Curtailment Contracts for Flexible Loads

Marie-Louise Arlt* Dirk Neumann † Ram Rajagopal‡

July 6, 2020

Abstract

Load aggregation can be an effective approach to bundle residential loads and market their flexibility and central markets. In this work, we propose a business model for an aggregator with uncertain generation. By deploying flexible demand resources, the aggregator is able to balance short-term fluctuations of generation and avoid balancing costs. We design a contract menu for customers with HVAC systems and private temperature sensitivities and show that an incentive-compatible contract menu is feasible. This extends existing literature to time-interdependent operations and specific electricity-based services. Furthermore, we use a numerical example to demonstrate that customers are better off under a curtailment contract than under a fixed retail tariff, that low-comfort customers are better off than under real-time pricing, and that the bundling of generation and flexible load realizes synergies for the aggregator as opposed to the separate marketing of generation and flexible load.

Keywords: *Local electricity market, Demand flexibility, Time-interdependent demand, Smart home systems, Automation*

*Albert Ludwig University of Freiburg, Stanford University, SLAC National Accelerator Laboratory
†Albert Ludwig University of Freiburg
‡Stanford University

1 Introduction and Related Literature

Growing concerns about climate change have motivated communities worldwide to commit to a transformation of electricity systems toward renewable energies and sustainable system operation. To operate such systems effectively and efficiently, it is necessary to leverage the available flexibility on the demand side (e.g. IRENA, 2018). Fortunately, new technologies and the digitization in particular enable a more active involvement of demand, opening up new opportunities to include flexible loads into markets and system operations.

We identify two major strands of proposals for the design of a new economic framework. First, time-dependent retail prices and other price-based programs such as demand response (DR) can incentivize system-friendly load behavior (e.g. Bohn et al., 1984; Borenstein, 2005). In particular, 1RTP reflects the temporal scarcity of electricity and confront customers with the real costs of consumption (Bohn et al., 1984). As a result, consumption and investment will be efficient (Borenstein, 2005). However, the literature has also pointed out that demand elasticity is limited (Faruqui and Sergici, 2010); that customers exhibit risk aversion and behavioral biases with regard to the valuation of electricity (Schneider and Sunstein, 2017); and closed loops in price building might lead to price oscillations and instabilities in market and system operations (Roozbehani et al., 2010; Cho and Meyn, 2010).

Alternatively, concepts to pool and centrally schedule flexible demand resources through *aggregation* have been proposed. A wide range of the literature has developed algorithms for optimal scheduling of different types of flexible loads, without or in combination with uncertain supply, including Papavasiliou and Oren (2010), Subramanian et al. (2013), O'Brien and Rajagopal (2013), Bitar and Xu (2017), and Nayyar et al. (2016). Among them, Yu et al. (2013) have analyzed the scheduling of home appliances and 1HVAC systems in particular. Furthermore, suitable and incentive-compatible contract designs for curtailment or quality-differentiated power procurement were suggested, including Chao (1987), Tan and Varaiya (1993), and Campaigne and Oren (2016).

In our work, we built upon the above mentioned contract literature and design multi-period curtailment contracts, specify them for users of HVAC systems with heterogeneous comfort preferences for temperature, and analyze the implications for customers and aggregators. Thereby, we extend previous work in the following important aspects regarding the *modeling of curtailment contracts*: first, we propose multi-period curtailment contracts with repeated job scheduling. Previous work has concentrated on single-period or single-job contracts. Second, the contract literature has largely considered the value of electricity or the constraints for operations as given. In our work, we specifically model the application of HVAC systems which are a major electricity load for residential and commercial customers. We derive contract decisions endogenously from customers' utility and private comfort preference for HVAC operations. Third, we show that, under the given assumptions, an incentive-compatible and individually rational contract is possible. Furthermore, we propose the following contributions regarding the implications of curtailment contracts on *customer surplus and the profits of the aggregator:* first, we demonstrate that curtailment contracts improve customer surplus, as compared to a fixed retail rate. Furthermore, for customers with a low comfort preference, a curtailment contract is superior to an RTP scheme. Second, the marketing of flexible load increases profits for the aggregator. Moreover, the bundling of uncertain generation and flexible demand increases the aggregator's profits as compared to their separate marketing. Third, we can show that curtailment contracts are more valuable to the aggregator under a higher uncertainty of generation, higher imbalance costs, and a higher price variance.

The chapter is structured as follows. In Section 2, we introduce the model of the aggregator and the customers. We characterize the optimal curtailment contract and solve the end-to-end operations problem of the aggregator. In Section 3, we describe our numerical solution approach. We conduct numerical experiments and present the resulting contract menu as well as the results for customer surplus and aggregator profits in Section 4. We discuss our results and conclude in Section 5.

2 The Model

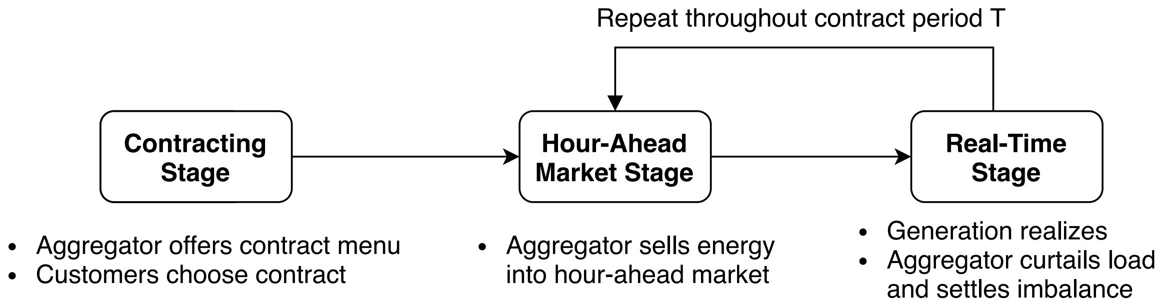

Figure 1: Sequence of actions: contract choice and hourly operations

We consider an aggregator who operates variable renewable energy generation and flexible load resources. To acquire these resources, the aggregator signs curtailment contracts with customers who, in return for an *ex ante* contract payment, grant limited load control to the aggregator ('contracting stage'). During operations, the aggregator sells electricity in the hour-ahead market ('hour-ahead market stage'). At the time of delivery, the aggregator's variable generation realizes and potentially reveals an imbalance between the aggregator's commitment in the hour-ahead market and actual generation. The aggregator must subsequently settle his balance by curtailing flexible load resources and/or paying imbalance payments ('real-time stage'). Figure 1 summarizes the process.

We formalize the aggregator's profit maximization problem using the following objective function Eq. (1),

$$\max_{F,\overrightarrow{q},\overrightarrow{DR}} -F + \mathbb{E}\sum_{t=0}^{T-1}\{p_t q_t + a_t(DR_t + s_t - q_t)^+ - b_t(q_t - DR_t - s_t)^+\}. \tag{1}$$

The components of the aggregator's profit correspond to the problem stages as described in Fig. 1. First, in the contracting stage, the aggregator pays an aggregate *ex ante* contract payment F to all customers who sign a curtailment contract for a duration of T periods. Second, in the hour-ahead market stage, the aggregator generates a profit from selling electricity q in the hour-ahead market at the wholesale-market price p. In the real-time stage, the aggregator learns about his real-time generation s. The price p in the hour-ahead market stage as well as generation s in the real-time stage are random variables and associated with a real-time state of the world ω_{RT} which follows the distribution function $h(\omega_{RT})$. Finally, the aggregator needs to settle any imbalances between his sales q in the hour-ahead stage and the generation s. There are two possible ways to resolve an imbalance: one, the aggregator can curtail part of the flexible generation under his control and achieve a demand reduction DR. Two, for any further imbalances of size $|DR + s - q|$, the aggregator can settle by paying imbalance fees. In case that there is an oversupply of resources and $DR + s > q$, the aggregator only receives a reduced price $a \leq p$. In the case of undersupply, the aggregator pays a fee $b \geq p$ for each unit of missing supply. The hour-ahead market stage and the real-time stage repeat throughout the contract duration of T periods. Table 1 presents an overview of all variables and parameters.

Symbol	Description	Symbol	Description
Customer variables and parameters		***Aggregator variables and parameters***	
u	Net utility	F	Aggregate contract payments
d	Stage-wise comfort loss	q	Hour-ahead market sales
τ	Comfort preference	p	Wholesale market price
$\underline{\tau}$	Lowest comfort preference	a	Imbalance price for oversupply
$\overline{\tau}$	Highest comfort preference	b	Imbalance cost of undersupply
$g(\tau)$	Distribution function of types τ	s	Generation
$G(\tau)$	Cumulative distribution function of types τ	DR	Aggregate curtailments
c	Curtailment	τ'	Contract type
D	Total comfort loss	$f(\tau')$	Contract payments
D_T	Disposal comfort loss after the end of the contract period	$k(\cdot)$	Curtailment function
P	Rated power of flexible load	T	Contract length
x	General state variable of flexible load	$\widetilde{\tau}$	Marginal customer of curtailment
$m(\cdot)$	General transition function of state variable	ϕ	Marginal contract cost of curtailment
		C	Settlement cost
HVAC-specific variables and parameters			
θ	Internal temperature	***System variables and parameters***	
θ^{set}	Setpoint temperature	t	Time index
θ^{out}	Outside temperature	λ	Fixed retail rate
β	Speed of temperature convergence	ω_{RT}	State of the world
γ	HVAC efficiency	Ω	State space
P^{el}	HVAC rated power	$h(\omega_{RT})$	Distribution function of states

Table 1: Variables and parameters

In the following sections, we sequentially solve the aggregator's optimization problem stage by stage. We start with the contracting stage in Section 2.1. We will then solve the optimization problem in the hour-ahead market and the real-time stage by backward induction. Therefore, we will first present the solution to the aggregator's optimization problem in the real-time stage in Section 2.2 and will then provide the solution to the hour-ahead market stage in Section 2.3.

2.1 Contracting Stage

This section describes the construction of a profit-maximizing contract. First, we describe the requirements of a general curtailment contract (Section 2.1.1). Then, we will specify the contract for HVAC systems which are a major residential load and can be operated in a flexible way (Section 2.1.2).

2.1.1 General Contract Description

Customer characteristics. The aggregator procures demand flexibility from customers with flexible load resources. Customers are utility-maximizing agents and can be characterized by their comfort preference τ ('customer of type τ'). Customers with a low comfort preference experience less comfort loss from curtailments than customers with high comfort preference. The comfort preference τ is private information and the aggregator is only aware of the minimum and maximum boundaries of τ, i.e. $\tau \in [\underline{\tau}, \overline{\tau}]$, and the fact that τ is distributed according to the distribution function $g(\tau)$.

In the contracting stage, the aggregator offers a curtailment contract menu to customers. We are focusing on contracts of the following form: in the contracting stage, the customer τ who chooses a contract τ' receives an *ex ante* contract payment according to the chosen contract type. During operations over a contract period of T periods, the customer is curtailed according to the curtailment policy associated with contract τ'. More curtailments are in general associated with higher comfort losses for the customer.

