer ought to realize (at least) that the state of nature i she reacts to and the price she will pay, $\widehat{p}_i$, are correlated; for example, an American consumer should realize that the use of air conditioning in a hot weather condition is correlated with high electricity prices. Let us first motivate our analysis of rational consumers by a couple of examples in which the consumers' information structure is exogenous, and then build a general theory in which consumer information acquisition is costly and optimized.

Example 1: Suppose that the state of nature is (ij) where i and j each belong to $[0,1]$. The joint density is denoted f_{ij}. The wholesale price is p_{ij}. The consumer observes i (the local weather), but not j (the weather elsewhere, or the availability of the transmission lines or generators).[13] The observable and unobservable components of uncertainty may be correlated. The consumer's gross surplus $S_i(q)$ depends only on the observable part of the state of nature. Let $\widehat{p}_i = E_j\left[\widehat{p}_{ij}\right]$ denote the average marginal price when the observable component is i. Thus, a rational consumer chooses his consumption q_i when observing event i so as to solve:

$$\max_{q_i}\left\{S_i\left(q_i\right) - \widehat{p}_i q_i\right\},$$

defining a demand function $q_i = D_i\left(\widehat{p}_i\right)$.

The Ramsey optimum consists in choosing a retail price vector $\widehat{p} = \{\widehat{p}_{ij}\}$ with marginals $\{\widehat{p}_i\}$ and a fixed fee A so as to solve:

$$\max_{\{\widehat{p},A\}}\left\{E\left[S_i\left(D_i\left(\widehat{p}_i\right)\right) - \widehat{p}_i D_i\left(\widehat{p}_i\right)\right] - A\right\}$$

s.t.

$$E\left[\left(\widehat{p}_i - p_{ij}\right)D_i\left(\widehat{p}_i\right)\right] + A \geq 0,$$

[13] We here take this information structure as given. Presumably, it results from some optimization as in the more general model considered below.

or

$$\max_{\{\widehat{p}\}} \{E[S_i(D_i(\widehat{p}_i)) - p_i D_i(\widehat{p}_i)]\}$$

where $p_i \equiv E_j[p_{ij}]$ is the average wholesale price and $\widehat{p}_i = E_j[\widehat{p}_{ij}]$ is the average retail price when the observable component is i. The optimal policy is therefore a passthrough of the wholesale price: $\widehat{p}_i = p_i$, which can for example be obtained by:

$$\widehat{p}_{ij} = p_{ij},$$

and to charge no fixed fee: $A = 0$. The same proof as in Proposition 2 furthermore shows that LSE competition delivers this optimal passthrough.

Example 2: Let us next give an example in which the consumer does not observe the state of nature, yet his consumption is state-dependent. Consider equipment (e.g., space heater, air conditioning, pool heater) that, for a given quality of service s (e.g., indoor temperature set once and for all by the consumer) consumes a state-contingent amount of electricity. The real-time price profile of electricity affects the quality s (for example, an increase in winter prices lowers the indoor temperature chosen by the consumer or induces the consumer to switch to oil heat).

More formally, letting j be the full description of the state of nature, the consumer, who does not observe j, sets s so as to maximize his net surplus, equal to the gross surplus $S(s)$ minus the electricity bill:

$$E_j[S(s) - \widehat{p}_j D_j(s)] - A$$

where $D_j(s)$ is the state-contingent consumption needed to reach level s (for example a given swimming pool temperature requires a higher consumption of electricity when the weather is cold). Let $s(\widehat{p}_\cdot)$ denote the selected setting. The Ramsey optimum then solves:

$$\max_{\{\widehat{p}_\cdot, A\}} \{E[S(s(\widehat{p}_\cdot)) - \widehat{p}_j D_j(s(\widehat{p}_\cdot))] - A\}$$

s.t.

$$E[(\widehat{p}_j - p_j) D_j(s(\widehat{p}_\cdot))] + A \geq 0,$$

or

$$\max_{\{\widehat{p}_\cdot\}} \{E[S(s(\widehat{p}_\cdot)) - p_j D_j(s(\widehat{p}_\cdot))]\}$$

Again, the Ramsey optimum (or the LSEs' equilibrium offer for that matter) is obtained by a passthrough policy:

$$\widehat{p}_j = p_j.$$

Let us now consider a more general environment and further allow the consumer to choose his degree of awareness of the real time price. Namely, let ω denote the state of nature (for instance, $\omega = (ij)$ in Example 1). Let $\mathcal{P}$ denote the consumer's partition (for example, $\mathcal{P}((ij)) = i$ in Example 1). That is, the consumer observes that ω belongs to an event $\mathcal{P}(\omega)$. Let $C(\mathcal{P})$ denote the total transaction cost associated with partition $\mathcal{P}$; one has in mind that choosing a finer partition $\mathcal{P}$ (for example, keeping informed of the real time price) is costly, although we will not need to make this assumption.

