Regulating Retail Electricity Prices: A Review of Theoretical and Empirical Issues

Graziano ABRATE

Working Paper 7, 2003

Real Collegio Carlo Alberto
Via Real Collegio, 30
10024 Moncalieri (To)
011 640 27 13 - 642 39 70
info@hermesricerche.it
http://www.hermesricerche.it

Regulating Retail Electricity Prices: A Review of Theoretical and Empirical Issues[♣]

Graziano Abrate

(*Università di Pavia*, *HERMES*)

November 2003

Abstract

At present, demand is almost completely unresponsive to price in most power markets, since consumers usually face fixed retail electricity prices, which do not reflect the time-varying marginal wholesale cost of production. This is a source of inefficiency, in particular in a deregulated power market, where utilities are exposed to a competitive wholesale market. This work describes what is meant by demand-side participation programs and the different ways they can be implemented to promote demand responsiveness. The objective is to highlight their effectiveness and their effects on consumers' welfare. The theoretical advantages of dynamic pricing are discussed together with the technological, cultural and regulatory barriers that they face in practice. Benefits from such programs depend crucially on the possibility to shift consumption across different time-periods. Different empirical studies have provided estimates of this substitutability, and here I present a survey of results and techniques. There is agreement over the customer's ability to respond to price signals, but the extent of such a response varies widely across users. This can raise equity issues when implementing time-varying retail prices, that must be assessed together with the expected benefits in terms of efficiency brought by an increased demand responsiveness.

Key words: electricity market, time-varying retail prices, demand responsiveness, efficiency, equity

JEL: C1, D6, L10, L5, L94

[♣] This study has been carried out within the HERMES research project *Liberalizing public utilities in Europe: the case of water, energy and local transport in Italy, France and Switzerland*. The paper comes from the dissertation prepared for the MSc in Economics at the University College of London (UCL). I am grateful to my supervisor Ian Preston for his clear comments and helpful suggestions. I also would like to thank Giovanni Fraquelli and Diego Piacentino for some useful hints.

1. Introduction

The marginal cost of producing electricity varies considerably over time, since demand is highly variable, whereas production is subject to rigid short term capacity constraint. During off-peak times, there is plenty of capacity and the cost of producing an additional kilowatt-hour only reflects fuel and some operating and maintenance costs. On the other hand, during peak periods, the capacity constraint will be binding and the incremental cost can increase greatly. Generally, the end-use consumer faces a fixed retail price, which does not give a signal of the actual system load, and demand does not play an active role in determining prices.

Promoting demand responsiveness is becoming an important objective especially since many countries have undergone structural reforms moving from a model in which there was a vertically integrated unit with monopoly power towards a deregulated market. In such restructured models, utilities are exposed to a competitive wholesale market, where prices vary on an hourly (or half-hourly) basis, reflecting the interaction between demand and supply. However, liberalisation did not always bring the expected benefits. The classical example of failure is California, which in summer 2000 experiences rapidly increasing prices in its liberalised wholesale market. Lack of demand participation has been identified as one of the main causes of the crisis, and in this direction were the suggested solutions.

There are many ways a demand response program can be implemented and they involve the introduction of time-varying retail prices and demand-side bidding. This work aims at describing the advantage and the weakness of such programs, highlighting in particular their possible effect on welfare and their effectiveness in the sense of actually achieving demand responsiveness.

First, I will compare the different possibilities of time varying prices that can be used, explaining the effect of having dynamic versus static prices. I will focus in particular on the differences between adopting Real Time Pricing (RTP) and Time of Use Pricing (TOU). The first term refers to any system that charges different electricity retail prices for different hours of the days and for different days. Under TOU, instead, prices vary in a preset way within certain blocks of time. The key difference is that under RTP prices adjust frequently according to the actual balance between demand and supply, while TOU provide preset tariffs, and so they are less likely to reflect the prices in the wholesale market.

Though RTP is more efficient, TOU have been more widely used and accepted, in part because it is easier and less costly (in term of metering) to implement. In section 3, I will analyse some frequent concerns about implementing RTP. In particular, I will describe the way of hedging against the risk for the customer, the distributional concerns (who win and who loose from the introduction of such a program?) and the mandatory versus voluntary programs.

Then, I will move to the analysis of the empirical literature that has dealt with the measurement of demand responsiveness (section 4). There is a wide literature available over the estimation of price elasticities of demand. While both techniques and results are quite variable, there is uniform agreement that industrial, residential and commercial electricity consumers can, and will, respond to the price signals they face. However, only few works have analysed real time programs and so I will focus on them, after briefly reviewing the main results and techniques used in the basic models and in the one concerning TOU programs.

Finally, I will give some ideas for a future research, in particular concerning the points of interest that could be analysed in an empirical investigation.