The customer of type τ maximizes his total expected net utility from comfort payments and comfort losses by the choice of a contract. The contracts of the contract menu are indexed by type τ',

$$\max_{\tau'} \mathbb{E}\, u(\tau'|\tau) = \max_{\tau'} f(\tau') + \mathbb{E}\, D(\tau'|\tau). \tag{2}$$

The expected net utility under a curtailment contract τ' corresponds to the sum of the *ex ante* contract payment $f(\tau')$ and the expected comfort loss from curtailments $\mathbb{E}\, D(\tau'|\tau)$ during the contract duration of T periods, $\mathbb{E}\, D(\tau'|\tau) < 0$. The latter corresponds to the total utility change as compared to the situation when the consumer constantly receives firm electricity supply from a retailer. For a specific state of the world ω_{RT}, net utility for the contract duration T is described by,

$$D(\tau, \omega_{RT}, \vec{c}^{\tau'}) = \sum_{t=0}^{T-1} \{d(\tau, x_t) + \lambda_t P c(\tau', \omega_{RT}, t)\} + D_T(\tau, x_T) \tag{3}$$

$$s.t. x_{t+1} = m(x_t, c(t)).$$

Eq. (3) consists of two components. First, the customer saves electricity costs whenever the flexible load is curtailed. We assume that $c_t = 0$ if there is no curtailment in t and 1 if there is a curtailment in t. λ_t is the retail price and $\lambda_t P$ describes the total savings in a period t. Second, the customer suffers comfort losses. The comfort loss covers comfort losses $d(\tau, x_t) \leq 0$ during the contract duration as well as comfort losses after the end of the contract period, represented by $D_T(\tau, x_T) \leq 0$ but caused by curtailment during the contract duration. Comfort losses increase with more curtailments, i.e. $\frac{dD}{dc_t} < 0, \forall t \in \{0, ..., T-1\}$. x_t describes the state of service provision by the flexible load. This can be, for instance, the temperature achieved through the operation of an HVAC system or the number of miles an electric vehicle is able to travel because of charging. States across periods are coupled by the general transition function $m(\cdot)$ and are influenced by curtailment decisions of the aggregator.

Direct mechanism. Optimally, customers with a low comfort parameter and less comfort losses from curtailments provide more load flexibility than customers with a high comfort parameter. However, given that the comfort preference τ is private information, the aggregator cannot directly discriminate between the different customer types. In such a situation, incentive-compatible mechanisms can help to maximize profits. A mechanism is incentive-compatible if a customer with a certain comfort preference τ self-selects into a contract $\tau' = \tau$ because it is utility-maximizing to do so,

$$\tau \in \arg \max_{\tau' \in [\underline{\tau}, \overline{\tau}]} \mathbb{E}\, u(\tau'|\tau)\ \forall \tau. \tag{4}$$

The aggregator therefore aims to specify an incentive-compatible direct mechanism according to the following Definition 2.1.

Definition 2.1 (Direct mechanism) *We define the direct mechanism for a curtailment contract menu by the ex ante payment function,*

$$[\underline{\tau}, \overline{\tau}] \ni \tau \to f(\tau) \in \mathbb{R}, \tag{5}$$

and the curtailment function,

$$[\underline{\tau}, \overline{\tau}] \ni \tau \to k(\tau, \omega_{RT}, t) \in \{0, 1\}. \tag{6}$$

The *ex ante* payment function describes how much a customer receives as an *ex ante* payment if he chooses a contract which is intended for customers with the comfort preference τ. The curtailment function describes the curtailment which a customer experiences under a contract τ during the contract period T, as a function of the state of the world ω_{RT}.

Furthermore, the mechanism must be individually rational for participating customers, i.e. a customer τ signing a contract $\tau' = \tau$ must be at least as well off as under the outside option (no curtailment contract),

$$u(\tau) \geq 0 \ \forall \tau \in [\underline{\tau}, \overline{\tau}]. \tag{7}$$

We can leverage the envelope theorem as generalized by Milgrom and Segal (2002) to construct an incentive compatible net utility profile ($\tau' = \tau$) over the type spectrum.

Lemma 2.2 (Incentive-compatible net utility of a contract menu) *An contract menu which provides the following net utility to customers,*

$$\mathbb{E}\, u(\tau) = \mathbb{E}\, u(\overline{\tau}) - \int_{\tau}^{\overline{\tau}} \frac{\partial}{\partial z} \mathbb{E}\, D(\tau'|z)|_{\tau'=z} dz, \tag{8}$$

is incentive-compatible if $\frac{\partial}{\partial z} \mathbb{E}\, D(\tau'|z)|_{\tau'=z}$ *is continuous and differentiable.*

Proof 1 *See Section 6.*

The contract is built from the customer with the highest comfort preference, $\overline{\tau}$. For this customer, the second term becomes zero as the integration interval has a length of 0. $\mathbb{E}\, u(\overline{\tau})$ is subject to the choice of the aggregator. If he sets $\mathbb{E}\, u(\overline{\tau}) < 0$, the customer $\overline{\tau}$ will not consider the contract as the menu does not comply with the contraint of individual rationality. If the aggregator sets $\mathbb{E}\, u(\overline{\tau}) > 0$, he grants additional utility to the customers. $\mathbb{E}\, u(\overline{\tau}) = 0$ is, therefore, the profit-maximizing choice for the aggregator. For customers with comfort preferences $\tau < \overline{\tau}$, the second terms becomes non-zero and, as $\frac{\partial}{\partial z} \mathbb{E}\, D(\tau'|z)|_{\tau'=z} < 0$, $\mathbb{E}\, u(\tau)$ becomes positive. The result shows that the menu binds ‘at the top’. For all other types, expected net utility becomes positive. This reflects the cost of asymmetric information for the aggregator which is called the ‘information rent’. If the aggregator was aware of the comfort preference of each customer, he could discriminate directly and all customers would receive a zero net utility. The aggregator would keep the information rent, increasing his profits.

Based on Theorem 2.2, we can characterize the incentive-compatible *ex ante* contract payments.

Theorem 2.3 (Contract payments) *Contract payments $f(\tau)$ can be calculated as follows,*

$$f(\tau) = -\,\mathbb{E}\,D(\tau'|z)|_{\tau'=z}. - \int_{\tau}^{\overline{\tau}} \frac{\partial}{\partial z}\,\mathbb{E}\,D(\tau'|z)|_{\tau'=z} dz \tag{9}$$

Proof 2 *The result can be obtained from inserting Theorem 2.2 into the definition $f(\tau) = \mathbb{E}\,u(\tau'|\tau) - \mathbb{E}\,D(\tau'|\tau)$ and requiring $\mathbb{E}\,u(\overline{\tau}) = 0$ for profit-maximization.*

Note that both parts of the sum are negative and, therefore, the contract payment is positive, i.e. $f(\tau) > 0$. Equation 9 illustrates that the contract payment consists of two parts. According to the first part of the payment, the aggregator compensates customers for their expected comfort loss from curtailment. By doing so, it is ensured that consumers participate and choose a contract other than the outside option ('individual rationality'). The second part of the sum is an additional 'information rent' which the aggregator pays to incentivize customers to reveal their true type τ ('incentive compatibility'). The customer with the highest comfort preference $\overline{\tau}$ is only compensated for his comfort loss and does not get any information rent. The lower the comfort preference of a customer, the higher is the information rent he receives as the information rent is integrated over the space $z \in [\tau, \overline{\tau}]$.

Lemma 2.4 describes the aggregate contract payment which can be derived by integration over the type space $[\underline{\tau}, \overline{\tau}]$.

Lemma 2.4 *Aggregate contract payments total the sum of individual contract payments,*

$$F = -\int_{\underline{\tau}}^{\overline{\tau}} g(\tau)[\int_{\tau}^{\overline{\tau}} \frac{\partial}{\partial z}\,\mathbb{E}\,D(\tau'|z)|_{\tau'=z} dz + \mathbb{E}\,D(\tau'|z)|_{\tau'=z}] d\tau. \tag{10}$$

Proof 3 *The result is found by inserting Theorem 2.3 into the definition of aggregate payments, i.e. $F = \int_{\underline{\tau}}^{\overline{\tau}} g(\tau) f(\tau) d\tau$.*

Our business model and the described contract design can be applied to different kinds of application. In this work, we focus on HVAC systems which are a major load in residential households and commercial facilities and specify the respective contract in the following section.

2.1.2 HVAC-Specific Contract Description

Customers are operating HVAC systems for their personal comfort. We specify Eq. (3) to describe the total utility change of a customer with an HVAC system,

$$D(\tau, \overrightarrow{c}^{\tau'}) = \sum_{t=0}^{T-1} \{d(\tau, \theta_t) + \lambda_t P^{el} c_t^{\tau'}\} + D_T(\tau, \theta_T). \tag{11}$$

Similar to the general in Eq. (3), the customer saves electricity costs of $\lambda_t P^{el}$ whenever the HVAC system is curtailed. P^{el} is the rated power of the HVAC system. Second, the customer suffers comfort losses from temperature deviations when his HVAC system is curtailed. We choose the following utility function to describe the comfort changes due to curtailments under a curtailment contract,

$$d(\tau, \theta_t) = -\tau(|\theta_t - \theta^{set}|)^2. \tag{12}$$

The description is inspired by Fanger (1970). The customer's utility decreases in the difference $\Delta\theta_t = |\theta_t - \theta^{set}|$. The comfort preference τ can be interpreted as the customer's sensitivity toward temperature deviations: customers with a low comfort preference experience a smaller comfort loss from temperature deviations than customers with a high comfort preference, i.e. $\frac{dd(\tau,\theta_t)}{d\tau} < 0$. We further assume that utility decreases over-proportionally in the temperature difference to the setpoint $\Delta\theta_t$, i.e. $\frac{d^2 d}{d\Delta\theta^2} < 0$.

The temperature profile of customers is linked by the following transition function (Mathieu et al., 2013),

$$\theta_{t+1} = \beta\theta_t + (1-\beta)[\theta^{out} - (1-c_t)\gamma P^{el}]. \tag{13}$$

β describes the speed of convergence between the outdoor temperature θ^{out} and the indoor temperature θ_t. γ indicates the efficiency and P^{el} the rated capacity of the HVAC system. We define γP^{el} such that the temperature can be held constant at θ^{set} when the HVAC system is continuously operating, i.e. $\theta^{set} = \beta\theta^{set} + (1-\beta)(\theta^{out} - \gamma P^{el})$. As a result, net HVAC power is characterized by $\gamma P^{el} = \theta^{out} - \theta^{set}$. We require that the technical characteristics are identical across customers.

