Second-best Pricing for Incomplete Market Segments: Applications to Electricity Pricing

Nicolas Astier*

January 3, 2020

Abstract

Due to technological, political or practical considerations, simple rate structures prevail in many market segments. This reality contrasts with the fundamental theorems of welfare economics where prices are fully differentiated by time, location and contingency of delivery. This paper develops a tractable framework to design simple rate schedules under a large family of exogenous constraints. One can then easily assess the relative efficiency gains from using more complex price schedules and speculate about the absolute value of the benefits that may arise from R&D or lobbying efforts to remove existing technological or political barriers. Conveniently, implementation relies on basic machine learning techniques and typically available information. Retail electricity pricing in both France and California are used as example applications.

Keywords: second-best, imperfect pricing, incomplete markets, utility pricing
JEL: H21, L94, L95, Q41

*I am grateful to Michel Le Breton, Estelle Cantillon, Richard Green, Charles Kolstad and seminar participants at the Toulouse School of Economics, IFPEN Paris, Stanford, and IFN Stockholm for helpful comments and suggestions. Remaining errors are mine. The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement number 609642.

*Stanford University. Contact: nicolas.astier[at]stanford.edu

1 Introduction

As end consumers, most of the goods and services we purchase are priced according to relatively simple rules. Some services such as phone plans are even sold at a fixed monthly fee, (almost) independently of usage. More generally, many market segments clear according to price schedules that only partially reflect the structure of supply costs. This widespread reliance on simple rates sharply contrasts with the fundamental theorems of welfare economics, which define "commodities" as goods or services differentiated by their location, date and contingency of delivery, and traded accordingly at state-contingent prices.

The practical challenges raised by a strict application of the Arrow-Debreu theorems have of course been discussed in many occasions. For example, in his debate with R. Coase on marginal cost pricing, W. Vickrey acknowledges the practical necessity of using simple rates: "*it is at least doubtful whether any advocate of marginal cost pricing has ever seriously proposed that prices should slavishly follow marginal costs in every detail, without making some allowance for administrative costs involved in such detailed rate structures and for the fact that, beyond a certain point, the consumer may become so confused that the more intricate rate schedule would cease to function effectively as a guide to consumer choice, thus losing its* raison d'être" (Vickrey, 1948). Relatedly, Radner (1968) notes a few years later that the "*Arrow-Debreu world has been criticized as a model of reality, or even as a normative model of planning, for requiring the existence of too many markets.*"

These observations raise an important question which has received surprisingly little attention: *how simple should simple price schedules be?* Increasing rate complexity indeed often requires costly investments (technology upgrades, lobbying, etc.), and such costs may at some point outweigh the expected benefits from using more sophisticated price schedules.

This paper develops a simple framework to address this issue. It proposes a very tractable method to design simple rate schedules under a large family of exogenous constraints, which makes it possible to assess the relative efficiency gains from using more sophisticated rates. If one is willing to make quantitative assumptions about demand characteristics, the absolute value of the opportunity cost induced by the exogenous constraints enforced may be estimated. This cost may then be compared to the cost of the efforts (e.g. R&D, lobbying) needed to remove technological or political barriers. Conveniently, the proposed framework proves easily implementable using basic machine learning techniques and typically available data. As an example of application, we discuss retail electricity pricing in both France and

California. It is shown that simple rates known as "time-of-use" only grasp a modest share of possible efficiency gains, and that significantly increasing their complexity does not pay off. Indeed, the inherent inefficiencies embedded in this family of tariffs prevent them from improving welfare beyond a certain level.

While other applications of our framework can easily be envisioned (e.g. Pigouvian taxation or financial product design), public utility pricing is probably the most relevant one. It indeed combines all the common rationales for using simple rate structures. First, rates are faced by end consumers who may not want to incur the cognitive costs of understanding complex price schedules.[1] Second, metering and billing water, natural gas or electricity with a fine granularity require non-trivial investments.[2] Third, political considerations usually impose a variety of exogenous constraints on rate design. For instance, one may enforce identical prices across a given geographical area, or forbid price variations upon certain contingencies. Among public utility sectors, the electricity industry is a particularly relevant area to apply our framework. It indeed plays a key role in the fight against climate change and is experiencing significant transformations both in the structure of supply costs (e.g. increasing share of wind and solar generation) and in demand characteristics (e.g. electric vehicles, batteries). Failing to adapt to this new environment may induce very costly inefficiencies, which is giving birth to many debates, such as the relevance of rolling out smart-meters (Joskow, 2012; Léautier, 2014) or the inefficiencies created by policies such as net-metering or increasing-block tariffs.

Not surprisingly, this paper builds on a huge literature to which it is hard to do justice. This paragraph provides a tentative – and necessarily incomplete – overview of related work. First, the design of second-best rates must rely on a well-defined measure of welfare. Starting with Dupuit (1844), a large literature has considerably improved the definition of such a metric, with notably contributions from Walras, Hicks, Hotelling, Samuelson, or Debreu (1954). While theoretically appealing, the obtained measures typically involves unrealistic information requirements, compromising empirical applications. The derivation of local approximations, either in a partial (Ramsey, 1927) or a general (Boiteux, 1951b; Kolm, 1969) equilibrium framework,[3] paved the way to empirical applications, a path taken by Harberger

[1] Joskow and Tirole (2005) however note that rational inattention is not *per se* a valid rationale for using simple rate structures.

[2] When Georges Westinghouse started selling electricity using an alternative-current system, the technology to meter consumption did not even exist.

[3] Diamond and McFadden (1974) and Kay (1980) later revisited these contributions through the lens of the expenditure function, while also discussing local approximations of welfare in the neighborhood of suboptimal

(1964, 1971) and many followers since then (Chetty, 2009). Second, one can envision several reasons why the first-best outcome cannot be achieved, which are explored in the literature on second-best policies. Unfortunately, the analysis of second-best problems is not, and arguably cannot be, as holistic as the theory of the first-best (Lipsey and Lancaster, 1956; Davis and Whinston, 1965; Bohm, 1967). In particular, one should carefully define the set of feasible policies as well as additional constraints that prevent the achievement of the first-best outcome.[4] Many different settings may then be envisioned, for which the merits of piecemeal policies have to be assessed.[5] Incomplete markets are a class of second-best problems of particular interest here. This literature takes note of the fact that real-life assets traded in financial markets typically consist in bundles of atomistic "state claims" or Arrow-Debreu commodities.[6] If the dimension of the vector space generated by the contingent returns on tradable assets is lower than the number of underlying Arrow-Debreu commodities, markets are incomplete (Sharpe, 2011; Carvajal et al., 2012). Relatedly, Radner (1968, 1972) studies a distinct class of problems where market incompleteness arises because of cognitive limitations, which impose restrictions on agents' information structure. Finally, the present paper relates to a large literature on imperfect taxation and pricing, and notably to the framework developed by Jacobsen et al. (2019), which is described in more details below. Within this literature, imperfect pricing of electricity has recently attracted particular attention,[7] although relevant theoretical insights can be traced back to the debates on marginal cost pricing (Coase, 1946; Vickrey, 1948) and the subsequent development of the theory of peak-load pricing (Boiteux, 1949, 1951a; Steiner, 1957; Turvey, 1968). Indeed, in the wake of the California electricity crisis, several studies have looked at the efficiency of various types of residential rates.[8] One approach has been to rely on simulations to assess the

equilibria as in Dixit (1975).

[4]Guesnerie (1975) lists three reasons why the first-best outcome may not be reached. First, some markets may not be organized or cleared (e.g. incomplete markets, externalities, etc.). Second, lump-sum transfers may not be feasible. Third, market power distortions may be present. This paper is motivated by the first rationale.

[5]For example, Ramsey (1927) and Boiteux (1956) study a situation where an entity must raise a given amount of revenue using linear taxes instead of lump sum transfers. More generally, the taxation literature has studied different situations where a subset of commodities may not be taxable, either at the individual (Corlett and Hague, 1953; Diamond and Mirrlees, 1971a), industry (Boiteux, 1956; Guesnerie, 1975) or country (Viner, 1950; Meade, 1955) level.

[6]For example, in the absence of contingent markets, a given good may be seen as a bundle of all the Arrow-Debreu commodities under all contingencies that may materialize (Newbery and Stiglitz, 1982).

[7]Interestingly, this fact somewhat echoes the historical development of the other areas of the literature we reviewed. Indeed, important contributions to the theory of second-best were motivated by problems faced by the electricity industry (Drèze, 1964). This industry also happens to face incomplete markets issues (Newbery, 2016; De Maere d'Aertrycke et al., 2017).

[8]On the supply side, Wolak (2011b) also estimates the decrease in short-term generation costs enabled by more granular pricing on the spatial dimension.

welfare consequences of moving away from flat rates (Borenstein and Holland, 2005; Holland and Mansur, 2006; Borenstein, 2013). Another has been to use field or natural experiments to measure consumers' response to various shades of dynamic pricing (Faruqui and Sergici, 2010; Wolak, 2011a; Jessoe and Rapson, 2013; Ito et al., 2013; Ito, 2014). Several recent papers discuss the inefficiencies arising from imperfect electricity pricing (Borenstein and Bushnell, 2018; Mcrae and Wolak, 2019), or suggest ways to optimize second-best policies (Blonz, 2016).

This paper studies the problem faced by a given entity (e.g. a utility) that must design, for a given market segment, a linear pricing schedule while relying on a limited number of prices and facing additional exogenous constraints. While this problem has been hinted at by Kolm (1969) or Guesnerie (1980),[9] it has not – to the best of my knowledge – received a systematic treatment. This gap in the literature on second-best problems is thus addressed. This work also contributes to the literature on incomplete markets by endogenously designing sets of composite commodities sharing a common average price, while the existing literature usually considers that the composition of security bundles is exogenous. On the empirical front, this paper provides complementary insights to the work by Jacobsen et al. (2019). In particular, the structure of second-best price schedules is made endogenous to the optimization problem, which enables us to consider many different exogenous constraints in a systematic manner. As example applications, we document the achievable welfare gains from various approaches to refine residential rates in both France and California. In California, our framework also provides a salient illustration of the on-going massive shift in the optimal second-best rates notably induced by the increasing share of solar generation.

The rest of the paper is organized as follows. Section 2 develops our theoretical framework and shows how second-best prices may be computed using simple machine learning techniques. Section 3 uses this framework to design second-best retail electricity rates in France. Section 4 carries a similar analysis for California. Section 5 concludes.

2 From theory to practical implementation

This section defines and solves the second-best problem studied in this paper. In order to clarify how our framework relates to the existing literature, we first rely on a fairly general

[9]Guesnerie (1980) concludes his study by noting that "*in practice there exist only a few differentiated tax rates associated with a few groups of commodities. Taxation concerns aggregate commodities when pricing policies are concerned with elementary goods. This feature, which has been ignored in this paper, should be incorporated in further studies on the coordination of pricing and taxation policies.*"

problem statement and then specialize to partial equilibrium, a simplification likely to be made in most empirical applications.

2.1 Theoretical framework

Let $j \in \{1, .., J, J+1, ..., \bar{J}\}$ index the Arrow-Debreu commodities of the economy. We consider the problem faced by a social planner who can set the vector of prices $\mathbf{p} \equiv \{p_1, ..., p_J\}$ for the first J commodities. These prices are faced by consumers in the market segment of interest. Either by reference to a general equilibrium model (Boiteux, 1951b; Debreu, 1954; Kay, 1980) or to an explicit objective function (Ramsey, 1927; Harberger, 1964) one can define a vector of welfare-maximizing prices $\mathbf{p}^*_{\bar{J}} \equiv \{p^*_1, ..., p^*_{\bar{J}}\}$.

Since we have mind applications to market segments representing only a fraction of the total market, we assume constant returns-to-scale in production. While not absolutely needed,[10] this simplifying assumption is commonly made (see for example Corlett and Hague (1953); Harberger (1964); Baumol and Bradford (1970); Diamond and Mirrlees (1971a,b); Diamond and McFadden (1974)) since it allows to focus only on the price system faced by consumers. Assuming optimal lump-sum redistribution and small deviations from optimal prices, a second-order Taylor approximation of the deadweight-loss arising because of the use of suboptimal prices p_1, ..., $p_{\bar{J}}$ is (Kay, 1980):

$$-\frac{1}{2}\sum_{i=1}^{\bar{J}}\sum_{j=1}^{\bar{J}}(p_i - p^*_i)(p_j - p^*_j)\frac{\partial x_i}{\partial p_j} \tag{1}$$

where $x_i(\mathbf{p}, \mathbf{u})$ is the compensated demand for commodity i. Welfare measures proposed in the literature vary regarding the exact point at which $\frac{\partial x_i}{\partial p_j}$ should be evaluated, or remain ambiguous about it.[11] This ambiguity usually does not matter in practice since compensated demand functions are generally assumed to be linear, either to get global results (Ramsey, 1927; Harberger, 1964), or as a local approximation driven by limitations in the statistical power achievable with available data.

To simplify notations, we assume that $\bar{J} = J + 1$ so that there is a single commodity in the rest of the economy. Unless indicated otherwise, this good is assumed to be the numeraire in a partial equilibrium setting. Although our welfare measure can have deeper

[10] See for example Boiteux (1951b) or Kay and Keen (1988) for a symmetric treatment of producers and consumers.

[11] Kay (1980) recommends to use the prices and utility levels of the distorted situation considered.

foundations, it is worth focusing on the partial equilibrium case for two reasons. First, it constitutes a simplification relevant for most empirical applications. Second, it will enable us to draw connections more easily with the rest of the literature.

