Target Tracking for Contextual Bandits:
Application to Demand Side Management

Margaux Brégère [1 2 3]   **Pierre Gaillard** [3]   **Yannig Goude** [1 2]   **Gilles Stoltz** [2]

Abstract

We propose a contextual-bandit approach for demand side management by offering price incentives. More precisely, a target mean consumption is set at each round and the mean consumption is modeled as a complex function of the distribution of prices sent and of some contextual variables such as the temperature, weather, and so on. The performance of our strategies is measured in quadratic losses through a regret criterion. We offer $T^{2/3}$ upper bounds on this regret (up to poly-logarithmic terms)—and even faster rates under stronger assumptions—for strategies inspired by standard strategies for contextual bandits (like LinUCB, see Li et al., 2010). Simulations on a real data set gathered by UK Power Networks, in which price incentives were offered, show that our strategies are effective and may indeed manage demand response by suitably picking the price levels.

1. Introduction

Electricity management is classically performed by anticipating demand and adjusting accordingly production. The development of smart grids, and in particular the installation of smart meters (see Yan et al., 2013; Mallet et al., 2014), come with new opportunities: getting new sources of information, offering new services. For example, demand-side management (also called demand-side response; see Albadi & El-Saadany, 2007; Siano, 2014 for an overview) consists of reducing or increasing consumption of electricity users when needed, typically reducing at peak times and encouraging consumption of off-peak times. This is good to adjust to intermittency of renewable energies and is made possi-

ble by the development of energy storage devices such as batteries or even electric vehicles (see Fischer et al., 2015; Kikusato et al., 2019); the storages at hand can take place at a convenient moment for the electricity provider. We will consider such a demand-side management system, based on price incentives sent to users via their smart meters. We propose here to adapt contextual bandit algorithms to that end, which are already used in online advertising. Other such systems were based on different heuristics (Shareef et al., 2018; Wang et al., 2015).

The structure of our contribution is to first provide a modeling of this management system, in Section 2. It relies on making the mean consumption as close as possible to a moving target by sequentially picking price allocations. The literature discussion of the main ingredient of our algorithms, contextual bandit theory, is postponed till Section 2.4. Then, our main results are stated and discussed in Section 3: we control our cumulative loss through a $T^{2/3}$ regret bound with respect to the best constant price allocation. A refinement as far as convergence rates are concerned is offered in Section 4. A section with simulations based on a real data set concludes the paper: Section 5. For the sake of length, most of the proofs are provided in the supplementary material.

Notation. Without further indications, $\|x\|$ denotes the Euclidean norm of a vector x. For the other norms, there will be a subscript: e.g., the supremum norm of x is is denoted by $\|x\|_\infty$.

2. Setting and Model

Our setting consists of a modeling of electricity consumption and of an aim—tracking a target consumption. Both rely on price levels sent out to the customers.

2.1. Modeling of the Electricity Consumption

We consider a large population of customers of some electricity provider and assume it homogeneous, which is not an uncommon assumption, see Mei et al. (2017). The consumption of each customer at each instance t depends, among others, on some exogenous factors (temperature, wind, season, day of the week, etc.), which will form a context vector

---

[1]EDF R&D, Palaiseau, France [2]Laboratoire de mathématiques d'Orsay, Université Paris-Sud, CNRS, Université Paris-Saclay, Orsay, France [3]INRIA - Département d'Informatique de l'École Normale Supérieure, PSL Research University, Paris, France. Correspondence to: Margaux Brégère <margaux.bregere@edf.fr>.

$x_t \in \mathcal{X}$, where $\mathcal{X}$ is some parametric space. The electricity provider aims to manage demand response: it sets a target mean consumption c_t for each time instance. To achieve it, it changes electricity prices accordingly (by making it more expensive to reduce consumption or less expensive to encourage customers to consume more now rather than in some hours). We assume that $K \geqslant 2$ price levels (tariffs) are available. The individual consumption of a given customer getting tariff $j \in \{1, \ldots, K\}$ is assumed to be of the form $\varphi(x_t, j) +$ white noise, where the white noise models the variability due to the customers, and where φ is some function associating with a context x_t and a tariff j an expected consumption $\varphi(x_t, j)$. Details on and examples of φ are provided below. At instance t, the electricity provider sends tariff j to a share $p_{t,j}$ of the customers; we denote by p_t the convex vector $(p_{t,1}, \ldots, p_{t,K})$. As the population is rather homogeneous, it is unimportant to know to which specific customer a given signal was sent; only the global proportions $p_{t,j}$ matter. The mean consumption observed equals

$$Y_{t,p_t} = \sum_{j=1}^{K} p_{t,j}\, \varphi(x_t, j) + \text{noise}\,.$$

The noise term is to be further discussed below; we first focus on the φ function by means of examples.

Example 1. The simplest approach consists in considering a linear model per price level, i.e., parameters $\theta_1, \ldots, \theta_K \in \mathbb{R}^{\dim(\mathcal{X})}$ with $\varphi(x_t, j) = \theta_j^{\mathrm{T}} x_t$. We denote $\theta = (\theta_j)_{1 \leqslant j \leqslant K}$ the vector formed by aggregating all vectors θ_j.

This approach can be generalized by replacing x_t by a vector-valued function $b(x_t)$. This corresponds to the case where it is assumed that the $\varphi(\cdot, j)$ belong to some set $\mathcal{H}$ of functions $h : \mathcal{X} \to \mathbb{R}$, with a basis composed of $b_1, \ldots, b_q$. Then, $b = (b_1, \ldots, b_q)$. For instance, $\mathcal{H}$ can be given by histograms on a given grid of $\mathcal{X}$. $\quad\square$

Example 2. Generalized additive models (Wood, 2006) form a powerful and efficient semi-parametric approach to model electricity consumption (see, among others, Goude et al., 2014; Gaillard et al., 2016). It models the load as a sum of independent exogenous variable effects. In our simulations, see (13), we will consider a mean expected consumption of the form $\varphi(x_t, j) = \varphi(x_t, 0) + \xi_j$, that is, the tariff will have a linear impact on the mean consumption, independently of the contexts. The baseline mean consumption $\varphi(x_t, 0)$ will be modeled as a sum of simple $\mathbb{R} \to \mathbb{R}$ functions, each taking as input a single component of the context vector:

$$\varphi(x_t, 0) = \sum_{i=1}^{Q} f^{(i)}(x_{t,h(i)})\,,$$

where $Q \geqslant 1$ and where each $h(i) \in \{1, \ldots, \dim(\mathcal{X})\}$. Some components $h(i)$ may be used several times. When the considered component $x_{t,h(i)}$ takes continuous values, these functions $f^{(i)}$ are so-called cubic splines: $\mathcal{C}^2$–smooth functions made up of sections of cubic polynomials joined together at points of a grid (the knots). Choosing the number