The consumer in state ω takes a decision s that is measurable with respect to partition $\mathcal{P}$. This decision can be his electricity consumption as in Example 1, but can be different from the consumption, as illustrated by Example 2. Let $D\left(s\left(\mathcal{P}(\omega)\right), \omega\right)$ denote the associated consumption. Letting $S(s, \omega)$ denote the consumer's gross surplus, and $\widehat{p}_{\omega}$ the usage price charged to the consumer, for event P in the partition, $s(P)$ is given by the social planner's solution to the following program:

$$V(P) = \max_{s} \left\{E\left[S(s,\omega) - \widehat{p}_{\omega} D(s,\omega) \mid \omega \in P\right]\right\}$$

and $\mathcal{P}$ is given by:

$$\max_{\mathcal{P}} \left\{E_{P\in\mathcal{P}}\left[V(P)\right] - C(\mathcal{P}) - A\right\}.$$

The budget constraint writes:

$$E\left[\left(\widehat{p}_{\omega} - p_{\omega}\right) D\left(s\left(\mathcal{P}(\omega)\right), \omega\right)\right] + A \geq 0.$$

Hence, the consumer's utility is

$$\max_{\mathcal{P}} \left\{E_{P\in\mathcal{P}}\left[\max_{s} E\left[S\left(s,\omega\right) - p_{\omega} D\left(s,\omega\right) \mid \omega \in P\right]\right] - C(\mathcal{P})\right\}. \tag{7}$$

Note that this utility, and the concomitant notion of a Ramsey outcome, includes the costs of information acquisition. This utility is maximized when the consumer is confronted with the wholesale prices: $\widehat{p}_{\omega} = p_{\omega}$.

Proposition 4 With real-time meters and imperfectly reactive, but rational consumers:
(i) the Ramsey optimum (consumption decision, consumer's information) is obtained when the consumer pays the real time wholesale price associated with her actual consumption pattern;
(ii) retail competition delivers the Ramsey optimum.

3.2 Real-time meters but uniform pricing

Finally, let us assume that for some unspecified reason, real-time pricing of retail consumers is prohibited. And so the marginal p is not state-contingent.

Retail competition leads LSEs to maximize profit subject to the necessity of attracting the consumers:

$$\max_{\{p,A\}} \{E[(p - p_i) D_i(p)] + A\}$$

subject to

$$E[S_i(D_i(p)) - pD_i(p)] - A \geq \overline{U}.$$

Eliminating A and maximizing with respect to p yields

$$E\left[(p^* - p_i) D_i'(p^*)\right] = 0.$$

Thus, the competitive marginal uniform price is the Ramsey price under load profiling!

3.3 Summing up

Table 2 recaps the main results obtained so far. The results obtained in BH are indicated in the shaded areas. Recall that:

- p_i is the state-contingent wholesale power cost;
- p^*, a uniform (i.e., non state-contingent) price, is given by:
$$p^* = \frac{E[p_i D_i'(p^*)]}{E[D_i'(p^*)]};$$
- $\widehat{p}$, a uniform price, is given by:
$$\widehat{p} = \frac{E[p_i D_i(\widehat{p})]}{E[D_i(\widehat{p})]}.$$

Table 2 shows that the BH retail competition outcome is suboptimal due to the two constraints that BH imposes. First, it is inefficient to charge uniform prices with RT meters (compare the second and third columns of Table 2). Second, even under the constraint of uniform pricing, it is suboptimal to charge a linear price; put differently, BH would obtain the constrained Ramsey optimum uniform price p^* if they allowed for two-part tariffs.

INSERT TABLE 2 ABOUT HERE

4 Non-scale heterogeneity and competitive screening

For expositional simplicity, we have assumed that consumers are homogeneous (perhaps up to a size factor σ). This section briefly investigates the implications of consumer heterogeneity for retail competition.