Consumer k chooses a bundle $(x_1, x_2, ..., x_{J+1})$ so as to maximize her utility function given her available income w_k and prices $(p_1, p_2, ..., p_J)$:

$$\max_{(x_1,x_2,...,x_{J+1})} U^k(x_1, x_2, ..., x_J) + x_{J+1}$$
$$\text{s.t.}$$
$$\sum_{j=1}^{J} p_j x_j + x_{J+1} \leq w_k$$

The system of first-order conditions $\partial x_j U^k(x_1, x_2, ..., x_J) = p_j$ defines consumer k indirect demand:

$$\mathbf{x^k}(\mathbf{p}) \equiv (x_1^k(\mathbf{p}), ..., x_J^k(\mathbf{p})) \equiv \operatorname*{argmax}_{(x_1,x_2,...,x_J)} U^k(x_1, x_2, ..., x_J) - \sum_{j=1}^{J} p_j x_j \tag{2}$$

The aggregate indirect demand in the market segment of interest is then the sum of individual indirect demands:

$$\mathbf{x}(\mathbf{p}) = \sum_k \mathbf{x^k}(\mathbf{p}) \tag{3}$$

Producers are assumed to have locally constant returns-to-scale with marginal costs $\mathbf{p}^*$. Pareto optimality then requires the maximization of:

$$W(\mathbf{p}) \equiv \underbrace{\sum_k U^k(\mathbf{x^k}(\mathbf{p})) - \mathbf{p}\mathbf{x}.(\mathbf{p})}_{\text{Consumers' surplus}} + \underbrace{(\mathbf{p} - \mathbf{p}^*).\mathbf{x}(\mathbf{p})}_{\text{Producers' surplus}} = \sum_k U^k(\mathbf{x^k}(\mathbf{p})) - \mathbf{p}^*.\mathbf{x}(\mathbf{p}) \tag{4}$$

Using the first-order conditions of consumers' problem, the first-order condition with respect to p_j for surplus maximization can be written:

$$\frac{\partial W}{\partial p_j}(\mathbf{p}) = \sum_{i=1}^{J} (p_i - p_i^*) \frac{\partial x_i}{\partial p_j}(\mathbf{p}) = 0 \tag{5}$$

These conditions are indeed met when $p_j = p_j^*$ for all j. Second-order derivatives are:

$$\frac{\partial^2 W}{\partial p_{j'} \partial p_j}(\mathbf{p}) = \frac{\partial x_{j'}}{\partial p_j}(\mathbf{p}) + \sum_{i=1}^{J} (p_i - p_i^*) \frac{\partial^2 x_i}{\partial p_{j'} \partial p_j}(\mathbf{p}) \tag{6}$$

Focusing on small deviations in the neighborhood of $\mathbf{p}^*$ (or assuming linear indirect demand functions) we finally get:

$$W(\mathbf{p}) \simeq W(\mathbf{p}^*) + \frac{1}{2}\sum_{i=1}^{J}\sum_{j=1}^{J}(p_i - p_i^*)(p_j - p_j^*)\frac{\partial x_i}{\partial p_j}(\mathbf{p}^*) \tag{7}$$

The second term is indeed the deadweight-loss measure of equation (1).

Hence, our second-best problem of interest has the following structure:

$$\max_{\mathbf{p}} \frac{1}{2}\sum_{i=1}^{J}\sum_{j=1}^{J}(p_i - p_i^*)(p_j - p_j^*)\frac{\partial x_i}{\partial p_j}(\mathbf{p}^*) \equiv \frac{1}{2}(\mathbf{p}-\mathbf{p}^*)^T H(\mathbf{p}-\mathbf{p}^*)$$
$$\text{s.t.}$$
$$\text{constraint } \mathcal{C}$$

where $H \equiv \left(\frac{\partial x_i}{\partial p_j}(\mathbf{p}^*)\right)_{ij}$ denotes the matrix of indirect demand derivatives and constraint $\mathcal{C}$ is discussed below. Three points are worth emphasizing. First, the matrix H is symmetric as the Hessian matrix of a well-behaved function and a special case of the Slutsky matrix. Second, for convex economies, an assumption likely to be made to be able to define $\mathbf{p}^*$ in the first place, H will be semi-definite negative meaning the objective function takes non-positive values. Third, the formulation of the problem assumes linear pricing, implicitly restricting the class of second-best problems studied. Extension to non-linear pricing is discussed below.

The theoretical framework discussed so far is fairly general and must be further specialized. One may for example assume constraint $\mathcal{C}$ to be $(\mathbf{p}-\mathbf{p}^*).\mathbf{x}(\mathbf{p}) \geq R$ which is the second-best problem studied by Ramsey (1927) and Boiteux (1956). A second-best problem closer to the problem studied in this paper can be found in Jacobsen et al. (2019). Their paper presents an elegant approach to "*quantify the efficiency costs of constraints on the design of externality-correcting tax schemes, or more generally the costs of imperfect pricing, using simple regression statistics*". They consider a situation where the social planner faces restrictions in the vector of taxes that he may choose. In their framework, a tax t_j on commodity j corresponds to the price wedge $p_j - p_j^*$ considered here. They further define an exogenously-given vector $\mathbf{f} \equiv \{f_1, ..., f_J\}$ which they interpret as an observable characteristic that is used in practice to proxy for the externality. Given these definitions, they study the situation where the constraint $\mathcal{C}$ requires the tax schedule to be a linear combination of $\mathbf{f}$:

$$t_j = \alpha + \beta f_j$$
(constraint $\mathcal{C}$ of Jacobsen et al. (2019))

where α and β are two constants. If one denotes $\mathbf{e} \equiv \{1, ..., 1\}$, this constraint may be rewritten as:

$$(\mathbf{p} - \mathbf{p}^*) \in \mathbb{R}\mathbf{e} + \mathbb{R}\mathbf{f}$$

that is $(\mathbf{p} - \mathbf{p}^*)$ must belong to the two-dimensional vector space generated by $\mathbf{e}$ and $\mathbf{f}$. Jacobsen et al. (2019) approach thus naturally extends to the family of constraints $\mathcal{C}$:[12]

$$(\mathbf{p} - \mathbf{p}^*) \in E \text{ where } E \text{ is a vector space of dimension } n < J$$

In Jacobsen et al. (2019), the set E is exogenously given and consists in the vector space generated by the observables that may be used to enforce the public policy of interest. The methodology developed in the present paper provides complementary insights by allowing the social planner to select the optimal vector space E. We further enable the social planner to enforce a large family of exogenous constraints on the set within which E may be picked. The rest of this section formalizes these statements.

2.2 Second-best problem of interest (unconstrained case)

We study a situation where the social planner cannot use more than an exogenously given number N of distinct prices in his rate schedule. We assume N to be much lower than J, meaning markets are incomplete for the considered segment. A typical situation where such a constraint $\mathcal{C}$ arises is retail utility pricing. In this context, consumption is usually aggregated over long periods of time and is charged a constant per-unit price for each period.[13] Many other applications fit this framework: small businesses infrequently update their rate schedules; insurance contracts usually specify families of contingencies instead of atomistic ones; most public transportation systems use homogenous prices across zones instead of charging on a point-to-point basis, etc. Importantly, because we are studying a given market segment, the use of coarse prices does not prevent the underlying Arrow-Debreu markets to clear. For example, a wholesale market, a reinsurance market, or competing modes of transportation may ensure market clearing.

[12] Note that this family of constraints is formally similar to the assumption of incomplete markets in the finance literature.

[13] Joskow (1976) notes that defining such periods is a complex and important issue for electricity rates.

Incomplete markets are usually considered to arise because of rigidities in quantities, in the sense that traded assets are composite commodities defined as fixed bundles of underlying Arrow-Debreu commodities or "state claims". By contrast, the approach proposed here consists in fixing the *average prices* of composite commodities that are defined as any bundle chosen from endogenously defined sets of underlying Arrow-Debreu commodities. Indeed, since each Arrow-Debreu state has a positive probability to materialize, a rate schedule implicitly maps each state to one of the N prices $\{\bar{p}_1, ..., \bar{p}_N\}$. Formally, we may represent an unconstrained rate as a function ι such that:

$$\begin{array}{rccc} \iota: & \{1,...,J\} & \rightarrow & \mathbb{R} \\ & j & \mapsto & p_j \end{array}$$

where p_j is the price that consumers face in state j. Using this notation, the second-best problem of interest is then:

$$\begin{gathered} \max_{\iota(.)} \frac{1}{2} \sum_{i=1}^{J} \sum_{j=1}^{J} (\iota(i) - p_i^*)(\iota(j) - p_j^*) \frac{\partial x_i}{\partial p_j}(\mathbf{p}^*) \\ \text{s.t.} \\ \#\{\iota(\{1,...,J\})\} = N \end{gathered} \tag{8}$$

where $\#\{\iota(\{1,...,J\})\}$ denotes the cardinal of the image of the set $\{1,...,J\}$ through function ι.

While the formulation in (8) is useful to clarify how our setting relates to second-best problems in general, a more intuitive description is to divide the problem in two steps. First, the social planner must define the sets of "composite commodities" that will be sold at a constant average price. They are defined as mutually exclusive subsets of $\{1,...,J\}$. Second, the average price of each set of composite commodities must be optimized. Formally, if one denotes $\mathcal{S}_N^J$ the set of N-set partitions of $\{1,...,J\}$, which we characterize as the set of the injunctive functions s mapping $\{1,...,J\}$ to $\{1,...,N\}$,[14] our problem of interest is:

$$\max_{s \in \mathcal{S}_N^J} \left(\max_{\bar{p}_1,...,\bar{p}_N} \frac{1}{2} \sum_{i=1}^{J} \sum_{j=1}^{J} (\bar{p}_{s(i)} - p_i^*)(\bar{p}_{s(j)} - p_j^*) \frac{\partial x_i}{\partial p_j} \right) \tag{9}$$

Note in particular that for a given s that partitions $\{1,...,J\}$ into N sets S_1, S_2, ... S_N, the inner maximization problem may be rewritten:

[14] $s(j) = n$ then means that state j belongs to the nth set of the corresponding partition.

$$\max_{\mathbf{p}} \frac{1}{2} \sum_{i=1}^{J} \sum_{j=1}^{J} (p_i - p_i^*)(p_j - p_j^*) \frac{\partial x_i}{\partial p_j}$$
$$\text{s.t.}$$
$$\mathbf{p} \in \textstyle\sum_{n=1}^{N} \mathbb{R}\mathbf{1}_{S_n}$$

where $\mathbf{1}_{S_n}$ is a vector whose ith coordinate takes the value 1 if $i \in S_n$ and 0 otherwise.

In other words, up to an affine transformation, the inner maximization problem is the second-best problem studied in Jacobsen et al. (2019). This observation clarifies in which sense this paper provides complementary results to theirs. Having made this clarification, the rest of the paper will use the formulation given in equation (9), which proves more convenient to handle.

2.3 Computing second-best prices

We start by showing how our problem may be solved when the underlying Arrow-Debreu commodities are independent. Doing so, we note that practical implementations can rely on basic machine learning techniques. We then propose a default heuristic for situations where significant substitution or complementarity patterns exist.

2.3.1 Independent Arrow-Debreu commodities

This paragraph makes the following assumption:

Assumption 1 (independent commodities) *Underlying Arrow-Debreu commodities are independent, that is* $\frac{\partial x_i}{\partial p_j} = 0$ *for* $i \neq j$.

Under Assumption 1, expression (9) simplifies to:

$$\max_{s \in \mathcal{S}_N^J} \left(\max_{\bar{p}_1, \ldots, \bar{p}_N} \frac{1}{2} \sum_{j=1}^{J} (\bar{p}_{s(j)} - p_j^*)^2 \frac{\partial x_j}{\partial p_j} \right)$$

Proposition 1 (unconstrained independent case)
The second-best price schedule $(\bar{p}_1, \ldots, \bar{p}_N)$ *is given by the N-step function that best approximates the inverse of the cumulative distribution function of first-best prices* p_j^* *weighted by* $|\frac{\partial x_j}{\partial p_j}|$, *when errors are penalized in a quadratic fashion.*

From an implementation perspective, second-best prices are obtained by applying a weighted k-means clustering algorithm to the distribution $\{p_j^*\}_j$ *of first-best prices, with* N *clusters, weights given by* $\{|\frac{\partial x_j}{\partial p_j}|\}_j$ *and using the Euclidian distance.*

Proof.
See Appendix A.1. ∎

The result of Proposition 1 may look familiar to readers used to the literature on non-linear pricing.[15] This literature relies on mechanism design theory to compute price schedules that vary continuously with the quantity sold. A well-known result is that approximating this price schedule with a handful of two-part tariffs is usually enough to grasp most of the gains achievable through non-linear pricing. The similarity between this result and Proposition 1 is not spurious. Indeed, Wilson (1993) describes how one may think of non-linear pricing as monopoly pricing applied to each quantity increment of demand. The slope of the optimal continuous tariff then provides a distribution of optimal marginal prices for each infinitesimal quantity increment. These quantity increments can be seen as analogous objects to the Arrow-Debreu commodities considered in the present paper. Non-linear pricing however involves either a different objective function for the optimizing entity (profit maximization) or an additional exogenous constraint of different nature (e.g. budget balance), not to mention significantly higher information requirements.

Quite conveniently, Proposition 1 can easily be applied to very large datasets, and thus used in situations with a large number of Arrow-Debreu commodities. Indeed, individual observations p_j^* may be first aggregated into a distribution with only a few thousand weighted bins, where weights W_i are given by the sum of the $\{|\frac{\partial x_j}{\partial p_j}|\}_j$ for the p_j^* falling into a given bin. A weighted k-means algorithm may then be applied to the obtained distribution.

2.3.2 General case

Given the difficulty of estimating cross-elasticities, the assumption of independent Arrow-Debreu commodities is likely to be a necessary approximation in many empirical applications. In some cases, however, one may have an idea about – or be willing to test the consequences of – credible substitution and complementarity patterns. Going back to expression (9), one may first note that solving the inner problem should not raise any difficulty.

Lemma 1 (inner problem)
For a given partition s in $\mathcal{S}_N^J$, the optimal price vector $(\bar{p}_1, ..., \bar{p}_N)$ of the composite commodities defined by s is the solution of a linear system.