In the following analysis, we focus on the use case of cooling. In this case, the following specifications apply. The outdoor temperature is higher than the setpoint temperature, $\theta^{out} > \theta^{set}$ and $\frac{dd(\tau,\theta_t)}{d\theta_t} \leq 0$. The internal temperature is bounded between them, i.e. $\theta^{out} \geq \theta_t \geq \theta^{set}$. If the HVAC system does not operate at all over multiple periods, the internal temperature approaches the outdoor temperature as a consequence of Eq. (13). *Vice versa*, if it operates continuously, the internal temperature approaches the setpoint temperature. Furthermore, as a condition for constructing the contract menu according to Theorem 2.2, we need to analyze the properties of $\frac{\partial}{\partial\tau}\mathbb{E}\,D(\tau'|\tau)$.

Lemma 2.5 (Continuity and differentiability) *Equilibrium utility ($\tau' = \tau$) is continuous and differentiable with respect to its true type τ. The first derivative of $\mathbb{E}\,D(\tau'|\tau)$ with respect to the true type τ equals,*

$$\frac{\partial}{\partial\tau}\mathbb{E}\,D(\tau'|\tau) = -\,\mathbb{E}(\sum_{t=0}^{T-1}\{|\theta_t - \theta^{set}|^2\} + \frac{1}{1-\beta^2}|\theta_T - \theta^{set}|^2). \tag{14}$$

$\frac{\partial}{\partial\tau}\mathbb{E}\,D(\tau'|\tau)$ is bounded by $0 \geq \frac{\partial}{\partial\tau}\mathbb{E}\,D(\tau'|\tau) \geq -|\theta^{out} - \theta^{set}|^2(T + \frac{\beta^2}{1-\beta^2})$.

Proof 4 *See Section 6.*

We find that for a given contract type τ', the first derivative of the expected comfort loss with respect to the true type $\frac{\partial}{\partial \tau} \mathbb{E}\, D(\tau'|\tau)$ is a bounded real number. This number only depends on the expected temperature changes $\Delta\theta$ during the contract duration which are independent of the true type τ and only dependent on the contract type τ'. Therefore, the expected comfort loss is continuous and differentiable with respect to the true type and we can apply Theorem 2.2. The result describes how much net utility a customer of type τ experiences under an incentive-compatible contract.

Furthermore, we can insert the expression Eq. (14) in Eq. (9) and Theorem 2.4 to derive the type-dependent contract payments $f(\tau)$ and the aggregate contract payment F. However, the specification of the contract requires the expected temperature path of each customer. In the next two sections, we therefore derive the optimal curtailment of customers which determines the temperature path $\theta_0, ..., \theta_{T-1}$.

2.2 Real-Time Stage

The optimal curtailment of load is determined by its marginal cost. During operations, curtailing load does not cause any short-term costs for the aggregator but only to customers. However, in the long-run, curtailment increases costs for the aggregator through the *ex ante* contract payments. The cost of curtailing the marginal customer $\widetilde{\tau}$ are therefore defined by the first derivative of aggregate contract payments of Theorem 2.4. Theorem 2.6 describes the marginal cost of curtailment ϕ.

Lemma 2.6 (Marginal cost of curtailment) *The marginal cost $\phi_t(\widetilde{\tau})$ of curtailing a customer $\widetilde{\tau}$ are described by,*

$$\phi_t(\widetilde{\tau}) \triangleq \frac{\partial F}{\partial c_t(\widetilde{\tau})} = -\left(\frac{\partial}{\partial c_t(\widetilde{\tau})} D_t(\widetilde{\tau}, \theta_t) + \frac{G(\widetilde{\tau})}{g(\widetilde{\tau})} \frac{\partial^2}{\partial c_t \partial \tau} D_t(\tau, \omega_{RT})\right)\Big|_{\tau=\widetilde{\tau}}. \tag{15}$$

Proof 5 *See Section 6.*

Like the contract payments $f(\tau)$ described by Theorem 2.3, the marginal costs ϕ_t consist of two parts. First, the marginal customer $\widetilde{\tau}$ must be compensated for the comfort loss he experiences because of a curtailment. Again, this includes possible comfort losses in t as well as in subsequent periods. This compensation ensures that the constraint of individual rationality is met. Second, the aggregator needs to pay an information rent to all customers with a lower comfort preference than the marginal customer, i.e. to all customers of type $\tau < \widetilde{\tau}$. The total amount of this payment depends on number of customers and is represented by $\frac{G(\widetilde{\tau})}{g(\widetilde{\tau})}$. The information rent ensures that the contract menu is incentive-compatible. Furthermore, we find that the marginal cost of curtailment increases in the temperature θ_t of a customers $\widetilde{\tau}$, see Eq. (12) and Eq. (14).

In the real-time stage of time t, the aggregator determines the marginal costs for curtailing each of the contracted customers. The marginal cost curve of curtailment $C'(DR)$ can subsequently be derived from sorting $\phi_t(\widetilde{\tau})$ in an ascending order, i.e. types with low marginal cost of curtailment would be curtailed before those with high marginal cost of curtailment. One special case applies if the internal temperature θ_t is equal across customers. In that case, ϕ_t is monotonously increasing in τ and customers with a low comfort preference will be curtailed first.

In the real-time stage, the only remaining decision variable of the customer is curtailment DR: the aggregator has already decided on sales q_t in the hour-ahead market and the renewable generation s_t has already realized and is known to the aggregator. Taking the first derivative of Eq. (1) with respect to curtailment DR_t produces the FOC provided by Theorem 2.7.

Lemma 2.7 (FOC of optimal curtailment) *In the real-time stage, the FOC of optimal curtailment DR is described by,*

$$\frac{\partial F}{\partial DR_t} + \mathbb{E}\sum_{t'=t+1}^{T-1}[p_{t'}\frac{dq'_t}{dDR_t} + \frac{\partial C_{t'}}{\partial q_{t'}}\frac{dq_{t'}}{dDR_t} - \frac{\partial C_{t'}}{\partial DR_t}] = \begin{cases} a, & for \quad DR_t + s_t \geq q_t; \\ b, & for \quad DR_t + s_t < q_t. \end{cases} \tag{16}$$

The optimal curtailment is a trade-off between the marginal cost of curtailment (LHS of the FOC) and the expected savings from curtailment (RHS of the FOC). With regard to costs, curtailments cause contract fees to rise ($\frac{\partial F}{\partial DR_t} > 0$). Furthermore, curtailments increase the temperature of a customer and, therefore, the cost of curtailments in the future, as we have explained for Theorem 2.6. As a consequence, curtailment in t will increase balancing costs $C(DR)$ in the future ($\frac{\partial \mathbb{E} C_{t'}}{\partial DR_t} > 0$) and, as we will explain in the following section, decrease future sales on the wholesale market $q_{t'}$ for $t' > t$ and hour-ahead market income. Both effects interact in decreased expected settlement cost ($\frac{\partial \mathbb{E} C_{t'}}{\partial q_{t'}} \frac{dq_{t'}}{dDR_t} < 0$). In total, the cost of curtailment is positive which results in a trade-off between savings from additional curtailment DR and the cost of doing so. The savings achieved by a curtailment depend on the balance of electricity sales q, generation s, and curtailments DR. If the system is in oversupply and $DR_t + s_t \geq q_t$, additional curtailment will increase oversupply even further which are associated with an additional income $a \leq p$ to the aggregator. If the system is in undersupply and $DR_t + s_t < q_t$, additional curtailment will decrease the amount of undersupply and saves the aggregator imbalance fees of size $b \geq p$. If the expected cost of curtailment is smaller than the imbalance payment a, it is optimal to curtail even further and increase the oversupply. *Vice versa*, if the cost of curtailment exceeds b, it is optimal to not increase curtailment but pay the imbalance fee b instead.

2.3 Hour-Ahead Market Stage

We can now determine optimal sales q at the hour-ahead market. At this stage, the generation s is not known yet. Again, based on the objective function 1, we derive the FOC with respect to q.

Lemma 2.8 (FOC of optimal hour-ahead sales) *In the hour-ahead market stage, the FOC of optimal hour-ahead sales q are characterized by,*

$$p_t = \frac{d\,\mathbb{E}\,C_t}{dq_t} - \mathbb{E} \sum_{t'=t+1}^{T-1} [p_{t'} \frac{dq^*_{t'}}{dq_t} - \frac{\partial C_{t'}}{\partial q^*_{t'}} \frac{dq_{t'}}{dq_t} - \frac{\partial C_{t'}}{\partial DR^*_t} \frac{\partial DR^*_t}{\partial q_t}], \tag{17}$$

where

$$\frac{d\,\mathbb{E}\,C_t}{dq_t} = Prob(s_t + DR_t - q_t > 0)a + Prob(s_t + DR_t - q_t < 0)b \\ + \int_{\underline{\tau}}^{\overline{\tau}} Prob(s_t + DR_t(s) - q_t = 0)\phi_t(s)ds. \tag{18}$$

Lemma 2.8 states that, in the optimum, the marginal income from sales, i.e. the hour-ahead market price p_t, equals the marginal expected cost of curtailment. According to Eq. (17), the latter consist of the following components: first, higher sales q increase the expected amount of curtailment DR in the real-time stage ($\frac{d\,\mathbb{E}\,C_t}{dq_t} > 0$). As higher DR implies that customers with a higher marginal cost of curtailment ϕ will be curtailed, the marginal cost of curtailment increases, i.e. $\frac{d\,\mathbb{E}\,C_t}{dq_t} > 0$. Second, analogous to the argument in Section 2.2, higher curtailment increases the temperature of customers and, therefore, increases the future curtailment cost $C(DR)$, with consequences for future sales q.

Eq. (18) details the expected marginal curtailment cost for an increase in q. If the optimal curtailment decision DR in the real-time stage leads to an oversupply, the applicable marginal cost is a. In contrast, if the optimal decision DR leads to an undersupply, the applicable marginal cost is b. Finally, if the optimal curtailment decision DR leads to a balancing of sales and supply, the marginal cost of curtailment apply, i.e. $\phi_t(\widetilde{\tau})$. All the cost components are weighted according to their probability of occurrence. In our model, probability of each case realizing is driven by the probability distribution of the uncertain generation.

Eq. (18) further illustrates why the cost of curtailment increases for future periods $t' > t$ through curtailment in t. We know from the previous Section 2.2 that the marginal cost of curtailment increases with the temperature of customers. Accordingly, the expected marginal cost of curtailment represented in the last part of Eq. (18) increase. Therefore, it also becomes more likely that the aggregator pays imbalance fees b. This leads to an increase in the marginal expected cost of curtailment $\frac{d\,\mathbb{E}\,C_t}{dq_t}$. As a consequence, hour-ahead sales q must decrease to comply with the FOC stated in Eq. (17).

3 Numerical Solution Approach

The model is not easily solvable for two reasons: first, the problem is not analytically tractable because an HVAC system cannot be continuously controlled. Second, the problems suffers from the curse of dimensionality, introduced by the possible sequences of uncertain generation and prices. To address the problem, we propose to discretize customer types (Section 3.1), simplify the aggregator's intertemporal decision taking (Section 3.2), and solve for the equilibrium contract by convergence (Section 3.3).