q_i of knots (points at which the sections join) and their locations is sufficient to determine (in closed form) a linear basis $(b_1^{(i)}, \ldots, b_{q_i}^{(i)})$ of size q_i, see Wood (2006) for details. The function $f^{(i)}$ can then be represented on this basis by a vector of length q_i, denoted by $\theta^{(i)}$:

$$f^{(i)} = \sum_{j=1}^{q_i} \theta_j^{(i)} b_j^{(i)}\,.$$

When the considered component $x_{t,h(i)}$ takes finitely many values, we write $f^{(i)}$ as a sum of indicator functions:

$$f^{(i)} = \sum_{j=1}^{q_i} \theta_j^{(i)} \mathbf{1}_{\{v_j^{(i)}\}}\,,$$

where the $v_j^{(i)}$ are the q_i modalities for the component $h(i)$. All in all, $\varphi(x_t, j)$ can be represented by a vector of dimension $K + q_1 + \ldots + q_Q$ obtained by aggregating the ξ_j and the vectors $\theta^{(i)}$ into a single vector. $\quad\square$

Both examples above show that it is reasonable to assume that there exists some unknown $\theta \in \mathbb{R}^d$ and some known transfer function ϕ such that $\varphi(x_t, j) = \phi(x_t, j)^{\mathrm{T}}\theta$. By linearly extending ϕ in its second component, we get

$$Y_{t,p_t} = \phi(x_t, p_t)^{\mathrm{T}}\theta + \text{noise}\,.$$

We will actually not use in the sequel that $\phi(x, p)$ is linear in p: the dependency of $\phi(x, p)$ in p could be arbitrary.

We now move on to the noise term. We first recall that we assumed that our population is rather homogeneous, which is a natural feature as soon as it is large enough. Therefore, we may assume that the variabilities within the group of customers getting the same tariff j can be combined into a single random variable $\varepsilon_{t,j}$. We denote by ε_t the vector $(\varepsilon_{t,1}, \ldots, \varepsilon_{t,K})$. All in all, we will mainly consider the following model.

Model 1: tariff-dependent noise. When the electricity provider picks the convex vector p, the mean consumption obtained at time instance t equals

$$Y_{t,p} = \phi(x_t, p)^{\mathrm{T}}\theta + p^{\mathrm{T}}\varepsilon_t\,.$$

The noise vectors $\varepsilon_1, \varepsilon_2, \ldots$ are ρ–sub-Gaussian[1] i.i.d. random variables with $\mathbb{E}[\varepsilon_1] = (0, \ldots, 0)^{\mathrm{T}}$. We denote by $\Gamma = \mathrm{Var}(\varepsilon_1)$ their covariance matrix. No assumption is made on Γ in the model above (real data confirms that Γ typically has no special form, see 5.1). However, when it is proportional to the $K \times K$ matrix $[1]$, the noises associated with each group can be combined into a global noise, leading to the following model. It is less realistic in practice, but we discuss it because regret bounds may be improved in the presence of a global noise.

Model 2: global noise. When the electricity provider picks the convex vector p, the mean consumption obtained at time instance t equals

$$Y_{t,p} = \phi(x_t, p)^{\mathrm{T}}\theta + e_t\,.$$

---

[1] A d–dimensional random vector ε is ρ–sub-Gaussian, where $\rho > 0$, if for all $\nu \in \mathbb{R}^d$, one has $\mathbb{E}[e^{\nu^{\mathrm{T}}\varepsilon}] \leqslant e^{\rho^2 \|\nu\|^2/2}$.

The scalar noises $e_1, e_2, \ldots$ are ρ–sub-Gaussian i.i.d. random variables, with $\mathbb{E}[e_1] = 0$. We denote by $\sigma^2 = \mathrm{Var}(e_1)$ the variance of the random noises e_t.

2.2. Tracking a Target Consumption

We now move on to the aim of the electricity provider. At each time instance t, it picks an allocation of price levels p_t and wants the observed mean consumption Y_{t,p_t} to be as close as possible to some target mean consumption c_t. This target is set in advance by another branch of the provider and p_t is to be picked based on this target: our algorithms will explain how to pick p_t given c_t but will not discuss the choice of the latter. In this article we will measure the discrepancy between the observed Y_{t,p_t} and the target c_t via a quadratic loss: $(Y_{t,p_t} - c_t)^2$. We may set some restrictions on the convex combinations p that can be picked: we denote by $\mathcal{P}$ the set of legible allocations of price levels. This models some operational or marketing constraints that the electricity provider may encounter. We will see that whether $\mathcal{P}$ is a strict subset of all convex vectors or whether it is given by the set of all convex vectors plays no role in our theoretical analysis.

As explained in Section 3.1, we will follow a standard path in online learning theory: to minimize the cumulative loss suffered we will minimize some regret.

2.3. Summary: Online Protocol

After picking an allocation of price levels p_t, the electricity provider only observes Y_{t,p_t}: it thus faces a bandit monitoring. Because of the contexts x_t, the problem considered falls under the umbrella of contextual bandits. No stochastic assumptions are made on the sequences x_t and c_t: the contexts x_t and c_t will be considered as picked by the environment. Finally, mean consumptions are assumed to be bounded between 0 and C, where C is some known maximal value. The online protocol described in Sections 2.1 and 2.2 is stated in Protocol 1. We see that the choices x_t, c_t and p_t need to be $\mathcal{F}_{t-1}$–measurable, where

$$\mathcal{F}_{t-1} \stackrel{\text{def}}{=} \sigma(\varepsilon_1, \ldots, \varepsilon_{t-1}).$$

2.4. Literature Discussion: Contextual Bandits

In many bandit problems the learner has access to additional information at the beginning of each round. Several settings for this side information may be considered. The adversarial case was introduced in Auer et al. (2002, Section 7, algorithm Exp4): and subsequent improvements were suggested in Beygelzimer et al. (2011) and McMahan & Streeter (2009). The case of i.i.d. contexts with rewards depending on contexts through an unknown parametric model was introduced by Wang et al. (2005b) and generalized to the non-i.i.d. setting in Wang et al. (2005a), then to the multi-

---

Protocol 1 Target Tracking for Contextual Bandits

Input
    Parametric context set $\mathcal{X}$
    Set of legible convex weights $\mathcal{P}$
    Bound on mean consumptions C
    Transfer function $\phi : \mathcal{X} \times \mathcal{P} \to \mathbb{R}^d$