Proof. For a given partition $s \in \mathcal{S}_N^J$, the problem:

[15] I am grateful to Estelle Cantillon for pointing to this connection.

$$\max_{\bar{p}_1,...,\bar{p}_N} \frac{1}{2}\sum_{i,j}(\bar{p}_{s(i)} - p_i^*)(\bar{p}_{s(j)} - p_j^*)\frac{\partial x_i}{\partial p_j}$$

yields the first-order conditions:

$$\text{For } n \in \{1, ..., N\}, \sum_{i\in s^{-1}(n)} \sum_j (\bar{p}_{s(j)} - p_j^*)\frac{\partial x_i}{\partial p_j} = 0$$

which is a linear system of N equations and N unknowns. ■

A greedy approach would thus be to apply Lemma 1 to all the possible partitions of $\{1, ..., J\}$ and pick the one with the lowest implied deadweight loss. Unfortunately, the number $S(J, N)$ of ways to partition a set of J objects into N non-empty subsets, which defines a sequence known as Stirling numbers of the second kind, asymptotically grows as $\frac{N^J}{N!}$, that is much too fast to use a greedy approach.

Depending on the specific structure of the problem at hand, different heuristics may be used to reduce the subset of partitions over which one should optimize. In particular, the matrix $\{\frac{\partial x_i}{\partial p_j}\}_{ij}$ will be sparse, and diagonal elements are likely to dominate.[16] This suggests a default two-step heuristic described in the following proposition.

Proposition 2 (heuristic for the unconstrained general case)
When the Arrow-Debreu commodities cannot be considered as independent, a default heuristic may be defined as:

- *Apply Proposition 1 replacing the matrix $\{\frac{\partial x_i}{\partial p_j}\}_{ij}$ by the diagonal matrix whose i^{th} diagonal element is $\frac{\partial x_i}{\partial p_i}$ (i.e. ignore non-diagonal elements);*
- *Apply Lemma 1 to the full matrix $\{\frac{\partial x_i}{\partial p_j}\}_{ij}$, taking as given the partition obtained in the first step.*

Testing this heuristic on random matrices with dominant diagonal elements yields encouraging results (see Appendix A.2).

In the absence of additional structure, Propositions 1 and 2 may yield impractical results. Indeed, since no restrictions have been put so far on acceptable partitions of $\{1, ..., J\}$, the obtained sets of composite commodities may exhibit very complex patterns. For example,

[16] "*In dealing with practical problems, the presumptive dominance of the diagonal elements in the matrices of reaction coefficients can be put to good use*" (Harberger, 1964).

assuming commodities are only differentiated by time of delivery, and that the rate designer sets $N = 3$ in the hope to get a simple rate structure, nothing prevents the obtained price schedule from changing every hour in a pattern $\{\bar{p}_1, \bar{p}_2, \bar{p}_3, \bar{p}_1, \bar{p}_2, \bar{p}_3, ...\}$. Arguably, in many applications, such a pattern may not be suitable. The next section thus shows how a broad family of additional exogenous constraints may seamlessly be enforced.

2.4 Enforcing additional constraints on second-best prices

In most applications, additional constraints on feasible second-best prices exist. For example, one may want prices to remain constant during long enough periods not to confuse customers. The availability and/or cost of the technology needed to enforce a given price schedule may add further constraints. Finally, political considerations sometimes impose another layer of constraints. While there may be other situations of interest, we now show how our framework can be extended to account for constraints of the type "*commodity i must be sold at the same price as commodity j*". Many constraints of practical relevance fall into this category. For example, one may want prices to remain constant for all working hours of a given working day, or to be identical within a given county. Formally, this family of constraints translates into the existence of a finest partition $\underline{s} \equiv \{\underline{S}_1, ..., \underline{S}_M\}$ that must be a possible refinement of the partition used to define the sets of composite commodities. In other words, the set of feasible partitions is restricted to the set $\underline{\mathcal{S}}_N^J$ defined as:

$$\underline{\mathcal{S}}_N^J \equiv \{s \in \mathcal{S}_N^J \,|\, \forall m \in \{1, ..., M\},\, \{i_1, i_2\} \in \underline{S}_m \Rightarrow s(i_1) = s(i_2)\} \subset \mathcal{S}_N^J$$

Assumption 2 (finest partition) *The additional constraints on feasible second-best prices may be formalized as the existence of a finest partition* $\underline{s} \equiv \{\underline{S}_1, ..., \underline{S}_M\}$ *that must be a possible refinement of the partition that ends up defining the optimal sets of composite commodities.*

The problem of interest then becomes:

$$\max_{s \in \underline{\mathcal{S}}_N^J} \left(\max_{\bar{p}_1, \dots, \bar{p}_N} \frac{1}{2} \sum_{i=1}^{J} \sum_{j=1}^{J} (\bar{p}_{s(i)} - p_i^*)(\bar{p}_{s(j)} - p_j^*) \frac{\partial x_i}{\partial p_j} \right) \tag{10}$$

Again, we start by looking at the situation where Assumption 1 holds, that is we assume that Arrow-Debreu commodities are independent. For each set $\underline{S}_m$ belonging to the finest partition $\{\underline{S}_1, ..., \underline{S}_M\}$ we define:

$$\hat{p}^*_m \equiv \frac{\sum_{i \in \underline{S}_m} \frac{\partial x_i}{\partial p_i} p^*_i}{\sum_{i \in \underline{S}_m} \frac{\partial x_i}{\partial p_i}}$$

Without loss of generality, the subsets $\{\underline{S}_1, ..., \underline{S}_j\}$ are assumed to be indexed such that $\hat{p}^*_1 \leq \hat{p}^*_2 \leq ... \leq \hat{p}^*_M$. We further denote:

$$W_0 \equiv 0 \text{ and } W_m \equiv \sum_{j \in \underline{S}_m} |\frac{\partial x_j}{\partial p_j}|$$

Finally, we construct the function $\hat{G}^{-1}$ as:

$$\hat{G}^{-1}(z) = \sum_{m=1}^{M} \hat{p}^*_m \mathbf{1}_{\sum_{k=0}^{m-1} W_k \leq z < \sum_{k=0}^{m} W_k}$$

Proposition 3 (constrained independent case)
The second-best price schedule $(\bar{p}_1, ..., \bar{p}_N)$ is given by the N-step function that best approximates $\hat{G}^{-1}$, when errors are penalized in a quadratic fashion.

From an implementation perspective, second-best prices are obtained by applying a weighted k-means clustering algorithm to the distribution of prices $\{\hat{p}^*_m\}_m$, with N clusters, weights $\{W_m\}_m$, and using the Euclidian distance.

In addition, welfare losses may be decomposed as the sum of two terms:

- *A first term $\frac{1}{2} \sum_{m=1}^{M} \sum_{j \in \underline{S}_m} \left(\hat{p}^*_m - p^*_j\right)^2 \frac{\partial x_j}{\partial p_j}$ measures the welfare losses arising because of the exogenous constraint of enforcing a finest partition of Arrow-Debreu states. In a sense, this term measures the social cost of the finest partition used, that is the minimal deadweight loss that must be incurred because of this exogenous constraint.*
- *A second term, consisting in the remaining welfare losses, measures the additional inefficiencies arising because of the limited number of prices used in the rate schedule.*

Proof.
See Appendix A.3. ■

As illustrated by our empirical applications, Proposition 3 is likely to be a very helpful tool in practice. Indeed, when Assumptions 2 is met, one may apply the exact same approach as in the absence of additional constraints by replacing the underlying Arrow-Debreu commodities with the composite commodities built from the enforced finest partition. These composite commodities should be considered to have "first-best" prices $\hat{p}^*_m$ and "indirect demand slopes" $-W_m$. Appendix B discusses in details a toy example to provide more intuition about the logic behind Proposition 3.

When Assumption 1 is not met, combinatorial issues arise again unless the finest partition is so coarse that $\#\{\underline{\mathcal{S}}^N_X\}$ is small. The discussion of paragraph 2.3.2 applies equally.

2.5 Presence of distortions elsewhere in the economy

In some applications, ignoring the distortions that exist in other markets may significantly bias welfare estimations, and thus the obtained rates. Goulder and Williams (2003) notably highlight the possibility of significant interactions between a commodity tax and distortions in the labor market. To take such general equilibrium interactions into account, we use expression (1) as our welfare measure in this paragraph. Let $J+1$ be a commodity in an external market (e.g. labor) subject to an exogenous distortion τ_{J+1}. We assume commodities 1 to J to be independent. The welfare measure to be optimized is then:

$$-\frac{1}{2}\sum_{j=1}^{J}(p_j - p_j^*)^2\frac{\partial x_j}{\partial p_j} - \sum_{j=1}^{J}(p_j - p_j^*)\tau_{J+1}\frac{\partial x_j}{\partial p_{J+1}} - \frac{1}{2}\tau_{J+1}^2\frac{\partial x_{J+1}}{\partial p_{J+1}}$$

Since the considered entity optimizes on $\{p_1, ..., p_J\}$, it can ignore the last term in the above equation and rewrite it as:

$$-\frac{1}{2}\sum_{j=1}^{J}\left(p_j - p_j^* + \tau_{J+1}\frac{\frac{\partial x_j}{\partial p_{J+1}}}{\frac{\partial x_j}{\partial p_j}}\right)^2\frac{\partial x_j}{\partial p_j} + \frac{1}{2}\sum_{j=1}^{J}\frac{\left(\tau_{J+1}\frac{\partial x_j}{\partial p_{J+1}}\right)^2}{\frac{\partial x_j}{\partial p_j}}$$

Since $\frac{\partial x_j}{\partial p_j}$ is evaluated at p_j^*, the second term may be ignored which leaves us with:

$$-\frac{1}{2}\sum_{j=1}^{J}\left(p_j - p_j^* + \tau_{J+1}\frac{\frac{\partial x_j}{\partial p_{J+1}}}{\frac{\partial x_j}{\partial p_j}}\right)^2\frac{\partial x_j}{\partial p_j} \tag{11}$$

The presence of significant distortions elsewhere in the economy may thus be conveniently handled through a shift in the distribution of $\mathbf{p}^*$, replacing p_j^* by $p_j^* - \tau_{J+1}\frac{\frac{\partial x_j}{\partial p_{J+1}}}{\frac{\partial x_j}{\partial p_j}}$.

The rest of the paper applies our framework to retail electricity pricing in both France and California.

3 Application 1 - residential electricity pricing in France

3.1 Background

By the laws of physics, supply and demand of electricity must be balanced in real time. To complicate things further, storing large amounts of electricity is very costly and power plants are subject to unplanned outages. Real-time balancing hence relies on assets labelled as “dispatchable”, meaning that they have the ability to change their power production/consumption both on demand and relatively quickly. Power plants represent the vast majority of such assets, and have schematically specialized into “high capital cost/low variable cost” and “low capital cost/high variable cost” plants in order to cost-efficiently match demand (Boiteux, 1949). Connecting diversified loads to a large electricity network also allows to get smoother demand profiles, but grid expansion is constrained by high capital costs and low acceptability from local populations. As a consequence, the marginal cost of supplying a given kWh of electricity may fluctuate by several orders of magnitude depending on the time, location and contingency of delivery.

In France, wholesale electricity prices are not differentiated by location. Most trades take place in a country-wide zonal “spot” market[17] where prices vary by the hour and can span from negative values to several thousands of euros per MWh. Residential consumers however do not face spot prices. They are instead billed according to simpler rates, to be chosen among a handful of options which are to a large extent designed by a single entity. Indeed, although retail markets are open to competition, most households have so far chosen to stick to a regulated tariff set by the national energy regulator.[18] Hence, the vast majority of residential consumers currently face either a flat rate tariff called “base” with a marginal price of 147 €/MWh,[19] or a two-period time-of-use (TOU)[20] tariff called “heures pleines/heures creuses” with an off-peak price of 123 €/MWh and a peak price of 158 €/MWh. As smart meters are being rolled-out nationwide since 2015,[21] the French retail market is thus entering an interesting period where, although the technical infrastructure to use complex price schedules is becoming available, the existing *status quo* does not take advantage of these new degrees of freedom. The framework developed in this paper can thus

[17] Technically speaking, the “spot” market is actually a day-ahead forward market.

[18] As of December 31, 2018, 25.3 millions households out of 32.7 millions were billed according to a regulated tariff (Commission de Régulation de l’Energie, 2019).

[19] As of May 2019. More precisely, the marginal price is 145.2 €/MWh for customers who subscribed 6 kVA or less, and 147 €/MWh for customers who subscribed higher power connections.

[20] A TOU rate defines *ex ante* several time periods. Electricity is then charged at period-specific constant prices.

[21] A full roll-out is expected to be completed by 2021.

be used to assess where the highest efficiency gains are likely to come from.

3.2 Data

The data we use come from two sources. First, hourly spot prices between 2012 and 2016 are recovered from EpexSpot. Second, hourly gross electricity consumption is downloaded from the open data platform of French network operators. Table 1 provides relevant summary statistics.

Variable	Spot price (€/MWh)	Gross consumption (GW.h)
Mean	40.00	54.61
Standard deviation	23.93	12.08
Min	-200	29.68
Max	1938.5	101.65
Number obs.	43848	

Table 1: Summary statistics of data used for France (period 2012-2016)

3.3 Assumptions

Since national aggregate consumption has been roughly flat during the period 2012-2016, we consider each year as a different random realization of possible spot prices. We thus define underlying Arrow-Debreu commodities as being indexed by $\{h, y\}$ where the hour h is a proxy for the time dimension, and the year y a proxy for the contingency dimension.

We then use the distribution of spot prices as a proxy for the distribution of first-best prices $\mathbf{p}^*$. The assumption of exogenous spot prices is motivated by two reasons. First, we are interested in rates that only apply to a market segment, and not to the whole market. Second, in a similar application for PJM, Jacobsen et al. (2019) found that accounting for convexities in the cost function did not have a big impact on their results. A more accurate proxy for social marginal costs would of course account for network costs and unpriced externalities as in Borenstein and Bushnell (2018). Since our main goal is to illustrate how our framework may be applied in practice, we abstract from these considerations.