3.1 Discretization of Customer Types

By the discretizing the continuously defined type space $\underline{\tau}, \overline{\tau}$, we serve two purposes: first, a limited number of contract classes seems to be closer to reality than the continuously defined contract sets typically proposed by mechanism design. Second, we make the problem numerically tractable. As the marginal cost of curtailment $\phi_t(\widetilde{\tau})$ are determined by the combination of the comfort preference type τ and the current temperature θ_t, the temperature of each customer and, accordingly, the curtailment cost function $C(DR)$ must be recalculated at each time step t.

To provide an example, at the beginning of the contract period, θ_0 is equal across customers. Then, the marginal cost of curtailment $\phi_0(\tau)$ is increasing in the comfort preference τ. However, once θ_t diverges because of past curtailments, the marginal utility loss by additional curtailments is not increasing in τ anymore as θ_t introduces an additional heterogeneity. Higher types τ^H might then be curtailed rather than lower types τ^L if $\theta_t^H < \theta_t^L$.

For the discretization, we use the following approoach. As stated in the theoretical model in Section 2, the aggregator knows the distribution of customers' comfort parameter $g(\tau)$ and the number of customers N. The aggregator uses this information to construct N discrete contract classes of equal probability. For instance, if $g(\tau)$ describes a uniform distribution, the type space is divided into equal sub-spaces of length $\Delta\tau = \frac{\overline{\tau}-\underline{\tau}}{N}$. Each contract class n is denoted by its medium type, e.g. $\widehat{\tau} = \underline{\tau} + \frac{\Delta\tau}{2}$ for the lowest class. We furthermore require incentive compatibility to apply across the whole class, i.e. a customer of class n can only be curtailed during operations if it is also worth for the highest type $\overline{\tau}_n$ of this class.

3.2 Approximation of Optimal Decisions by Aggregator

In the hour-ahead stage, the aggregator determines the optimal hour-ahead sales q according to Lemma 2.8. It is not trivial to specify the stated FOC as the problem of determining the optimal sales strategy suffers from the curse of dimensionality. Furthermore, aggregators and customers are usually not able to correctly forecast possible scenarios and their probability of realization. Therefore, in practice, they often rely on heuristics or simplified decision frameworks.

For the numerical experiments in the following Section 4, we will use the heuristic of a myopic aggregator without sophisticated instruments of forecasting and approximate the FOC by,

$$p_t \approx \frac{d\,\mathbb{E}\,C_t}{dq_t}. \tag{19}$$

This approximation systematically underestimates the cost implications of a curtailment. We apply the same logic for the real-time stage and use the following approximation of increasing DR,

$$\frac{\partial F}{\partial DR_t} \approx \begin{cases} a, & \text{for} \quad DR_t + s_t \geq q_t; \\ b, & \text{for} \quad DR_t + s_t < q_t. \end{cases} \tag{20}$$

Again, the aggregator generally underestimates the marginal cost of curtailment and curtails loads too frequently than it would be profit-maximizing. We will re-visit the implications of this assumption in Section 4.

3.3 Equilibrium Contract Payments

The derivation of the equilibrium contract payments requires the weighting of the outcome under all possible realizations of the state of the world ω_{RT}. This is not feasible for longer contract durations due to the curse of dimensionality. Alternatively, we can take advantage of the law of large numbers, simulate the operations under a curtailment contract for an extended time, and use the resulting curtailments and temperatures to approximate the contract payments according to Theorem 2.5 and Theorem 2.3.

4 Numerical Experiments

In the following section, we use numerical experiments to illustrate our results. Section 4.1 describes the parametrization of our experiments. We then analyze the resulting contract menu and the effects for customers, the aggregator, and market efficiency in Section 4.2. Finally, in Section 4.3, we illustrate how external parameters impact our conclusions.

4.1 Parametrization

We describe customers by the following parametrization. We consider five customers with uniformly distributed comfort parameters $\tau \in [0.15, 0.25]$. The setpoint temperature θ^{set} is $70°F$, the outdoor temperature θ^{out} is $80°F$. The insulation parameter β is 0.5 and the rated power of the HVAC system is normalized to 1. Therefore, in order to maintain the setpoint temperature constant if the HVAC system is operating without curtailment, we require HVAC efficiency γ to be 10 (see Section 2.1.2). The relevant parameters of the default case are summarized in Table 2. We drop the physical units for better readability.

The aggregator is characterized as follows. He owns an installed uncertain generation capacity s^{max} of 5 which equals the number of available flexible customers, i.e. $s^{max} = N$. The possible realizations of generation s are discrete with $s \in \{0, ..., s^{max}\}$. Generation follows a process which is determined by the probability $Prob(s_t = s_{t-1})$ that generation is

Table 2: Parametrization of consumers and market

Parameter	Specification	Parameter	Specification
Customer parameters		***Aggregator parameters***	
N	5	s^{max}	5
$\underline{\tau}$	0.15	s	$\{0, 1, ..., s^{max}\}$
$\overline{\tau}$	0.25	$Prob(s_t = s_{t-1})$	0.8
$g(\tau)$	uniform	T	24
θ_0	70	***System parameters***	
θ^{set}	70	a	0
θ^{out}	80	b	10
β	0.5	p	AR(1)
γ	10	σ^2_{price}	1
P^{el}	1	λ	5

equal in two subsequent periods. For the remaining possible generation realizations, probabilities are symmetric for equally distanced possible outcomes, i.e. $Prob(s_t = s_{t-1} + \Delta s) = Prob(s_t = s_{t-1} - \Delta s)$, and exponentially decreasing with distance, i.e. $Prob(s_t = s_{t-1} + \Delta s) = 2 \cdot Prob(s_t = s_{t-1} + 2 \cdot \Delta s)$. The probabilities other than $Prob(s_t = s_{t-1})$ are scaled such that $\sum_{s=0}^{s^{max}} Prob(s_t = s) = 1$. The aggregator offers curtailment contracts of a duration of 24 hours.

Finally, on the system side, the hour-ahead market price p is bounded by the imbalance prices, i.e. $p \in [a, b]$. The imbalance price in a situation of oversupply a is defined as 0. For undersupply, we define an imbalance payment of 10. The price p furthermore follows an AR(1) process with a variance of 1. The fixed retail tariff λ corresponds to the expected average of the wholesale market price, i.e. $(a + b)/2$.

We simulate operations over 100,000 time steps (simulation hours) to approximate the equilibrium contract menu by convergence. This corresponds to more than 4,000 contract periods of 24 hours.

4.2 Results

In this section, we report the results of our numerical experiment. First, we present the resulting contract menu (Section 4.2.1). Then, we explain the implications for customers (Section 4.2.2). Finally, we analyze the implications for the aggregator (Section 4.2.3).

4.2.1 Contract Menu

We first present the resulting contract menu. To improve the interpretability and comparability of our numerical experiments, we normalize our calculations by the utility experienced by customers if they have not signed a contract. Under firm supply, we define the customer surplus to be equal to their payments for electricity. For instance, if they consume electricity worth 1 USD, they remain with a surplus of 1 USD which is the net utility from HVAC operations.

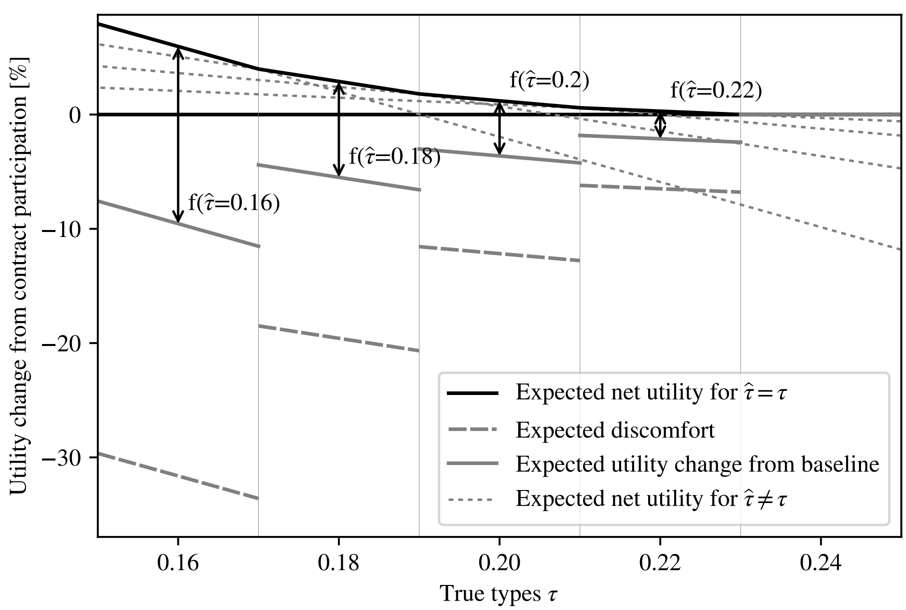

Figure 2: Default contract menu, parametrizations according to Table 2

Fig. 2 illustrates the normalized contract menu. The solid black line represents the net utility change of customers under a curtailment contract. For the customer class with the lowest medium comfort preference $\hat{\tau} = 0.16$, signing a curtailment contract can increase net utility by 4.0% (if $\tau = 0.17$) to 7.9% (if $\tau = 0.15$). This advantage drops with increasing contract types, down to 0.0%. The customer $\tau = 0.23$ gets included into operations and is compensated for his comfort loss but he does not receive an information rent. Therefore, the resulting net utility is 0 and he is indifferent between a curtailment contract and firm supply at the fixed retail tariff. Higher types with comfort preferences $\tau > 0.23$ are not included in the curtailments and only get offered a zero contract with $f(\hat{\tau} = 0.24) = 0$.

The net utility is composed of the following components. The dashed grey line in Fig. 2 illustrates the *comfort loss* experienced from curtailments. The lowest customer type class with $\tau \in [0.15, 0.17]$ experiences significant curtailments, causing a comfort loss of -29.6% to -33.6% compared to net utility under the benchmark scenario. The comfort loss is partially compensated by the bill savings of the customers as illustrated by the solid grey line. In sum, the comfort loss and the bill savings evaluate to -7.6% to -11.5% for the customers with a comfort parameter of $\tau \in [0.15, 0.17]$. Finally, the aggregator pays customers an *ex ante* contract payment $f(\hat{\tau})$. For the contract $\hat{\tau} = 0.16$, this equals to 15.5% of the baseline utility. The payment corresponds to the distance between the net utility change between the grey and the black solid lines and is indicated by the arrows. According to Theorem 2.3, first, the payment compensates the customers for the experienced comfort loss to a net utility change of 0% and maintain 'individual rationality'. This is reflected by the distance between the grey line and the x axis. Second, the payment covers the information rent to achieve incentive compatibility. This corresponds to the distance between the x axis and the black net utility line.