Suppose that there are different classes of consumers $h \in [0,1]$ with state-contingent demands $D_i^h(p)$ and state-contingent surplus $S_i^h\left(D_i^h(p)\right)$. Let n^h denote the frequencies of consumers of type h, and $E_h\left[\cdot\right]$ denote the expectations with respect to consumer types (the expectations with respect to the state of nature are now labeled $E_i\left[\cdot\right]$). Let us begin with a few general remarks.

The first is that under *load profiling*, the analysis of retail competition is a simple generalization of that in Section 2. The retailers charge a linear price $\widehat{p}$ given by

$$\widehat{p} = \frac{E_i\left[E_h\left[p_i D_i^h\left(\widehat{p}\right)\right]\right]}{E_i\left[E_h\left[D_i^h\left(\widehat{p}\right)\right]\right]}$$

and fail to achieve the (second-best) Ramsey optimum. In particular, the retailers face no adverse selection problem to the extent that they pay a per-kWh price $a = \widehat{p}$ that is independent of the type of consumers they end up attracting.

Second, neither do LSEs face an adverse selection problem when dealing with rational consumers on real-time meters as in Section 3.1:[14] Proposition 4 above showed that it is optimal for LSEs to pass the wholesale price through to the consumer. And so at the optimal contract, the LSE breaks even on usage in each state of nature. Its profit is therefore unaffected by the consumer's actual load profile. Consumer heterogeneity then has no impact on competitive outcomes.

Third, competitive screening issues arise only if consumers are equipped with real time meters but non-uniform pricing is prohibited (Section 3.2). This is the case because passthrough of wholesale prices is then in general suboptimal, consumers differ in their load profiles and LSEs need to be careful of whom they attract. This situation is reminiscent of Rothschild and Stiglitz's celebrated treatment of insurance markets (1976). A complete analysis of competitive screening with uniform pricing and consumers with real time meters in retail electricity markets lies out of the scope of the paper. We however built an example[15] with two classes of consumers with different load profiles, and showed that retail competition is not optimal even among allocations that must satisfy uniform (non-RT) pricing.

[14] They may face forms of adverse selection unrelated to the consumer's load profile. For example, LSEs may try to obtain superior information about the probability of consumer default.

[15] Available as supplementary material for this paper at http://idei.fr/member.php?i=3.

Proposition 5 With heterogeneous consumers:

(i) Adverse selection does not arise when consumers are either on traditional meters and load profiled, or on real-time meters and rational. The analysis of Sections 2 and 3.1 thus generalizes to heterogeneous consumers.

(ii) By contrast, with consumers on real-time meters meters but under a regulatory prohibition of real-time retail pricing, adverse selection and the concomitant competitive screening can prevent retail competition from achieving the Ramsey constrained outcome, unlike in the case of homogeneous consumers.

5 Incentives to install real-time meters and communication equipment

Let us investigate the consequences of the previous analysis of the case where there are traditional meters, load profiling and retail competition for retailers' incentives to install real-time meters with or without communication, starting with the Ramsey incentives. Suppose that consumers have the same load profile but differ in the size σ of their demand: Consumer of type σ has demand $q_i = \sigma D_i(p)$ and surplus $\sigma S_i\left(q_i/\sigma\right)$. There is a continuous distribution of consumers σ on $[0, \infty)$.

Consumers initially have traditional meters and thus cannot react to the RTP. Two types of equipment can be added to a traditional meter:[16]

- *a real-time meter*, costing $m > 0$, that measures and makes verifiable the consumer's RT consumption, but makes this consumption imperfectly reactive to the RTP as in Section 3;
- *communication* (plus real-time metering), costing $M > m$, that furthermore makes it possible for consumers to perfectly react to the RT prices through remote control of appliances and equipment.

Ramsey benchmark.

Consider a rational consumer with type $\sigma = 1$. Let U^{FB} be the utility that this consumer could obtain if her consumption could adjust efficiently to variations in the real time price (see (1) above). Let U^* be the second-best utility that could be achieved by a Ramsey social planner

[16] Note that we assume that there are no returns to scale in installing equipments. In practice, LSEs incur costs, such as wireless bay stations enabling remote real time recording, that are common across consumers in a neighborhood. Such costs give rise to non-convexities and inefficiencies unless they are shared among LSEs.

for the consumer with a traditional meter (see (3) above). And let $U^{**} \in \left(U^{*}, U^{FB}\right)$ denote her utility when endowed with a real-time meter without communication (see (7) above). The utilities of a consumer with type σ are equal to σ times these utilities. The Ramsey planner (or a monopoly retailer) would endow consumers in $(\sigma^{*}, \sigma^{**})$ with a real-time meter, and those in (σ^{**}, ∞) with real time meters plus communication, where:[17]

$$\sigma^{*} = \frac{m}{U^{**} - U^{*}} \quad \text{and} \quad \sigma^{**} = \frac{M - m}{U^{FB} - U^{**}}.$$

Load profiling.

