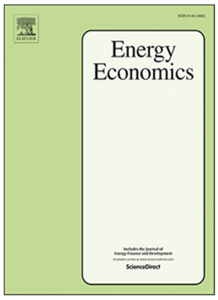


Identifying optimal capacity expansion and differentiated capacity payments under risk aversion and market power: A financial Stackelberg game approach

Maxim Bichuch, Benjamin F. Hobbs, Xinyue Song

PII: S0140-9883(23)00065-8
DOI: https://doi.org/10.1016/j.eneco.2023.106567
Reference: ENEECO 106567

To appear in: *Energy Economics*

Received date : 4 January 2022
Revised date : 9 January 2023
Accepted date : 2 February 2023

# Identifying Optimal Capacity Expansion and Differentiated Capacity Payments under Risk Aversion and Market Power: A Financial Stackelberg Game Approach

Maxim Bichuch[a] Benjamin F. Hobbs[b] Xinyue Song[c]

February 17, 2023

**Abstract**

We investigate how capacity payments in combination with scarcity pricing of energy can ensure resource adequacy in electricity markets, defined as the ability of supply and other resources to provide enough energy and capacity to meet demand under steady-state operating conditions. This work generalizes models for determining capacity payments by deriving second-best discriminatory payments by resource type that account not only for the "missing money" market failure that arises from energy price caps, but also for market power in the capacity market and differences in risk tolerance among resource types that can arise from failures in risk and capital markets. A bi-level equilibrium-constrained optimization model is proposed to define second-best capacity payments in a static long-run setting, considering the impacts of those payments on the mix and cost of generation investment and energy outputs. The lower-level suppliers play a Nash game to determine the generation mix under a capacity payment scheme, while the upper-level regulator considers consumer welfare and resource adequacy. We introduce an equivalent formulation via a variational inequality approach, and find conditions for the solution to exist. Discriminatory payments are found to be second-best when there is market power in the investment game, price caps in energy markets and imperfections in risk markets that lead to diverse risk attitudes.

**Keywords:** Electricity, Capacity Payment, Resource Adequacy, Reliability, Nash Equilibrium, Variational Inequality, Mathematical Program with Equilibrium Constraints (MPEC)

**JEL Classification codes:** C6, C7, Q4

**Highlights:**

- Bi-level decision model for second-best capacity payments with reliability constraint
- Proof of capacity planning solvability with price cap, risk aversion, and market power
- Numerical examples of second-best capacity solutions with demand uncertainty

[a]Department of Applied Mathematics & Statistics, Johns Hopkins University, Baltimore, MD 21218, USA; email: mbichuch@jhu.edu.

[b]Department of Environmental Health & Engineering, Johns Hopkins University, Baltimore, MD 21218, USA; email: bhobbs@jhu.edu.

[c]Department of Applied Mathematics & Statistics, Johns Hopkins University, Baltimore, MD 21218, USA; email: xsong11@jhu.edu.

Declarations of interest: none.

Partial support provieded by NSF grant DMS-1736414. The work was presented at INFORMS 2022, Indianapolis. We thank an anonymous reviewer for helpful comments.

# 1 Introduction

Electric power markets around the world are confronted with environmental mandates, rapidly changing generation mixes, and extreme environmental events that challenge system reliability. Since the establishment of spot power markets in the 1980s, a recurring theoretical and empirical question has been: are energy and operating reserve spot prices sufficient to incentivize sufficient investment in generation capacity, storage, and demand-side management in order to meet demand with an acceptably low probability of shortfalls under steady-state operating conditions? This is the well-known challenge of *resource adequacy*, which is distinguished from the separate but also crucial challenge of *system security*, which considers the dynamic ability of a system to withstand disturbances without interrupting supply.

The theory of spot pricing, as developed in [15], proved theorems that indicate that spot prices can be sufficient to support a net market surplus-maximizing amount of investment not only in generation, but also in transmission. Economists, power engineers, and policy makers have identified several reasons, however, why this first-best solution to the resource adequacy problem is unlikely to hold in practice (e.g., [21]). One reason is the market failure of "missing money", in which regulatory price caps or other limitations prevent energy prices from reaching high enough levels in enough hours to cover the cost of peaking capacity. For instance, if peaking plants have an annual cost of \$100,000/MW/year,[1] and a reliability target of 2.4 hours/year of shortage is sought (which is one interpretation of the industry's common standard of one-day-in-10-years), then gross margins have to be several tens of thousands of dollars per MWh during periods of scarcity, while price caps are typically an order of magnitude smaller than that. Another reason is the failure due to lack of market participation by consumers in price formation. Obstacles to their participation include missing or distorted price signals and inability to respond to those signals. In particular, there is a lack of metering equipment and retail pricing policies that would make real-time variations in system marginal cost visible to consumers, and unavailability of control equipment that would facilitate demand adjustments in response to those variations. Some markets attempt to substitute administrative scarcity pricing mechanisms for consumer response in order to set high prices to signal scarcity conditions. But such mechanisms may yield prices that are only remotely related to consumer marginal-willingness-to-pay for power during shortages, and they do not promote efficient consumer management of shortages. A third reason is missing markets for multi-year or -decadal contracts or other hedging instruments that risk-averse resource owners would prefer when financing construction of facilities that could have 40-year lifetimes.

In response, many – but not all – power markets have instituted markets or other mechanisms that provide payments to installed generation capacity in addition to spot market revenues. These so-called "resource adequacy" mechanisms take various forms. The literature (e.g., [2][12][20][22][30]) gives a full account of key design features of capacity markets, including capacity price formation, contract duration, adjustments for capacity reliability during times of system stress, incentives to improve capacity reliability, and relationships to power price caps and net margins. Resource adequacy mechanisms have been subjected to many theoretical analyses to understand under what conditions they can provide incentives that support efficient amounts, mixes, and locations of investment in generation, storage, and demand-side management. There has been some success in Europe using energy-only markets without supplementary capacity markets. Various reasons for this have been suggested. For instance, the missing money problem is identified as mostly misallocation of money in [33] which can be solved via energy price correction policies, and capacity renumeration is believed to cause problems such as over-procurement ([54]). Existing

[1] As an example, we estimate the investment cost of a combustion turbine peaking plant to be \$102.1/kW/year according to data from US Energy Information Administration, see Table 1, Footnote [5].

reliability procurement practices include demand-side management ([48]), cross-border trade ([9]), and strategic reserves ([53]), among others.

In many regions including the US, capacity payments are one of the prevailing methods of regulatory intervention to encourage capacity adequacy, where capacity owners receive a payment per MW of installed capacity. There are many designs for such payment-based mechanisms. Basic forms include fixed payment schedules that are announced by regulators, and another is market prices that emerge from centralized or bilateral markets for tradable capacity obligations in which entities that contract to serve demand are obliged to acquire a certain amount of rights from suppliers of those rights based on the peak demands of their consumers ([62]). Such payments, according to [55], act as a form of insurance to cover the risk of high spot prices resulting from insufficient generation, and can be supplemented by a secondary market to offer correct signals to producers. Recognizing that capacity mechanisms are in part motivated by a desire to insure consumers against price risks and market power ([32]), we study how second-best capacity payments can be defined in a basic electricity spot market that also includes the complications of demand-side uncertainty, supplier market power, and risk aversion.

The basic question of interest is how capacity payments can be designed in a second-best world to provide market signals that promote optimal mixes of technology investment that guarantee reliability, defined as a target probability that demand has to be involuntarily curtailed (loss-of-load-probability). To answer that question, it is necessary to address the related question about how generators decide their optimal capacity, given a capacity payment scheme, which is typically answered in a game setting. In [49], the optimal capacity mix was investigated in a risk-neutral world with a fixed quantity being demanded ("load", in power system parlance), and the competitive equilibrium was verified to be indeed welfare-maximizing and affected by spot price caps. Meanwhile, [3][60] took risk aversion into consideration, where the authors sought a Nash-Bertrand equilibrium among risk-averse generators. They related the market price to production decisions by clearing supply against a demand function, but depending on the rationing scheme adopted, market clearing may fail in the presence of load shedding. In addition, we add another source of price randomness by considering load uncertainty. Such additional stochasticity was considered in [56], where the stochastic Nash equilibrium was solved by a two-stage approximation algorithm in a very general context. Here, we explore the exact solutions arising in the induced closed loop among producers, in order to specifically address the equilibrium capacity investment for risk-averse players with market power in an electricity market combining price-capped energy markets with capacity payments.

Then we investigate our basic question by introducing a regulator or market designer which sets payments to optimize net market benefits. One approach to analyzing capacity payment-based systems and defining optimal payments was conceptualized in [62] for dispatchable, perfectly reliable capacity that is not subject to intertemporal constraints (such as limited energy storage or ramp limits) but can be built in continuous amounts. If an energy price cap is imposed, the probability distribution of the load is known and static, and suppliers are risk-neutral, then the equilibrium capacity price in a competitive market would yield the same least-cost solution as the ideal uncapped pure energy market described in [15]. This load-duration curve approach conceptualized in [62] has been elaborated upon in later work to consider random forced outages of thermal capacity [32], and storage together with variable renewable generation [17]. In related work, [35] took a classical economic view to study how capacity payments can be used to ensure a target level of reliability.