We can draw the following sights from Fig. 2. First, in our illustration, customers with lower comfort preferences experience higher comfort losses from curtailment. This is driven by frequent curtailments. However, this is not a general finding. As customers with low comfort preferences experience less comfort loss from the same curtailments, more frequent curtailment could still result in less comfort loss. Second, in line with our theoretical results, lower comfort types receive higher information rents. Third, Fig. 2 demonstrates that the contract menu is incentive-compatible. The dotted lines indicate the net utility of a type τ under alternative contracts. For the true type $\tau = 0.16$, for instance, the intended contract $\widehat{\tau} = 0.16$ leads to a net utility increase of 6.0%, while contracts $\widehat{\tau} = 0.18$ and $\widehat{\tau} = 0.2$ would only lead to 5.1% and 3.6%, respectively. It is, thus, utility maximizing to choose contract $\widehat{\tau} = 0.16$ as net utility describes the outer hull of all contracts $\widehat{\tau}$, evaluated for all possible true types τ.

4.2.2 Implications for Customers

	Curtailment contract	**RTP**
Net utility [%]	101.9	101.9
Comfort [%]	93.4	90.8
Average temperature [%]	101.4	101.8
Temperature std.dev. [$^\circ F$]	1.7	1.9
Electricity cost [%]	84.9	79.6
kWh price [%]	94.7	91.0

Table 3: Welfare effects for customers

Table 3 displays the aggregate consumer surplus change as compared to firm supply at the fixed retail rate. Additionally, we introduce the direct participation of customers at the RTP as a benchmark into the analysis. At the same time, the behavior of HVAC systems under an RTP characterizes the efficient dispatch behavior (Bohn et al., 1984; Borenstein, 2005, e.g.). In general, we find that, in the aggregate, curtailment contracts and RTP lead to the same net utility level of 101.9% as compared to a fixed retail tariff (100%). Detailing the components of this net utility change, however, reveals differences between the two flexible pricing schemes. First, while the comfort from HVAC operations deteriorates in both cases, comfort is higher in the curtailment contract case. Under a curtailment contract, the average comfort level of customers drops to 93.4% of the comfort level under a fixed retail tariff. In the case of RTP, comfort decreases further to a level of 90.8%. The drop in comfort levels is driven by the temperature of consumers. Under a curtailment contract, the average temperature of customers increases by 1.4% as compared to an increase of 1.8% under RTP. This is reflected by the higher standard deviation in the temperature ($1.7°F$ versus $1.9°F$). Second, electricity costs for customers decrease substantially under both flexible pricing schemes. To ensure comparability, we use a fixed retail rate which equals the realized average wholesale market price of the simulation ($\lambda = 4.974$). Then, under a curtailment contract, electricity costs drop to a level of 84.9%. The reduced cost are driven by savings of electricity costs during curtailment and the *ex ante* contract payment. If the customer is subject to the RTP, costs are further decreased to 79.6%, as a results of less frequent operations and lower dispatch prices. Both pricing schemes result in lower average wholesale market procurement costs per kWh of 94.7% and 91.0%, respectively.

Table 4 specifies the results for the different customer types. We find that, for the lowest customer type $\tau = 0.16$, comfort under a curtailment contract is less than under an RTP. The opposite is true for higher types which experience higher comfort under a curtailment contract. For the highest contract type 0.24, there is no change of comfort under a contract but some curtailment under an RTP. Moreover, electricity costs generally increase with the type. They are lowest for a customer with the comfort preference of 0.16 and, for this particular customer, lower under a curtailment contract than under wholesale market participation. The reverse is true for customers of type 0.18 or higher. Furthermore, in line with the theory, net utility is consistently higher under both flexible pricing schemes than under a fixed retail rate. The only exception is for type 0.24 for which the net utility is equal to the net utility under firm supply as it is not profitable for the aggregator to include this customer into the curtailments.

The results presented in Table 4 provide the following key insights. First, given the information on comfort and electricity cost, we can conclude that the lowest customer type 0.16 is curtailed more often under a curtailment contract than under RTP. This indicates that this customer is curtailed more often than it would be efficient. Second, the opposite is true for customers with higher comfort preference. Under a curtailment contract, they get

	Comfort		Electricity cost		Net utility	
Comfort preference τ	**Contract**	**RTP**	**Contract**	**RTP**	**Contract**	**RTP**
0.16	85.14	86.57	64.67	69.74	105.60	103.40
0.18	90.80	88.70	78.87	75.10	102.72	102.29
0.20	94.28	91.35	87.43	80.81	101.13	101.90
0.22	96.95	92.59	93.62	84.05	100.28	101.13
0.24	100.00	94.56	100.00	88.27	100.00	100.85

Table 4: Welfare effects across customer types

curtailed less than it would be efficient. In particular, this is true for type 0.24 which does not get curtailed at all under a curtailment contract. Finally, we find that, for the lowest type, a curtailment contract is preferable over RTP. The opposite is true for types of 0.18 and higher. However, this assumes that no other costs apply to customers, e.g. information or other transaction costs.

4.2.3 Implications for the Aggregator

In this section, we present the implications of curtailment contracts for the aggregator. We compare our findings to the results for an aggregator with the same installed capacity of uncertain generation but without flexible demand resources. First, we find that the profits of the aggregator increase by 5.0% per generation capacity unit installed. This is achieved by two effects. First, flexible demand is a valuable resource which can be additionally marketed on the wholesale market. As a result, the aggregator can increase the amount of electricity sales in the hour-ahead market by 9.6% per generation capacity unit installed. This increases the hour-ahead income. Second, the aggregator can decrease the balancing costs by 36.0%. This demonstrates that flexible demand resources can be deployed to match generation uncertainty. As a result, the average of imbalances decreases by 54.4% and the variance of imbalances by 29.4%. Ultimately, this decreases the overall imbalance in the market and increases market transparency.

Fig. 3 plots the profits from an increased procurement of flexible demand resources. While we keep the installed generation capacity of the aggregator unchanged, we vary the

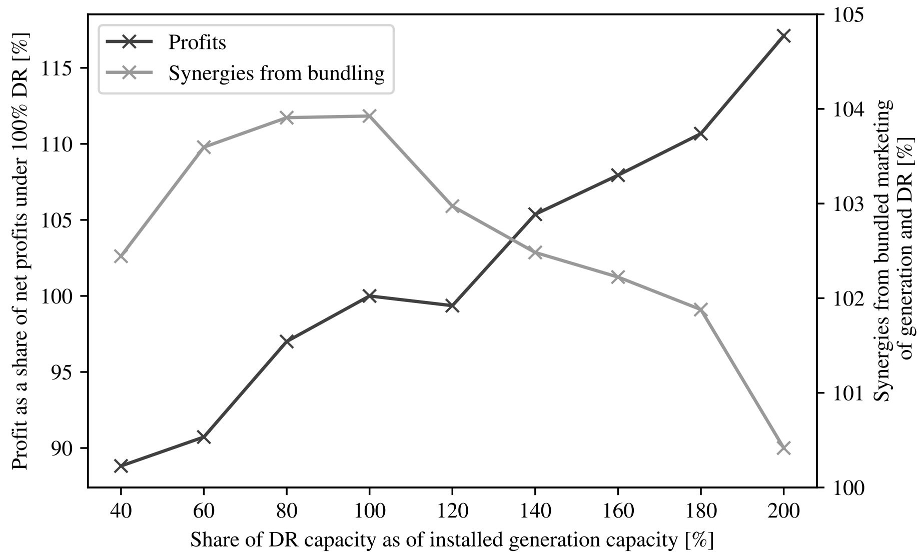

Figure 3: Synergies from bundling uncertain generation and curtailment contracts [%]

number of customers. The x axis reports the resulting share of the amount of flexible load capacity as a percentage of the installed generation. We find that profits increase with more customers signing a curtailment contract. Compared to profits with 100% flexible customers as a ratio to installed generation capacity, profits steadily increase from a level of 88.8% for 40% of flexible demand to a level of 117.1% for 200% of flexible demand. This supports our previous finding that the availability of flexible demand resources is *per se* valuable and profitable for the aggregator. To, however, enable comparability, we relate our subsequent numerical results to a situation when the aggregator markets generation and flexible load separately. Figure 3 therefore illustrates the synergies from bundling uncertain generation and flexible load. The synergies are calculated as the ratio of profits under a curtailment contract versus the sum of profits of an aggregator if both generation and flexible load were separately marketed. We find that profits reach a level of 103.9% when flexible demand and generation gets marketed in a bundle compared to a situation when they are marketed separately. Synergies of bundling are maximized if generation and flexible load match each other in a ratio of close to 1:1, i.e. if available DR reached 80% to 100% of installed generation capacity. If flexible demand is only 40%, synergies drop to 102.4%. *Vice versa*, if flexible demand reaches 200% of installed generation capacity, synergies drop to 100.4%.

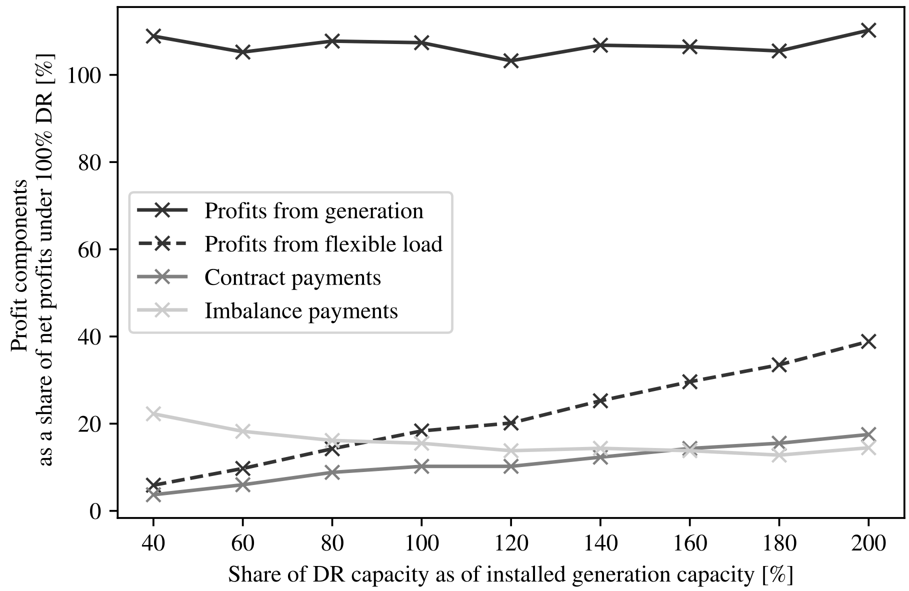

Figure 4: Profit components as a share of separately marketed profits [%]

Figure 4 details the different profit components as a share of profits under a flexible load availability of 100%. First, we find that the major value is generated by generation. The income from generation accounts, on average, for 106.8%. We calculate this 'pure' generation effect based on the results for wholesale market operations of an aggregator without access to flexible load. Additionally, the aggregator creates value from marketing flexible load and bundling it with generation. We derive this value from the difference between the total hour-ahead market income under bundling and the pure generation income without flexible demand. Our results demonstrate that, for a 40% share of flexible load of generation capacity, the flexible load contributes 5.9% to hour-ahead market income. This number increases to 38.9% if the flexible load share reaches a level of 200%. The aggregator profits from two effects: first, with more flexible demand available, more flexible load is available to market and curtail. As a results, they contribute a higher share to overall profits. Second, the available demand potential is not as easily exhausted as the temperature of individual customers has more periods to recover, even if curtailments are necessary in multiple subsequent periods. On the cost side, contract and imbalance costs apply. Aggregate contract payments increase with the share of flexible demand. For a share of 40%, contract costs account for 3.7% of profits and increase to 17.5% for a share of 200%. Imbalance cost, on the other hand, decrease with more flexible demand resources, from 22.3% to 14.5% for flexible demand shares of 40% to 200%, respectively.