We assume that the Arrow-Debreu commodities are independent. This assumption is likely to prove relatively weak in our context. Indeed, our goal is to compute second-best tariffs consisting in only a handful of prices. As a consequence, consumers will face a constant price during extended periods of time, meaning substitution and complementarity patterns will have less opportunity to play a significant role compared to a situation where households would face many different prices. Two specifications for aggregate demand are used. The first one assumes that $\frac{\partial x_j}{\partial p_j}$ is constant across states, meaning that the indirect demand function is locally linear with a constant slope. Variations in demand are then driven by shocks shifting the level of the intercept.[22] The second specification assumes that $\frac{\partial x_j}{\partial p_j}$ is proportional to gross consumption, which corresponds to a linear approximation of an isoelastic demand. Indeed, in the neighborhood of p_0,[23] we have for an isoelastic demand with elasticity ϵ:

$$\frac{\partial x_j}{\partial p_j} \simeq \frac{\epsilon}{p_0} x_j$$

Since we do not directly observe the gross consumption x_j of the market segment of interest, we assume it is proportional to national gross consumption.

The final set of assumptions relates to the definition of the finest admissible partition. In what follows, it is assumed as an illustration that tariffs may only discriminate between:

- Seasons: January to March, April to June, July to September, October to December;
- Types of day: working day and weekend;[24]
- Hours of the day: blocks of three consecutive hours (0 am to 3 am, 3 am to 6 am, ... , 9 pm to 0 am).

When designing time-of-use tariffs, we further assume that price differentiation along the contingency dimension is not possible.

3.4 Results

While an optimal flat rate is found to charge 40 €/MWh (resp. 42.5 €/MWh) under the assumption of a constant-slope (resp. isoelastic) demand,[25] a two-period TOU tariff meeting the constraints enforced by our finest partition is found to charge a rate of 29.1 €/MWh (resp. 31.4 €/MWh) off-peak and of 48.3 €/MWh (resp. 50.9 €/MWh) on-peak. Interestingly, the ratio of on-peak to off-peak prices is found to be 1.6-1.7 while the ratio used in current tariffs in only of 1.3. This milder ratio is likely to be driven by the other components of the electricity bill, namely network fees and taxes. Figure 1 depicts the obtained price schedule

[22] This specification is for example used in Léautier (2014).

[23] Consumers currently face either a constant price p_0, or two prices mildly differentiated around p_0

[24] We assume working days to be Monday to Friday and do not label explicitly every single French national holiday.

[25] Energy procurement indeed only represents about a third of residential consumers' bills, the rest consisting in taxes and grid access fees.

under the assumption of an isoelastic demand.[26] Consistently with Figure 1, the current TOU rate defines the off-peak period as night time. In addition, a recent innovation in electricity rates has been to offer tariffs where the off-peak period extends to the whole weekend.

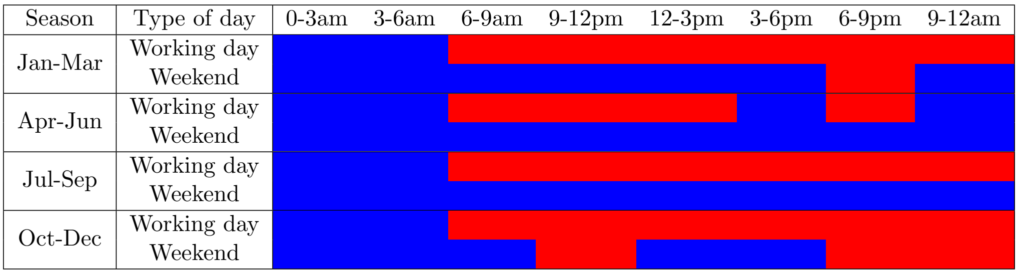

Figure 1: Obtained rate schedule (isoelastic demand) for a two-period time-of-use tariff (blue: off-peak period; red: peak period)

The relative efficiency gains to move from a flat rate to a two-period time-of-use tariff is estimated to be 16% for a constant-slope demand (resp. 12% for an isoelastic demand). Figure 2 shows however that the incremental benefits from further increasing the number of TOU periods vanish quickly: using more than four or five periods provides virtually no additional efficiency gains. Applying Proposition 3, it thus appears that most ($\sim$ 80%) of the inefficiency losses incurred are driven by the finest partition we chose.

Where do the additional efficiency gains to move from a two-period to a three-period TOU tariff come from? The rate schedule obtained for a three-period tariff (Figure 3) shows that they would mainly arise from the ability to reflect seasonal variations in spot prices, and more specifically higher prices during winter working days. Converting these relative efficiency gains into absolute numbers requires to assess the order of magnitude of the flat-rate benchmark deadweight loss. This exercise relies more heavily on the assumptions made about $\frac{\partial x_j}{\partial p_j}$, and is thus inherently more speculative. To provide a rough estimate, if one assumes an isoelastic demand with $\epsilon = -0.1$,[27] and notes that the residential sector only accounts for about 35% of the national gross consumption (RTE, 2019), the order of magnitude of the (short-run) inefficiencies arising because of the use of a flat rate is a few tens of million euros per year.[28] This estimate is consistent with the findings of Holland and Mansur

[26] TOU periods obtained under the assumption of a constant-slope demand are almost identical.

[27] This value is commonly used as an estimate of the short-run price elasticity for electricity (Borenstein and Holland, 2005; Holland and Mansur, 2006).

[28] The point estimate of the flat-rate deadweight loss computed over the period 2012-2016 is 90 M€/year.

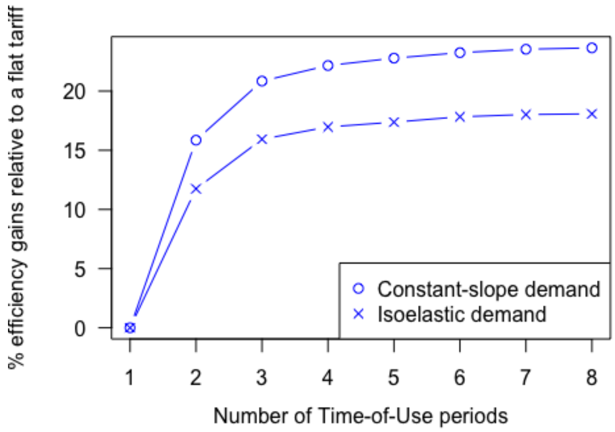

Figure 2: Efficiency gains from increasing the number of Time-of-Use periods (France)

(2006), but is however likely to understate long-term inefficiencies (Borenstein, 2005; Borenstein and Holland, 2005).

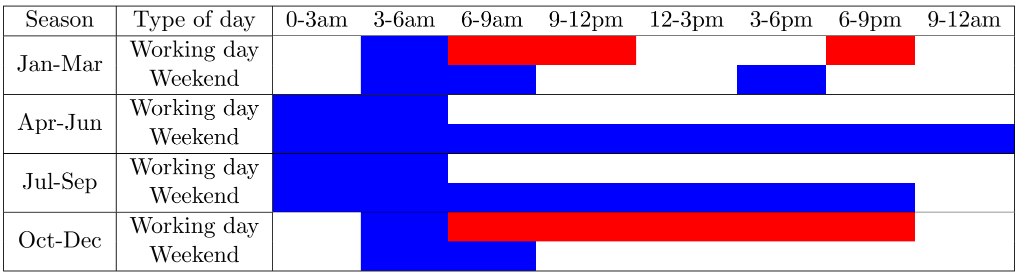

Figure 3: Obtained rate schedule (isoelastic demand) for a three-period time-of-use tariff (blue: off-peak period; white: shoulder period; red: peak period)

Figure 2 may look to some extent discouraging. It indeed suggests that even sophisticated time-of-use tariffs are unlikely to get more than 20% of the possible efficiency gains relative to a flat-rate benchmark. The failure of TOU tariffs to grasp a higher share of achievable efficiency gains is intrinsic to their *ex ante* nature. Indeed, they cannot signal to consumers the high scarcity events that follow very rare contingencies. By contrast, the

This value is however very sensitive to tail events (see below). Reporting the order of magnitude rather than an exact number hence seems more appropriate.

electricity system is usually most under stress because of stochastic shocks such as the outage of a big power plant or a cold wave. In the move from two to three price instruments, it is thus worth exploring the relative merit of a critical-peak-pricing (CPP) tariff. CPP tariffs allow the utility to raise the price of electricity *ex post*, typically on a day-ahead notice, but only for a limited number of days per year. This is not a new idea: such a tariff, called "Effacement Jours de Pointe" (EJP) was launched in France as early as 1982 by EDF. A more recent tariff called "Tempo" is built around the same idea. However, the take-up rate of CPP tariffs is now low and they are no more advertised to consumers.

To estimate the potential efficiency gains from a CPP price schedule, we proceed in two steps. First, we rank first-best prices p_j^* in decreasing order and truncate the ordered distribution at the price $p_\dagger^*(N_h)$ such that no more than N_h hours in a given year have a price $p_j^* > p_\dagger^*(N_h)$. This constraint reflects the fact that, although a CPP tariff is able to discriminate between different contingencies *ex post*, the number of critical events in a given year will be capped to some threshold N_h.[29] The critical-peak price $\bar{p}_{CPP}$ is obtained as a $\frac{\partial x_j}{\partial p_j}$-weighted average of the p_j^* greater than $p_\dagger^*(N_h)$. Second, a two-period time-of-use rate is computed for the subset of states with $p_j^* < p_\dagger^*(N_h)$. These two steps can be iterated to adjust N_h. Somewhat surprisingly, because the distribution of spot prices is very skewed to the right due to a few very high prices over the period 2012-2016, the highest efficiency gains are found for CPP tariffs with a very low N_h. For example, a CPP tariff that may be called at most 5 hours per year is found to provide over 50% efficiency gains compared to the flat-rate benchmark. However, the corresponding event price exceeds 1000 €/MWh, that is 25 times the flat rate tariff. In addition, such a tariff is very sensitive to the tail of the distribution of $\mathbf{p}^*$, which is inevitably poorly captured with only five years of data. We thus report results with $N_h = 200$ hours. This choice would translate into a critical-peak price of 120 €/MWh (resp. 124 €/MWh) under the assumption of a constant-slope demand (resp. isoelastic demand). Efficiency gains of about 15-16% relative to a two-period TOU. By contrast, a three-period TOU tariff only achieves 5-6% efficiency gains relative to the same benchmark.

Our application to retail electricity pricing has shown how our framework can be used to compute second-best pricing schedules under two main exogenous constraints: (i) using simple rates; and (ii) having a limited ability to reflect contingencies. Our next application to

[29] Additional constraints about critical-event periods, such as imposing a high price for the whole day, can easily be enforced in the same spirit as in Proposition 3 (one would them trim the right tail of "first-best" prices for the finest partition sets $\hat{p}_m^*$). Since we are only interested in the order of magnitude of the potential gains here, we do not add such constraints to simplify the exposition.

California will further illustrate (iii) the extent to which important changes in the generation mix can impact the stability over time of optimal TOU rates; and (iv) how our framework may be used in the context of spatially differentiated first-best prices.

4 Application 2 - residential electricity pricing in California

4.1 Background

Since April 2009, California has adopted a nodal design for wholesale electricity markets. The California Independent System Operator (CAISO) routinely computes locational marginal prices (LMPs) for several thousand transmission grid nodes. However, the retail market segment has not been liberalized, and three investor-owned utilities (IOUs) – Pacific Gas and Electric, Southern California Edison and San Diego Gas and Electric – are granted regional monopolies to supply the vast majority of consumers. California retail consumers thus do not face LMPs for two main reasons. First, utilities themselves do not directly face LMPs. Their load is instead aggregated into three "Default Load Aggregation Points" (DLAP) and bid in and settled at DLAP prices. Second, retail rates are fixed through an administrative process and do not simply pass through wholesale prices.

Load aggregation across broad geographical areas has motivated a long debate between the Federal Energy Regulatory Commission (FERC) and the CAISO. In September 2006, FERC Order on the Market Redesign and Technology Upgrade (MRTU) accepted the DLAP approach as a simplified temporary measure but requested that the number of Load Aggregation Point (LAP) zones would be increased after three years of experience with the new market. Subsequent studies and stakeholder processes led by the CAISO in 2010 and 2013 concluded that defining relevant geographical areas smaller than DLAPs but larger than single nodes would be challenging and involve significant costs. FERC accepted to delay the deadline to disaggregate DLAPs, but at first refused to completely waive the obligation to define smaller zones. As new studies reached similar conclusions, FERC finally accepted on October 2015 CAISO's request to maintain the *status quo*. The concerns expressed during this debate provide significant motivation for the present work. First, enabling further load disaggregation was argued to require significant investments.[30] Second, benefits were deemed hard to assess in a reliable manner, and were proxied by looking at several *ad hoc*

[30] "*Cost estimates total $21.7 million in one-time implementation costs with $2.5 million annually for slight disaggregation, and $147 million in one-time implementation costs with $12.3 million annually for fully nodal disaggregation*" (CAISO, 2015).

metrics capturing the dispersion of LMPs.[31]

Retail electricity rates in California are set by the California Public Utilities Commission (CPUC), along with other Local Regulatory Authorities. While TOU rates have existed since the 1970s, they were mainly designed to address summer peak demand and did not foresee the recent massive increase in production from renewable resources. On July 3, 2015, the CPUC decided to move completely California residential customers towards updated TOU tariffs by 2019-2020. In this context, CAISO (2016) investigates what such a tariff may look like from a system perspective, and proposes a four-period TOU tariff with (i) a super-peak period between 4pm and 9pm on July and August weekdays; (ii) a peak period between 4pm and 9pm for all other weekdays, and from 12pm to 4pm during July and August weekdays; (iii) a super-off-peak period between 10am and 4pm on weekdays in March and April, and on weekends/holidays in all months except July and August; (iv) an off-peak period for all other hours (see Appendix C). Their study provides further motivation for the present work. First, simplicity constraints prevailed: "*As counseled by CPUC staff, the CAISO sought to minimize complexity and time-period variations when evaluating potential TOU periods and structures*" (CAISO, 2016). Second, in the absence of a formalized framework, the analysis by CAISO focused successively on how net load varies with seasons, then with geographical areas, then by day of the week, etc. While this approach definitely makes sense, our framework provides a tool to screen all these features simultaneously.