More sophisticated modeling approaches have addressed the dynamics of investment in a non-steady state environment (for a general treatment of such problems, see [6]). More specifically, analytical approaches have been used to derive general results concerning the properties of capacity mechanisms in dynamic environments for simplified market settings. For example, [16] obtained analytical results in a multi-stage Nash setting with either price-elastic or inelastic demand. For more complicated and realistic dynamic settings, agent-based approaches have provided insights. For instance, [31][57] used a dynamic agent-based simulation method to help the US PJM market to design its "reliability pricing model", which is based on clearing capacity offers against an administrative demand curve for capacity. As in the present paper, agents are risk-averse, but unlike here, agents use heuristics to determine their levels of investment. An agent-based model has also been used to explore impacts of cross-border capacity limits, demand-response, and storage on capacity markets in Central Europe [13][36]. A highly comprehensive and parameterized model to compute optimal capacity payments was introduced in [8], where the capacity decisions influence the game through a set of average prices. However, the cornerstone of this model was that all suppliers have equal profitability at equilibrium in a risk-neutral world, while in our analysis we introduce heterogeneous risk attitudes and diverse technologies, albeit in a static long-run model rather than a dynamic multi-stage setting.

Another distinction of our analysis relative to the above analyses is that we consider the exercise of market power through strategic decisions about how much capacity to install in the investment stage and make available to the spot market. The capacity is then operated in competitive spot markets (such as the Bertrand spot markets in [39] or the regulated cost-based markets in [64]). Such markets exist, as an example, in South America where capacity investments are unregulated but spot market offers must be cost-based [51]. Meanwhile, stringent market power mitigation procedures in the US, since the 2000-01 California crisis, have made short-run marginal cost pricing a reasonable approximation in spot markets under normal conditions. (For instance, the California system's market monitor has concluded that spot prices in its markets have generally been at competitive (i.e., short-run marginal cost) levels in recent years, even during the shortage conditions of August 2020, although tightening market conditions may change that in the future [1].) In contrast, Nash equilibrium-based analyses by, for example, [16], assume that there are no regulatory constraints on the exercise of market power in spot markets.

Central to our calculation of equilibria are dynamic spot prices and risk-averse utility maximization by market parties. In [66], similar concepts were used to compute oligopoly strategies on a related question: how should a firm invest on a continuum of technologies to maximize profit? In absence of risk aversion and capacity payments, the author found the suppliers (firms) have incentives to overinvest in baseload technologies but the total capacity suffers from underinvestment, which tells a story that we hope to tell from another point of view. Due to market power, the spot market alone is insufficient to provide incentives for enough capacity to guarantee reliability and the system regulator's interference is necessary (which comes as capacity payments in our story). In [46], a model with two technologies and uniform load was considered to answer the same question, and to show how the cost structure of a technology influences investment. (The author also incorporated production cost risk and suppliers' risk aversion in [47].) In our discussion, we analyze a technology's cost efficiency in a similar manner and show its role in the second-best dispatch, while adding the complications of nonstationary and random demand.

There has also been a good deal of recent attention paid to how rapidly growing renewable energy penetration impacts capacity payment designs and vice versa, e.g., [10][28]. But variable renewable resources will not be our focus. Instead, we are interested in using a simply structured benchmark market to address the problem of optimal (second-best) capacity prices in the presence of three market failures: heterogeneity of risk attitudes among market participants in the absence of efficient markets for risk; caps on spot energy prices; and market power in the capacity market. We also consider how imposition of a spot price cap comes into play and affects incentives for capacity investment. The discriminatory capacity payments are set at levels to provide optimal (second-best) incentives to producers who face uncertain market demand to willingly build an adequate amount and mix of capacity, while satisfying a reliability constraint and energy price caps so that consumers are not frequently exposed to very high electricity prices or loss-of-load events.

Bi-level modeling is also central to our analysis. In general, such models have been widely used in the fields of economics (e.g., [4][52][67]), operations research (e.g., [27][65]), and finance (e.g., [7][59]) to explore questions about optimal electric utility regulation and contracting in the face of market imperfections. However, the model in this paper is the first to consider how second-best capacity payment schemes can be designed considering Nash equilibria in capacity investment games, in which risk-averse market participants with heterogeneous technologies face price risks because of demand uncertainties. To make the models tractable and facilitate analytical results, we make several simplifying assumptions concerning market structure, and treat risk aversion, cost and price parameters as inputs to the model, which a more general analysis could instead treat as endogenous. For example, system operators sometimes consider modifying energy price caps in order to improve market functioning, but current practices in the US and EU typically involve a uniform regional price cap, and in the present paper we take the price cap as fixed. Also, the investment cost and its payment schedule may depend on the riskiness and debt/asset of producer and vary across time. In [45], authors proposed a method to endogenize capital costs with consideration of revenue uncertainty, under certain assumptions about what hedging instruments are traded, and estimation of risk attitudes of market participants implied by observed generation mixes. This work is a very recent contribution to a small but very interesting literature that attempts to combine engineering-economic modeling (such as power plant operations) with models of incomplete financial markets to examine how decisions about the amount and type of power generation technology interacts with available hedging instruments (e.g., [63]). The combined analysis of power and incomplete financial markets is a very interesting problem, but it introduces an enormous amount of complications that we prefer to leave to future research.

The paper is organized as follows. Section 2 gives a full description of our bi-level optimization model for the optimal capacity payment determination and resulting long-run capacity investment equilibrium. Section 3 gives mathematical results on the existence of a solution to this model, and an equivalent semi-open-loop format of the bi-level problem, which hints at practical methods for computation. (Details on some of the proofs are contained in Appendix A.) Example numerical results are presented in Section 4 that show how second-best payments can discriminate among technology types, and how the results depend on the distribution of risk aversion, level of price caps, whether capacity payments can discriminate among capacity types, and finally whether reliability levels are allowed to be determined endogenously rather than being subject to a regulatory constraint. A summary of our results and some policy implications is provided in Section 5.

# 2 Model Description

In this work, we use the following mathematical notation: $\mathbb{R}^N$ denotes the N-dimensional Euclidean space and $\mathbb{R}^N_+$ its nonnegative quadrant (and when $N = 1$, it is omitted). For $a \in \mathbb{R}$, we denote $a^+ := \max\{a, 0\}$, $a \wedge b := \min\{a, b\}$. For set $\mathcal{S} \subset \mathbb{R}^N$, its closure is denoted by $\mathrm{cl}\,(\mathcal{S})$. Vectors are denoted by bold letters, and their entries by superscripts in brackets, e.g., a $N$-dimensional vector $\boldsymbol{x} := [x^{(1)}, \dots, x^{(N)}]$. The notation $\boldsymbol{x}_1 \geq \boldsymbol{x}_2$ means each entry in $\boldsymbol{x}_1$ are no less than the corresponding one in $\boldsymbol{x}_2$. We also denote the $(N-1)$-dimensional $\boldsymbol{x}^{(-i)} := [x^{(1)}, \dots, x^{(i-1)}, x^{(i+1)}, \dots, x^{(N)}]$. When the dimension $N$ is clear, we define $\mathbf{1}_n$ such that $\mathbf{1}_n^{(i)} = \begin{cases} 1 & \text{if } 1 \leq i \leq n \\ 0 & \text{if } n < i \leq N \end{cases}$. For an event $A$, we write $\mathbb{I}_A := \begin{cases} 1 & \text{if } A \\ 0 & \text{else} \end{cases}$.

## 2.1 Model Overview

We consider an electricity market with three types of participants: consumers, suppliers, and the system operator. The consumers have time-varying demand for power, and individually purchase electricity from the market. The suppliers, with different technologies, generate electricity and sell to consumers at the market price. Their power generation is subject to their capacity at that time, and a supplier invests in capacity that is beneficial from the point of view of its objective (expected utility, based on a risk-averse utility function). Still, the total capacity in the market may fail to supply the entire load in some hours, and consumers are then exposed to load rationing and high shortage prices. A static situation is assumed, where capacity is built at the start and is in place for the whole assessment period. For instance, in our numerical example we consider an entire year, in which case builders of capacity incur an annualized investment cost.

The System Operator (SO) is a non-profit organization that operates and clears the market, determining the second-best dispatch cost and rationing energy if supply is less than demand. As mentioned, we assume effective market power mitigation in the spot market, such that energy offers in that market reflect actual short-run marginal costs of supply (in \$/MWh), unless there is scarcity, in which case the SO sets the price at the scarcity level (in other words, merit order dispatch plus scarcity pricing). The SO also selects the level of annual capacity payment, in \$/MW/yr to provide to each type of generation capacity.