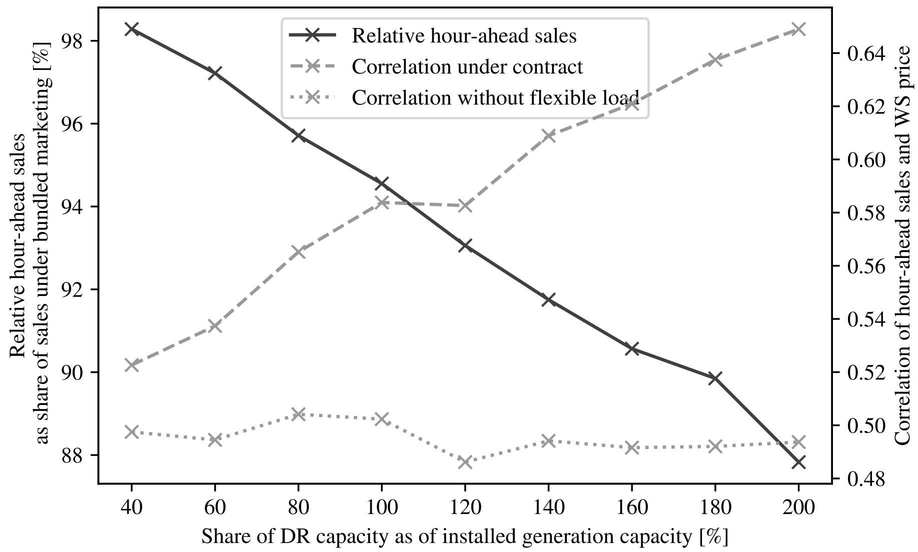

Figure 5: Hour-ahead sales per generation capacity installed

Finally, the availability of flexible load impacts the aggregator's behavior in the hour-ahead market. Figure 5 shows that the hour-ahead sales under bundled marketing are less than the sales under separate marketing of generation and flexible demand. Specifically, the relative hour-ahead sales decrease from a level of 98.3% for a 40% share of flexible load level to 87.8% for a 200% share. In combination with the decreased imbalances under a curtailment contract, this indicates that curtailment contracts decrease the need for risk taking in the hour-ahead market. On the other hand, as also illustrated by Figure 5, the bundling of uncertain generation with flexible demand resources increases the correlation of sales and prices. While, without flexible demand and curtailment contracts, the average correlation of sales and prices is 0.50, the availability of flexible load increases the correlation slightly to 0.52 for 40% of flexible resources and to even 0.65 for 200% of flexible resources. This illustrates that, with the help of flexible demand resources, sales can be increasingly decoupled from the availability of uncertain generation. As a consequence, while less electricity is sold than under separate marketing, the aggregator can take advantage of higher prices, increasing overall income from the hour-ahead market.

4.3 Dependency on external parameters

The following section illustrates the dependency of the results on external parameters. Fig. 6 plots the synergies of bundling for different degrees of auto-correlation of uncertain generation, i.e. $Prob(s_t = s_{t-1})$. First, we find that bundled marketing of uncertain generation and flexible load resources has a consistent profit advantage over separate marketing. Furthermore, the synergies reach the highest level of 112.1% for the lowest auto-correlation of 0.5 and the lowest synergies of 100.6% for the highest auto-correlation of 0.9. This finding indicates that curtailment contracts are particularly valuable in situations when real-time generation is more volatile and less predictable.

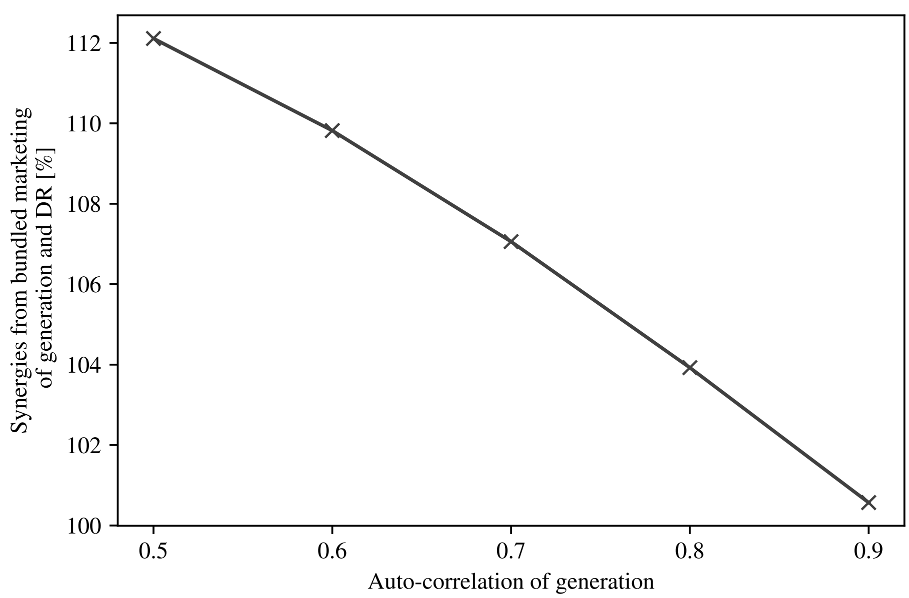

Figure 6: Profit advantage of curtailment contracts as a function of auto-correlation of uncertain generation

Similarly, Fig. 7 documents the synergies under a curtailment contract for different imbalance prices b, as a ratio of the imbalance price in the default scenario. The advantage of curtailment contracts is persistent over varying imbalance prices. Synergies are highest for high imbalance costs. For a 20% increase, the synergies of bundling reach a level of 107.4%. They descrease for lower imbalace cost and reach synergies of 101.1% for a 20% reduction in imbalance payments. Our result demonstrates that, if balancing is costly, the relative value of contracts increases as they provide an increasingly attractive alternative to settlement payments.

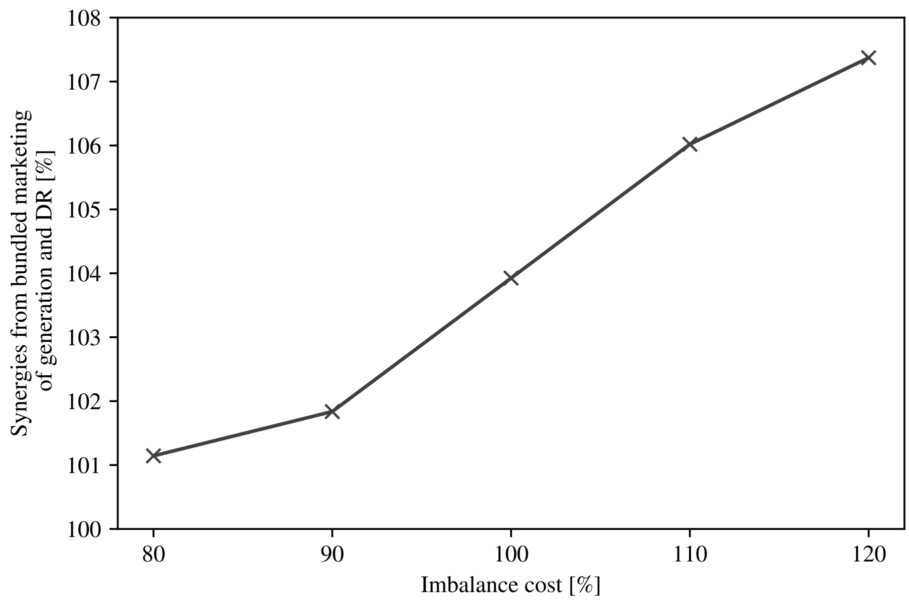

Figure 7: Synergies from curtailment contracts as a function of the relative imbalance price b

Finally, we investigate the impact of the price variance. Fig. 8 shows that the synergies of bundling generation and flexible load are largely higher than 100%. However, the synergies from bundling decrease with an increasing price variance and even drop below 100% for a price variance of 4 or higher. This fact likely reflects shortcomings in the myopic optimization solution of the aggregator as introduced in Section 3.2 as the theoretical lower bound for synergies is 100%: the unbundled solution can always be implemented under bundling and must therefore be at least as good as the solution under separate marketing. This finding indicates that the results of this section can be understood as a lower bound to the theoretically possible synergies and profit increases if the aggregator adopts an optimized approach to curtailment cost estimation.