We keep the assumption that the consumption of retail consumers with traditional meters is load profiled using the load profile of the consumers in that class. Under perfect retail competition with load profiled consumers, the consumer obtains σU^{RC} when keeping a traditional meter, $\sigma U^{**} - m$ when equipped with a real-time meter, and $\sigma U^{FB} - M$ when equipped with communication.

Simple derivations yield:

Proposition 6 (i) Under *pure retail competition with load profiling*:
• Consumers with type $\sigma \geq \sigma^{**}$are equipped with communication, where σ^{**} is the Ramsey level.
• Consumers with type $\sigma \in \left[\sigma^{RC}, \sigma^{**}\right]$ are equipped with real-time meters,when
$\sigma^{RC} = m / \left[U^{**} - U^{RC}\right] < \sigma^{*}$, the Ramsey level.
(ii) Consequently, there is more investment in meters that measure real-time consumption than in the Ramsey optimum. Given the inefficiencies introduced by the combination of load profiling and retail competition, however investments are socially optimal.

The constrained efficiency of market-determined investment in metering equipment (part (ii) of the proposition) deserves some comment. There are really two Ramsey benchmarks, one unconstrained by retail competition and the other constrained by retail competition. The investments are socially optimal given the inefficiencies created by retail competition with load profiling. By contrast, BH found that the market produces too much or too little installation of communication equipment (recall that BH assumes that the actual real time consumption of a competitive LSE's customers can be measured accurately — there is no load profiling. This can only be the case with retail competition if all of the LSE's customers have RT meters.

[17] Assuming $\left(U^{**} - U^{*}\right) M \geq \left(U^{FB} - U^{*}\right) m$. Otherwise, it is not optimal to install real-time meters without communication.

So, BH implicitly assume that retail customers are effectively already equipped with real time meters.) Because BH impose constraints on what retailers can charge (uniform and linear tariffs: see Section 3.3), the retail competition outcome is inefficient and so public intervention in the market for communication equipment is an optimal response to inefficient consumption in their absence.

6 The joint interruptibility problem

In our companion paper (Joskow-Tirole 2005) we derive the efficient prices and investment program for an electricity market with demand uncertainty, price insensitive consumers, and LSEs that can choose any level of rationing they prefer contingent on the real time price. We then identify the assumptions required for a competitive wholesale and retail market equilibrium to achieve this efficient price and investment program. One of the key assumptions is that different users can choose and the system operator can implement different levels of priority in rationing that reflect users' individual preferences. The validity of this assumption requires the system operator to be able physically to cut off individual retail consumers.

There is no theoretical reason why individual customers cannot be rationed. It requires installing communications and control equipment between the customer's connection to the network and the control center. However, this equipment is costly. As a practical matter, except for very large customers that have direct control equipment, most directed interruptions must occur at points on the network ("zones") that can be controlled by the distribution network operator.[18] The affected zone has (a) customers served by multiple LSEs that compete with one another (so every house on a street can be "served" by a different LSE) and (b) customers with heterogeneous preferences.

An optimal dispatch when zones but not individual consumers are controlled by the system operator must elicit each zone's *aggregate* willingness to pay for being served. From the point of view of the set of LSEs and industrial users in a given zone, reliability is a *public* good.

In principle, one can make use of the theory of public goods in order to design incentive-compatible mechanisms of elicitation of individual preferences for reliability.[19] For instance, one could use the Clarke (1971)-Groves (1973) scheme. Suppose that, due to a shortage in supply,

[18] In reality, system operators generally try to squeeze out all of the price sensitive demand first before they start rolling blackouts. This may not be optimal of course. There is also some priority rationing in that circuits with hospitals and fire stations, etc. will often be placed on a "do not blackout list." In this case, all customers on the same circuit get the benefit of being near a fire station or hospital. This example illustrates the fact that different consumers may have different values of lost load, and that furthermore the dispatcher cannot fine-tune the intensity of rationing.