The SO is a Stackelberg leader which knows that suppliers possess potentially heterogenous risk preferences, would make investment decisions to their own benefit, and have market power in that they recognize that their investment decisions will ultimately affect spot market prices. The operator considers consumer welfare (expected utility) as its objective function. It is concerned about consumers' utility with respect to power prices and supply adequacy, and achieves this goal by offering (second-best) incentives such that the suppliers will make overall better decisions.

In this paper we model the suppliers' expansion decisions, and the SO's second-best decisions about incentives for capacity investment. In the rest of this section, we first provide a mathematical characterization of the utility maximization problems of the suppliers (which are Stackelberg followers), and then give the decision model for the SO (the Stackelberg leader).

## 2.2 Load and Supply Dispatch

We focus on the generation side, and treat the load as exogenous. A total number of $N$ consumers are grouped into $N_c$ types according to their risk structure and demand dynamics, each containing $N^{(i)}$ individuals with identical characteristics. A consumer of type $1 \le i \le N_c$, has risk aversion $\gamma_c^{(i)} > 0$ and discount factor $r_c^{(i)} > 0$. Consumer $i$'s demand evolves by the stochastic process $d_t^{(i)}$, which we assume to be normally distributed $d_t^{(i)} \sim \mathcal{N}(\mu_t^{(i)}, (\sigma_t^{(i)})^2)$, with $0 < \mu_{min}^{(i)} \le \mu_t^{(i)} \le \mu_{max}^{(i)}, 0 < \sigma_{min}^{(i)} \le \sigma_t^{(i)} \le \sigma_{max}^{(i)}$ for all $t \ge 0$. The demand may be correlated across time and consumer types. We denote the instant correlation across types by $\rho_t^{(i_1 i_2)} := \mathrm{Corr}(d_t^{(i_1)}, d_t^{(i_2)})$ and suppose without loss of generality $|\rho_t^{(i)}| < 1$ for some $t$. (Otherwise the two types can be merged into one type, with a new mean and number of individuals.) The total MW load in the market at time $t$ is $D_t := \sum_{i=1}^{N_c} d_t^{(i)} \sim \mathcal{N}(\mu_t, \sigma_t^2)$, with $\mu_t := \sum_{i=1}^{N_c} \mu_t^{(i)}$, $\sigma_t := \big(\sum_{i=1}^{N_c} \rho_t^{(l_1 l_2)} \sigma_t^{(l_1)} \sigma_t^{(l_1)}\big)^{\frac{1}{2}}$ both bounded. When $D_t$ exceeds the total existing capacity, the resulting load curtailment is assumed to be spread among all consumers according to an *ex ante* proportion $\alpha^{(i)}$, $1 \le i \le N_c$ reflecting each consumer's average demand, and normalized such that $\sum_{i=1}^{N_c} N^{(i)} \alpha^{(i)} = 1$. (We use the same simple allocation rule for assigning responsibility for capacity payments among consumers. More complex rules that account for realized loads are also possible, for example by individual demand during peak periods, but the present simple assumption contributes to the transparency of our results without changing the fundamental insights.)

The suppliers are grouped into $N_s$ types, differentiated by their investment and marginal operating costs and risk aversion. All generation types are assumed to be 100% reliable and available in each hour, which is a simplification for both thermal generators (which are subject to random forced outages in reality) and variable renewable generators (whose hourly availability depends on the wind or sunshine). A further simplification is that we consider that there is one firm of each type. More general assumptions are of course possible and desirable for actual application, but our simple case suffices to illustrate the model structure and the type of numerical results and insights it can yield. The supplier of type $1 \le j \le N_s$ has risk aversion $\gamma^{(j)} > 0$ and discount factor $r^{(j)} > 0$. They are ordered by their marginal costs of generation $0 \le G^{(1)} \le G^{(2)} \le \cdots \le G^{(N_s)} < \infty$. We denote existing capacity in the system by $\boldsymbol{K}_0 \ge \boldsymbol{0}$ (in MW), and suppliers decide the optimal amount of new capacity $\boldsymbol{K} \ge \boldsymbol{0}$, also in MW, to be added at the start of the decision period. The marginal cost for building capacity is $\boldsymbol{F} \ge \boldsymbol{0}$ (expressed as annualized cost, in \$/MW/yr). In other words, supplier $j$ has ${K_0}^{(j)}$ amount of capacity, and then at time 0, it decides to build new capacity of amount $K^{(j)}$ at a marginal cost of $F^{(j)}$. By maximizing the expected utility as a function of annualized investment cost, annualized capacity payments, and gross margins from the spot market, this static model implicitly assumes that the amount of capacity installed will be in place indefinitely, and that the operating costs and demand distributions will remain the same in subsequent years.

The generation is dispatched according to merit order: the cheapest suppliers are dispatched first to meet the demand; in case that the total capacity is insufficient, the market price is set to the scarcity price (price cap) $M$, in \$/MWh, which can be set by the SO to be equal to an estimated VOLL, or at lower levels, as in most US markets. Without loss of generality, we assume $M > G^{(N_s)}$, otherwise the suppliers with marginal costs of at least $M$ will never build capacity, and can be removed from the problem. The marginal supplier, i.e., the most expensive supplier being dispatched, is given by the index function $J(D_t; \boldsymbol{K}) := \min\{j \in \{1, \ldots, N_s\} : \mathbf{1}_j^T(\boldsymbol{K}_0 + \boldsymbol{K}) > D_t\}$, with the convention $\min\{\emptyset\} = \infty$. Under our assumptions about effective market power mitigation in the spot market, the market price is defined by the stack price function which directly reflects the capacities and marginal operating costs of the suppliers:

$$s(D_t; \boldsymbol{K}, M) := G^{(J(D_t;\boldsymbol{K}))}, \text{ with } G^{(\infty)} := M.$$

**Remark 1.** *Actual spot market scarcity pricing schemes are more sophisticated. For instance, some schemes increase price more gradually as load approaches capacity, reflecting the fact that as operating reserves (short run spare capacity) become shorter in supply, this exposes the system to more risk if an unexpected equipment outage, demand spurt, or reduction in renewable output occurs. One implementation of this idea is the "operating reserve demand curve", first suggested by Stoft [62], and now adopted by the Electric Reliability Council of Texas. Our model could be extended to implement this and other scarcity pricing designs.*

### 2.3 Energy Market Settlements and Profits

The decisions that market participants make are based on their anticipated cash flow distributions and their expected utility. Given the installed capacity $\boldsymbol{K}$, supplier $j$'s rate of profit from energy sales at time $t$ is, in \$/hr:

$$\Pi_t^{(j)}\big(K^{(j)}; \boldsymbol{K}^{(-j)}, M\big) := \big(s(D_t; \boldsymbol{K}, M) - G^{(j)}\big)\big((D_t - \mathbf{1}_{j-1}^T(\boldsymbol{K}_0 + \boldsymbol{K}))^+ \wedge (K_0^{(j)} + K^{(j)})\big),$$

i.e., selling the dispatched amount of electricity (upper bounded by its total capacity) at the market price (which may hit the price cap).

$d_t^{(i)}$, the demand of consumer $1 \leq i \leq N_c$, may be fulfilled or not, depending on whether the total capacity is adequate. In case of capacity inadequacy, the total supply shortage of amount $\left(D_t - \mathbf{1}_{N_s}^T(\boldsymbol{K}_0 + \boldsymbol{K})\right)^+$ is shed among all $N$ consumers, each suffering a proportion of $\alpha^{(i)}$. This lost load is monetarized through VOLL, $V^{(i)} \geq 0$. Given $\boldsymbol{K}$, consumer $i$'s rate of expenditure is:

$$P_t^{(i)}(\boldsymbol{K}; M) := s\left(D_t; \boldsymbol{K}, M\right)\left(d_t^{(i)} - \alpha^{(i)}\left(D_t - \mathbf{1}_{N_s}^T(\boldsymbol{K}_0 + \boldsymbol{K})\right)^+\right)^+ + V^{(i)}\alpha^{(i)}\left(D_t - \mathbf{1}_{N_s}^T(\boldsymbol{K}_0 + \boldsymbol{K})\right)^+,$$

i.e., fulfilling part of the demand at the market price, and losing the remaining part at a cost of VOLL.

Note the electricity price is completely determined by the demand-supply pair, and bounded by the price cap. The existence of the cap might, depending on its level, provide an inadequate incentive to build enough capacity to meet the reliability target. The SO solves the issue through a resource adequacy mechanism that provides a payment to suppliers for installed capacity, such that building new capacity is more profitable. The SO sets the payments at a level that induces compliance with a reliability standard (expressed as an expected length of time when capacity is inadequate to meet demand), and decides the capacity payment $\boldsymbol{C} = [C^{(j)}]_{j=1,\ldots,N_s}$ that maximizes an assumed welfare measure. For simplicity, this payment is collected from all consumers by the allocation rule (for load shedding) mentioned in Section 2.2, but other cost allocation schemes can be modeled if desired.