Unknown parameters
    Transfer parameter $\theta \in \mathbb{R}^d$
    Covariance matrix Γ of size $K \times K$ (Model 1)
    Variance σ^2 (Model 2)

for $t = 1, 2, \ldots$ **do**
    Observe a context $x_t \in \mathcal{X}$ and a target $c_t \in (0, C)$
    Choose an allocation of price levels $p_t \in \mathcal{P}$
    Observe a resulting mean consumption

$$Y_{t,p_t} = \phi(x_t, p_t)^{\mathrm{T}}\theta + p_t^{\mathrm{T}}\varepsilon_t \qquad \text{(Model 1)}$$
$$Y_{t,p_t} = \phi(x_t, p_t)^{\mathrm{T}}\theta + e_t \qquad \text{(Model 2)}$$

    Suffer a loss $(Y_{t,p_t} - c_t)^2$
end for

Aim
    Minimize the cumulative loss $L_T = \sum_{t=1}^{T}(Y_{t,p_t} - c_t)^2$

---

variate and nonparametric case in Perchet & Rigollet (2013). Hybrid versions (adversarial contexts but stochastic dependencies of the rewards on the contexts, usually in a linear fashion) are the most popular ones. They were introduced by Abe & Long (1999) and further studied in Auer (2002). A key technical ingredient to deal with them is confidence ellipsoids on the linear parameter; see Dani et al. (2008), Rusmevichientong & Tsitsiklis (2010) and Abbasi-Yadkori et al. (2011). The celebrated UCB algorithm of Lai & Robbins (1985) was generalized in this hybrid setting as the LinUCB algorithm, by Li et al. (2010) and Chu et al. (2011). Later, Filippi et al. (2010) extended it to a setting with generalized additive models and Valko et al. (2013) proposed a kernelized version of UCB. Other approaches, not relying on confidence ellipsoids, consider sampling strategies (see Gopalan et al., 2014) and are currently extended to bandit problems with complicated dependency in contextual variables (Mannor, 2018). Our model falls under the umbrella of hybrid versions considering stochastic linear bandit problems given a context. The main difference of our setting lies in how we measure performance: not directly with the rewards or their analogous quantities Y_{t,p_t} in our setting, but through how far away they are from the targets c_t.

3. Main Result, with Model 1

This section considers Model 1. We take inspiration from LinUCB (Li et al., 2010; Chu et al., 2011): given the form of the observed mean consumption, the key is to estimate the parameter θ. Denoting by I_d the $d \times d$ identity matrix

and picking $\lambda > 0$, we classically do so according to

$$\widehat{\theta}_t \overset{\text{def}}{=} V_t^{-1} \sum_{s=1}^{t} Y_{s,p_s} \phi(x_s, p_s) \qquad (1)$$

where $V_t \overset{\text{def}}{=} \lambda \mathrm{I}_d + \sum_{s=1}^{t} \phi(x_s, p_s)\phi(x_s, p_s)^{\mathrm{T}}$.

A straightforward adaptation of earlier results (see Theorem 2 of Abbasi-Yadkori et al., 2011 or Theorem 20.2 in the monograph by Lattimore & Szepesvári, 2018) yields the following deviation inequality; details are provided in the supplementary material (Appendix A).

Lemma 1. *No matter how the provider picks the p_t, we have, for all $t \geqslant 1$ and all $\delta \in (0, 1)$,*

$$\sqrt{(\widehat{\theta}_t - \theta)^{\mathrm{T}} V_t (\widehat{\theta}_t - \theta)} \overset{\text{def}}{=} \left\| V_t^{1/2}(\widehat{\theta}_t - \theta) \right\|$$
$$\leqslant \sqrt{\lambda}\|\theta\| + \rho\sqrt{2\ln\frac{1}{\delta} + d\ln\frac{1}{\lambda} + \ln\det(V_t)},$$

with probability at least $1 - \delta$.

Actually, the result above could be improved into an anytime result ("with probability $1 - \delta$, for all $t \geqslant 1$, ...") with no effort, by applying a stopping argument (or, alternatively, Doob's inequality for super-martingales), as Abbasi-Yadkori et al. (2011) did. This would slightly improve the regret bounds below by logarithmic factors.

3.1. Regret as a Proxy for Minimizing Losses

We are interested in the cumulative sum of the losses, but under suitable assumptions (e.g., bounded noise) the latter is close to the sum of the conditionally expected losses (e.g., through the Hoeffding–Azuma inequality). Typical statements are of the form: for all strategies of the provider and of the environment, with probability at least $1 - \delta$,
$$L_T = \sum_{t=1}^{T}(Y_{t,p_t} - c_t)^2$$
$$\leqslant \sum_{t=1}^{T} \mathbb{E}\big[(Y_{t,p_t} - c_t)^2 \,\big|\, \mathcal{F}_{t-1}\big] + O\big(\sqrt{T\ln(1/\delta)}\big).$$

All regret bounds in the sequel will involve the sum of conditionally expected losses $\overline{L}_T$ above but up to adding a deviation term to all these regret bounds, we get from them a bound on the true cumulative loss L_T. Now, the choices x_t, c_t and p_t are $\mathcal{F}_{t-1}$–measurable, where $\mathcal{F}_{t-1} = \sigma(\varepsilon_1, \ldots, \varepsilon_{t-1})$. Therefore, under Model 1,

$$\mathbb{E}\big[(Y_{t,p_t} - c_t)^2 \,\big|\, \mathcal{F}_{t-1}\big]$$
$$= \mathbb{E}\Big[\big(\phi(x_t, p_t)^{\mathrm{T}}\theta + p_t^{\mathrm{T}}\varepsilon_t - c_t\big)^2 \,\Big|\, \mathcal{F}_{t-1}\Big]$$
$$= \big(\phi(x_t, p_t)^{\mathrm{T}}\theta - c_t\big)^2 + \mathbb{E}\big[(p_t^{\mathrm{T}}\varepsilon_t)^2 \,\big|\, \mathcal{F}_{t-1}\big]$$
$$\qquad + \mathbb{E}\Big[2\big(\phi(x_t, p_t)^{\mathrm{T}}\theta - c_t\big)p_t^{\mathrm{T}}\varepsilon_t \,\Big|\, \mathcal{F}_{t-1}\Big]$$
$$= \big(\phi(x_t, p_t)^{\mathrm{T}}\theta - c_t\big)^2 + p_t^{\mathrm{T}}\Gamma p_t, \qquad (2)$$

that is, after summing,

$$\overline{L}_T = \sum_{t=1}^{T}\big(\phi(x_t, p_t)^{\mathrm{T}}\theta - c_t\big)^2 + p_t^{\mathrm{T}}\Gamma p_t.$$

We therefore introduce the (conditional) regret

$$\overline{R}_T = \sum_{t=1}^{T}\big(\phi(x_t, p_t)^{\mathrm{T}}\theta - c_t\big)^2 + p_t^{\mathrm{T}}\Gamma p_t$$
$$- \sum_{t=1}^{T}\min_{p \in \mathcal{P}}\Big\{\big(\phi(x_t, p)^{\mathrm{T}}\theta - c_t\big)^2 + p^{\mathrm{T}}\Gamma p\Big\}.$$

This will be the quantity of interest in the sequel.