4.2 Data

As in CAISO (2016), we focus on the service area of the three main IOUs. Day-ahead hourly prices are recovered for their respective DLAP over the period 2011-2018 from CAISO website.[32] Hourly consumption over the same period for the three Transmission Access Charge (TAC) zones comes from the same data source and are used to proxy for gross consumption at the three DLAPs. Finally, day-ahead hourly LMPs for the 23 sub-load aggregation points (SLAPs) within IOUs service territories are recovered for the period 2011-

[31] "*Quantifying benefits in a meaningful manner is challenging and would require several contentious assumptions to be made. Therefore given the expected negligible benefits, the ISO does not intend to conduct extensive quantitative studies but rather will speak to the benefits qualitatively*" (CAISO, 2015).

[32] While real-time (RT) LMPs at the node level would represent a better proxy for the underlying social marginal cost of power, we use DA prices for simplicity since utilities themselves do not face real-time node-level LMPs. If one assumes that DA prices are close enough to be an indirect-demand-weighted average of RT prices, a natural extension of Proposition 3 shows that using DA prices instead of RT yields the correct TOU rate but underestimates welfare losses by an additive term which captures the discrepancy between DA and RT prices.

2016 to explore the spatial dimension.[33] In the absence of load data at a finer granularity than TAC zones and of geographical information about nodes locations, the SLAP level was preferred to physical nodes to illustrate one possible application of our framework in the presence of spatial variations. Table 2 provides some relevant summary statistics.

Variable		Mean (std)	Min	Max	# SLAPs
PG&E	DLAP price ($/MWh)	36.2 (18.8)	-17.3	946.4	16
	TAC load (GW.h)	11.5 (1.9)	7.8	21.3	
SCE	DLAP price ($/MWh)	37.0 (21.9)	-28.6	1000.0	6
	TAC load (GW.h)	11.9 (2.6)	7.5	25.8	
SDG&E	DLAP price ($/MWh)	38.1 (23.0)	-71.2	1007.5	1
	TAC load (GW.h)	2.3 (0.5)	1.4	4.7	
Number obs.	70128				

Table 2: Summary statistics of data used for California (period 2011-2018)

4.3 Assumptions

Between 2011 and 2018, utility-scale solar photovoltaic has grown from a negligible share to about 12% of total electricity generation in California (CAISO, 2019). To illustrate the impact of this evolution on second-best TOU rates, we divide the data in two samples, 2011-2014 and 2015-2018. One part of the analysis consists in comparing the obtained results for each sample, while the other part focuses on the period 2015-2018. Spatial differentiation is discussed in two steps. First, we estimate the magnitude of the inefficiencies arising from implementing a zonal TOU tariff instead of IOU-specific tariffs. Second, we assess the relatively efficiency gains from using SLAPs as distinct pricing zones.

Again, we use the distribution of spot prices as a proxy for the distribution of first-best prices $\mathbf{p}^*$ and assume that Arrow-Debreu commodities are independent. When computing second-best prices using IOUs as the finest geographical unit, we test the same two specifications of demand as in the previous application (constant-slope and isoelastic demand). However, since we lack disaggregated load data, only a constant-slope demand specification is implemented at the SLAP level.

Finally, in order to assess the full potential of TOU tariffs, we enforce a finest partition that can discriminate between months, types of day (weekends vs working days) and hours

[33]Due to changes in the definition of some SLAPs in 2017, we restrict the time window to 2011-2016.

of the day.

4.4 Results

Using data from 2015 to 2018, Figure 4 plots, for each IOU, the relative efficiency gains achievable by sophisticated TOU tariffs (as defined by the enforced finest partition). Despite the freedom left in the definition of TOU periods, achievable gains are found not to exceed 30% of the inefficiencies arising under a flat-rate tariff.[34] Assessing the order of magnitude of these inefficiencies proves again more speculative. Assuming retail consumers represent a third of the total load and currently respond to a price of $180/MWh (EIA, 2019) with a price elasticity $\epsilon = -0.1$, the order of magnitude of these short-term inefficiencies is a few tens of millions of dollars per year for the three IOUs combined.[35]

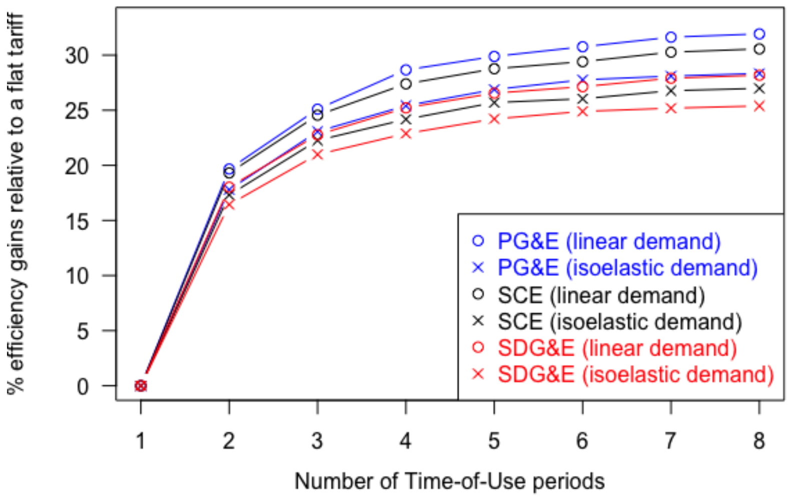

Figure 4: Efficiency gains from increasing the number of Time-of-Use periods (2015-2018, California)

Since four periods appear to be enough to grasp the majority of achievable efficiency gains, we focus on four-period TOU tariffs. For now, we enforce the additional constraint

[34]Hogan (2014) finds a similar result for the PJM service area. Achievable gains using data from 2011 to 2014 are found not to exceed 35-40%.

[35]The point estimate is 45 M$/year.

that the tariff should be the same for all three IOUs. Figure 5 represents the obtained four-period TOU tariff based on data from 2011 to 2014 (assuming an isoelastic demand). While the obtained tariff may not be simple enough to be directly implemented as such, one may in practice gradually impose further structure on the TOU rate (as we for example did in the previous application) and iterate our approach to get simpler rates. The corresponding prices for the different periods are \$27.4/MWh, \$36.8/MWh, \$44.8/MWh, and \$55.3/MWh.[36] Somewhat reassuringly, the obtained rate structure is consistent with the historical rationale for having TOU tariffs in California, which was to address a relatively spread out summer peak demand. Figure 6 depicts the obtained outcome when using data from 2015 to 2018. Corresponding prices are \$21.6/MWh, \$33.7/MWh, \$50.9/MWh, and \$81.6/MWh.[37] One can observe very sharp changes with notably the emergence of (i) a "super-peak" period in July and August; and (ii) solar off-peak hours in the spring and during the weekends, as advocated by CAISO (2016). The narrowness of the "super-peak" period is however driven by the important price spikes that were experienced in 2017 and 2018.[38] If the probability of occurrence of such price spikes is actually less than what is observed in our sample, a TOU tariff built from a forward-looking probabilistic distribution of spot prices would exhibit a more spread out peak period.

On Figure 4, the incremental efficiency gain to move from three to four periods appears to be relatively modest. As in the previous application, one may thus wonder to want extent a three-period TOU coupled with critical peak events may reach a more efficient outcome. Similarly, estimates based on only a few years of data should be considered with caution since we use yearly outcomes to proxy for contingencies. Having this important caveat in mind, we find that while a four-period TOU rate increases efficiency by about 3% relative to a three-period TOU rate, making use of critical peak events instead of a fourth time-of-use period yields about 40% gains relative to the same benchmark.[39] Relative to the flat-rate tariff benchmark, these efficiency gains are estimated to reach about 55%. Additional gains

[36]Respective prices for a constant-slope demand specification are \$24.7/MWh, \$33.4/MWh, \$42.1/MWh, and \$52.9.

[37]Respective prices for a constant-slope demand specification are \$20.6/MWh, \$31.9/MWh, \$46.5/MWh, and \$72.7.

[38]To illustrate this, we computed the incompressible part of the deadweight loss (the first term in Proposition 3) on a yearly basis. While this term keeps a similar order of magnitude for the years 2011 to 2016, it increases 10 to 30-fold in 2017 and 2018.

[39]As in the previous application, we report the results for a CPP tariff where peak events are capped at 200 hours per year. The corresponding critical peak price is \$218/MWh (resp. \$209/MWh) under the assumption of an isoelastic (resp. constant-slope) demand. By contrast, the most efficient design is found to limit critical events to about 60 hours a year, at the political cost of a higher critical peak price of \$358/MWh (resp. \$354/MWh).

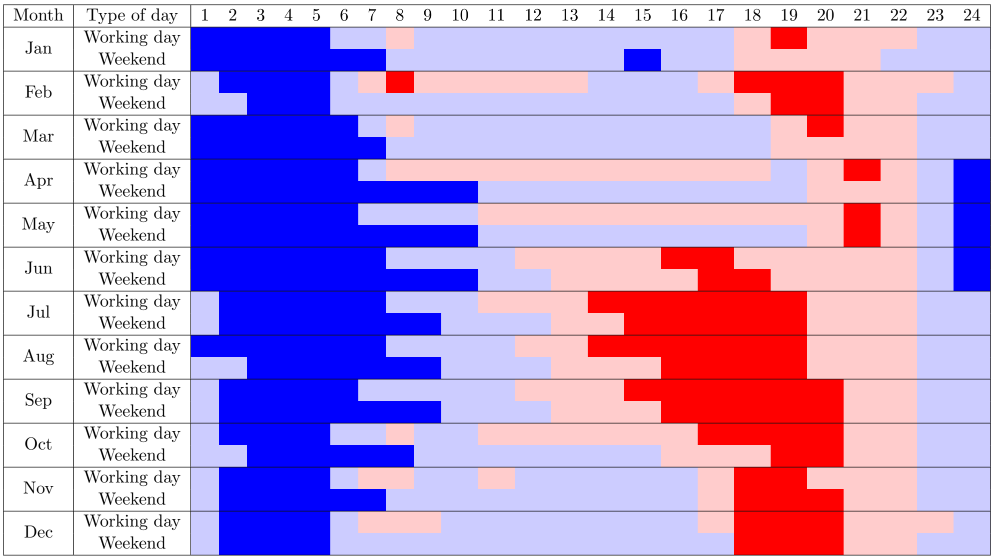

Figure 5: Obtained California-wide TOU tariff (isoelastic demand, 2011-2014 data)

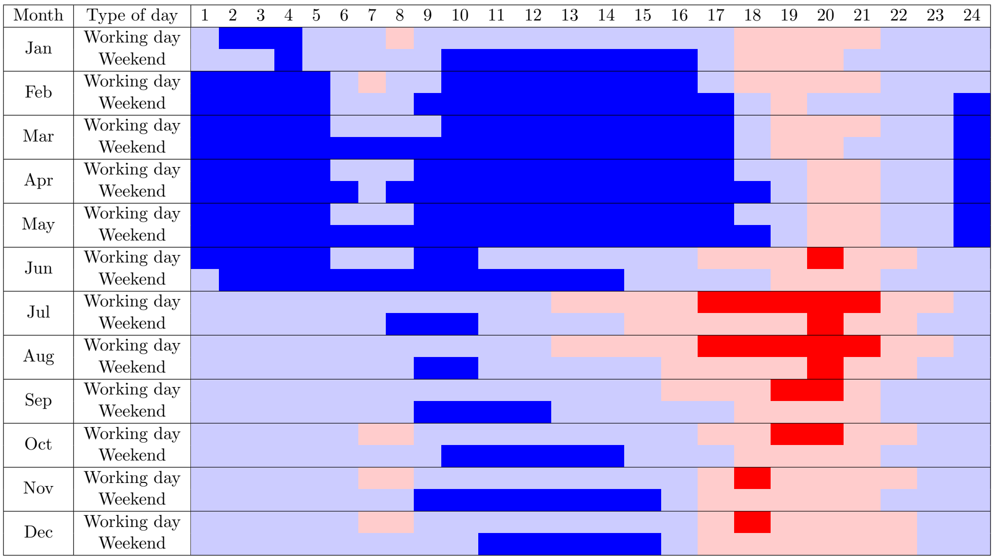

Figure 6: Obtained California-wide TOU tariff (isoelastic demand, 2015-2018 data)

from a CPP tariff are also likely to arise in the long run as less generation capacity will be needed to meet capacity adequacy requirements (Blonz, 2016).

Finally, the framework developed in this paper equally applies to the spatial dimension.

In their discussion with the FERC, the CAISO (2015) argued that, short of a nodal approach, a refinement of load aggregation zones was unlikely to yield significant benefits. This claim may be quantified at two levels of spatial disaggregation. First, at the IOU level, Appendix D shows the obtained IOU-specific four-period TOU tariffs. The graphical similarity between IOU-specific rates is confirmed quantitatively. Moving from a California-wide to an IOU-specific tariff is found to yield efficiency gains of about 1% in both demand specifications. Second, one can investigate whether SLAPs would constitute a more suitable geographical disaggregation unit. In the absence of load information at the SLAP level, we assume a constant-slope specification of the indirect demand. Using data from 2011 to 2016 (resp. 2015 to 2016), we find that tailoring SLAP-specific four-tier TOU tariffs yields efficiency gains of about 2% (resp. 7%) relative to a situation where a uniform TOU is implemented over all SLAPs. Our results are thus consistent with CAISO (2015) analysis that, except for physical nodes, the efficiency gains to use smaller administrative zones for rate design purposes might be relatively modest in California.