### 2.4 Bi-Level Decision Model

We consider the utility of market participants in a financial (von Neumann-Morgenstern type) perspective, by evaluating a concave (risk-averse) function of a monetary attribute. The supplier's utility is directly based on the net cash flows to be received, i.e., revenue from power sales and capacity payments minus costs of power generation and capacity investment. The consumer's utility is computed based on energy bill amounts, capacity payments and monetarization of supply shortfall relative to existing demand.[2] The utility of supplier $1 \leq j \leq N_s$ (consumer $1 \leq i \leq N_c$) is evaluated through the constant risk-aversion (negative exponential) utility function $\mathcal{U}_s^{(j)}(x) := -e^{-\gamma_s^{(j)}x}$ ($\mathcal{U}_c^{(i)}(x) := -e^{-\gamma_c^{(i)}x}$).

We propose a bi-level model to represent the relationship between the SO (upper-level Stackelberg leader) which chooses capacity payments and clears the energy market, and the suppliers (lower-level Stackelberg followers) which optimize their investments to maximize the expected utility of their profits from capacity investments and energy sales.

The lower-level game is a pure-strategy Nash game among suppliers, in which the strategic variable is their investment in new capacity (analogous to, e.g., [39][64]). Supplier $j$'s utility over time horizon $T$ depends on its profit and investment on new capacity, and is given by taking the expectation of the utility function discounted over the period.

$$u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M) := \mathbb{E}\Big[\int_0^T e^{-r_s^{(j)}t}\mathcal{U}_s^{(j)}\Big((C^{(j)} - F^{(j)})K^{(j)} + \Pi_t^{(j)}\big(K^{(j)}; \boldsymbol{K}^{(-j)}, M\big)\Big)dt\Big].$$

[2]Economically, this utility can be viewed as a quantification of monetized loss of consumer surplus, compared with the ideal situation where consumer's demand is fulfilled at zero price.

Given the the leader's choice of capacity payment, each supplier observes others' capacity decisions, and the lower-level game is settled at the equilibrium $\boldsymbol{K}^*$ such that

$$u^{(j)}(K^{*(j)};\boldsymbol{K}^{*(-j)},\boldsymbol{C},M) \geq u^{(j)}(K^{(j)};\boldsymbol{K}^{*(-j)},\boldsymbol{C},M) \text{ for any } K^{(j)} \geq 0, \quad j=1,\ldots,N_s. \qquad (P_s)$$

This is a Nash equilibrium in capacity, which is dependent on both the capacity payment and price cap, and may not be unique. We denote all possible equilibria by the set function $\boldsymbol{K}^*(\boldsymbol{C};M)$, which may take values at a singleton or null set. We remind the reader that the spot energy market is not a Nash game, as we assume that generators are required to disclose their true variable costs in their offers.

The upper-level problem is the optimization problem of the SO, whose goal is to maximize the aggregate consumer welfare while maintaining grid reliability in the assessment period $[0,T]$, in order to decide the optimal payment $\boldsymbol{C}^*$, which is collected from consumers by the allocation rule mentioned above. The SO realizes that the suppliers play a Nash game in the capacity market, and aims at maximizing the total expected utility of consumers. With fixed load, this calculation maximizes the sum across consumers of the expected utility of payments made by consumers for energy as well as capacity, and also accounts for VOLL losses due to curtailment, and is given by the weighted sum of expected discounted utility.

$$u^{(SO)}(\boldsymbol{C},\boldsymbol{K};M) := \sum_{i=1}^{N_c} N_c^{(i)}\mathbb{E}\left[\int_0^T e^{-r_c^{(i)}t}\mathcal{U}_c^{(i)}\left(-P_t^{(i)}(\boldsymbol{K};M)-\alpha^{(i)}\boldsymbol{C}^T\boldsymbol{K}\right)dt\right]. \qquad (1)$$

**Remark 2.** *In most of the literature, the regulator maximizes total welfare, i.e., producers' surplus plus consumers' surplus, which is equivalent to minimizing total costs in a risk-neutral world with inelastic demand. The total welfare approach is also used by the US Federal Energy Regulatory Commission (FERC) as an economic translation of the 1935 US Federal Power Act's requirement for "just and reasonable prices". In our model we use the weighted consumers' utility* (1) *as the SO's objective. One of the reasons for not including producer utility is that risk-averse utility functions only provide a ranking of solutions, and it is widely recognized in welfare economics that simple summations of nonlinear utility functions of market parties are not meaningful. And even if the SO does not consider suppliers' welfare, they will not suffer a net loss by participating in the capacity market, because* $\boldsymbol{K}=\boldsymbol{0}$ *is always a feasible point in* $(P_s)$ *and any optimal solution will have nonnegative profits. We note that it is sometimes argued that utility regulation should pursue maximization of consumer welfare in part because the purpose of regulation is to protect consumers from producers with market power; several normative models for power system planning and operations have therefore been formulated using that objective, e.g., [42][61]. In Section 4.6, we give an illustrative example of a SO who considers both consumers' and suppliers' utility by using a total-utility function as an alternative objective, and show the possible impacts on optimal capacity payments and generation mix.*

We restrict the decision space for payment to $\boldsymbol{C} \in \Omega^{\epsilon}_{\boldsymbol{C}} := \prod_{j=1}^{N_s}[0,F^{(j)}-\epsilon]$ for some small $\epsilon > 0$, because otherwise the suppliers may decide to build infinite capacity, which is an invalid value, and the lower-level player (supplier) response to upper-level (SO) capacity payment decisions would lose compactness. The SO solves

$$(P_\epsilon(\rho)) := \sup_{(\boldsymbol{C},\boldsymbol{K})\in\Omega^{\epsilon}_{\boldsymbol{C}}\times\mathbb{R}^{N_s}_+} u^{(SO)}(\boldsymbol{C},\boldsymbol{K};M) \quad \text{s.t. } \mathbb{E}\left[\int_0^T \frac{1}{T}\mathbb{I}_{\{D_t>\boldsymbol{1}_{N_s}(\boldsymbol{K}_0+\boldsymbol{K})\}}dt\right] \leq \rho,\ \boldsymbol{K}\in\boldsymbol{K}^*(\boldsymbol{C};M). \qquad (P_{SO})$$

The first constraint states the reliability standard: the *a priori* expected time of load shedding should not exceed some pre-specified portion $\rho$, e.g., 1 day per 10 years. The second constraint represents the lower level response, by forcing the suppliers' investment decisions to be at equilibrium when the SO decides on the capacity payment.

**Remark 3.** *$(P_{SO})$ is effectively a best-case scenario if there are multiple lower-level equilibria for a given set of capacity payments: the SO is optimistic and anticipates that suppliers will choose the one that is most beneficial for consumers. In contrast, the worst-case scenario analysis, sometimes called pessimistic bi-level optimization, solves a min-max problem:*

$$\sup_{\boldsymbol{C}\in\Omega^{\epsilon}_{\boldsymbol{C}}} \inf_{\boldsymbol{K}\in\mathbb{R}^{N_s}_{+}} u^{(SO)}(\boldsymbol{C},\boldsymbol{K};M) \quad s.t.\ \mathbb{E}\left[\int_0^T \frac{1}{T}\mathbf{1}_{\{D_t>\mathbb{I}_{N_s}(\boldsymbol{K}_0+\boldsymbol{K})\}}dt\right] \le \rho,\ \boldsymbol{K}\in\boldsymbol{K}^*(\boldsymbol{C};M).$$

*The pessimistic problem can be approached via a sequence of optimistic problems [24][41]. Here we focus on $(P_{SO})$, and leave exploration of the pessimistic framing to future research.*

We illustrate the information and decision flows by Figure 1. Each market participant and its decision is grouped in a square box, with decision variables in a circle. The model inputs and outputs are shown in round-cornered boxes, with the optimal outcome denoted by a superscript $*$.