3.2. Optimistic Algorithm: All but the Estimation of Γ

We assume that in the first n rounds an estimator $\widehat{\Gamma}_n$ of the covariance matrix Γ was obtained; details are provided in the next subsection. We explain here how the algorithm plays for rounds $t \geqslant n + 1$. We assumed that the transfer function ϕ and the bound $C > 0$ on the target mean consumptions were known. We use the notation $[x]_C = \min\{\max\{x, 0\}, C\}$ for the clipped part of a real number x (clipping between 0 and C). We then estimate the instantaneous losses (2)

$$\ell_{t,p} \overset{\text{def}}{=} \mathbb{E}\big[(Y_{t,p} - c_t)^2 \,\big|\, \mathcal{F}_{t-1}\big] = \big(\phi(x_t, p)^{\mathrm{T}}\theta - c_t\big)^2 + p^{\mathrm{T}}\Gamma p$$

associated with each choice $p \in \mathcal{P}$ by:

$$\widehat{\ell}_{t,p} = \Big(\big[\phi(x_t, p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C - c_t\Big)^2 + p^{\mathrm{T}}\widehat{\Gamma}_n p.$$

We also denote by $\alpha_{t,p}$ deviation bounds, to be set by the analysis. The optimistic algorithm picks, for $t \geqslant n + 1$:

$$p_t \in \arg\min_{p \in \mathcal{P}}\big\{\widehat{\ell}_{t,p} - \alpha_{t,p}\big\}. \qquad (3)$$

Comment: In linear contextual bandits, rewards are linear in θ and to maximize global gain, LinUCB (Li et al., 2010) picks a vector p which maximizes a sum of the form $\phi(x_t, p)^{\mathrm{T}}\widehat{\theta}_{t-1} + \tilde{\alpha}_{t,p}$. Here, as we want to track the target, we slightly change this expression by substituting the target c_t and taking a quadratic loss. But the spirit is similar.

3.3. Optimistic Algorithm: Estimation of Γ

The estimation of the covariance matrix Γ is hard to perform (on the fly and simultaneously) as the algorithm is running. We leave this problem for future research and devote here the first n rounds to this estimation. We created from scratch the estimation of Γ proposed below and studied in Lemma 2, as we could find no suitable result in the literature.

For each pair

$$(i, j) \in E \overset{\text{def}}{=} \big\{(i, j) \in \{1, \ldots, K\}^2 : 1 \leqslant i \leqslant j \leqslant K\big\}$$

we define the weight vector $p^{(i,j)}$ as: for $k \in \{1, \ldots, K\}$,

$$p_k^{(i,j)} = \begin{cases} 1 & \text{if } k = i = j, \\ 1/2 & \text{if } k \in \{i, j\} \text{ and } i \neq j, \\ 0 & \text{if } k \notin \{i, j\}. \end{cases}$$

These correspond to all weights vectors that either assign all the mass to a single component, like the $p^{(i,i)}$, or share

the mass equally between two components, like the $p^{(i,j)}$ for $i \neq j$. There are $K(K+1)/2$ different weight vectors considered. We order these weight vectors, e.g., in lexicographic order, and use them one after the other, in order. This implies that in the initial exploration phase of length n, each vector indexed by E is selected at least

$$n_0 \stackrel{\text{def}}{=} \left\lfloor \frac{2n}{K(K+1)} \right\rfloor$$

times. At the end of the exploration period, we define $\widehat{\theta}_n$ as in (1) and the estimator

$$\widehat{\Gamma}_n \in \underset{\widehat{\Gamma} \in \mathcal{M}_K(\mathbb{R})}{\arg\min} \sum_{t=1}^{n} \left(\widehat{Z}_t^2 - p_t^{\mathrm{T}} \widehat{\Gamma} p_t \right)^2, \quad (4)$$

where $\widehat{Z}_t \stackrel{\text{def}}{=} Y_{t,p_t} - \left[\phi(x_t, p_t)^{\mathrm{T}} \widehat{\theta}_n \right]_C$. Note that $\widehat{\Gamma}_n$ can be computed efficiently by solving a linear system as soon as K is small enough.

3.4. Statement of our Main Result

Theorem 1. *Fix a risk level $\delta \in (0, 1)$ and a time horizon $T \geqslant 1$. Assume that the boundedness assumptions (5) hold. The optimistic algorithm (3) with an initial exploration of length $n = O(T^{2/3})$ rounds satisfies*

$$\overline{R}_T = O\left(T^{2/3} \ln^2 \left(\tfrac{T}{\delta} \right) \sqrt{\ln \tfrac{1}{\delta}} \right)$$

with probability at least $1 - \delta$.

When the covariance matrix Γ is known, no initial exploration is required and the regret bound improves to $O(\sqrt{T} \ln T)$ as far as the orders of magnitude in T are concerned. These improved rates might be achievable even if Γ is unknown, through a more efficient, simultaneous, estimation of Γ and θ (an issue we leave for future research, as already mentioned at the beginning of Section 3.3).

3.5. Analysis: Structure

Assumption 1: boundedness assumptions. They are all linked to the knowledge that the mean consumption lies in $(0, C)$ and indicate some normalization of the modeling:

$$\|\phi\|_\infty \leqslant 1, \qquad \|\theta\|_\infty \leqslant C, \qquad \phi^{\mathrm{T}}\theta \in [0, C]. \quad (5)$$

As a consequence, $\|\theta\| \leqslant \sqrt{d}\, C$ and all eigenvalues of V_t lie in $[\lambda, \lambda + t]$, thus $\ln\big(\det(V_t)\big) \in \big[d \ln \lambda, d \ln(\lambda + t)\big]$.

The deviation bound of Lemma 1 plays a key role in the algorithm. We introduce the following upper bound on it:

$$B_t(\delta) \stackrel{\text{def}}{=} \sqrt{\lambda d}\, C + \rho \sqrt{2 \ln \tfrac{1}{\delta} + d \ln \left(1 + \tfrac{t}{\lambda} \right)}. \quad (6)$$

Finally, we also assume that a bound G is known, such that

$$\forall p \in \mathcal{P}, \qquad p^{\mathrm{T}}\Gamma p \leqslant G.$$

A last consequence of all these boundedness assumptions is that $L \stackrel{\text{def}}{=} C^2 + G$ upper bounds the (conditionally) expected losses $\ell_{t,p} = \big(\phi(x_t, p)^{\mathrm{T}}\theta - c_t\big)^2 + p^{\mathrm{T}}\Gamma p$.

Structure of the analysis. The analysis exploits how well each $\widehat{\theta}_t$ estimates θ and how well $\widehat{\Gamma}_n$ estimates Γ. The regret bound, as is clear from Proposition 1 below, also consists of these two parts. The proof is to be found in the supplementary material (Appendix B).