5 Conclusion

As end consumers, the rate schedules we face every day are much simpler than state-contingent prices over a full set of Arrow-Debreu commodities. This paper provides a very tractable framework to assess the magnitude of the trade-offs involved when designing such simple rates, and thus to better understand the simplified rate structures that are likely to emerge in different environments.

We study the second-best pricing problem of an entity (e.g. a utility) that must design a linear pricing schedule for a given market segment relying on a fixed number of prices, while facing exogenous "simplicity" constraints. These constraints may be driven by technological, political or practical considerations. After having clearly formalized this problem and how it relates to the vast literature on second-best pricing, we show that it may be solved using basic machine learning techniques and typically available information. Empirical applications prove in turn easy to implement and fast to run. As a consequence, one can explore the relative merits of many different ways to increase rate sophistication. When simplicity constraints derive from technological or political barriers, one can also speculate about the benefits that may be achieved through R&D or lobbying efforts.

Retail electricity pricing in both France and California are used as example applications. We show that time-of-use tariffs can only grasp a modest share of potential efficiency gains, and that the marginal benefits from increasing tariff complexity are very limited beyond three to four time-of-use periods. In France, moving from a two-tier to a three-tier TOU rate would allow to reflect seasonal variations, with higher prices during the winter. In California, the massive increase in installed capacities of photovoltaic generation is progressively shaping "super-peak" periods in July and August, as well as off-peak solar hours in the winter and the spring, as well as during the weekends. In both applications, using critical-peak pricing instead of more sophisticated TOU rates yields much higher efficiency gains, while still falling significantly short of the first-best real-time-pricing benchmark. The issue of spatial aggregation of consumers into "zones" is also discussed for California. Our results suggest that, for this specific example, load disaggregation at the IOU or SLAP level provides little efficiency gains when residential consumers are to be charged a four-period time-of-use tariff.

While our applications have focused on retail electricity pricing, our theoretical framework is fairly general and may prove useful in other areas such as Pigouvian taxation or financial product design.

Appendices

A Proofs

A.1 Proof of Proposition 1

We start by rewriting expression (9) as:

$$\min_{\bar{p}_1,...,\bar{p}_N} \left(\min_{s \in \mathcal{S}_N^J} \frac{1}{2} \sum_{j=1}^{J} (\bar{p}_{s(j)} - p_j^*)^2 |\frac{\partial x_j}{\partial p_j}| \right)$$

In words, we tackle the two steps of the problem in a reverse order. First, for a given price vector $(\bar{p}_1, ..., \bar{p}_N)$, we compute the optimal sets of composite commodities. Second, we optimize the prices $(\bar{p}_1, ..., \bar{p}_N)$.

Step 1: optimal partition for given average prices

Consider an exogenously given price vector $(\bar{p}_1, ..., \bar{p}_N)$. Without loss of generality, we assume $\bar{p}_1 < ... < \bar{p}_N$. If we map the Arrow-Debreu commodity j to the price $\bar{p}_n$, the corresponding efficiency loss is:

$$\frac{\partial x_j}{\partial p_j}(p_j^* - \bar{p}_n)^2$$

It is thus optimal to map the Arrow-Debreu commodity j to the closest $\bar{p}_n$, that is to a set S_n such that $|p_j^* - \bar{p}_n|$ is minimal.

Step 2: optimal prices for the sets of composite commodities

Without loss of generality, we assume that $p_1^* < ... < p_J^*$ (if $p_i = p_j$ for some i and j, one can collapse them into a single commodity with indirect demand derivative $\frac{\partial x_i}{\partial p_i} + \frac{\partial x_j}{\partial p_j}$).

We define the cumulative frequency distribution function of $\{p_j^*\}_j$ weighted by $\frac{\partial x_j}{\partial p_j}$ as:

$$G(p) \equiv \sum_{p_j^* \; s.t. \; p_j^* \leq p} |\frac{\partial x_j}{\partial p_j}| = \sum_{j=1}^{\sup\{i \,|\, p_i^* \leq p\}} |\frac{\partial x_j}{\partial p_j}|$$

In particular, with the notation $G(p_0^*) \equiv 0$, we have:

$$|\frac{\partial x_j}{\partial p_j}| = G(p_j^*) - G(p_{j-1}^*)$$

The "inverse" cumulative frequency distribution may then be defined as:

$$G^{-1}(\omega) \equiv \sum_{j=1}^{J} p_j^* \mathbf{1}_{G(p_{j-1}^*)<\omega\leq G(p_j^*)}$$

In words, $G^{-1}(.)$ is a step function that takes the value p_j^* on the interval $\left]G(p_{j-1}^*), G(p_j^*)\right]$. This interval is of length $|\frac{\partial x_j}{\partial p_j}|$ by definition of G.

From Step 1, and denoting $\bar{p}_0 \equiv 2\inf_j(p_j^*) - \bar{p}_1$ and $\bar{p}_{N+1} \equiv 2\sup_j(p_j^*) - \bar{p}_N$, the optimization problem may be written:

$$\min_{\bar{p}_1,\dots,\bar{p}_N} \frac{1}{2}\sum_{n=1}^{N} \sum_{\frac{\bar{p}_{n-1}+\bar{p}_n}{2}\leq p_j^*<\frac{\bar{p}_n+\bar{p}_{n+1}}{2}} \left(\bar{p}_n - p_j^*\right)^2 |\frac{\partial x_j}{\partial p_j}| \tag{12}$$

Using the objects previously defined, the objective function may be rewritten:

$$\frac{1}{2}\sum_{n=1}^{N} \sum_{\frac{\bar{p}_{n-1}+\bar{p}_n}{2}\leq p_j^*<\frac{\bar{p}_n+\bar{p}_{n+1}}{2}} \int_{G(p_{j-1}^*)}^{G(p_j^*)} \left(\bar{p}_n - G^{-1}(\omega)\right)^2 d\omega$$

Since the p_j^* are ordered and strictly increasing, for any price π there exists an index $\underline{j}(\pi)$ such that:

$$p_{\underline{j}(\pi)}^* \leq \pi < p_{\underline{j}(\pi)+1}^*$$

and by definition of G we have: $G(\pi) = G(p_{\underline{j}(\pi)}^*)$. The above expression may then be rewritten:

$$\frac{1}{2}\sum_{n=1}^{N} \int_{G(p^*_{\underline{j}(\frac{\bar{p}_{n-1}+\bar{p}_n}{2})})}^{G(p^*_{\underline{j}(\frac{\bar{p}_n+\bar{p}_{n+1}}{2})})} \left(\bar{p}_n - G^{-1}(\omega)\right)^2 d\omega$$

which proves the first part of Proposition 1.

It is worth noting that the proof is much shorter if one is willing to take a less rigorous approach. Indeed, assuming that G is differentiable, one can directly rewrite from Step 1 the expression for the deadweight loss as:

$$\min_{\bar{p}_1,\dots,\bar{p}_N} \frac{1}{2}\sum_{n=1}^{N} \int_{\frac{\bar{p}_{n-1}+\bar{p}_n}{2}}^{\frac{\bar{p}_n+\bar{p}_{n+1}}{2}} (\bar{p}_n - p)^2 \, dG(p) = \frac{1}{2}\sum_{n=1}^{N} \int_{G(\frac{\bar{p}_{n-1}+\bar{p}_n}{2})}^{G(\frac{\bar{p}_n+\bar{p}_{n+1}}{2})} \left(\bar{p}_n - G^{-1}(\omega)\right)^2 d\omega$$

which yields the first-order condition:

$$\int_{\frac{\bar{p}_{n-1}+\bar{p}_n}{2}}^{\frac{\bar{p}_n+\bar{p}_{n+1}}{2}} (p - \bar{p}_n) \, dG(p) = 0 \tag{13}$$

k-means implementation

The first-order conditions of equation (13) may be rewritten:

$$\bar{p}_n = \frac{\int_{\frac{\bar{p}_{n-1}+\bar{p}_n}{2}}^{\frac{\bar{p}_n+\bar{p}_{n+1}}{2}} pdG(p)}{\int_{\frac{\bar{p}_{n-1}+\bar{p}_n}{2}}^{\frac{\bar{p}_n+\bar{p}_{n+1}}{2}} dG(p)}$$

In words, $\bar{p}_n$ is the weighted average of the first-best prices p_j^* of the states in S_n, the weights being given by $|\frac{\partial x_j}{\partial p_j}|$. Building on this intuition, for a given partition $\{S_1, ..., S_N\}$, we can define μ_n as the $|\frac{\partial x_j}{\partial p_j}|$-weighted mean of the prices p_j^* of the states that belong to S_n. From:

$$\min_{s \in \mathcal{S}_N^J} \left(\min_{\bar{p}_1,\ldots,\bar{p}_N} \frac{1}{2} \sum_{j=1}^{J} (\bar{p}_{s(j)} - p_j^*)^2 |\frac{\partial x_j}{\partial p_j}| \right)$$

we can then write:

$$\min_{s \in \mathcal{S}_N^J} \left(\min_{\bar{p}_1,\ldots,\bar{p}_N} \frac{1}{2} \sum_{j=1}^{J} (\bar{p}_{s(j)} - \mu_n + \mu_n - p_j^*)^2 |\frac{\partial x_j}{\partial p_j}| \right)$$

which by definition of μ_n simplifies to:

$$\min_{s \in \mathcal{S}_N^J} \left(\frac{1}{2} \sum_{n=1}^{N} \sum_{p_j^* \in S_n} (p_j^* - \mu_n)^2 |\frac{\partial x_j}{\partial p_j}| + \min_{\bar{p}_1,\ldots,\bar{p}_N} \frac{1}{2} \sum_{j=1}^{J} \sum_{p_j^* \in S_n} (\bar{p}_n - \mu_n)^2 |\frac{\partial x_j}{\partial p_j}| \right)$$

It is then obviously optimal to pick $\bar{p}_n = \mu_n$ which simplifies the problem to:

$$\min_{s \in \mathcal{S}_N^J} \frac{1}{2} \sum_{n=1}^{N} \sum_{p_j^* \in S_n} (p_j^* - \mu_n)^2 |\frac{\partial x_j}{\partial p_j}|$$

This latter formulation is precisely the definition of a weighted k-means clustering problem.

A.2 Simulations testing the heuristic of Proposition 2

To test the heuristic described in Proposition 2, we generated random 10×10 matrices with dominant diagonal elements. More specifically, a random symmetric matrix was first drawn, and its diagonal elements were then replaced by minus the absolute value of the sum of the other coefficient on their respective rows. We tested two polar cases where all the non-diagonal elements had the same sign, either positive or negative. Arguably, such situations are likely to represent worst-case scenarios in empirical applications.

We then computed second-best prices in two different manners:

- A greedy approach testing all partitions and keeping the optimal one;
- The heuristic of Proposition 2.

Out of 500 draws, the proposed heuristic found the optimal solution in 20% of the cases. The inefficiency of the heuristic solution relative to the optimal one was lower than 10% in more than 85% of runs, and the worst inefficiency due to having chosen a suboptimal partition was about 50%.

A.3 Proof of Proposition 3

We follow the same two steps as in the proof of Proposition 1.

Step 1: optimal partition for given average prices

Let $(\bar{p}_1, ..., \bar{p}_N)$ be a given vector of average prices for the sets of composite commodities. We want to construct the partition $s \in \underline{\mathcal{S}}_N^J$ that solves:

$$\max_{s \in \underline{\mathcal{S}}_N^J} \frac{1}{2} \sum_{m=1}^{M} \sum_{j \in \underline{S}_m} \left(\bar{p}_{s(j)} - p_j^*\right)^2 \frac{\partial x_j}{\partial p_j} \tag{14}$$

In other words, we want to allocate each $\underline{S}_j$ to a given $\bar{p}_n$ for an exogenously given vector of composite goods' prices $(\bar{p}_1, ..., \bar{p}_N)$.

For a given $m \in \{1, ..., M\}$, the function $Z \mapsto \sum_{j \in \underline{S}_m} \left(Z - p_j^*\right)^2 \frac{\partial x_j}{\partial p_j}$ is a degree-two polynomial which reaches its maximum at $\hat{p}_m^*$. Indeed, we purposely defined $\hat{p}_m^*$ as:

$$\hat{p}_m^* \equiv \frac{\sum_{j \in \underline{S}_m} \frac{\partial x_j}{\partial p_j} p_j^*}{\sum_{j \in \underline{S}_m} \frac{\partial x_j}{\partial p_j}}$$

By symmetry around this maximum, the optimal partition is obtained by mapping each set $\underline{S}_m$ of the finest partition to the price $\bar{p}_n$ that minimizes $|\hat{p}_m^* - \bar{p}_n|$.