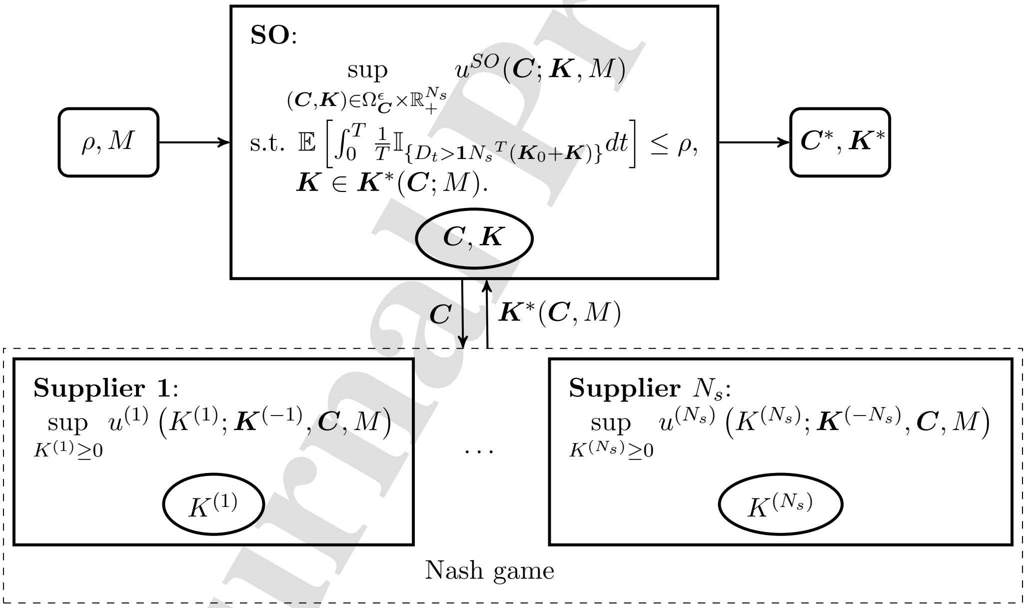


Figure 1: Model Diagram

## 3 Mathematical Results

$(P_{SO})$ is a bilevel optimization that is cast as a Mathematical Program with Equilibrium Constraints (MPEC) problem. The authors are not aware of universal solvers of a MPEC, except in simple cases that can be transformed into a linear complementarity-constrained linear problem which can be solved by mixed integer linear program (e.g., [26][43]). In this section, we introduce an equivalent, but more explicit format (i.e., as a variational inequality) of the lower-level equilibrium constraint specified by $(P_s)$, and proceed from this transformation to address the solvability of the bi-level problem $(P_{SO})$.

### 3.1 Suppliers' Equilibrium

We start with a formal definition of the suppliers' decision function $\boldsymbol{K}^*(\boldsymbol{C}; M)$, i.e., the Nash equilibrium of $(P_s)$.

**Definition 1** (Suppliers' Nash Equilibrium). *Let $\mathcal{S} \in \mathbb{R}_+^{N_s}$ be a closed set. Given $\boldsymbol{C}, M$, a Nash equilibrium of the suppliers' game on $\mathcal{S}$ is a strategy $\boldsymbol{K}^* \in \mathcal{S}$, such that*

$$\begin{aligned}
&u^{(j)}(K^{*(j)}; \boldsymbol{K}^{*(-j)}, \boldsymbol{C}, M) \geq u^{(j)}(K^{(j)}; \boldsymbol{K}^{*(-j)}, \boldsymbol{C}, M),\\
&\forall K^{(j)} \ s.t.\ [K^{*(1)}, \ldots, K^{*(j-1)}, K^{(j)}, K^{*(j+1)}, \ldots, K^{*(N_s)}] \in \mathcal{S}, \ \text{for all } 1 \leq j \leq N_s.
\end{aligned}$$

*The set of all such Nash equilibria is denoted by $\boldsymbol{K}^*_{\mathcal{S}}(\boldsymbol{C}; M)$.*

*The Nash equilibria of $(P_s)$ is defined to be $\boldsymbol{K}^*(\boldsymbol{C}; M) := \boldsymbol{K}^*_{\mathbb{R}_+^{N_s}}(\boldsymbol{C}; M)$.*

A classical way to study the Nash equilibrium is the fixed point approach: it is a fixed point of the optimal response function, which profiles for each $j$, the maximizers of $u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M)$ as function of $\boldsymbol{K}^{(-j)}$ given $\boldsymbol{C}, M$. Often, iterative Gauss-Seidel methods are used to search for such a fixed point, or a closed-form analytical solution is derived. But in consideration of the complexity of our function $\boldsymbol{u}$, such an optimal response can be intractable, especially when we use it as a constraint in the upper level (leader's) game. Instead, we will use Variational Inequalities (VI), a popular and powerful tool from convex optimization, to give equilibrium certificates of the suppliers' game.

**Definition 2** (Suppliers' VI Problem). *Let $\mathcal{S} \in \mathbb{R}_+^{N_s}$ be a closed set and denote*

$$\nabla \boldsymbol{u}(\boldsymbol{K}; \boldsymbol{C}, M) := \left[\frac{\partial}{\partial K^{(j)}} u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M)\right]_{j=1,\ldots,N_s}.$$

*The solution set to the VI problem related to $\nabla \boldsymbol{u}(\cdot; \boldsymbol{C}, M)$ and $\mathcal{S}$ is defined as*

$$\mathit{Sol}(\boldsymbol{C}, M, \mathcal{S}) := \{\boldsymbol{K}^\dagger \in \mathcal{S} : \left\langle \nabla \boldsymbol{u}(\boldsymbol{K}^\dagger; \boldsymbol{C}, M), \boldsymbol{K} - \boldsymbol{K}^\dagger \right\rangle \leq 0, \forall \boldsymbol{K} \in \mathcal{S}\}.$$

Further, the above VI problem restricted to a compact region can be proven to have solutions.

**Lemma 3.1.** *Given $\boldsymbol{C} \geq \boldsymbol{0}, M > G^{(N_s)}$, $u^{(j)}(\cdot; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M)$ is smooth for any $\boldsymbol{K}^{(-j)} \geq \boldsymbol{0}$. For any $\mathcal{S}$ compact, Sol$(\boldsymbol{C}, M, \mathcal{S})$ is nonempty and closed.*

*Proof.* See Appendix A. □

The question is whether we can relate the Nash equilibrium of $(P_s)$, whose decision space is unbounded, to the compact VI problem. In general VI problems are closely related to maximization of concave functions. Being a (gradient) VI solution is necessary for a maximizer, and this is also sufficient under reasonable concavity conditions (e.g., [38]). If $(P_s)$ possesses the desired concavity, then we can use this linkage to transform the Nash equilibrium problem to a VI problem and show existence results by investigating the latter, i.e., $\boldsymbol{K}^*(\boldsymbol{C}; M) = \mathrm{Sol}(\boldsymbol{C}, M, \mathbb{R}_+^{N_s}) \neq \emptyset$.

The roadmap of the rest of this subsection is as follows: First we introduce Lemma 3.2 which guarantees certain compactness in solving $(P_s)$. The desired concavity is given by Lemma 3.4, which leads to Lemma 3.5 that draws the equivalence between the suppliers' game and the VI problem when restricted to a compact region. Finally Theorem 3.6 extends the equivalence to the whole $\mathbb{R}_+^{N_s}$ to conclude on the Nash equilibrium of $(P_s)$.

**Lemma 3.2.** *Let $\epsilon > 0$. For any $M > G^{(N_s)}$, we can compute $\bar{\boldsymbol{K}}_\epsilon < \infty$ such that for any $\boldsymbol{C} \in \Omega^\epsilon_{\boldsymbol{C}}$,*

$$\boldsymbol{K}^*(\boldsymbol{C}, M) \subset \mathcal{S}(\bar{\boldsymbol{K}}_\epsilon) := [0, \bar{K}_\epsilon{}^{(1)}) \times \cdots \times [0, \bar{K}_\epsilon{}^{(N_s)}).$$

*And on the boundary we have $\frac{\partial}{\partial K^{(j)}} u^{(j)}(\bar{K}_\epsilon{}^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M) < 0$, $\forall \boldsymbol{K}^{(-j)} \geq \boldsymbol{0}$.*

*Proof.* See Appendix A. □

That is to say, we can restrict the search for the Nash equilibrium of $(P_s)$ to a compact region $\mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)$. We give some further description of this region that is useful for later results.

**Proposition 3.3.** *$\bar{\boldsymbol{K}}_\epsilon$ has the following properties:*

1. *For any $0 < \epsilon_1 < \epsilon_2 < \min\{F^{(j)}\}_j$, $\bar{\boldsymbol{K}}_{\epsilon_1} \leq \bar{\boldsymbol{K}}_{\epsilon_2}$.*
2. *For any finite $\boldsymbol{K}'$, there is $\epsilon > 0$ such that $\bar{\boldsymbol{K}}_\epsilon \geq \boldsymbol{K}'$.*
3. *For a fixed $\epsilon$, there is $\tilde{M}_\epsilon$ such that if $M \geq \tilde{M}_\epsilon$, $\bar{\boldsymbol{K}}_\epsilon$ is independent of $M$.*

*Proof.* See Appendix A. □

With help of Lemma 3.2, we claim the concavity required for translating $(P_s)$ into a VI problem.