Proposition 1. *Fix a risk level $\delta \in (0, 1)$ and an exploration budget $n \geqslant 2$. Assume that the boundedness assumptions (5) hold. Consider an estimator $\widehat{\Gamma}_n$ of Γ such that $\sup_{p \in \mathcal{P}} \big| p^{\mathrm{T}}(\Gamma - \widehat{\Gamma}_n)p \big| \leqslant \gamma$ with probability at least $1 - \delta/2$, for some $\gamma > 0$.*

Then choosing $\lambda > 0$ and

$$a_{t,p} = \min\left\{ L,\ 2C\, B_{t-1}(\delta t^{-2}) \left\| V_{t-1}^{-1/2} \phi(x_t, p) \right\| \right\},$$
$$\alpha_{t,p} = \gamma + a_{t,p}, \quad (7)$$

the optimistic algorithm (3) ensures that w.p. $1 - \delta$,

$$\sum_{t=n+1}^{T} \ell_{t,p_t} - \sum_{t=n+1}^{T} \min_{p \in \mathcal{P}} \ell_{t,p} \leqslant 2 \sum_{t=n+1}^{T} \alpha_{t,p_t}.$$

Comment: Li et al. (2010) pick $\alpha(t, p)$ proportional to $\left\| V_{t-1}^{-1/2} \phi(x_t, p) \right\|$ only, but we need an additional term to account for the covariance matrix.

We are thus left with studying how well $\widehat{\Gamma}_n$ estimates Γ and with controlling the sum of the $a_{t,p}$. The next two lemmas take care of these issues. Their proofs are to be found in the supplementary material (Appendices C and D).

Lemma 2. *For all $\delta \in (0, 1)$, the estimator (4) satisfies: with probability at least $1 - \delta$,*

$$\sup_{p \in \mathcal{P}} \left| p^{\mathrm{T}}\big(\widehat{\Gamma}_n - \Gamma\big)p \right| \leqslant (K + 8)\kappa_n \sqrt{n}/n_0 = O(\kappa_n/\sqrt{n})$$
$$= O\left(\frac{1}{\sqrt{n}} \ln^2(n/\delta) \sqrt{\ln(1/\delta)} \right),$$

where we recall that $n_0 = \lfloor 2n/(K(K+1)) \rfloor$ and where $\kappa_n = \big(C + 2M_n\big) B_n(\delta/3) + M'_n$ with $M_n = \rho/2 + \ln(6n/\delta)$ and $M'_n = M_n^2 \sqrt{2\ln(3K^2/\delta)} + 2\sqrt{\exp(2\rho)\delta/6}$.

Comment: We derived the estimator of Γ as well as Lemma 2 from scratch: we could find no suitable result in the literature for estimating Γ in our context.

Lemma 3. *No matter how the environment and provider pick the x_t and p_t,*

$$\sum_{t=n+1}^{T} a_{t,p_t} \leqslant \sqrt{\left(2C\overline{B}\right)^2 + \tfrac{L^2}{2}} \sqrt{dT \ln \tfrac{\lambda+T}{\lambda}}$$
$$= O\left(\sqrt{T \ln T \ln(T/\delta)} \right),$$

where $\overline{B} \stackrel{\text{def}}{=} \sqrt{d\lambda}\, C + \rho \sqrt{2\ln(T^2/\delta) + d\ln(1 + T/\lambda)}$.

Comment: This lemma follows from a straightforward adaptation/generalization of Lemma 19.1 of the monograph by Lattimore & Szepesvári (2018); see also a similar result in Lemma 3 by Chu et al. (2011).

We are now ready to conclude the proof of Theorem 1. Using for the first $n \geqslant 2$ rounds that $L = C^2 + G$ upper bounds the (conditionally) expected losses ℓ_{t,p_t}, Proposition 1 and Lemmas 2 and 3 show that, w.p. $1 - \delta$

$$\overline{R}_T \leqslant nL + T\gamma + \sum_{t=n+1}^{T} a_{t,p_t}$$

$$\leqslant nL + O\Big(T\ln^2\big(\tfrac{n}{\delta}\big)\sqrt{\tfrac{\ln(1/\delta)}{n}} + \sqrt{T\ln T\ln(T/\delta)}\Big).$$

Picking n of order $T^{2/3}$ concludes the proof.

Case of known covariance matrix Γ: We then have $\gamma = 0$ in Proposition 1 and we may discard Lemma 2. Taking $n = 2$, the obtained regret bound is $2L + \sqrt{T\ln T\ln(T/\delta)}$.

Comment: The algorithm of Theorem 1 depends on δ via the tuning (7) of α. But we can also define a regret with full expectations $\mathbb{E}\big[\ell_{t,p_t}\big]$ and $\min\mathbb{E}\big[\ell_{t,p}\big]$—remember from Sections 3.1 and 3.2 that the losses $\ell_{t,p}$ are conditional expectations. In that case the algorithm can be made independent of δ. Only Step 3 of the proof of Proposition 1 is to be modified. The same rates in T are obtained.

4. Fast Rates, with Model 2

In this section, we consider Model 2 and show that under an attainability condition stated below, the order of magnitude of the regret bound in Theorem 1 can be reduced to a poly-logarithmic rate. This kind of fast rates already exist in the literature of linear contextual bandits (see, e.g., Abbasi-Yadkori et al., 2011, as well as Dani et al., 2008) but are not so frequent. We underline in the proof the key step where we gain orders of magnitude in the regret bound. Before doing so, we note that similarly to Section 3.1,

$$\mathbb{E}\big[(Y_{t,p} - c_t)^2\big] = \big(\phi(x_t,p)^{\mathrm{T}}\theta - c_t\big)^2 + \sigma^2, \quad (8)$$

which leads us to introduce a regret $\overline{R}_T$ defined by $\overline{R}_T =$

$$\sum_{t=1}^{T}\big(\phi(x_t,p_t)^{\mathrm{T}}\theta - c_t\big)^2 - \sum_{t=1}^{T}\min_{p\in\mathcal{P}}\big\{\big(\phi(x_t,p)^{\mathrm{T}}\theta - c_t\big)^2\big\}.$$

Thus, as far as the minimization of the regret is concerned, Model 2 is a special case of Model 1, corresponding to a matrix Γ that can be taken as the null matrix $[0]$. Of course, as explained in Section 2.1, the covariance matrix Γ of Model 2 is $\sigma^2[1]$ in terms of real modeling, but in terms of regret-minimization it can be taken as $\Gamma = [0]$. Therefore, all results established above for Model 1 extend to Model 2, but under an additional assumption stated below, the $T^{2/3}$ rates (up to poly-logarithmic terms) obtained above can be reduced to poly-logarithmic rates only.