[19] See Green-Laffont (1979a,b) for the general theory of public goods.

the ISO must shut down one of cities A,B,C,... To simplify computations, cities demand the same load. Within city A, say, there are n users, each demanding 1 unit of load and having valuations v_i, which are private information. These users can either be price-sensitive, industrial users or LSEs serving price-insensitive users. Let the ISO shut down the city with the lowest total declared willingness to pay. That is, city A is served if and only if

$$\widehat{V}_A \equiv \sum_{i \in A} \widehat{\mathrm{v}}_i \geq \widehat{V}$$

where $\hat{V}$ is the lowest total declared willingness to pay among other cities. City A then pays $\hat{V}$. The problem then boils down to a standard public good problem (the cost of getting the public good is $\hat{V}$-possibly unknown to members of city A, but this does not matter as this value is revealed through the aggregate bids in other cities).
In particular, use can be made of Clarke-Groves mechanisms : Member i of city i pays

$$\begin{cases} \widehat{V} - \Sigma_{j \neq i} \widehat{\mathrm{v}}_j & if \;\; \widehat{\mathrm{v}}_i + \Sigma_{j \neq i} \widehat{\mathrm{v}}_j \geq \widehat{V} \\ 0 & \text{otherwise.} \end{cases}$$

Telling the truth ($\widehat{\mathrm{v}}_i = \mathrm{v}_i$) is then a dominant strategy. [The Clark-Groves mechanism does not balance the ISO's budget, but a variant of it (the d'Aspremont-Gerard Varet (1979) scheme) does so in expectation.]

Besides transaction costs, there is under retail competition a major snag with such zonal voting mechanisms. While large industrial users' on real time meters willingness to pay for reliability is not distorted by competition for the final consumer,[20] competing retailers' profit in a given zone depends only on the *relative* quality of their offer as compared with their competitors'. A retailer that bids for reliability increases the quality of service to its retail consumers, but it also increases its rivals' quality of service by the same amount, bringing no extra profit. This is best seen when considering the following timing under Bertrand competition: First, LSEs bid for reliability ($\widehat{\mathrm{v}}_k^z$ for LSE k in zone z). Second, given the resulting reliability in each zone z, they compete for retail consumers. Given that they make no profit at stage 2, LSEs aim at minimizing expenditure at state 1 (they have de facto willingness to pay $\mathrm{v}_k^z = 0$ in reference to our previous discussion).

Proposition 7 Zonal rationing implies that the demand for rationing in a given zone is an aggregated demand.
(i) In the absence of transaction costs, the constrained optimum can be obtained through

[20] Unless two large industrial users both compete on the product market and produce in the same zone.

standard public goods mechanisms if consumers are (non-competing) industrial users and retail consumers served by a monopoly distributor.
(ii) By contrast, the elicitation of consumers' willingness to pay for non-interruptibility is problematic under retail competition. In particular, if LSEs bid for reliability and then compete for retail consumers, no information can be obtained from LSEs concerning the consumers' demand for non-interruptibility.

Remark (retail competition prices). An important item on the research agenda is to find practical mechanisms that will enable social rationing in the presence of retail competition. When the probabilities α_i of rationing in state of nature i are exogenous, then it is straightforward to generalize the analysis of section 2.2. For simplicity, let us focus here on the special case of perfectly foreseen outages associated with rolling blackouts. The consumers' gross surplus and demand in state i are then $\alpha_i S_i(D_i)$ and $\alpha_i D_i$. Assume that LSEs take α_i as exogenous, that rationing is zonal and the retail competition with load profiling corresponds to competition within a zone. LSEs, as earlier, maximize the joint surplus:

$$\max_p \left\{ E\left[\alpha_i \left[S_i\left(D_i(p)\right) - aD_i(p)\right]\right]\right\}$$

yielding:

$$E\left[\alpha_i (p-a) D_i'(p)\right] = 0.$$

Again, it is optimal for LSEs to pass the average wholesale price a onto consumers:

$$p = a.$$

And so

$$p = \frac{E\left[p_i \alpha_i D_i(p)\right]}{E\left[\alpha_i D_i(p)\right]}.$$

7 Conclusion

Let us first summarize the paper's main insights. The retail consumer's lack of responsiveness to the RT wholesale price may stem from (a) the absence of RT meter, (b) transaction costs associated with monitoring and reacting to RT prices, (c) load profiling, or (d) joint interruptibility. When retail consumers are on traditional meters which measure their aggregate consumption over relatively long time periods rather than in real time, neither retail consumers nor, under retail competition, the LSEs responsible for purchasing the power required to serve their demand, face the real time wholesale prices associated with the power they consume from the system. We show that under these conditions the competitive retail market equilibrium involves linear average-wholesale-cost pricing rather than more efficient two-part tariffs. By contrast, retail competition delivers efficient pricing and equipment installation when consumer non-responsiveness is due to transaction costs rather than a lack of recording of RT consumption. Finally, when the system operator is physically unable to cut off individual customer loads, and instead must ration on a zonal basis, individual retail customers cannot obtain their preferred priority for rationing by the system operator. Given this constraint, the Ramsey social planner could turn to standard public goods mechanisms to determine the relative priorities of the different zones that are physically capable of being cut off by the system operator and use this information to establish a second-best priority cutoff schedule. By contrast, in the presence of retail competition LSEs may misreport their consumers' demand for non-interruptibility because they would prefer to free ride on the other LSEs serving consumers in the same zone.

Our paper leaves open many questions relative to retail competition. For example, we already noted that robust mechanisms allowing consumers to bid for reliability (or priority servicing) should be designed so as to allow an efficient zonal rationing under the joint-interruptibility constraint and retail competition. Second, the welfare conclusions of this paper rest on the assumption of competitive wholesale prices. While we feel that generator market power is most naturally addressed through regulation of the supply side, some demand-side policies may also help curb market power; for example, subsidizing equipments that make consumers more responsive to prices may alleviate market power at peak and improve welfare,[21] while we have seen that such subsidies are unwarranted when generators are competitive. Third, we have assumed that retailers are perceived by consumers to supply homogenous services and that the retail market is competitive. Yet, in practice, competing retail suppliers often seek to create brand identities and to differentiate their products in a variety of ways. In addition, it is increasingly evident that small retail consumers face significant switching costs and are viewed by retailers as being "sticky." This suggests that further analysis of the attributes of imperfectly competitive retail markets would be productive.

[21] See Joskow-Tirole (2005).

References

[1] Borenstein, S. and Holland, S. “On the Efficiency of Competitive Electricity Markets With Time-Invariant Retail Prices,” (2003). Working Paper no 116, Center for the Study of Energy Markets, University of California, Berkeley (*Rand Journal of Economics*, forthcoming).

[2] Clarke, E. “Multipart Pricing of Public Goods.” *Public Choice*, Vol. 2 (1971), pp. 19–33.

[3] D’Aspremont, C. and Gerard Varet, L.A. “Incentives and Incomplete Information.” *Journal of Public Economics,* Vol. 11 (1979), pp. 25– 45.

[4] Green, J. and Laffont, J.J. *Incentives in Public Decision Making.* Amsterdam: North Holland, 1979a.

[5] —– and —– eds. *Aggregation and Revelation of Preferences.* North-Holland, 1979b.

[6] Groves, T. “Incentives in Teams.” *Econometrica*, Vol. 41 (1973), pp. 617–631.

[7] Joskow, P. and Tirole, J. “Reliability and Competitive Electricity Markets.” mimeo, MIT and IDEI, 2005.

[8] Rothschild, M. and Stiglitz, J. “Equilibrium in Competitive Insurance markets: An Essay in the Economics of Imperfect Information.” *Quarterly Journal of Economics*, Vol. 90 (1976), pp. 629–650.

LSE's energy purchase cost corresponds to:

		customer's RT consumption	customer's load profiled consumption
Customer's metering equipment and sensitivity to RTP:	✓ RT meter ✓ Full responsiveness	Textbook treatment	
	✓ RT meter ✓ No/partial responsiveness	Section 3.1	
	Traditional meter (⟹ no responsiveness)	Section 2.1	Section 2.2

Table 1

	Ramsey	Profit-maximizing LSEs	
		Non-linear pricing	Linear pricing constraint
Traditional meter	p^*	Monopoly p^* Competition $\hat{p}$	$\hat{p}$
RT meter (with limited, but rational consumer reactivity)	p_i	p_i	p_i
RT meter, but uniform pricing constraint	p^*	p^*	$\hat{p}$

Table 2: Marginal retail price[22]

[22] Last two columns indicate marginal prices under LSE profit maximization, either under retail competition or under a regulated, breaking even retail monopolist (the distinction between the two is indicated only when the results differ).