**Lemma 3.4.** *Let $\epsilon > 0$. There exists $\hat{M}_\epsilon$ such that if $M \geq \hat{M}_\epsilon$, $(P_s)$ is pseudo-concave for any given $\boldsymbol{C} \in \Omega^\epsilon_{\boldsymbol{C}}$.*

*Proof.* See Appendix A. □

The following equivalence between the suppliers' Nash game and VI problem, both restricted to $\mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)$, is a direct result of the equivalence of Nash equilibria and VI solutions under pseudo-concavity [18][38].

**Lemma 3.5.** *Let $\epsilon > 0$, $M > \max\{G^{(N_s)}, \hat{M}_\epsilon\}$. Then for any $\boldsymbol{C} \in \Omega^\epsilon_{\boldsymbol{C}}$,*

$$\boldsymbol{K}^*_{cl\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)}(\boldsymbol{C}, M) = Sol(\boldsymbol{C}, M, cl\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)).$$

In fact, the above restricted suppliers' Nash game and VI problem can be extended to $\mathbb{R}_+^{N_s}$ without changing the solution set. This results in the existence of the suppliers' Nash equilibrium.

**Theorem 3.6** (Existence of Suppliers' Equilibrium)**.** *Let $\epsilon > 0$, $M > \max\{G^{(N_s)}, \hat{M}_\epsilon\}$. Then $(P_s)$ is equivalent to the VI problem in the sense that $\boldsymbol{K}^*(\boldsymbol{C}, M) = Sol(\boldsymbol{C}, M, \mathbb{R}_+^{N_s})$, and admits at least one Nash equilibrium for any $\boldsymbol{C} \in \Omega^\epsilon_{\boldsymbol{C}}$.*

*Proof.* We first inspect the problems restricted to $\mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)$. Naturally $\boldsymbol{K}^*_{\mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)}(\boldsymbol{C}, M) \subset \mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)$. In fact we can further claim $\boldsymbol{K}^*_{\mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)}(\boldsymbol{C}, M) \subset \mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)$, since at any point on $\mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right) \setminus \mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)$, at least one supplier has a negative utility gradient according to Lemma 3.2, and can increase its utility by reducing its capacity. Now by Lemma 3.1 and Lemma 3.5, we have for sufficiently large $M$,

$$\emptyset \neq \mathrm{Sol}(\boldsymbol{C}, M, \mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)) = \boldsymbol{K}^*_{\mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)}(\boldsymbol{C}, M) \subset \mathcal{S}(\bar{\boldsymbol{K}}_\epsilon). \tag{2}$$

By definition of the solution set, $\mathrm{Sol}(\boldsymbol{C}, M, \mathbb{R}_+^{N_s}) \subset \mathrm{Sol}(\boldsymbol{C}, M, \mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right))$. For the reverse inclusion, suppose $\boldsymbol{K}^\dagger \in \mathrm{Sol}(\boldsymbol{C}, M, \mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right))$, i.e., $\left\langle \nabla u(\boldsymbol{K}^\dagger; \boldsymbol{C}, M), \boldsymbol{K} - \boldsymbol{K}^\dagger \right\rangle \leq 0$ for any $\boldsymbol{K} \in \mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)$. By (2) we know $\boldsymbol{K}^\dagger \in \mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)$, so we can write any $\boldsymbol{K'} \geq \boldsymbol{0}$ as $\boldsymbol{K'} = \boldsymbol{K}^\dagger + \lambda(\boldsymbol{K} - \boldsymbol{K}^\dagger)$ for some $\lambda \geq 0$ and $\boldsymbol{K} \in \mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)$. Hence $\left\langle \nabla u(\boldsymbol{K}^\dagger; \boldsymbol{C}, M), \boldsymbol{K'} - \boldsymbol{K}^\dagger \right\rangle \leq 0$. As a result,

$$\mathrm{Sol}(\boldsymbol{C}, M, \mathbb{R}_+^{N_s}) = \mathrm{Sol}(\boldsymbol{C}, M, \mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)). \tag{3}$$

By definition of the Nash equilibrium, $\boldsymbol{K}^*(\boldsymbol{C}, M) = \boldsymbol{K}^*_{\mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)}(\boldsymbol{C}, M)$, because $\boldsymbol{K}^*(\boldsymbol{C}, M) \subset \mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)$ by Lemma 3.2. Combining this with (2)(3), we have that for such sufficiently large $M$,

$$\boldsymbol{K}^*(\boldsymbol{C}, M) = \mathrm{Sol}(\boldsymbol{C}, M, \mathbb{R}_+^{N_s}) \neq \emptyset.$$

□

### 3.2 SO's Optimization

Once $M$ is large enough, the suppliers' Nash game can be translated into the equivalent VI problem, and we may rewrite the MPEC $(P_{SO})$ as a bi-level problem with a VI constraint:

$$\sup_{(\boldsymbol{C},\boldsymbol{K}) \in \Omega^\epsilon_{\boldsymbol{C}} \times \mathbb{R}_+^{N_s}} \; u^{(SO)}(\boldsymbol{C}, \boldsymbol{K}; M) \tag{$P_{VI}$}$$

$$\text{s.t. } \mathbb{E}\left[\int_0^T \frac{1}{T} \mathbb{I}_{\{D_t > \boldsymbol{1}_{N_s}(\boldsymbol{K}_0 + \boldsymbol{K})\}} dt\right] \leq \rho, \; \boldsymbol{K} \in \mathrm{Sol}(\boldsymbol{C}, M, \mathbb{R}_+^{N_s}).$$

**Remark 4.** *From the proof of Theorem 3.6 and the constructive proof in [44], the set Sol*$(\boldsymbol{C}, M, \mathbb{R}_+^{N_s})$ *is exactly the set of all fixed points of the mapping* $\boldsymbol{K} \mapsto \mathcal{P}(\boldsymbol{K} - \nabla \boldsymbol{u}(\boldsymbol{K}; \boldsymbol{C}, M); \mathit{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right))$, *where* $\mathcal{P}(\cdot; \mathcal{S})$ *denotes the projection mapping onto set* $\mathcal{S}$. *This proposes fixed-point iterations that a computer can use to solve the VI.* [3]

With two auxiliary lemmas, the existence of solution to $(P_{VI})$ and hence $(P_{SO})$, is given by Theorem 3.9.

**Lemma 3.7.** *Given* $M > G^{(N_s)}$ $\boldsymbol{K}^*(\boldsymbol{C}; M)$ *is upper hemi-continuous in* $\boldsymbol{C}$.

*Proof.* See Appendix A. □

**Lemma 3.8.** *Given* $M > G^{(N_s)}$, $u^{(SO)}(\boldsymbol{C}, \boldsymbol{K}; M)$ *is continuous in* $(\boldsymbol{C}, \boldsymbol{K})$.

*Proof.* See Appendix A. □

**Theorem 3.9** (Existence of SO's Decision)**.** *For fixed* $\epsilon > 0, 0 < \rho \leq 1$, $M > \max\{G^{(N_s)}, \hat{M}_\epsilon, \tilde{M}_\epsilon\}$, $(P_{SO})$ *has a bounded optimal solution if it is feasible.*

*Proof.* Denote $\Gamma(\boldsymbol{K}) := \frac{1}{T}\int_0^T \left(1 - \Phi\left(\frac{\boldsymbol{1}^T_{N_s}(\boldsymbol{K}_0 + \boldsymbol{K}) - \mu_t}{\sigma_t}\right)\right) dt = \mathbb{E}\left[\int_0^T \frac{1}{T} \mathbb{I}_{\{D_t > \boldsymbol{1}_{N_s}(\boldsymbol{K}_0 + \boldsymbol{K})\}} dt\right]$, which is continuous in $\boldsymbol{K}$. Denote $\Delta(\Omega) := \{\boldsymbol{C} \in \Omega^\epsilon_{\boldsymbol{C}} : \boldsymbol{K}^*(\boldsymbol{C}, M) \cap \Omega \neq \emptyset\}$ for $\Omega \subset \mathbb{R}^{N_s}$, which is nontrivial

[3] Recall that intrinsically a Nash equilibrium can also be written as a fixed point of the best response function. But the Nash iteration is a closed loop, since the response function is not always easily solved (except in very simple cases). In contrast, the VI solution as a fixed point is an open-loop problem.

since $\boldsymbol{K}^*(\boldsymbol{C}, M) \neq \emptyset$ by Theorem 3.6. Then with $\Omega^\epsilon_{\boldsymbol{K}}(\rho) := \Gamma^{-1}([0,\rho]) \bigcap \mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)$, $(P_{SO})$ is equivalent to

$$(P_\epsilon(\rho)) = \sup_{\boldsymbol{C},\boldsymbol{K}} u^{(SO)}(\boldsymbol{C}, \boldsymbol{K}; M) \quad \text{s.t. } \boldsymbol{C} \in \Omega^\epsilon_{\boldsymbol{C}} \cap \Delta\left(\Omega^\epsilon_{\boldsymbol{K}}(\rho)\right),\, \boldsymbol{K} \in \boldsymbol{K}^*(\boldsymbol{C}; M). \qquad (P'_{SO})$$