Assumption 2: attainability. For each time instance $t \geqslant 1$, the expected mean consumption is attainable, i.e.,

$$\exists p \in \mathcal{P}: \quad \phi(x_t,p)^{\mathrm{T}}\theta = c_t. \quad (9)$$

We denote by $p_t^\star$ such an element of $\mathcal{P}$. In Model 2 and under this assumption, the expected losses $\ell_{t,p}$ defined in (8)

are such that, for all $t \geqslant 1$ and all $x_t \in \mathcal{X}$,

$$\min_{p\in\mathcal{P}}\ell_{t,p} = \ell_{t,p_t^\star} = \sigma^2. \quad (10)$$

As in Model 2 the variance terms σ^2 cancel out when considering the regret, the variance σ^2 does not need to be estimated. Our optimistic algorithm thus takes a simpler form. For each $t \geqslant 2$ and $p \in \mathcal{P}$ we consider the same estimators (1) of θ as before and then define

$$\widetilde{\ell}_{t,p} = \big(\phi(x_t,p)^{\mathrm{T}}\widehat{\theta}_{t-1} - c_t\big)^2$$

(no clipping needs to be considered in this case). We set

$$\beta_{t,p} = B_{t-1}(\delta t^{-2})^2 \big\|V_{t-1}^{-1/2}\phi(x_t,p)\big\|^2 \quad (11)$$

and then pick: $\qquad p_t \in \arg\min_{p\in\mathcal{P}}\big\{\widetilde{\ell}_{t,p} - \beta_{t,p}\big\} \quad (12)$

for $t \geqslant 2$ and p_1 arbitrarily. The tuning parameter $\lambda > 0$ is hidden in $B_{t-1}(\delta t^{-2})^2$. We get the following theorem, whose proof is deferred to Appendix E and re-uses many parts of the proofs of Proposition 1 and Lemma 3. Without the attainability assumption (9), a regret bound of order $\sqrt{T}$ up to logarithmic terms could still be proved.

Theorem 2. *In Model 2, assume that the boundedness and attainability assumptions (5) and (9) hold. Then, the optimistic algorithm (12), tuned with $\lambda > 0$, ensures that for all $\delta \in (0,1)$,*

$$\overline{R}_T \leqslant d\big(4\overline{B}^2 + \tfrac{C^2}{2}\big)\ln\tfrac{\lambda+T}{\lambda} = O\big(\ln^2(T)\big),$$

w.p. at least $1 - \delta$, where $\overline{B}$ is defined as in Lemma 3.

5. Simulations

Our simulations rely on a real data set of residential electricity consumption, in which different tariffs were sent to the customers according to some policy. But of course, we cannot test an alternative policy on historical data (we only observed the outcome of the tariffs sent) and therefore need to build first a data simulator.

5.1. The Underlying Real Data Set / The Simulator

We consider open data published[2] by UK Power Networks and containing energy consumption (in kWh per half hour) at half hourly intervals of a thousand customers subjected to dynamic energy prices. A single tariff (among High–1, Normal–2 or Low–3) was offered to all customers for each half hour and was announced in advance. The report by Schofield et al. (2014) provides a full description of this experimentation and an exhaustive analysis of results. We only kept customers with more than 95% of data available (980 clients) and considered their mean[3] consumption. As far as

---

[2]*SmartMeter Energy Consumption Data in London Households* – see https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households

[3]Only such a level of aggregation allows a proper estimation (individual consumptions are erratic); Sevlian & Rajagopal, 2018.

contexts are concerned, we considered half-hourly temperatures τ_t in London, obtained from https://www.noaa.gov/ (missing data managed by linear interpolation). We also created calendar variables: the day of the week w_t (equal to 1 for Monday, 2 for Tuesday, etc.), the half-hour of the day $h_t \in \{1, \dots, 48\}$, and the position in the year: $y_t \in [0, 1]$, linear values between $y_t = 0$ on January 1st at 00:00 and $y_t = 1$ on December the 31st at 23:59.

Realistic simulator. It is based on the following additive model, which breaks down time by half hours:

$$\varphi(x_t, j) = \sum_{h=1}^{48} \left[f_h^\tau(\tau_t) + f_h^y(y_t) + \eta_h \right] \mathbf{1}_{\{h_t=h\}} + \sum_{w=1}^{7} \zeta_w \mathbf{1}_{\{w_t=w\}} + \xi_j , \quad (13)$$

where the f_h^τ and f_h^y are functions catching the effect of the temperature and of the yearly seasonality. As explained in Example 2, the transfer parameter θ gathers coordinates of the f_h^τ and the f_h^y in bases of splines, as well as the coefficients η_h, ζ_w and ξ_j. Here, we work under the assumption that exogenous factors do not impact customers' reaction to tariff changes (which is admittedly a first step, and more complex models could be considered). Our algorithms will have to sequentially estimate the parameter θ, but we also need to set it to get our simulator in the first place. We do so by exploiting historical data together with the allocations of prices picked, of the form $(0, 1, 0)$, $(1, 0, 0)$ and $(0, 0, 1)$ only on these data (all customers were getting the same tariff), and apply the formula (13) through the R–package mgcv (which replaces the λ identity matrix with a slightly more complex definite positive matrix S, see Wood, 2006). The deterministic part of the obtained model is realistic enough: its adjusted R-square on historical observations equals 92% while its mean absolute percentage of error equals 8.82%. Now, as far as noise is concerned, we take multivariate Gaussian noise vectors ε_t, where the covariance matrix Γ was built again based on realistic values. The diagonal coefficients $\Gamma_{j,j}$ are given by the empirical variance of the residuals associated with tariff j, while non-diagonal coefficients $\Gamma_{j,j'}$ are given by the empirical covariance between residuals of tariffs j and j' at times t and $t \pm 48$; this matrix Γ has no special form, see Appendix F for more details and its numerical expression.

5.2. Experiment Design: Learning Added Effects ξ_j

Target creation. We focus on attainable targets c_t, namely, $\varphi(x_t, 1) \leqslant c_t \leqslant \varphi(x_t, 3)$. To smooth consumption, we pick c_t near $\varphi(x_t, 3)$ during the night and near $\varphi(x_t, 1)$ in the evening. These hypotheses can be seen as an ideal configuration where targets and customers portfolio are in a way compatible.