Step 2: optimal prices for the sets of composite commodities

Given the optimal partition found in step 1, the objective function of the problem defined by expression (14), which we now want to optimize with respect to $\{\bar{p}_1, ..., \bar{p}_N\}$, can be written:

$$\frac{1}{2}\sum_{n=1}^{N}\sum_{\frac{\bar{p}_{n-1}+\bar{p}_n}{2}\leq\hat{p}_m^*<\frac{\bar{p}_n+\bar{p}_{n+1}}{2}}\sum_{j\in\underline{S}_m}\left(\bar{p}_n-p_j^*\right)^2\frac{\partial x_j}{\partial p_j} \tag{15}$$

We then note that the last sum may be rewritten:

$$\begin{aligned}
\sum_{j\in\underline{S}_m}\left(\bar{p}_n-p_j^*\right)^2\frac{\partial x_j}{\partial p_j} &= \sum_{j\in\underline{S}_m}\left(\bar{p}_n-\hat{p}_m^*+\hat{p}_m^*-p_j^*\right)^2\frac{\partial x_j}{\partial p_j}\\
&= \sum_{j\in\underline{S}_m}\left(\bar{p}_n-\hat{p}_m^*\right)^2\frac{\partial x_j}{\partial p_j}+\sum_{j\in\underline{S}_m}\left(\hat{p}_m^*-p_j^*\right)^2\frac{\partial x_j}{\partial p_j}+\underbrace{2\sum_{j\in\underline{S}_m}\left(\hat{p}_m^*-p_j^*\right)\frac{\partial x_j}{\partial p_j}}_{=0\text{ by definition of }\hat{p}_m^*}\\
&= \left(\bar{p}_n-\hat{p}_m^*\right)^2\sum_{j\in\underline{S}_m}\frac{\partial x_j}{\partial p_j}+\sum_{j\in\underline{S}_m}\left(\hat{p}_m^*-p_j^*\right)^2\frac{\partial x_j}{\partial p_j}
\end{aligned}$$

Defining $W_m \equiv \sum_{j\in\underline{S}_m}|\frac{\partial x_j}{\partial p_j}|$, expression (15) may be rewritten:

$$-\frac{1}{2}\sum_{n=1}^{N}\sum_{\frac{\bar{p}_{n-1}+\bar{p}_n}{2}\leq\hat{p}_m^*<\frac{\bar{p}_n+\bar{p}_{n+1}}{2}}\left(\bar{p}_n-\hat{p}_m^*\right)^2W_m+\frac{1}{2}\sum_{m=1}^{M}\sum_{j\in\underline{S}_m}\left(\hat{p}_m^*-p_j^*\right)^2\frac{\partial x_j}{\partial p_j} \tag{16}$$

This expression has to be optimized with respect to $\{\bar{p}_1,...,\bar{p}_N\}$. As a consequence, the second term can be ignored, leaving us with:

$$-\frac{1}{2}\sum_{n=1}^{N}\sum_{\frac{\bar{p}_{n-1}+\bar{p}_n}{2}\leq\hat{p}_m^*<\frac{\bar{p}_n+\bar{p}_{n+1}}{2}}\left(\bar{p}_n-\hat{p}_m^*\right)^2W_m$$

This expression is precisely the same as equation (12) if one replaces p_j^* by $\hat{p}_m^*$, and the weights $\frac{\partial x_j}{\partial p_j}$ by W_i. The same logic as in Proposition 1 hence applies.

Finally, one can note that Equation (16) decomposes welfare losses into two terms:

- An *exogenous* term arising because of the finest partition constraint:

  $$\tfrac{1}{2}\textstyle\sum_{m=1}^{M}\sum_{j\in\underline{S}_m}\left(\hat{p}_m^*-p_j^*\right)^2\frac{\partial x_j}{\partial p_j}$$

  This term represents the minimal amount of welfare losses to be incurred because of the enforced finest partition.

- An *endogenous* term arising because of the limited number of price instruments used:

$$-\frac{1}{2}\sum_{n=1}^{N}\sum_{\frac{\bar{p}_{n-1}+\bar{p}_n}{2}\leq\hat{p}_m^*<\frac{\bar{p}_n+\bar{p}_{n+1}}{2}}(\bar{p}_n-\hat{p}_m^*)^2\,W_m$$

B Toy example

Consider a retail business that sells apples in two cities $C \in \{A, B\}$. City B being further away from production, it happens to be more costly to serve. Supply cost further depends on the season $S \in \{H, L\}$, where H stands for the harvesting season. Finally, in a given year, yields $Y \in \{G, B\}$ may be either "good" or "bad" depending on random shocks such as weather.

Our simple example thus features eight Arrow-Debreu commodities $j \in \{A, B\}\times\{H, L\}\times\{G, B\}$. To fix ideas, we assume the marginal costs of supply are such that:

$$p^*_{A,H,G} < p^*_{B,H,G} < p^*_{A,H,B} < p^*_{B,H,B} < p^*_{A,L,G} < p^*_{B,L,G} < p^*_{A,L,B} < p^*_{B,L,B}$$

We hence assume that selling during the harvesting season drives down supply costs "more" than being in a good year, which itself drives down costs "more" than selling in the cheaper-to-serve city. For simplicity, we assume that Arrow-Debreu commodities are independent and that the slope of their inverse demand is the same for all commodities. Figure B.1 plots the inverse of the (weighted) cumulative distribution function of these first best prices.[40]

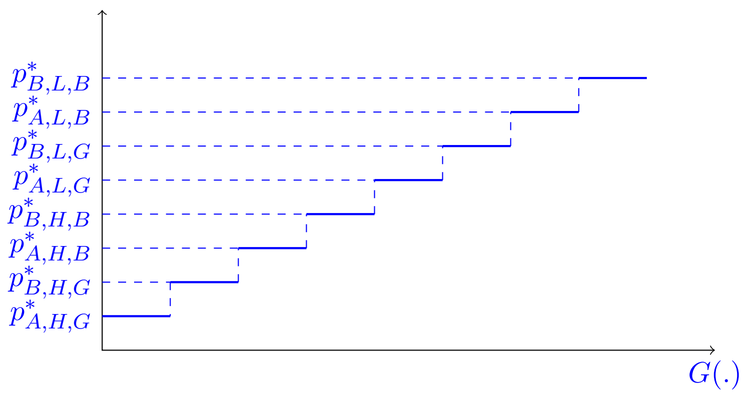

Figure B.1: Inverse cumulative distribution function of (weighted) first-best prices

[40] Assuming that both seasons have similar durations. Regarding contingencies, first-best prices already account for potential differences in probabilities of occurrence.

Assume that, in order to avoid having to handle too many price tags, the business owner considers using only two different prices so that $N = 2$. By Proposition 1, the second-best price schedule is then the best approximation of G^{-1} with a two-step function, which is depicted on Figure B.2.

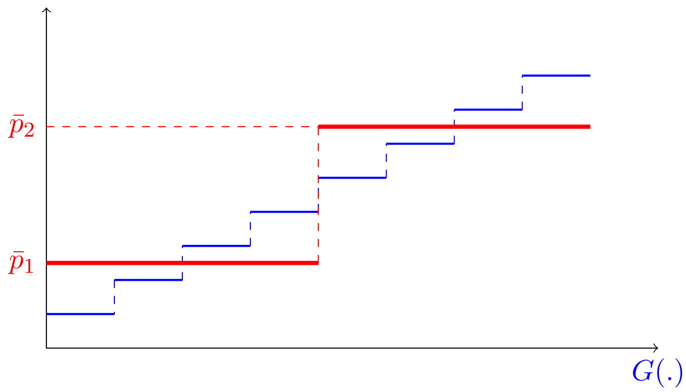

Figure B.2: Unconstrained case with $N = 2$

The obtained price schedule sets a price $\bar{p}_1$ during the harvesting season, and $\bar{p}_2$ during the rest of the year, capturing the fact that this is the most important dimension explaining variations in underlying supply costs.

Now assume that the government, genuinely concerned about its citizens' health, wants to encourage fruit consumption. Their investigations have revealed that the high price of apples during the non-harvesting season significantly discourages people from eating apples during this season. They thus enact a law stating that the price of apples should be constant across a given year. Within our framework, this exogenous constraint can be taken into account through the use of the finest partition $\{\underline{S}_{A,G}, \underline{S}_{B,G}, \underline{S}_{A,B}, \underline{S}_{B,B}\}$ where:

$$\underline{S}_{A,G} \equiv \{\{A, H, G\}, \{A, L, G\}\}; \ \underline{S}_{B,G} \equiv \{\{B, H, G\}, \{B, L, G\}\};$$
$$\underline{S}_{A,B} \equiv \{\{A, H, B\}, \{A, L, B\}\}; \ \underline{S}_{B,B} \equiv \{\{B, H, B\}, \{B, L, B\}\}$$

Following the steps of Proposition 3, one may build the objects $\hat{p}_m$ and $\hat{G}(.)$, which are depicted in green on Figure B.3.

These objects may then be used to build the final price schedule as shown on Figure B.4.

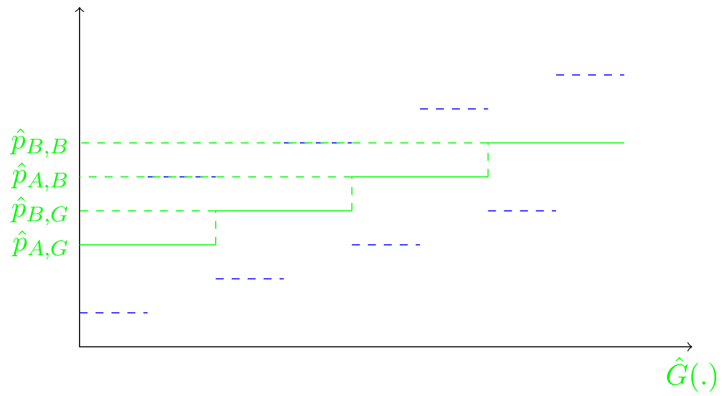

Figure B.3: Constrained case with $N = 2$: building $\hat{p}$ and $\hat{G}$

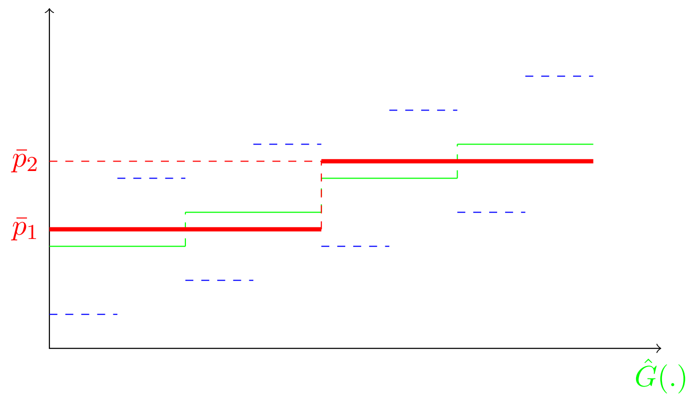

Figure B.4: Constrained case with $N = 2$: obtained price schedule

The obtained price schedule has intuitive features. First, the price gap between $\bar{p}_1$ and $\bar{p}_2$ has decreased because of the inability to reflect seasonal variations in underlying supply prices. Second, the price schedule has become a "good year/bad year" price schedule, where a higher price is charged when yields are low in a given year. This is consistent with the observation that this dimension is the second largest driver of underlying supply cost variations. Third, this price schedule is less efficient. Up to a multiplicative constant, its inefficiency may be measured by the sum of square errors between the applied price schedule (in red) and the supply cost of the underlying Arrow-Debreu commodities (in blue). Our framework thus allows to assess the welfare opportunity cost of the law enacted by the government, either in relative terms compared to the previous *status quo*, or in absolute terms if one makes quantitative assumptions about the slopes $\frac{\partial x_j}{\partial p_j}$.

C Time-of-use rate proposed by CAISO

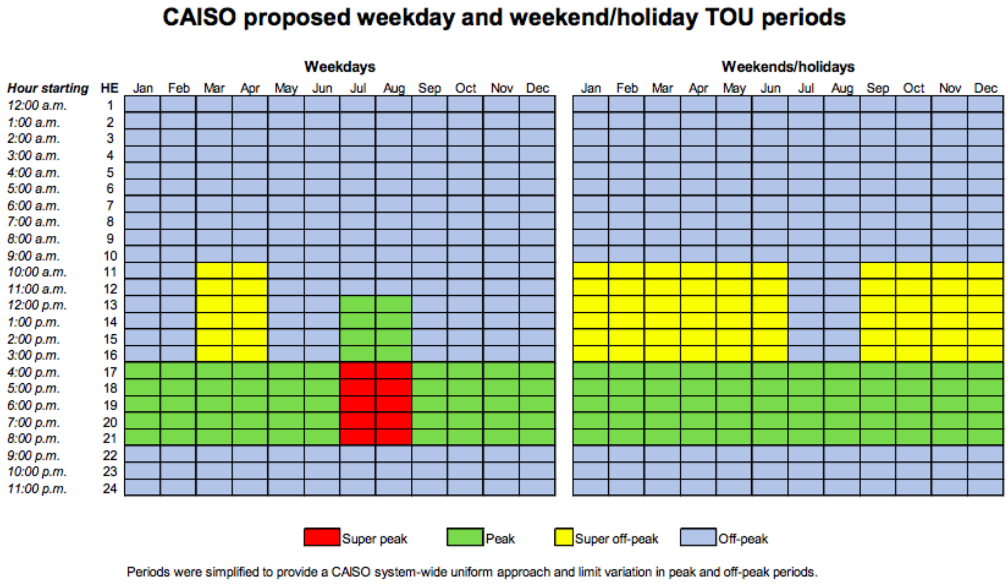

Figure C.1: Time-of-use rate proposed in CAISO (2016)

D Obtained four-period TOU for individual IOUs (isoelastic demand)

Figures D.1, D.2 and D.3 depicts the obtained TOU at the perimeter of each IOU.