By Lemma 3.7, $\boldsymbol{K}^*(\cdot; M)$ is upper hemi-continuous, so its lower inverse-image of a closed set is closed. Note $\Omega^\epsilon_{\boldsymbol{K}}(\rho)$ is closed as a union of closed sets, hence $\Delta(\Omega^\epsilon_{\boldsymbol{K}}(\rho))$ is closed and $\Omega^\epsilon_{\boldsymbol{C}} \cap \Delta(\Omega^\epsilon_{\boldsymbol{K}}(\rho))$ is compact. The compactness of $\boldsymbol{K}^*(\boldsymbol{C}; M)$ follows from Lemma 3.1 and Lemma 3.2. As a result, $(P'_{SO})$ is on a compact feasible region, and the objective function is continuous by Lemma 3.8. So $(P'_{SO})$, and equivalently $(P_{SO})$, admits a maximum, as soon as the feasible region is nonempty. □

Theorem 3.9 gives a first result on the existence of solution to $(P_{SO})$, and shows that the sup is indeed a max. But it remains to be decided when $(P_{SO})$ is feasible. This is partially answered by Theorem 3.10, which states that for any price cap $M$ that ensures pseudo-concavity, there is a minimal reliability standard $\rho_0$ that is realizable and guarantees the feasibility and solvability.

**Theorem 3.10** (Existence of SO's Decision, Minimal Reliability)**.** *Let $\epsilon > 0$, $M > \max\{G^{(N_s)}, \hat{M}_\epsilon, \tilde{M}_\epsilon\}$. There exists $0 < \rho_0 < 1$, such that $(P_{SO})$ has a bounded optimal solution for all $\rho \in [\rho_0, 1]$.*

*Proof.* We show this for the equivalent problem $(P'_{SO})$ as follows.

By Theorem 3.6, $\mathrm{Sol}(\boldsymbol{C}, M, \mathbb{R}^{N_s}_+) \neq \emptyset$ for any $\boldsymbol{C} \in \Omega^\epsilon_{\boldsymbol{C}}$, so $\Delta(\Omega^\epsilon_{\boldsymbol{K}}(1)) = \Delta(\mathbb{R}^{N_s}_+) = \Omega^\epsilon_{\boldsymbol{C}}$. So $(P_\epsilon(1))$ is feasible and Theorem 3.9, it has a bounded optimal solution, denoted by $(\hat{\boldsymbol{C}}^*, \hat{\boldsymbol{K}}^*) \in \Omega^\epsilon_{\boldsymbol{C}} \times \mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)$.

Note $\Omega^\epsilon_{\boldsymbol{K}}(\rho)$ changes continuously in $\rho$, so there is $\rho_0 \in (0,1)$ such that $\Omega^\epsilon_{\boldsymbol{K}}(\rho_0) \cap \mathcal{S}(\bar{\boldsymbol{K}}_\epsilon) \neq \emptyset$. For any $\rho \in [\rho_0, 1]$, we still have $\hat{\boldsymbol{K}}^* \in \Omega^\epsilon_{\boldsymbol{K}}(\rho)$. In other words, $(\hat{\boldsymbol{C}}^*, \hat{\boldsymbol{K}}^*)$ remains feasible for $(P_\epsilon(\rho))$. By Theorem 3.9, the problem $(P'_{SO})$ therefore has a bounded optimal solution. □

From Proposition 3.3 and the proof of Theorem 3.10, we see the minimal reliability $\rho_0$ monotonically decreases as $\epsilon \downarrow 0$. Still, we are interested in whether $\rho_0$ is sufficient to realize a given reliability. That is, can we guarantee a solution to $(P_\epsilon(\rho))$ given any $\rho > 0$, by altering other auxiliary parameters? The answer is given by Theorem 3.11, which states that the effect of minimal reliability $\rho_0$ can be achieved by adjusting the scarcity price $M$ and decision margin $\epsilon$.

**Theorem 3.11** (Existence of SO's Decision, Scarcity Pricing)**.** *For arbitrary $\rho \in (0,1]$, there exists $M_0 > G^{(N_s)}$, such that for all $M \geq M_0$, $(P_{SO})$ has a bounded optimal solution for all $\epsilon \leq \epsilon_0$ for some $\epsilon_0 > 0$.*

*Proof.* Fix $\rho \in (0,1]$. We show for the equivalent problem $(P'_{SO})$.

By the first two properties in Proposition 3.3, $\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)$ increases unboundedly as $\epsilon \downarrow 0$. So there is $\epsilon_0$ such that $\Omega^{\epsilon_0}_{\boldsymbol{K}}(\rho) \neq \emptyset$ and $(P_{\epsilon_0}(\rho))$ is well defined. We only show the existence result for $(P_{\epsilon_0}(\rho))$, and $(P_\epsilon(\rho))$ with any $\epsilon \leq \epsilon_0$ can be treated in the same manner.

Take $M > \max\{G^{(N_s)}, \hat{M}_{\epsilon_0}, \tilde{M}_{\epsilon_0}\}$ as in Proposition 3.3, Lemma 3.4, which guarantees pseudo-concavity and all previous results, and that $\mathcal{S}(\bar{\boldsymbol{K}}_{\epsilon_0})$ is independent of $M$. We now show that $(P_{\epsilon_0}(\rho))$ has a feasible point, and hence, by Theorem 3.9 again, has a bounded optimal solution.

In $(P'_{SO})$, Theorem 3.6 guarantees that $\boldsymbol{K}^*(\boldsymbol{C}, M) \neq \emptyset$, so $(P_{\epsilon_0}(\rho))$ has a feasible point if $\Omega^{\epsilon_0}_{\boldsymbol{C}} \cap \Delta(\Omega^{\epsilon_0}_{\boldsymbol{K}}(\rho)) \neq \emptyset$, or equivalently,

$$\mathrm{Sol}\left(\boldsymbol{C}, M, \mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_{\epsilon_0})\right)\right) \cap \Omega^{\epsilon_0}_{\boldsymbol{K}}(\rho) \neq \emptyset \text{ for some } \boldsymbol{C} \in \Omega^{\epsilon_0}_{\boldsymbol{C}}. \tag{4}$$

Note $0 \leq \Gamma(\boldsymbol{K}_1) < \Gamma(\boldsymbol{K}_2)$ if $\mathbf{1}_{N_s}^T \boldsymbol{K}_1 > \mathbf{1}_{N_s}^T \boldsymbol{K}_2$, so $\Gamma^{-1}([0,\rho]) = \{\boldsymbol{K} \in \mathbb{R}_+^{N_s} : \mathbf{1}_{N_s}^T \boldsymbol{K} \geq a_\rho\}$ for some $a_\rho \geq 0$. As a result, $\Omega_{\boldsymbol{K}}^{\epsilon_0}(\rho)$ is a full-dimensional compact polyhedron. Define the boundary set $\mathcal{B} := \{\boldsymbol{K} \in \mathbb{R}_+^{N_s} : \mathbf{1}_{N_s}^T \boldsymbol{K} = a_\rho\}$. Define $\boldsymbol{K}^\dagger(\boldsymbol{C}, M) := \mathrm{Sol}\big(\boldsymbol{C}, M, \mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_{\epsilon_0})\right) \cap \Omega_{\boldsymbol{K}}^{\epsilon_0}(\rho)\big)$, which is nonempty by Lemma 3.1. We will show for sufficiently large $M$,

$$\boldsymbol{K}^\dagger(\boldsymbol{C}, M) = \mathrm{Sol}\big(\boldsymbol{C}, M, \mathrm{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_{\epsilon_0})\right)\big), \tag{5}$$

which, combined with the fact $\tilde{\boldsymbol{K}}^\dagger(\boldsymbol{C}, M) \subset \Omega_{\boldsymbol{K}}^{\epsilon_0}(\rho)$, leads to (4).