$\mathcal{P}$ **restriction.** We assume that the electricity provider cannot send Low and High tariffs at the same round and that population can be split in $N = 100$ equal subsets. Thus, $\mathcal{P}$ is restricted to the grid

$$\left\{ \left(\tfrac{i}{N}, 1 - \tfrac{i}{N}, 0 \right), \; \left(0, \tfrac{i}{N}, 1 - \tfrac{i}{N} \right), \quad i \in \{0, \dots, N\} \right\}$$

Training period, testing period. We create one year of data using historical contexts and assume that only Normal tariffs are picked at first: $p_t = (0, 1, 0)$; this is a training period, which corresponds to what electricity providers are currently doing. As they can accurately estimate the covariance matrix Γ by ad-hoc methods, we assume that the algorithm knows the matrix Γ used by the simulator. Then the provider starts exploring the effects of tariffs for an additional month (a January month, based on the historical contexts) and freely picks the p_t according to our algorithm; this is the testing period. The estimation of θ in this testing period is still performed via the formula (13) and as indicated above (with the mgcv package), including the year when only $p_t = (0, 1, 0)$ allocations were picked. For learning to then focus on the parameters ξ_j, as other parameters were decently estimated in the training period, we modify the exploration term $\alpha_{t,p}$ of (3) into

$$\alpha_{t,p} = 2CB_{t-1}(\delta t^{-2}) \| \tilde{V}_{t-1}^{-1/2} p_t \| ,$$

with $\tilde{V}_{t-1} = \lambda I_d + \sum_{s=1}^{t-1} p_s p_s^{\mathrm{T}}$. We pick a convenient λ.

5.3. Results

Algorithms were run 200 times each. The simplest set of results is provided in Figure 3: the regrets suffered on each run are compared to the theoretical orders of magnitude of the regret bounds. As expected, we observe a lower regrets for Model 2. The bottom parts of Figures 1–2 indicate, for a single run, which allocation vectors p_t were picked over time. During the first day of the testing period, the algorithms explore[4] the effect of tariffs by sending the same tariff to all customers (the p_t vectors are Dirac masses) while at the end of the testing period, they cleverly exploit the possibility to split the population in two groups of tariffs. We obtain an approximation of the expected mean consumption $\varphi(x_t, p_t)$ by averaging the 200 observed consumptions, and this is the main (black, solid) line to look at in the top parts of Figures 1–2. Four plots are depicted depending on the day of the testing period (first, last) and of the model considered. These (approximated) expected mean consumptions may be compared to the targets set (dashed red line). The algorithms seem to perform better on the last day of the testing period for Model 2 than for Model 1 as the expected mean consumption seems closer to the target. However, in Model 1, the algorithm has to pick tariffs leading to the best bias-variance trade-off (the expected loss features a variance term). This is why the average consumption does not overlap the target as in Model 2. This results in a slightly biased estimator of the mean consumption in Model 1.

---

[4]Note that, over the first iterations, the exploration term for Model 2 is much larger than the exploitation term (but quickly vanishes), which leads to an initial quasi-deterministic exploration and an erratic consumption (unlike in Model 1).
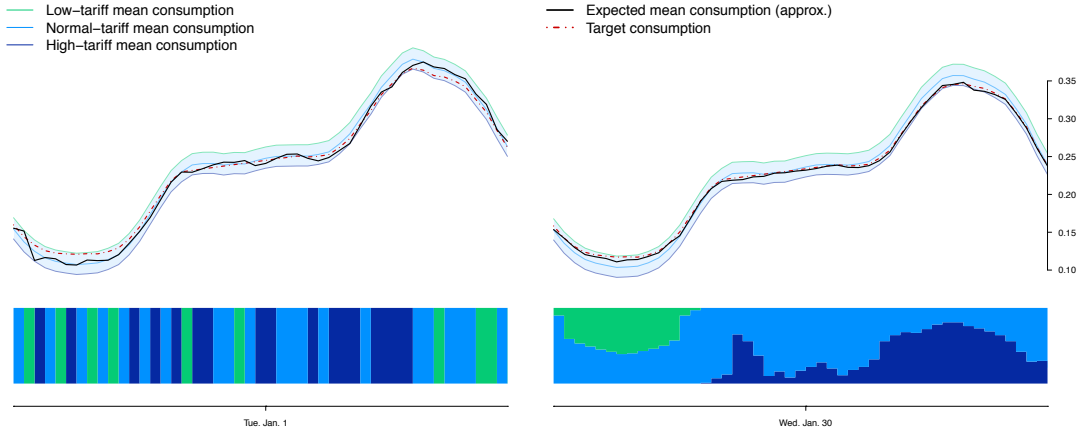
Figure 1.   *Left*: January 1st (first day of the testing set).   *Right*: January 30th (last day of the testing set).

Top: 200 runs are considered. Plot: average of mean consumptions over 200 runs for the algorithm associated with Model 1 (full black line); target consumption (dashed red line); mean consumption associated with each tariff (Low–1 in green, Normal–2 in blue and High–3 in navy). The envelope of attainable targets is in pastel blue.

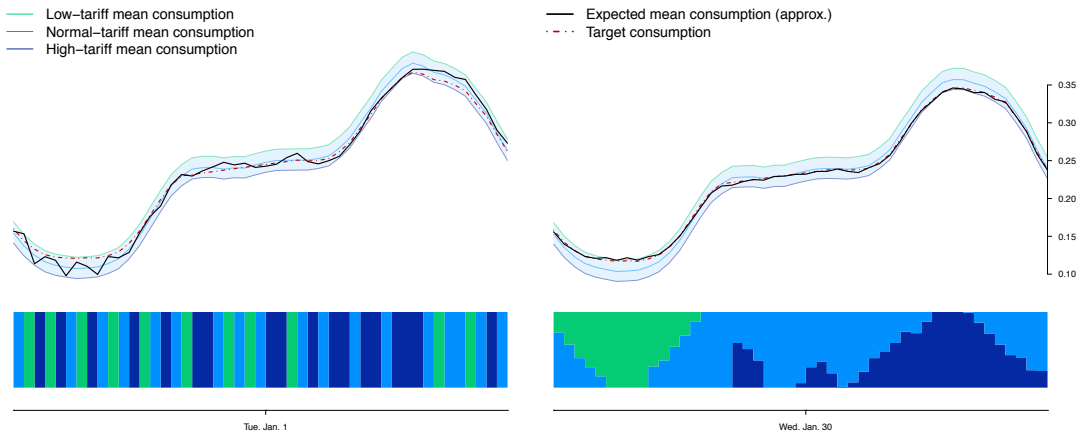*Bottom*: A single run is considered. Plot: proportions p_t used over time.