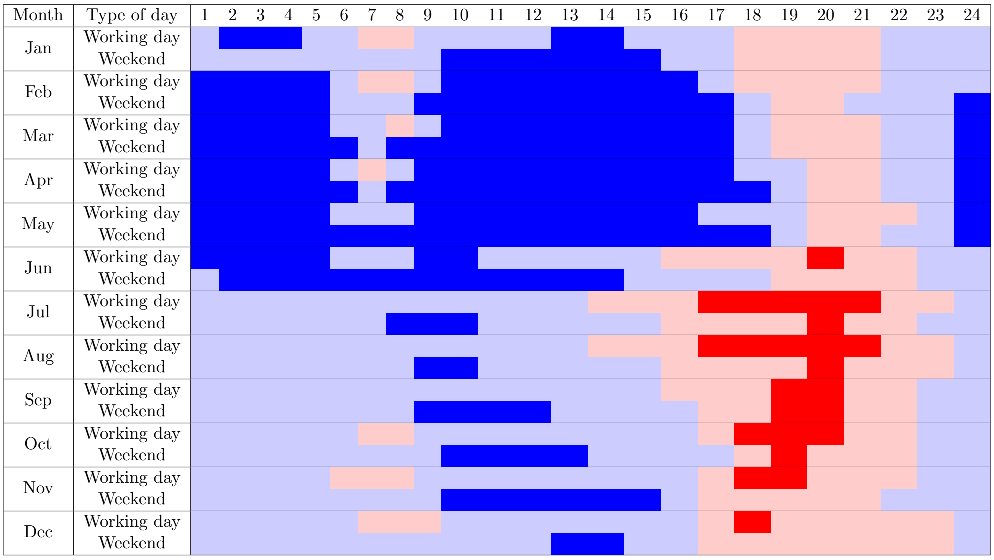

Figure D.1: Obtained rate schedule for a PG&E-only TOU (isoelastic demand, 2015-2018 data)

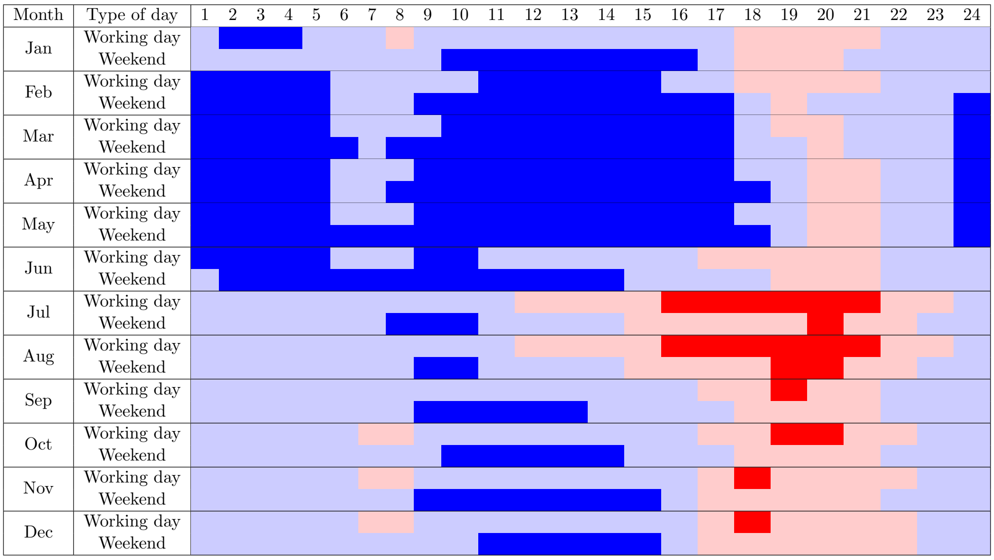

Figure D.2: Obtained rate schedule for a SCE-only TOU (isoelastic demand, 2015-2018 data)

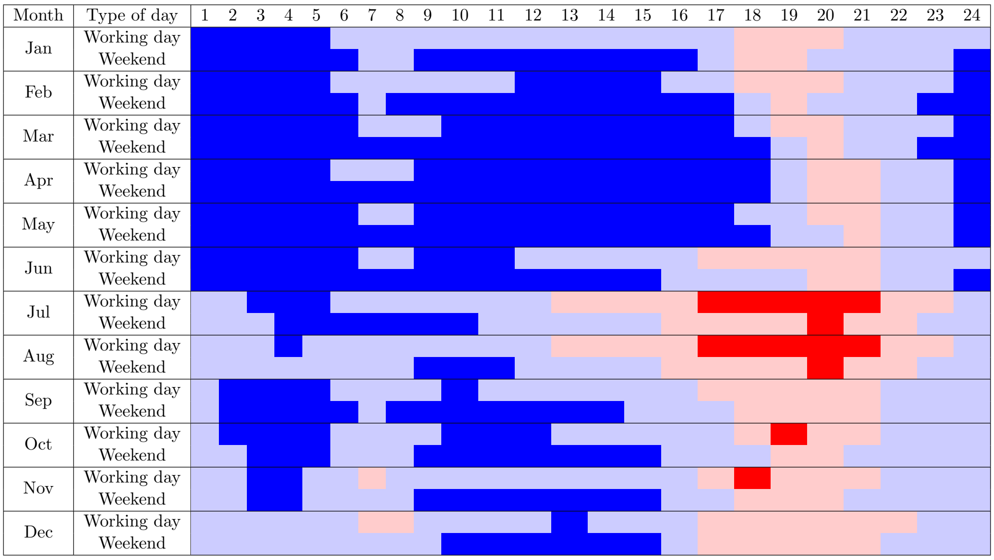

Figure D.3: Obtained rate schedule for a SDG&E-only TOU (isoelastic demand, 2015-2018 data)

References

William J Baumol and David F. Bradford. Optimal Departures From Marginal Cost Pricing. *The American Economic Review*, 60(3):265–283, 1970.

Joshua A Blonz. Making the best of the second-best: Welfare consequences of time-varying electricity prices. *EI@ Haas Working Papers, WP-275*, 2016.

Peter Bohm. On the theory of 'second best'. *Review of Economic Studies*, 34(3), 1967.

Marcel Boiteux. La tarification des demandes en pointe: application de la théorie de la vente au coût marginal. *Revue générale de l'Electricité*, LVIII(8), 1949.

Marcel Boiteux. La Tarification au coût marginal et les Demandes Aléatoires. *Cahiers du Séminaire d'Econométrie*, 1:56–69, 1951a.

Marcel Boiteux. Le "Revenu Distribuable" et les Pertes Economiques. *Econometrica*, 19(2): 112–133, 1951b.

Marcel Boiteux. Sur la gestion des monopoles publics astreints à l'équilibre budgétaire. *Econometrica*, pages 22–40, 1956.

Severin Borenstein. The Long-Run Efficiency of Real-Time Electricity Pricing. *The Energy Journal*, 26(3):93–116, 2005.

Severin Borenstein. Effective and Equitable Adoption of Opt-In Residential Dynamic Electricity Pricing. *Review of Industrial Organization*, 42(2):127–160, 2013.

Severin Borenstein and James B Bushnell. Do two electricity pricing wrongs make a right? cost recovery, externalities, and efficiency. 2018.

Severin Borenstein and Stephen Holland. On the efficiency of competitive electricity markets with time-invariant retail prices. *The RAND Journal of Economics*, 36(3):469–493, 2005.

CAISO. Load granularity refinements - pricing study results and implementation costs and benefits discussion. Technical report, CAISO, 2015.

CAISO. Caiso time-of-use periods analysis. Technical report, CAISO, 2016.

CAISO. 2018 annual report on market issues and performance. Technical report, CAISO, 2019.

Andrés Carvajal, Marzena Rostek, and Marek Weretka. Competition in Financial Innovation. *Econometrica*, 80(5):1895–1936, 2012. ISSN 0012-9682. doi: 10.3982/ecta9837.

Raj Chetty. Sufficient Statistics for Welfare Analysis : A Bridge Between Structural and Reduced-Form Methods. *Annual review of economics*, 1:451–487, 2009.

Ronald Harry Coase. The Marginal Cost Controversy. *Economica*, 13(51):169–182, 1946.

Commission de Régulation de l'Energie. Observatoire des marchés de détail du 4e trimestre 2018. Technical report, CRE, 2019.

W. J. Corlett and D. C. Hague. Complementarity and the Excess Burden of Taxation. *The Review of Economic Studies*, 21(1):21–30, 1953.

Otto A. Davis and Andrew B. Whinston. Welfare Economics and the Theory of Second Best. *The Review of Economic Studies*, 32(1):1–14, 1965.

Gauthier De Maere d'Aertrycke, Andreas Ehrenmann, and Yves Smeers. Investment with incomplete markets for risk : The need for long-term. *Energy Policy*, 105(January):571–583, 2017.

Gerard Debreu. A classical tax-subsidy problem. *Econometrica*, pages 14–22, 1954.

P.A. Diamond and D.L. McFadden. Some uses of the expenditure function. *Journal of Public Economics*, 3:3–21, 1974.

Peter a. Diamond and James a. Mirrlees. Optimal Taxation and Public Production I: Production Efficiency. *American Economic Review*, 61(1):8–27, 1971a.

Peter a. Diamond and James a. Mirrlees. Optimal Taxation II: Tax Rules. *American Economic Review*, 61(3):261–278, 1971b.

Avinash Dixit. Welfare effects of tax and price changes. *Journal of Public Economics*, 4: 103–123, 1975.

Jacques H Drèze. Some Postwar Contributions of French Economists to Theory and Public Policy: With Special Emphasis on Problems of Resource Allocation. *American Economic Review*, 54(4):2–64, 1964.

Jules Dupuit. De la mesure de l'utilité des travaux publics. 1844.

EIA. Electric power monthly with data for march 2019. Technical report, US Energy Information Association, 2019.

Ahmad Faruqui and Sanem Sergici. Household response to dynamic pricing of electricity: a survey of 15 experiments. *Journal of Regulatory Economics*, 38(2):193–225, August 2010.

Lawrence H. Goulder and Roberton C. Williams. The Substantial Bias from Ignoring General Equilibrium Effects in Estimating Excess Burden, and a Practical Solution. *Journal of Political Economy*, 111(4):898–927, 2003.

R. Guesnerie. Second-best pricing rules in the Boiteux tradition. Derivation, review and discussion. *Journal of Public Economics*, 13(1):51–80, 1980.

Roger Guesnerie. Production of the Public Sector and Taxation in a Simple Second Best Model. *Journal of Economic Theory*, 10:127–156, 1975.

Arnold Harberger. The Measurement of Waste. *The American Economic Review*, 54(3): 58–76, 1964.

Arnold C Harberger. Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay. *Journal of Economic Literature*, 9(3):785–797, 1971.

William W Hogan. Time-of-use rates and real-time prices. 2014.

Stephen P Holland and Erin T Mansur. The Short-Run Effects of Time-Varying Prices in Competitive Electricity Markets. *The Energy Journal*, 27(4):127–155, 2006.

Koichiro Ito. Do Consumers Respond to Marginal or Average Price ? Evidence from Nonlinear Electricity Pricing. *American Economic Review*, 104(2):537–563, 2014.

Koichiro Ito, Takanori Ida, and Makoto Tanaka. Using Dynamic Electricity Pricing to Address Energy Crises Evidence from Randomized Field Experiments. 2013.

Mark R Jacobsen, Christopher R Knittel, James M Sallee, and Arthur A Van Benthem. The Use of Regression Statistics to Analyze Imperfect Pricing Policies. *Journal of Political Economy*, 2019.

Katrina Jessoe and David Rapson. Knowledge is (Less) Power : Experimental Evidence from Residential Energy Use. *American Economic Review*, (April), 2013.

Paul Joskow. Contributions to the theory of marginal cost pricing. *The Bell Journal of Economics*, 7(1):197–206, 1976.

Paul Joskow. Creating a Smarter U.S. Electricity Grid. *Journal of Economic Perspectives*, 26(1):29–48, 2012.

Paul Joskow and Jean Tirole. Retail Electricity Competition. *The RAND Journal of Economics*, pages 1–25, 2005.

John Kay and Michael Keen. Measuring the inefficiences of tax systems. *Journal of Public Economics*, 35(3):265–287, 1988.

John A Kay. The deadweight loss from a tax system. *Journal of Public Economics*, 13(1): 111–119, 1980.

Serge-christophe Kolm. La vérité des prix dans un monde imparfait. *Revue économique*, 20 (4):727–740, 1969.

Thomas-Olivier Léautier. Is Mandating "Smart Meters" Smart ? *The Energy Journal*, 35 (4):135–158, 2014.

R. G. Lipsey and Kelvin Lancaster. The General Theory of Second Best. *The Review of Economic Studies*, 24(1):11–32, 1956.

Shaun D Mcrae and Frank a Wolak. Retail Pricing in Colombia to Support the Efficient Deployment of Distributed Generation and Electric Vehicles. working paper, 2019.

James Edward Meade. *The theory of customs unions*, volume 1. North-Holland Publishing Company, 1955.

David Newbery. Missing money and missing markets: Reliability, capacity auctions and interconnectors. *Energy Policy*, 94:401–410, 2016.

David Newbery and Joseph Stiglitz. The Choice of Techniques and the Optimality of Market Equilibrium with Rational Expectations. *Journal of Political Economy*, 90(2):223–246, 1982.

Roy Radner. Competitive Equilibrium Under Uncertainty. *Econometrica*, 36(1):31–58, 1968.

Roy Radner. Existence of Equilibrium of Plans, Prices, and Price Expectations in a Sequence of Markets. *Econometrica*, 40(2):289–303, 1972.

F.P. Ramsey. A contribution to the theory of taxation. *The Economic Journal*, 37(145): 47–61, 1927.

RTE. Bilan électrique 2018. Technical report, RTE, 2019.

William F Sharpe. *Investors and markets: portfolio choices, asset prices, and investment advice*. Princeton university press, 2011.

Peter O Steiner. Peak Loads and Efficient Pricing. *The Quarterly Journal of Economics*, 71 (4):585–610, 1957.

Ralph Turvey. Peak-Load Pricing. *Journal of Political Economy*, 76(1):101–113, 1968.

William Vickrey. Some Objections to Marginal-Cost Pricing. *Journal of Political Economy*, 56(3):218–238, 1948.

Jacob Viner. *The Customs Union Issue.* New York, Carnegie Endowment for International Peace, 1950.

Robert B Wilson. *Nonlinear pricing.* Oxford University Press, 1993.

Frank Wolak. Do Residential Customers Respond to Hourly Prices? Evidence from a Dynamic Pricing Experiment. *American Economic Review*, 101(3):83–87, 2011a.

Frank Wolak. Measuring the benefits of greater spatial granularity in short-term pricing in wholesale electricity markets. *American Economic Review*, 101(3):247–252, 2011b.