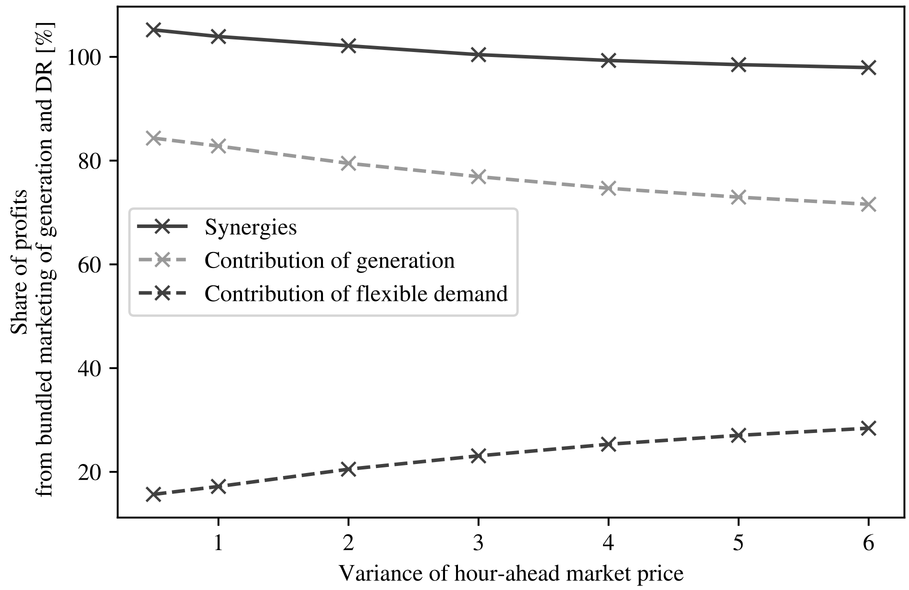

Figure 8: Synergies of curtailment contracts as a function of price variance

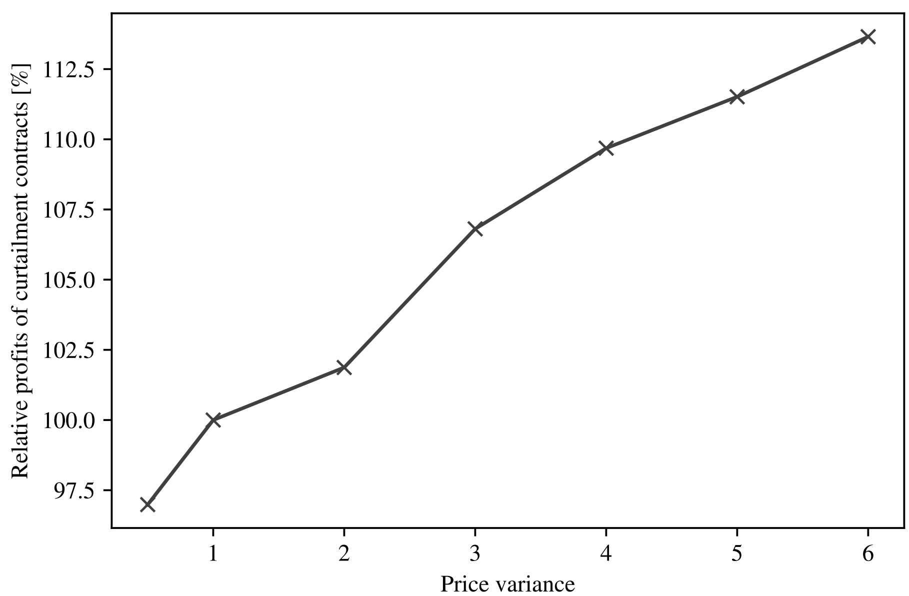

Figure 9: Profits of curtailment contracts as a function of price variance, relative to default parametrization

Fig. 9 furthermore shows the impact of an increasing price variance on the relative profit, as a ratio of the profit under a price variance of 1. We find that the relative profits are increasing with increasing price variance. The explanation is as follows: if the AR(1) process of the price has a low variance, prices are likely to be sticky and remain at high or low price levels for an extended time. For instance, during a high price period, the flexible load potential becomes exhausted (reflected by a temperature increase of customers) and the costs of curtailment increase. As a result, the aggregator cannot take advantage of the high price level anymore and generates less sales. Instead, if prices are more variable, the hour-ahead market prices are less sticky, stay at a high price level for a shorter amount of time but return to that high price level more frequently. As a result, the flexible demand potential can recover and HVAC systems can more often be used for balancing. As a consequence, hour-ahead sales increase by 9.1% if the price variance increases from 1 to 6. This observation is confirmed by Fig. 8 which shows that the importance of flexible load as a share of profits under separate marketing grows when the price variance increases.

5 Conclusion

In our work, we built upon the above mentioned contract literature and design multi-period curtailment contracts, specify them for users of HVAC systems with heterogeneous comfort preferences for temperature, and analyze the implications for customers and aggregators. In the following, we discuss our contributions, reflect on managerial and policy implications, and name opportunities for future research.

5.1 Discussion of Contributions

Modeling of curtailment contracts. In this work, we propose curtailment contracts for longer contract durations, e.g. for a month. Previous work has either abstracted from time-interdependencies altogether (e.g. Campaigne and Oren, 2016) or considered time-interdependencies for individual jobs (e.g. Bitar and Xu, 2017). These approaches, however, disregard the fact that electricity services are usually requested by customers on a repeated basis. This fact as well as information costs are probable reasons that, in practice, customers prefer medium- to long-term contracts. In our numerical experiments, we explore a contract duration of 24 hours, however, arbitrary contract durations are generally possible. The general multi-period curtailment contract framework can furthermore be extended to other services and appliances, for instance electric vehicle charging.

Second, we derive the flexibility potential endogenously from customers' utility and comfort from HVAC operations. HVAC systems are a major electricity load for residential and commercial customers and, due to the thermal storage of houses, provide opportunities for flexible operations. Previous work such as Campaigne and Oren (2016) relied on a generic valuation for electricity provision and have abstained from a specification for appliances. With our work, we also acknowledge that customers do not receive value from electricity consumption *per se* but from the service which is provided by it. Furthermore, based on our utility model for customer comfort from HVAC operations, we can endogenously derive customers' decisions for curtailment contracts of different qualities of services. This extends previous work which has taken the value of services or flexible load constraints as given (e.g. Campaigne and Oren, 2016; Bitar and Xu, 2017). We find that customers with a low comfort preference prefer curtailment contracts which exercise more frequent curtailments and are associated with a higher *ex ante* contract payment.

Third, we show that, under the given assumptions, an incentive-compatible and individually rational contract menu is feasible. Previous work has showed that curtailment contracts or quality-of-service contracts were incentive-compatible (Bitar and Xu, 2017), however, we are the first to model the customers' trade-off between the comfort provided by an electricity-based service and the associated contract payments. We demonstrate that, as a result, customers with a low comfort preference receive a high information rent to maintain incentive compatibility.

Implications for customer surplus and the profits of the aggregator. Our results demonstrate that curtailment contracts have positive implications for customer surplus. Customers which sign a non-zero contract experience a higher welfare than under firm supply with fixed retail rates. Despite curtailments leading to an increase in average temperature by 1.4% and decreased comfort, customers are compensated by *ex ante* contract payments, increasing their average net utility by 1.9% above the utility under firm supply. Net utility is highest for customers with the lowest comfort preference and corresponds to an increase of 7.9% compared to firm supply. This finding is in line with the predictions of the general literature on mechanism design. Additionally, we compare the results to the dispatch of customers under an RTP. To our knowledge, previous work has abstained from such a comparison. We find that lower type customers have higher net utility under curtailment contracts, and higher customer types under an RTP. We find that customers with a low comfort preference are curtailed more often under a curtailment contract than it would be efficient (e.g. under an RTP). The opposite is true for high types. This conclusion can change if potential additional costs such as information or transaction costs are considered.

Second, we find that the aggregator can increase his profits by 5% per unit of generation capacity installed if he includes flexible load into his portfolio. Moreover, bundling of uncertain generation with flexible load leads to positive synergies of up to 3.9% for the aggregator, compared to the separate marketing of both resources. This estimate is a lower bound for the potential of curtailment contracts as our numerical solution approach underestimates the cost of curtailment for the aggregator. The highest synergies can be achieved if the matching of flexible load and uncertain generation is 80% to 100%. Our analysis shows that the benefits of bundling are driven by reduced imbalance payments and a limited decoupling of uncertain generation and hour-ahead sales through the flexible demand resource. As a result, the aggregator can sell more at the hour-ahead market when prices are high.

Finally, we show that the value of curtailment contracts depends on the parameters of the system. We find that contracts are even more profitable and realize bundling synergies of up to 10% if uncertain generation is more difficult to forecast or imbalance costs are high. This reinforces the opportunities for avoiding imbalances and reducing settlement costs provided by flexible load.

5.2 Managerial and Policy Implications

Our work furthermore provides valuable insights for management and policy makers. First, curtailment contracts can be a valuable means to integrate demand flexibility. Especially on a residential level, demand is not sufficiently integrated and usually does not react flexibly to changing market and system conditions. Load aggregation can contribute to the solution because of several advantages. First, aggregators have better access to information and forecasting tools and are in a better position to optimally schedule dispatch in a volatile market environment. Second, by aggregating multiple loads, aggregators have access to more profitable value streams like wholesale or ancillary services markets which are not open to individual customers. Third, multi-period curtailment contracts can potentially overcome information costs and behavioral biases of customers. Curtailment contracts offer a fixed *ex ante* contract payment and require a single decision (sign/not sign). Therefore, from a policy perspective, existing regulation should be reviewed and obstacles for load aggregation removed to enable innovative and effective business models in this field.

Second, curtailment contracts increase the profits of aggregators and reduce imbalances. From a managerial perspective, aggregators should therefore investigate the bundling of uncertain generation with flexible load. Moreover, our results show that the bundling of both resources can create additional value as opposed to separate marketing by increasing sales in high-price periods and reducing balancing costs. From a policy perspective, reduced imbalances make forward markets more transparent (because the real-time allocation is more likely to coincide with the market result) and simplify the settlement process. However, the curtailment decisions of the aggregator interfere with the balancing efforts of the retailer of firm supply. Policy makers must therefore work on a market framework which aligns the incentives of the actors involved and reduces potential externalities.

Finally, from a managerial point of view, economic theory and mechanism design can be a powerful tool to differentiate between customers with low and high flexibility. Aggregators can furthermore use data collected during operations to tailor curtailment contracts according to the preferences of customers such as time-dependent temperature settings. Furthermore, data can help to inform contract design to incorporate further heterogeneity, e.g. with regard to the technical parameters of the HVAC system, the thermal characteristics of the house, and temperature uncertainty. However, while such an approach might enable a near-optimal deployment of flexible demand, it can lead to a reduction in customer surplus and have other adverse implications, e.g. on data privacy. Therefore, policy makers should investigate how reasonable data access can be structured and competition in the field enabled.

5.3 Future Research Opportunities

While we could demonstrate the benefits of curtailment contracts to both customers and aggregators, there are manifold opportunities for future research: first, future research should address the complexity of parameters present in real-world contract design, including different appliances and energy services, preferences, technical parameters of HVAC systems, and uncertainty. Such approaches can include more complex theoretical models and the use of data-driven tools for contract optimization. Furthermore, future research should investigate more complex market environments, e.g. with regard to a dynamic customer base and competition. The latter will also help regulators to evaluate if interventions by regulators are needed to ensure that customers enjoy a surplus and their data is protected. Finally, field experiments can be deployed to validate our business model as well as the underlying assumptions. The insights collected there can inform improvements to our proposal.

References

Bitar E, Xu Y (2017) Deadline Differentiated Pricing of Deferrable Electric Loads. *IEEE Transactions on Smart Grid* 8(1):13–25.

Bohn RE, Caramanis MC, Schweppe FC (1984) Optimal Pricing in Electrical Networks over Space and Time. *The RAND Journal of Economics* 15(3):360–376.

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

Campaigne C, Oren SS (2016) Firming renewable power with demand response: an end-to-end aggregator business model. *Journal of Regulatory Economics* 50(1):1–37.

Chao R Hung po; Wilson (1987) Priority Service: Pricing, Investment, and Market Organization. *The American Economic Review* 77(5):899–916.

Cho IK, Meyn S (2010) Efficiency and marginal cost pricing in dynamic competitive markets with friction. *Theoretical Economics* 5(2010):215–239.

Fanger P (1970) *Thermal comfort. Analysis and applications in environmental engineering.* (Copenhagen: Danish Technical Press).