2. Demand-side participation programs

The physical aspects of supply and demand must receive a great attention for understanding the fundamental economics of power markets. Stoft (2002) underlines the peculiar role played by the shifts in the level of demand that are *not* associated with price. Indeed, demand is highly variable between and within a day, and these hourly fluctuations determine the key long-run characteristics of supply. Traditionally, the demand for power can be described by a *load-duration curve*, which measures the number of hours per year the total load is at or above any given level of demand. Even if this curve does not include information on the sequence of the load levels[1], it gives information about the peak-level demand and its duration (say, the peak demand was 1,211 MW; the demand was above 1,100 MW for 122 hours in the year; and so on). A natural interpretation for such data is the probability that load will be at or above a certain level (in the previous example, 122 out of 8,760 hours in a year, i.e. 1.4 per cent of probability that demand will exceed 1,100 MW). These data are very important in

designing the productive structure, because since electricity is *not storable*, supply is equal to consumption at any time (ignoring losses)[2]. Therefore, peak demand must be satisfied by production from generators that are used as little as 1% of the time. The technology used to build such generators, so-called *peakers*, is a lot different from that used for the *baseload* generators, which run most of the time, and, in particular, the first ones generally imply a higher marginal cost of production. With a very broad approximation, it could be said that a higher load level is associated to a higher marginal cost, which can greatly increase when demand is at the highest level. It must also be noted that, even if supply always equals consumption, it may not equal demand, because supply is subject to rigid short term capacity constraints and so demand may be higher than the maximum possible supply in a certain moment[3].

As said before, the load-duration curve is independent from any consideration about *prices*, a dimension that must be added when talking about a market. Presently, demand is almost completely unresponsive to price in most power markets, and this happens also in bid-based markets. The problem, stated in Lafferty et al. (2001), is that wholesale buyers rarely submit price-sensitive bids; on the contrary, they typically submit bids stating only the quantity to be purchased. Actually, most of them are distribution utilities that have a legal obligation to provide electricity to their customers. Since the latter usually face fixed retail prices, so that they do not have any incentive to respond to hourly wholesale prices, also utilities bids cannot be price-sensitive. In other words, wholesale price fluctuations reflecting the supply-demand balance are not usually passed on to retail customers, and therefore their decisions are independent from the actual system load situation and from the marginal cost of production. According to Borenstein et al. (2002), a *demand-side participation program* is any method that can be used "to make the economic incentives of customers more accurately reflect the time-varying wholesale cost of electricity".

[1] So, for example, the same curve can describe wide daily swings in demand and little seasonal variation or wide seasonal variation and limited daily swings.
[2] To be precise, the amount stored is minuscule and cannot be utilized for trade.
[3] Technically, the difference between supply and demand cannot be indicated by flows of power, but must be measured in terms of voltage and frequency. Demand for power is defined as the amount of power that would be consumed if system frequency and voltage were equal to their target values for all consumers. If voltage or frequency are low, then customers consume less power than they would like so supply is less than demand. For a more detailed explanation see Stoft (2002), pag. 40-48 and pag. 373-388.

Table 1. Demand-side participation programs

	Definition	Signal of the actual supply/demand balance
Real Time Pricing (RTP)	Retail electricity prices that fluctuate with the real time wholesale prices	Accurate, depending on the lag time between the price announcement and the price implementation
Time-of-Use Pricing (TOU)	Retail electricity prices varying in a preset way within certain block of time	Approximate, since prices don't capture the price variation within a price block. Moreover, they are based on the average wholesale market variation and adjusted infrequently
Demand Charges	Instrument that allows a portion of the consumer's bill to be calculated on the basis of the consumer's maximum capacity usage	Approximate, since the charge is based on the individual peak and not on the system peak
Critical Peak Pricing (CPP)	System that usually starts with a TOU rate structure, and adds one more rate that applies to critical peak hours, which the system operator can call on short notice	Good, but less accurate than RTP for two reasons: first, the level of prices for the peak hours are preset; second, the number of peak hours that can be called in a year is limited.
Interruptible Demand Programs	System with a basic constant rate structure, with the option for the system operator to cut off supply to some customers.	Since the customers are not actually physically interrupted, but they retain an option to continue to consume at a greatly increased price, these programs can be viewed just as a crude form of CPP.
Real Time Demand-Reduction Programs (DRP)	System where certain customers are eligible to be paid to reduce their consumption at certain times.	Similar to interruptible demand programs

There are many different possibilities to achieve such a goal of a price-responsive demand. Table 1 provides a list of these methods and describes their capability to give an efficient signal of the real time demand-supply balance. This is clearly related to the possibility of varying the retail price on a short notice. RTP, which implies different retail prices for every hour of the day, varying every day, can achieve this goal almost

perfectly, depending on the lag between the price announcement and the price implementation. In its extreme (virtual) application, the real time price for each hour is announced at the beginning of the hour. However, where it has been implemented, the prices for all hours of a day are typically announced on the previous day, with the participants to the program informed via fax or/and internet (for example, on 24 July at 4 o'clock participants receive a fax containing the prices valid on 25 July from midnight to 1 o'clock, from 1 to 2, and so on). The more the lag increases, the more RTP becomes in a certain way similar to TOU, loosing the efficiency in reflecting the true variation in the wholesale market. Thus, a TOU structure entails preset prices based on the average wholesale variation, and for this reason it is not able to capture an unexpected shock. An empirical investigation for the summer of 2000 in California has shown that less than 20% of the variation in the wholesale market could have been reflected in a TOU structure, even setting the TOU prices ex-post (Borenstein et al., 2002)[4].