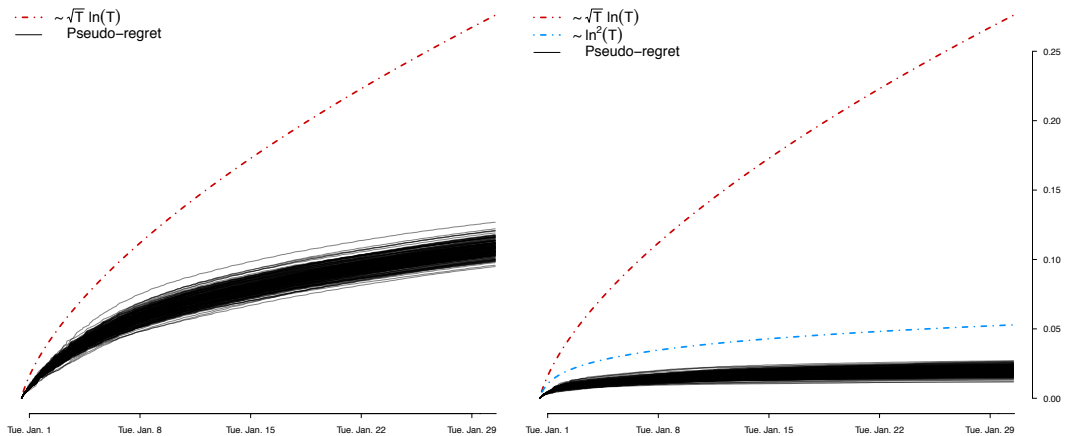*Figure 2.* Same legend, but with Model 2 (full black line).

Figure 3. Regret curves for each of the 200 runs for Model 1 (*left*) and Model 2 (*right*). We also provide plots of $c\sqrt{T}\ln T$ and $c'\ln^2(T)$ for some well-chosen constants $c, c' > 0$; these are the rates to be considered as the covariance matrix Γ is assumed to be known.

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems (NIPS'11)*, pp. 2312–2320, 2011.

Abe, N. and Long, P. M. Associative reinforcement learning using linear probabilistic concepts. In *Proceedings of the 16th International Conference on Machine Learning (ICML'99)*, pp. 3–11, 1999.

Albadi, M. H. and El-Saadany, E. F. Demand response in electricity markets: An overview. In *Proceedings of the IEEE Power Engineering Society General Meeting*, 2007.

Auer, P. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

Beygelzimer, A., Langford, J., Li, L., Reyzin, L., and Schapire, R. Contextual bandit algorithms with supervised learning guarantees. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS'11)*, pp. 19–26, 2011.

Cesa-Bianchi, N. and Lugosi, G. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Chu, W., Li, L., Reyzin, L., and Schapire, R. Contextual bandits with linear payoff functions. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS'11)*, pp. 208–214, 2011.

Dani, V., Hayes, T. P., and Kakade, S. M. Stochastic linear optimization under bandit feedback. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT'08)*, 2008.

Filippi, S., Cappé, O., Garivier, A., and Szepesvári, C. Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems (NIPS'10)*, pp. 586–594, 2010.

Fischer, D., Scherer, J., Flunk, A., Kreifels, N., Byskov-Lindberg, K., and Wille-Haussmann, B. Impact of HP, CHP, PV and EVs on households' electric load profiles. In *Proceedings of the IEEE PowerTech conference*, 2015.

Gaillard, P., Goude, Y., and Nedellec, R. Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. *International Journal of Forecasting*, 32(3):1038–1050, 2016.

Gopalan, A., Mannor, S., and Mansour, Y. Thompson sampling for complex online problems. In *Proceedings of the 31st International Conference on Machine Learning (ICML'14)*, pp. 100–108, 2014.

Goude, Y., Nedellec, R., and Kong, N. Local short and middle term electricity load forecasting with semi-parametric additive models. *IEEE Transactions on Smart Grid*, 5(1):440–446, 2014.

Kikusato, H., Mori, K., Yoshizawa, S., Fujimoto, Y., Asano, H., Hayashi, Y., Kawashima, A., Inagaki, S., and Suzuki, T. Electric vehicle charge-discharge management for utilization of photovoltaic by coordination between home and grid energy management systems. *IEEE Transactions on Smart Grid*, 10(3):3186–3197, 2019.

Lai, T. L. and Robbins, H. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

Lattimore, T. and Szepesvári, C. *Bandit algorithms*, 2018. Monograph to be published.

Li, L., Chu, W., Langford, J., and Schapire, R. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web (WWW'10)*, pp. 661–670, 2010.

Mallet, P., Granstrom, P. O., Hallberg, P., Lorenz, G., and Mandatova, P. Power to the people! european perspectives on the future of electric distribution. *IEEE Power and Energy Magazine*, 12(2):51–64, 2014.

Mannor, S. Misspecified and complex bandits problems, 2018. Talk at "50$^{\text{èmes}}$ Journées de Statistique", EDF Lab Paris Saclay, May 31st, 2018.

McMahan, H. B. and Streeter, M. J. Tighter bounds for multi-armed bandits with expert advice. In *Proceedings of the 22nd Conference on Learning Theory (COLT'09)*, 2009.

Mei, J., De Castro, Y., Goude, Y., and Hébrail, G. Nonnegative matrix factorization for time series recovery from a few temporal aggregates. In *Proceedings of the 34th International Conference on Machine Learning (ICML'17)*, pp. 2382–2390, 2017.

Perchet, V. and Rigollet, P. The multi-armed bandit problem with covariates. *The Annals of Statistics*, pp. 693–721, 2013.

Rusmevichientong, P. and Tsitsiklis, J. N. Linearly parameterized bandits. *Mathematics of Operations Research*, 35(2):395–411, 2010.

Schofield, J., Carmichael, R., Tindemans, S., Woolf, M., Bilton, M., and Strbac, G. Residential consumer responsiveness to time-varying pricing, 2014. Technical report.

Sevlian, R. and Rajagopal, R. A scaling law for short term load forecasting on varying levels of aggregation. *International Journal of Electrical Power & Energy Systems*, 98:350–361, June 2018.

Shareef, H., Ahmed, M. S., Mohamed, A., and Al Hassan, E. Review on home energy management system considering demand responses, smart technologies, and intelligent controllers. *IEEE Access*, 6:24498–24509, 2018.

Siano, P. Demand response and smart grids? A survey. *Renewable and Sustainable Energy Reviews*, 30:461–478, 2014.

Valko, M., Korda, N., Munos, R., Flaounas, I., and Cristianini, N. Finite-time analysis of kernelised contextual bandits, 2013. arXiv preprint arXiv:1309.6869.

Wang, C.-C., Kulkarni, S. R., and Poor, H. V. Arbitrary side observations in bandit problems. *Advances in Applied Mathematics*, 34(4):903–938, 2005a.

Wang, C.-C., Kulkarni, S. R., and Poor, H. V. Bandit problems with side observations. *IEEE Transactions on Automatic Control*, 50(3):338–355, 2005b.

Wang, Y., Chen, Q., Kang, C., Zhang, M., Wang, K., and Zhao, Y. Load profiling and its application to demand response: A review. *Tsinghua Science and Technology*, 20(2):117–129, 2015.

Wood, S. *Generalized Additive Models: An Introduction with R*. CRC Press, 2006.

Yan, Y., Qian, Y., Sharif, H., and Tipper, D. A survey on smart grid communication infrastructures: Motivations, requirements and challenges. *IEEE Communications Surveys Tutorials*, 15(1):5–20, 2013.