Faruqui A, Sergici S (2010) Household response to dynamic pricing of electricity: A survey of 15 experiments. *Journal of Regulatory Economics* 38(2):193–225.

IRENA (2018) Power system flexibility for the energy transition. Technical report.

Mathieu JL, Koch S, Callaway DS (2013) State estimation and control of electric loads to manage real-time energy imbalance. *IEEE Transactions on Power Systems* 28(1):430–440.

Milgrom P, Segal I (2002) Envelope Theorems for Arbitrary Choice Sets. *Econometrica* 70(2):583–601.

Nayyar A, Negrete-Pincetic M, Poolla K, Varaiya P (2016) Duration-differentiated energy services. *IEEE Transactions on Control of Network Systems* 3(2):182 – 191.

O'Brien G, Rajagopal R (2013) A method for automatically scheduling notified deferrable loads. *Proceedings of the 2013 American Control Conference (2013 ACC)* 5080–5085.

Papavasiliou A, Oren SS (2010) Supplying renewable energy to deferrable loads: Algorithms and economic analysis. *Proceedings of the 2010 General Meeting of the IEEE Power & Energy Society (2010 IEEE PES)* .

Roozbehani M, Dahleh M, Mitter S (2010) On the stability of wholesale electricity markets under real-time pricing. *Proceedings of the 2010 IEEE Conference on Decision and Control (2010 IEEE CDC)* 1911–1918.

Schneider I, Sunstein CR (2017) Behavioral considerations for effective time-varying electricity prices. *Behavioural Public Policy* 1(2):219–251.

Subramanian A, Garcia MJ, Callaway DS, Poolla K, Varaiya P (2013) Real-time scheduling of distributed resources. *IEEE Transactions on Smart Grid* 4(4):2122–2130.

Tan CW, Varaiya P (1993) Interruptible electric power service contracts. *Journal of Economic Dynamics and Control* 17(3):495–517.

Yu Z, Jia L, Murphy-Hoye MC, Pratt A, Tong L (2013) Modeling and Stochastic Control for Home Energy Management. *IEEE Transactions on Smart Grid* 4(4):2244–2255.

6 Proofs

Proof 6 (Theorem 2.2) *The original formulation of the envelope theorem by Milgrom and Segal (2002) is* $V(t) = V(0) + \int_0^t f_t(x^*(s), s)ds$. *In our setting with comfort losses, the incentive-compatibility constraints of the optimization problem bind into the other direction,* *i.e. customers with a lower comfort preference have an incentive to select contracts intended for customers with higher comfort preference. Incentive-compatibility therefore requires that customers with lower comfort parameters receive higher information rents and the incentive-compatible contract menu binds 'at the top', i.e. for customer $\overline{\tau}$ instead of $\underline{\tau}$.*

Furthermore, we can substitute the first derivative of net utility $\frac{\partial}{\partial \tau}\mathbb{E}\, u(\tau|\tau')$ by the first derivative of comfort $\frac{\partial}{\partial \tau}\mathbb{E}\, D(\tau|\tau')$ for the following reason,

$$
\begin{aligned}
\frac{\partial}{\partial \tau}\mathbb{E}\, u(\tau|\tau') &= \frac{\partial}{\partial \tau}\{\mathbb{E}\, D(\tau|\tau') + f(\tau')\} \\
&= \frac{\partial}{\partial \tau}\mathbb{E}\, D(\tau|\tau') \quad \blacksquare
\end{aligned} \tag{21}
$$

Proof 7 (Theorem 2.5) *First, we specify the comfort $D(\tau|\tau'))$ by the optimal temperature path under each state of the world ω_{RT} and take the first derivative with respect to τ,*

$$
\begin{aligned}
\frac{\partial}{\partial \tau}\mathbb{E}\, D(\tau|\tau') &= \frac{\partial}{\partial \tau}\int_{\Omega_{RT}} h(\omega_{RT}) \cdot D(\tau|\tau'))d\omega_{RT} \\
&= \frac{\partial}{\partial \tau}\int_{\Omega_{RT}} h(\omega_{RT}) \cdot (\sum_{t=0}^{T-1}\{-\tau|\theta_t - \theta^{set}|^2\} - \frac{\tau}{1-\beta^2}|\theta_T - \theta^{set}|^2)d\omega_{RT} \\
&= -\int_{\Omega_{RT}} h(\omega_{RT}) \cdot (\sum_{t=0}^{T-1}\{|\theta_t - \theta^{set}|^2\} + \frac{1}{1-\beta^2}|\theta_T - \theta^{set}|^2)d\omega_{RT}
\end{aligned} \tag{22}
$$

The expression demonstrates that the first derivative of expected comfort with respect to τ decreases in the temperature θ_t on the interval $[\theta^{set}, \theta^{out}]$, i.e. $\frac{d}{d\theta_t}\frac{\partial}{\partial \tau}\mathbb{E}\, D(\tau|\tau') < 0, \forall t \in \{0, ..., T\}$. Therefore, the expression is bounded. $\frac{\partial}{\partial \tau}\mathbb{E}\, u(\tau, \tau')$ has an upper bound of 0 at

$\theta_t = \theta_T = \theta^{set}\ \forall t$. *It furthermore has a lower bound at* $\theta_t = \theta_T = \theta^{out}\ \forall t$,

$$\begin{aligned}
\frac{\partial}{\partial \tau}\,\mathbb{E}\,u(\tau,\tau') &= -\,\mathbb{E}(\sum_{t=0}^{T-1}\{|\theta^{out}-\theta^{set}|^2\} + \frac{1}{1-\beta^2}|\theta^{out}-\theta^{set}|^2) \\
&= -\,\mathbb{E}(T|\theta^{out}-\theta^{set}|^2 + \frac{1}{1-\beta^2}|\theta^{out}-\theta^{set}|^2) \\
&= -\int_{\Omega_{RT}} h(\omega_{RT})\cdot(T+\frac{1}{1-\beta^2})|\theta^{out}-\theta^{set}|^2)d\omega_{RT} \\
&= -(T+\frac{1}{1-\beta^2})|\theta^{out}-\theta^{set}|^2)\int_{\Omega_{RT}} h(\omega_{RT})\cdot d\omega_{RT} \\
&= -(T+\frac{1}{1-\beta^2})|\theta^{out}-\theta^{set}|^2 \quad \blacksquare
\end{aligned} \tag{23}$$

Proof 8 (Theorem 2.6) *First, we reformulate the expression for the aggregate contract payment given in Theorem 2.6,*

$$\begin{aligned}
F &= \int_{\underline{\tau}}^{\overline{\tau}} g(\tau)f(\tau)d\tau \qquad (24)\\
&= -\int_{\underline{\tau}}^{\widetilde{\tau}} g(\tau)\{\mathbb{E}\,D(\tau'|z)|_{\tau'=z} + \int_{\tau}^{\widetilde{\tau}} \frac{\partial}{\partial z}\,\mathbb{E}\,D(\tau'|z)|_{\tau'=z}dz\}d\tau \\
&= -[\int_{\underline{\tau}}^{\widetilde{\tau}} g(\tau)\,\mathbb{E}\,D(\tau'|z)|_{\tau'=z}d\tau + \int_{\underline{\tau}}^{\widetilde{\tau}} g(\tau)\int_{\tau}^{\widetilde{\tau}} \frac{\partial}{\partial s}\,\mathbb{E}\,D(\tau'|z)|_{\tau'=z}dz\}d\tau].
\end{aligned}$$

Then, we analyze the two parts of the sum separately. The derivative of the first part of the sum is,

$$\begin{aligned}
\frac{dF_1}{d\widetilde{\tau}} &= -\int_{\underline{\tau}}^{\widetilde{\tau}} g(\tau)\,\mathbb{E}\,D(\tau'|z)|_{\tau'=z}d\tau \\
&= -g(\widetilde{\tau})\,\mathbb{E}\,D(\widetilde{\tau}).
\end{aligned} \tag{25}$$

Furthermore, we can re-formulate the second sum of the expression as follows, using inte-

gration by parts,

$$
\begin{aligned}
F_2 &= -\int_{\underline{\tau}}^{\widetilde{\tau}} g(\tau) \int_{\tau}^{\widetilde{\tau}} \frac{\partial}{\partial z} \mathbb{E}\, D(\tau'|z)|_{\tau'=z} dz\} d\tau \\
&= -\int_{\underline{\tau}}^{\widetilde{\tau}} g(\tau)\{\mathbb{E}\, D(\widetilde{\tau}) - \mathbb{E}\, D(\tau)\} d\tau \\
&= -[\mathbb{E}\, D(\widetilde{\tau}) \int_{\underline{\tau}}^{\widetilde{\tau}} g(\tau) d\tau - \int_{\underline{\tau}}^{\widetilde{\tau}} g(\tau)\, \mathbb{E}\, D(\tau) d\tau] \\
&= -[\mathbb{E}\, D(\widetilde{\tau}) \int_{\underline{\tau}}^{\widetilde{\tau}} g(\tau) d\tau - \int_{\underline{\tau}}^{\widetilde{\tau}} g(\tau)\, \mathbb{E}\, D(\tau) d\tau] \\
&= -[\mathbb{E}\, D(\widetilde{\tau}) G(\widetilde{\tau}) - \int_{\underline{\tau}}^{\widetilde{\tau}} g(\tau)\, \mathbb{E}\, D(\tau) d\tau]. \qquad (26)
\end{aligned}
$$

The cumulative distribution function $G(\tau)$ is zero for the type with the lowest comfort preference, i.e. $G(\underline{\tau}) = 0$. Taking the first derivative with respect to τ gives,

$$
\frac{dF_2}{d\widetilde{\tau}} = -[\frac{\partial}{\partial \widetilde{\tau}} \mathbb{E}\, D(\widetilde{\tau}) G(\widetilde{\tau}) + \mathbb{E}\, D(\widetilde{\tau}) g(\widetilde{\tau}) - g(\tau)\, \mathbb{E}\, D(\widetilde{\tau})]. \qquad (27)
$$

Therefore, the marginal effect of curtailment on the aggregate contract payments F is,

$$
\begin{aligned}
\frac{dF}{d\widetilde{\tau}} &= -g(\widetilde{\tau})\, \mathbb{E}\, D(\widetilde{\tau}) - [\frac{\partial}{\partial \widetilde{\tau}} \mathbb{E}\, D(\widetilde{\tau}) G(\widetilde{\tau}) + \mathbb{E}\, D(\widetilde{\tau}) g(\widetilde{\tau}) - g(\tau)\, \mathbb{E}\, D(\widetilde{\tau})] \\
&= -(\frac{\partial}{\partial \widetilde{\tau}} \mathbb{E}\, D(\widetilde{\tau}) G(\widetilde{\tau}) + g(\tau)\, \mathbb{E}\, D(\widetilde{\tau})) \quad \blacksquare \qquad (28)
\end{aligned}
$$