To summarise, the fundamental difference between TOU and RTP lies in a *static* versus *dynamic*[5] approach to retail pricing. It is also interesting to note that the other methods listed in Table 1 can be viewed either as an improvement of TOU (demand charges that are usually implemented together with TOU, and especially CPP), either as a particular form of CPP. The latter is a sort of a mixed system that uses a TOU static structure, but adding one more "dynamic" rate that can be called on a short notice to take into account of critical peak hours. Thus, if the matter is dynamic or static prices, we shall look at their theoretical and practical implications, and next chapter is devoted to this.

3. Implementing dynamic pricing: theory and practice

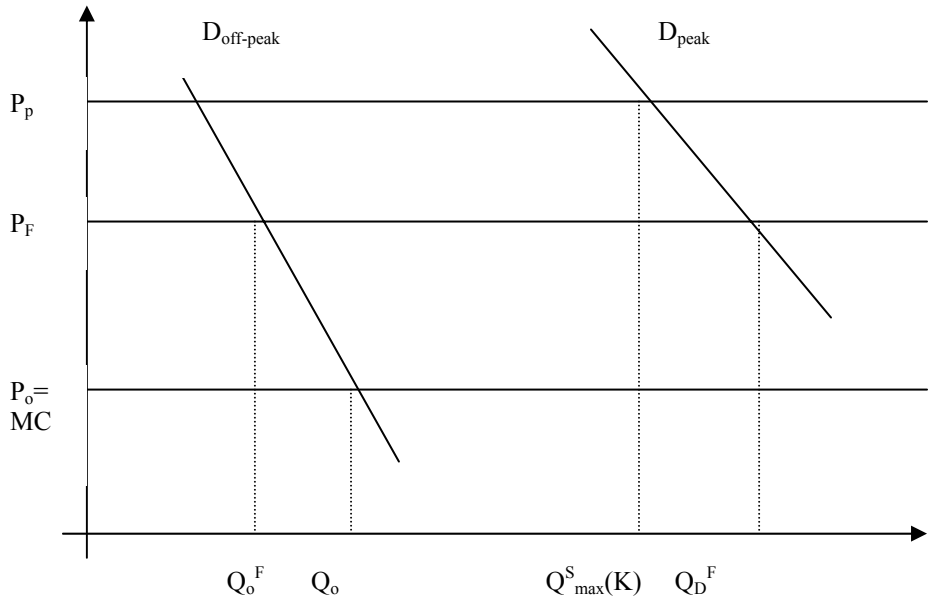
To illustrate the benefits of allowing dynamic pricing, consider the simple model in Borenstein (2003). Suppose that there are only a peak and an off-peak demand and that market is competitive. Thus, supposing an installed capacity of K, than, if time-varying

[4] This investigation was based on a regression of the hourly wholesale price on dummy variables for each of the TOU periods, and the R-squared of such a regression provides the share of price variation captured by using TOU periods rather than a single constant price.

[5] Here I use the adjective "static" to indicate a preset structure like TOU. Actually, TOU prices can also be periodically adjusted, but this usually happen just a few times a year. At the opposite end, I use "dynamic" to indicate that the adjustment is very frequent, even if it is never continuous.

rates are allowed, the prices will be Pp and Po during the peak and the off-peak period respectively, as shown in Figure 1. K represent an optimal capacity and there is no incentive to invest more.

Figure 1. Static against dynamic prices (from Borenstein, 2003)



If the price is constrained to be at the unique rate P_F, then the effects will be the following:

- an inefficient decrease in the off-peak consumption, causing a deadweight loss;
- a demand exceeding the supply at the peak rate, involving the need of some sort of rationing.

This second aspect would produce an incentive for firms to over-invest in capacity. Since in peak period they must sell at P_F and they cannot charge an higher price, there is an incentive to build new capacity to meet the additional demand. The author emphasizes the role of time varying prices, that encourage customers to consume less in peak periods avoinding this excess of capacity. Moreover, if the wholesale market is not competitive, with fixed retail price it is much more profitable for the wholesale seller to exercise *market power*. In fact, a raise in the wholesale price has no short-run impact on sales since end-use customers do not see a change in their bill.

Now suppose that a TOU structure is used and consider the effect on the simple model described above. In this case, we have an improvement because there will be two

rates, P_o^F and P_p^F, which however can only approximate the competitive unconstrained prices P_o and P_p, since they are fixed ex-ante. In the real world, since there are not only two time periods, and both peak demand and especially time when peaks occur are difficult to predict, the approximation can be very inaccurate with respect to RTP.

Though RTP is more efficient, TOU have been more widely used and accepted, in part because it is easier and less costly to implement. RTP benefits must be high enough to justify investments in metering, and needs efficient systems of communications. However technology is evolving fast, and can support the implementation of RTP at least in three directions: first, making available sophisticated metering technology at a reasonable cost; second, simplifying communication thanks to the internet; third, enhancing the ability to respond to frequent retail price signals, that sometimes could be achieved without the human intervention thanks to the use of "smart" energy management systems. Borenstein (2001) states that the cost of this investment may not be feasible for very small users, but would be certainly desirable for large users.

A part from the technological barriers, there are also cultural and regulatory barriers to RTP (Yoshimura, 2003). For example, it is a common belief that having electricity is a basic right, and that prices should be time invariant. Even if time-variant prices would produce savings. Moreover, policies usually support this belief, requiring the utilities to offer time invariant retail electricity prices. According to Borenstein (2003) the concerns about RTP typically involve three types of issues: the customer price risk, equity concerns and mandatory versus voluntary programs.

a) Hedging against the risk

Because the real-time or the day-ahead price of electricity is highly volatile, the customers are diffident towards RTP, for the risk of paying drastically increased prices during certain hours. This involve the need to create some form of insurance for the consumers, by purchasing some power on long term contracts in order to give a certain stability to their monthly bills. One approach is to implement a two-part RTP program with a Customer Baseline Load (CBL), that allows consumers to buy a certain amount of power according to standard TOU rates, while they face real-time rates when their consumption increase over a certain predefined level. However this raises difficulties on the definition of the CBL. Rather than assigning a certain baseline level, it seems more appropriate allowing the customer to purchase a baseline (with a forward contract) to

hedge as much he desires. The fact that incremental consumption decision are still subject to RTP ensures strong incentives to conserve at peak times.

b) Equity issues

Maybe the most important diffidence against RTP is the fact that such tariffs would necessarily involve an arbitrary redistribution among different types of customers. Of course, the most flexible consumers and those that usually tend to have a smoother consumption will be the first ones to gain from RTP, while customers with more "peaky" demand, unwilling to switch their consumption, will pay a high share of their power at the more expensive rates. However, the latter could expect to gain from positive externalities coming from the reduction of peak consumption by the most flexible consumers. In fact, lower peak demands mean less investment in excess capacity and therefore lower payments to the generators in the wholesale market. This is even more considerable if we consider the the total capacity is built on the basis of the system peak, but in order to minimise the risk of blackouts there are of course reserve requirements (usually set between 10 and 20 per cent of the peak demand). Price responsive demand will not only imply a lower system peak, but also a reasonable lower percentage of reserve requirement. This is because the increase in peak price will at least partially absorbe an unexpected system shortage. Moreover, RTP reduces the ability of sellers to exercise market power. The point is to understand the extent of these benefits in order to evaluate the feasibility of the program.

c) Mandatory or voluntary programs

If the gains from dynamic pricing depends crucially on the customer load curves, then one of the possibilities is to implement a voluntary program. This would allow the most inelastic users to stay at fixed rates. However, a voluntary approach can give raise to a problem of adverse selection, if its implementation generate a cross subsidisation from RTP users to the others. This could happen because the retailer will see a decrease in its revenues (since users will choose RTP only if they can save money). To keep its revenue at the same level he will decide to charge an adder on RTP, in order to equalise the average price between participants and non-participants. But this will clearly undermine the incentives to join the program. In order to be successful, a non-compulsory RTP program must have a commitment of no cross subsidisation.

4. Measurement of demand responsiveness in the economic literature

In the previous sections I described demand-side participation programs and their objectives. However, it is clear that the benefits of such programs, if any, depend crucially on the effective *price elasticity* of demand. Supposing that a customer is facing dynamic pricing rather than a fixed price, how much will he be willing to change its consumption? In this section I search an answer to this question in the available literature.

The interest in energy demand is early dated in economics, such that a first review of empirical works was already published in 1975. In general, the focus of these pioneer analyses was investigating the substitutability of energy with other factors. Only more recently the attention has been dedicated to the possibility of substitution between peak and non-peak usage. The majority of these studies consider a TOU static framework, whereas only a few works concern dynamic pricing.

4.1. Non-peak analysis

Energy does not yield utility in itself, rather it is desired as an input into other process. Ignoring the possibilities of substitution between peak and non-peak usage, demand response to higher prices typically involve substitution of other factors for energy. Different surveys are available concerning the econometric issues over the modelling of electricity demand[6]. A detailed analysis of these studies is out of the aims of this work. Here it is important to summarise some well-known characteristics of demand and some quantitative results[7].

a) Since energy is not desired in itself, energy demand is *derived* from demand for more basic end products (for example, light, warm and cold space, motive power).
b) Energy involves usage of durable goods, and therefore it is important to distinguish between short-run and long-run demand elasticity.
c) While both techniques and results are quite variable, with good approximation it can be stated that short-run elasticity is measured around -0.2, whereas long-run aggregate elasticity of demand is likely to be in the range of -0.4 to -0.9.
d) Energy is virtually used in all activities, but factor proportions vary widely. This leads to the implication that price elasticity vary widely across users.

[6] For detailed reviews see Taylor (1975) or Bohi et al. (1984)

e) Price elasticities can vary among different regions. This could lead to an overstatement in the estimation of long-run price elasticities in cross-section empirical analysis. Indeed, an increase in prices of a certain region can involve migration of industries to another region, reducing energy consumption in the initial location, but without influence on the total energy consumption.

4.2. Time-Of-Use

A number of TOU programs have been implemented in last 25 years, giving the opportunity to study the extent to which it is possible to induce a reallocation of energy consumption between different hours or different days. This kind of substitution does not involve a change in the total consumption. Since energy yields different levels of utility depending on when it is consumed (say, for example, the utility of heating in a cold or a warm day), energy consumed today can be considered as a different good with respect to energy consumed tomorrow. In other words, demand can be decomposed, in the more simple case, in peak and off-peak demand.

Generally, electricity is assumed to be separable from other goods[8], so that, following Mountain and Lawson (1992), it is possible to define a sub-utility function concerning kilowatt-hours consumption of electricity at different times of a certain basic unit of observation (a day, a week, a month). The representative consumer is therefore assumed to optimally choose the time allocation of electricity during the basic unit of observation. This leads to a system of demand equations (one for each TOU rate), that can be estimated in order to calculate own price and cross price elasticities. Alternatively, when firms are concerned, the starting point is the specification of a cost function where inputs are different times of use of energy.

To describe more in detail the theoretical framework, I will follow the notation used in Parks and Weitzel (1984) in their study about Wisconsin *residential* TOU price experiment. Suppose that the utility function can be written as follows:

$$u = U(x, z) = U[e(x), z] \qquad [1]$$

where $x = (x_1, x_2, \ldots x_h)$ is a vector of the quantities of electricity consumed during h time intervals, z is a vector of non-electricity goods, and e(x) is homogeneous of degree one in x. Thus, $U(\cdot)$ is homothetically separable in x[9], a necessary and sufficient

[7] See also Sweeney (1984), Lafferty et al. (2001)
[8] One significant reason is the lack of correspondingly accurate information on other commodities.
[9] For a non-homothetic approach see Mountain and Lawson (1992)

condition to validate a decentralised two-stage budgeting approach to the electricity demand approach. The first stage involves the allocation of total expenditure (y) between electricity and non-electricity goods. The maximisation process yields the indirect utility function, homothetically separable in the electricity prices.

$$\{\max_{x,z} U(x,z) \text{ subject to } px+qz=y\} \rightarrow V(p,q,y)=V[g(p)/y, q/y] \qquad [2]$$

where g(p), homogeneous of degree one, is a price index function for electricity.

The second stage concerns the maximisation of the sub-function e(x) subject to the constraint px=m, where m is the optimal total expenditure in electricity determined in stage one. This allows determining the optimal time allocation of electricity. By applying Roy's identity to [2] and after some algebraic steps[10], the demand functions for electricity in a particular time period are obtained:

$$x_i = \frac{m(\partial g/\partial p_i)}{\sum_{i=1}^{r}(\partial g/\partial p_i)p_i} \qquad [3]$$

This demand system gives all the necessary information to measure consumers' substitution response. The choice of the functional form g(p) is only restricted to be homogeneous of degree one, and it has been specified in a number of ways:

a) constant elasticity of substitution (CES)

$$g(p) = \left(\sum_{i=1}^{r}\alpha_i p_i^{-\rho}\right)^{-1/\rho} \qquad [4]$$

b) translogarithmic

$$\ln g(p) = \sum_{i=1}^{r}\alpha_i \ln p_i + 1/2\sum_{i=1}^{r}\sum_{j=1}^{r}\beta_{ij}\ln p_i \ln p_j \qquad [5]$$

c) generalised Leontief

$$g(p) = \sum_{i=1}^{r}\sum_{j=1}^{r}a_{ij}p_i^{1/2}p_j^{1/2} \qquad [6]$$

Of course the last two functional forms are more flexible, at the cost of a higher parameterisation of the model.

Many works also focused on welfare issues, attempting to evaluate if (eventual) gains from TOU rates are sufficient to justify investment in TOU metering technology. An analysis of welfare can be obtained by comparing the value of the individual electricity expenditure incurred during the experiment and the value of the expenditure

[10] See Parks and Weitzel (1984)

that would have allowed the consumer to achieve the same level of indirect utility under the standard tariff. Generally, consumers' welfare increases (and producer surplus declines) when the own-price elasticities are large in absolute values (Aigner et al., 1994).

According to Aigner (1984), the evidence on the TOU experimental rates (implemented in early '80s in USA) often seemed not to justify a mandatory program, sometimes not even for large users. However, the responsiveness varies seriously depending on the zone. For example, countries where the residential air conditioning is primarily responsible of the peak load conditions reveal a higher responsiveness. In other words, there is clear evidence that air conditioning activities are responsive to time-varying prices. Different peak load conditions (occurring for example in the winter) may be connected to activities that are not easily shifted over time. Voluntary implementation may raise selection bias problems, but seems to be more likely to have positive welfare effects.

4.3. Dynamic pricing

Now I focus on works that analysed a more complex tariff structure, involving some form of dynamic pricing. The main difference with respect to a TOU modelling framework is that a flexible tariff introduces uncertainty concerning future prices of electricity. The experimental pricing scheme analysed by Aubin et al. (1995) involved 60 French households and was an example of CPP. There was a standard TOU formulation, with 6 different rates fixed ex-ante: a daily peak and off-peak rate, varying among three different types of days (blue, white and red). The dynamic aspect was represented by the type of the day, which was announced to customers only the day before at 8 pm[11]. With respect to the analysis of TOU, the two stage maximisation procedure described before was then modified to take into account of the uncertainty about future prices.

In this case, in the first stage the consumer is supposed to determine a certain fixed amount of total annual electricity expenditure (y). Then, each day (t = 1,2,…T), he has to allocate y across the different periods, maximising an intertemporal value function. This day-to-day optimisation approach is based on the assumption that customer are not

[11] The number of days of each type was fixed ex-ante. Red days, corresponding to the highest tariffs, approximately corresponded to the periods when system supply was more constrained.

able to implement more complex strategies depending for example on the past observations. So the problem can be stated as follows:

$$V_t = V_t(y_t) = \max_{y_{t+1}, x_1, x_2} u_t(x_1, x_2) + \frac{1}{1+\rho} E_t[V_{t+1}(y_{t+1})] \quad [7]$$

subject to the constraint:

$$y_{t+1} = y_t - p_{1t}x_{1t} - p_{2t}x_{2t} \quad [8]$$

where y_t denotes total expenditure for electricity goods left over at the beginning of period t (with $y_T = 0$), $u_t(\cdot)$ is the within day utility function depending on peak and off-peak consumption, ρ is the discount rate. The Frisch demands of daily electricity consumption are derived from the first order condition of the maximisation problem.

$$x_{it}^{*} = \xi_i(p_{1t}p_{2t}, \lambda_t), i = 1,2 \quad [9]$$

This particular kind of demands does not depend on the level of expenditure, but just on prices and on the term λ_t (the Lagrange multiplier), which can be interpreted as the marginal utility of daily expenditure. This appears as an individual-specific time-varying effect, which can be estimated from repeated observation on each individual. From the first order condition it is also possible to derive the specification for λ_t :

$$\lambda_t = \frac{1}{1+\rho}\lambda_{t+1} \quad [10]$$

By taking logarithms and adding an error term, [10] becomes a random walk with drift. After deriving a parametric specification of [9], that requires making proper assumption on the utility function, on the price index and on the error term, [9] and [10] provide the system that must be estimated. The observation that this system has a state-space representation justify its estimation by means of the Kalman filter, nested in the maximum likelihood estimation[12].

The results indicate that customer responds to the price signal. In fact, even if the intertemporal elasticity of substitution takes low values, it depends crucially on the type of the day, with higher values in white and red days (those with higher rates). Own-price elasticities are quite high (-0.8), and the off-peak demand elasticity with respect to peak prices is much higher in absolute than the peak demand elasticity with respect to off-peak rates. The elasticity of substitution between peak and off-peak consumption is positive but small, indicating a low level of substitutability.

[12] See Aubin et al. (1995) for more details in the derivation of the econometric model

The study terminates with an evaluation of welfare impacts of the CPP tariff, found to be positive for the majority of the participants to the program. Another interesting finding is that customer who submitted to the tariff for a longer period had the best results, suggesting that people need time to get used to the new tariff, in order to better respond to the price signals.

A proper RTP experiment with a 24 hours dynamic tariff was conducted by Herriges and al. (1993). Their analysis relies on a particular set of data concerning large industrial users, most likely to generate RTP benefits that exceed metering and communications costs. In fact, during a first period, a number of customers were all facing a standard TOU. In the second period, some customers voluntarily enrolled to the RTP tariffs, and they were randomly assigned to either the control or to the treatment group. Thus, the control group, that continued to face the standard TOU rates, provides the experimental counterfactual for evaluating the program.

The model follow the standard cost-minimisation procedure, and a nested CES function is specified to define the aggregate monthly price index. This means that the functional form is [4], but here p_i is the daily price index, defined as follows:

$$p_i = \left(\sum_{h=1}^{24} \alpha_{i,h} p_{d,h}^{\ -\lambda} \right)^{-1/\lambda} \qquad [11]$$

In this type of representation, a price change in any hour causes daily price index changes through [11] and the p_i changes cause monthly price index changes through equation [4]. This means that the intraday and interday elasticity of substitution can be measured by estimating, respectively, the parameters $\sigma_H = 1 - \lambda$ and $\sigma_I = 1 - \rho$.

The energy demands (for each hour) are derived by applying the Shephard's lemma:

$$x_{i,h} = \frac{\partial C}{\partial g(p)} \cdot \frac{\partial g(p)}{\partial p_{d,h}} = C_p \cdot \frac{\partial g(p)}{\partial p_{d,h}} \qquad [12]$$

The use of the logarithm of the ratio between the demands in the test and baseline periods leads to a model that gives directly the estimates of the parameter σ_H and σ_I:

$$\ln \frac{x_{i,h}^{\ 1}}{x_{i,h}^{\ 0}} = A_m + \sigma_H \left(\ln \frac{p_{i,h}^{\ 1}}{p_{i,h}^{\ 0}} \right) + (\sigma_H - \sigma_I) \cdot \left(\ln \frac{p_i^{\ 1}}{p_i^{\ 0}} \right) + \varepsilon_{i,h} \qquad [13]$$

where the superscripts 1 and 0 indicate the test and baseload period, A_m is a monthly constant and $\varepsilon_{i,h}$ denotes the error term. The model was then modified in order to be estimated pooling observations from both test and control customers.

$$\left(\ln \frac{x_{i,h}^{1}}{x_{i,h}^{0}} \right)_j = A_{mj} + (\tau + J_j \sigma_H) \cdot \left(\ln \frac{p_{i,h}^{1}}{p_{i,h}^{0}} \right) + [\eta + J_j (\sigma_H - \sigma_I)] \cdot \left(\ln \frac{p_i^{1}}{p_i^{0}} \right)_j + \varepsilon_{j,i,h}$$

[14]

The subscript *j* denotes each customer and the dummy J is equal to 1 if the observation concerns a test customer and 0 otherwise.

The estimates of the intraday and interday elasticities of substitution were positive and in the order of 0.1 and 0.2 respectively, supporting a certain ability of firms to switch energy consumption. Also, the response was not uniform across firms, confirming that the possibilities of substitution depends crucially on the available technology and on the type of activity that is concerned.

Patrick and Wolak (2001) analyse more in depth this issue and find substantial heterogeneity across industries in the pattern of their within-day price responses, with reference to data about industrial and commercial customers purchasing electricity in the England and Wales electricity market, based on half-hourly prices. As in the other RTP designs, these customers are informed (by the distribution utility, i.e. the Regional Electricity Company), about the half-hourly energy price the day before their consumption occur. However, there are two peculiar characteristics in this tariff, which both introduce an higher dynamic uncertainty. First, the information received is only an *ex-ante forecast* of the actual price that will be paid[13]. Second, there is an additional demand charge levied on the average capacity used by each customer during the three half-hour load periods during the year corresponding to the system peak loads (*triads*)[14]. Clearly, the triads are known only at the end of the year.

To model this framework, the firm is assumed to minimise its daily *expected* variable cost, choosing optimally its 48 half-hourly energy consumption and the daily consumption of all other inputs that can be varied within the day, subject to the constraint of producing a certain planned level of output, and given the amount of within-day fixed inputs available. Formally,

$$\min_{x_1,\ldots x_{48}, z_1,\ldots,z_{48}} \sum_{i=1}^{48} [E(p_{id}) + 2/3 pr(DC_{id} = 1) \cdot p_D] \cdot x_i + \sum_{j=1}^{48} p_{jd} z_j$$

$$\text{subject to } Y_d = f(x_1,\ldots,x_{48}, z_1,\ldots z_{48}, F_d, W_d, U_d) \qquad [15]$$

[13] One component of the price (called UPLIFT) is determined ex-post and is known only 28 days after the consumption.

[14] Moreover, these triads must be separated each others by at least ten days.

where the subscript i denotes each half-hour period, the subscript d denotes each day, x is the quantity of energy, z the other half-hourly variable inputs, F the vector of fixed inputs, W is a vector containing information on the weather conditions, U is the unobservable. The price of energy consumption is composed by the expected value of the "regular" half-hourly rate (p_{id}), and the demand charge (p_D) multiplied by the probability that a certain period will be a triad[15]. The latter is modelled on the basis of publicly data available, in particular on the basis of the triad warnings given by the REC to its customers. A generalised McFadden price index is chosen in order to allow the half-hourly consumptions to be substitute or complement among each others.

The solution of this problem leads to a system of 48 electricity demands. Since specifying the own and cross price elasticities would lead to estimate 1,176 free parameters, prior restrictions on the form of the matrix of elasticities must be imposed. The work presents the estimates of this system for 5 different industrial sectors, classifying the data on customers according to the British Industry Classification (BIC). A price index based on the BIC too was used to approximate the unknown value of the price of the other variable materials.

The results indicate that the water supply industry is the most responsive to the price signals, with values of own-price elasticities up to -0.27, very large especially if taking into account the very high volatility of the wholesale prices. The other industries showed lower flexibility. Patrick and Wolak (2001) show that the estimate on the price elasticities can be used to build a demand-side bidding function for the distribution utility.

Summarising, even if literature provides a wide picture of techniques and estimates, it can be stated that industrial, commercial and residential electricity consumers can respond to price signals. The extent to which the benefits would exceed the costs of implementing time-varying tariffs, and which one would be the most appropriate, is an open issue. Aigner (1984) stated that probably benefits from TOU would not justify wide investments in metering. But in 20 years the technology has experienced substantial changes and dramatically reduced costs of metering and communication. Improvements in technology can support the implementation of more complex dynamic tariffs, making more efficient the customer response. Aubin et al. (1995) found that real time price signals can be welfare improving, even if the result is limited to a voluntary

[15] The 2/3 in the formula is due to two reasons: the three comes from the fact that there are three triad in a year, and the two from the requirement that 2 MW of capacity is necessary to produce 1MWH of energy during a half-hour period.

experiment and cannot be extended to the population. The sample of the experiment and the positive correlation of price responsiveness to factor such as the heating space suggest that high-income and high-electricity consumers may have the highest benefits, and this can raise equity concerns. However, one should not forget the long run implications of a wide implementation of dynamic pricing schemes, aiming at reducing the need of investment in new capacity. Such benefits spread out across all consumers.

5. Conclusions and suggestions for future research

Demand response programs provide incentives for retail customers to reduce demand for electricity during peak hours. The benefits are related to a more efficient use of resources, because customers can partially shift consumption to non-peak hours, thus reducing the excess capacity that should be built. Moreover, when utilities are exposed to a competitive wholesale market, demand responsiveness plays a key role in reducing the price volatility and the eventual market power of suppliers. Though all these effects are desirable, the best way to achieve them remains a discussion topic. Evolution in technology seems to suggest dynamic pricing as the most promising method, but a careful assessment of the benefits versus the cost of implementation is necessary. In this paper I reviewed several empirical works aiming at evaluating the effects of time-varying retail prices.

Generally, it can be stated that customers are able to respond to price signals, but the extent of such a response varies widely across users. This can raise equity issues and can imply the necessity to apply a different rate structure among different groups. Large industrial users are the ones more likely to have benefits above costs from the implementation of a RTP tariff. However, Patrick and Wolak (2001) have shown that price elasticities are very different across industries. A collection of data with firm-specific characteristics could be desirable to estimate the impact of a different cost structure of the firm (in terms of capital stock and labour) on the ability to shift consumption. From an econometric point of view, dynamic models have favoured a day-to-day optimisation, assuming that customers are not able to formulate more complex strategies. This is a convenient hypothesis especially to avoid that the matrix of elasticities of substitution to be estimated become extremely large. The exploration of

the validity of this assumption allowing a longer time horizon can provide another direction for future research.

When it comes to small retail consumers, probably a complex RTP tariff is just not feasible, whereas TOU or CPP can be the preferred options. However, another big question is whether to allow a voluntary or a compulsory program. Equity concerns addresses the choice towards the voluntary option, but an assessment of the long term benefits requires attention. First, any empirical application must take into account that customers need time to get used to a new tariff and to understand the potential savings that it may offer. Second, it must be clear that only a wide implementation of time-varying tariffs can produce a substantial price responsiveness in the aggregate demand, whose benefits will spread out across all users.

REFERENCES

Aigner D.J., Newman J. and Tishler A., "The Response of Small and Medium-Size Business Customers to Time-of-Use (TOU) Electricity Rates in Israel", *Journal of Applied Econometrics*, 9(3), 283-304.

Aigner, D.J. (ed.) (1984), "Welfare Econometrics of Peak-Load Pricing for Electricity", *Journal of Econometrics*, Annals 3, 26.

Aubin C., Fougere D., Husson E and Ivaldi M. (1995), "Real Time Pricing of Electricity for Residential Customers: Econometric Analysis of an Experiment", *Journal of Applied Econometrics*, 10, 171-191.

Bohi D.R. and Zimmerman M. (1984), "An Update to Econometric Studies of Energy Demand", *Annual Review of Energy* 1984(9), 105-154.

Borenstein S. (2001), *Frequently Asked Questions About Implementing Real-Time Electricity Pricing in California for Summer 2001*, University of California Energy Institute. Working paper

Borenstein S. (2003), "Time-Varying Retail Electricity Prices: Theory and Practice," in Griffin and Puller, eds., *Electricity Deregulation* (provisional title), Chicago, University of Chicago Press, forthcoming.

Borenstein S., Jaske M. and Rosenfeld A. (2002), *Dynamic Pricing, Advanced Metering, and Demand Response in Electricity Markets*. Center for the Study of Energy Markets. Working Paper CSEMWP-105

Herriges J.H., Baladi S.M, Caves D.W. and Neenan B.F. (1993), "The Response of Industrial Customers to Electric Rates Based upon Dynamic Marginal Costs", *The Review of Economics and Statistics*, 75(3), 446-454.

Lafferty R., Hunger D., Ballard J., Mahrenholz G., Mead D. and Bandera D. (2002), *Demand Responsiveness in Electricity Markets*, presented at FERC-DOE Demand Response Conference, February 2002.

Mountain, D.C. and Lawson E.L. (1992), "A Disaggregated Nonhomothetic Modeling of Responsiveness to Residential Time-Of-Use Electricity Rates", *International Economic Review*, 33, 181-207.

Parks, R.W. and Weitzel D. (1984), "Measuring the Consumer Welfare Effects of Time-Differentiated Electricity Prices", *Journal of Econometrics*, Annals 3, 26, 35-64.

Patrick R.H. and Wolak F.A. (2001), *Estimating the Customer-Level Demand for Electricity Under Real-Time Market Prices*, NBER Working Paper 8213.

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

Sweeney J.L. (1984), "The Response of Energy Demand to Higher Prices: What Have We Learned?", *American Economic Review*, 74(2), 31-37.

Taylor L.D. (1975), "The Demand for Electricity: A Survey", *The Bell Journal of Economics*, 6(1), 74-110.

Yoshimura H. (2003), *Making Demand Response Work in New England*, presented at the Northeast Energy and Commerce Association, January 2003.