Now we prove (5), which involves two VI problems on nested sets. In the same way as we prove (3), it suffices to show there is $\boldsymbol{C} \in \Omega_{\boldsymbol{C}}^{\epsilon_0}$ such that

$$\boldsymbol{K}^\dagger(\boldsymbol{C}, M) \cap \left(\Omega_{\boldsymbol{K}}^{\epsilon_0}(\rho) \setminus \mathcal{B}\right) \neq \emptyset. \tag{6}$$

Naturally $\boldsymbol{K}^\dagger(\boldsymbol{C}, M) \subset \Omega_{\boldsymbol{K}}^{\epsilon_0}(\rho)$, so the only case that (6) is violated, is that $\boldsymbol{K}^\dagger(\boldsymbol{C}, M) \subset \mathcal{B}$ for all $\boldsymbol{C} \in \Omega_{\boldsymbol{C}}^{\epsilon_0}$. We now show this cannot be true for all $M$. In fact, by increasing $M$, we can move $\boldsymbol{K}^\dagger(\boldsymbol{C}, M)$ off $\mathcal{B}$ for arbitrary $\boldsymbol{C} \in \Omega_{\boldsymbol{C}}^{\epsilon_0}$: By a similar proof to Lemma 3.7, we know $\boldsymbol{K}^\dagger(\boldsymbol{C}, M)$ is upper hemi-continuous in $M$, so $\{M > \max\{G^{(N_s)}, \hat{M}, \tilde{M}\} : \boldsymbol{K}^\dagger(\boldsymbol{C}, M) \cap \mathcal{B} \neq \emptyset\big) = \rho\}$ is closed, since it is the lower inverse-image of $\mathcal{B}$. That is to say, there is $M_0 > \max\{G^{(N_s)}, \hat{M}, \tilde{M}\}$, such that $\boldsymbol{K}^\dagger(\boldsymbol{C}, M) \subset \left(\Omega_{\boldsymbol{K}}(\rho) \setminus \mathcal{B}\right)$ for all $M > M_0$. For such large $M$, (6) holds and by all our earlier reasoning $(P_{\epsilon_0}(\rho))$ has a feasible solution.

□

# 4 Numerical Examples

In this section we show the numerical results from our model calibrated to PJM data. A brief summary of data and parameters is given in Section 4.1. The model outputs and analysis are shown in Section 4.2, with emphasis on the effect of risk aversion. In Section 4.3, we compare the results in Section 4.2 with the classical risk-neutral, competitive results, and show how market power distorts outcomes from the cost-efficient dispatch. Section 4.4 is devoted to investigation of price caps, where a number of price caps are applied, producing very different market outcomes. Finally, in Section 4.5, we provide a short discussion on allowing reliability to be endogenously determined rather than subject to a hard regulatory constraint.

## 4.1 Data and Parameter Description

In this section, we show some numerical results in our simulated market with $N_s = 5$ types of suppliers (renewable, nuclear, natural gas, coal, petroleum), and $N_c = 3$ types of consumers (residential, commercial and industrial market segments). We calibrate our parameters to the PJM area 2015-2016.[4] The total numbers of consumers are estimated to be $N^{(1)} = 2251579$, $N^{(2)} = 241964$, $N^{(3)} = 14787$. We assume a mean reverting process for individual demand: let $\boldsymbol{d}_t := \left[d_t^{(1)}, d_t^{(2)}, d_t^{(3)}\right]^T$, we suppose

$$\mathrm{d}\boldsymbol{d}_t = \boldsymbol{\kappa}_t(\boldsymbol{\theta}_t - \boldsymbol{d}_t)\mathrm{d}t + \boldsymbol{\Sigma}_t \boldsymbol{\rho} \mathrm{d}\boldsymbol{W}_t,$$

[4]The load was estimated in two steps: we first calibrated the individual demand process based on data collected and published by the Buildsmart DC project `http://www.newcityenergy.com/`; then we modified the data to fit PJM regional characteristics in `https://www.eia.gov/electricity/state/archive/2016/`. For full description of data feature and calibration, please see [14].

where $\boldsymbol{W}_t$ is a 3-dimensional Brownian motion, $\boldsymbol{\theta}_t := \left[\theta_t^{(1)}, \theta_t^{(2)}, \theta_t^{(3)}\right]^T$ is the long-term mean, $\boldsymbol{\kappa}_t := \begin{bmatrix} \kappa_t^{(1)} & 0 & 0 \\ 0 & \kappa_t^{(2)} & 0 \\ 0 & 0 & \kappa_t^{(3)} \end{bmatrix}$ the reversion speed, $\boldsymbol{\Sigma}_t := \begin{bmatrix} \sigma_t^{(1)} & 0 & 0 \\ 0 & \sigma_t^{(2)} & 0 \\ 0 & 0 & \sigma_t^{(3)} \end{bmatrix}$ the volatility, and $\boldsymbol{\rho} := \begin{bmatrix} 1 & \rho^{(12)} & \rho^{(13)} \\ \rho^{(12)} & 1 & \rho^{(23)} \\ \rho^{(13)} & \rho^{(23)} & 1 \end{bmatrix}$ the correlation coefficient across types. We approximate $\boldsymbol{d}_0$ by taking simple averages at time 0, and $\boldsymbol{\rho}$ by the correlation of different types' annual consumption. The process is discretized into 15-minute intervals, and we estimate the parameters $\{\boldsymbol{\kappa}_t, \boldsymbol{\theta}_t, \boldsymbol{\Sigma}_t\}_{t\geq 0}, \rho$ by maximum likelihood estimation. The marginal costs of generation and investment are shown in Table 1.[5] The price cap is taken to be \$2000/MWh, and the actual cost of curtailment experienced by the consumer (VOLL) is set to be \$6000/MWh.[6] The discount factors are taken to be $r_s^{(j)} = r_c^{(i)} = 7\%$ for all $i = 1, 2, 3$ and $j = 1, \dots, 5$. The assessment period is taken to be $T$ =1 year, with reliability $\rho = 0.027\%$ (1 day/10 years) and $\epsilon = 10^{-5}$ (\$/kW/yr). The allocation rule for load shedding and capacity allocation is proportional to the average energy consumption of individual consumers : $\alpha^{(1)} = 10^{-7}, \alpha^{(2)} = 1.27 \times 10^{-6}, \alpha^{(3)} = 3.15 \times 10^{-5}$ (corresponding to an average demand of 0.42 kW, 5.31 kW and 131.51 kW respectively for a residential, commercial, and industrial individual).

| | Renewable | Nuclear | Natural Gas | Coal | Petroleum |
|---|---|---|---|---|---|
| Generation Cost (\$/MWh) | 0 | 2.14 | 30.53 | 37.45 | 126.43 |
| Capacity Cost (\$/kW/yr) | 254.4 | 166.1 | 69.6 | 172.8 | 102.1 |

Table 1: Cost Parameters

**Remark 5.** *Note that the coal and petroleum generator types are dominated by natural gas, in that the latter has both a lower capacity cost and lower variable generation cost. Furthermore, the hypothetical renewable technology is also dominated by nuclear, even though the former has a lower variable cost, because even if renewables are run for 8760 hours, their total cost would exceed that of the nuclear facility. As a result, under these admittedly arbitrary assumptions, those dominated generation types would never enter a competitive investment market, as we will see in Section 4.3. However, in a Nash (oligopoly) investment market, inefficient generation types can be profitable, and indeed the leader (SO) might want to encourage them in order to have more players in the market. We conjecture that increasing the number of players, even if they are not all efficient, could lower prices. Future research can address under what circumstances this would be expected.*

[5]Date source: `https://www.eia.gov/todayinenergy/detail.php?id=31912`, `https://www.eia.gov/analysis/studies/powerplants/capitalcost/archive/2016/pdf/capcost_assumption.pdf`, `https://www.eia.gov/electricity/annual/html/epa_07_01.html`. The investment cost is annualized by a discount rate of 7% when applicable.

[6]Typically VOLL is estimated based on surveys and outage damage data, and in general ranges from \$3,000/MWh to \$10,000/MWh for the PJM market. For reference, we use \$6000/MWh, which was used in PJM's ORDC (Operating Reserve Demand Curve) proposal comments (`https://www.pjm.com/-/media/committees-groups/task-forces/epfstf/20181101/20181101-item-08-wilson-energy-economics-comments-on-pjm-ordc-proposal.ashx`). (MISO uses a VOLL of \$3500/MWh, we assume a higher value which is more robust in term of grid reliability.) In practice PJM does not use a price cap in spot markets, so we take the offer cap that is imposed on supply offers in the market (see `https://www.pjm.com/-/media/committees-groups/task-forces/epfstf/20181214/20181214-item-04-price-formation-paper.ashx`).

### 4.2 Optimal Capacity Design with Risk Aversion

In this subsection, we illustrate the insights that can be gained from our leader-follower framework for designing capacity payments by calculating the second-best differentiated payments for a small system. In particular, we illustrate their sensitivity to assumptions about risk attitudes by testing different levels of risk aversion. For simplicity, we assume that all suppliers (consumers) are equally risk averse, i.e., $\gamma_s^{(j)} = \gamma_s$ for $j = 1, 2, 3, 4, 5$ ($\gamma_c^{(i)} = \gamma_c$ for $i = 1, 2, 3$). The higher risk-aversion case is represented by an aversion coefficient of 0.1 (in units of \$/quarter-hour), and the lower risk-aversion case is indicated by a value of 0.001. We vary the degree of risk aversion separately for consumers and suppliers and thus obtain four cases. We assume zero existing capacity, and investigate the optimal pair of capacity amount and payment by technology with differently risk-averse players.

**Remark 6.** *The risk aversion levels 0.1 and 0.001 are used as a relative indicator of which market participant is more/less risk averse, and not correspond to actual conditions in some particular market. Estimation of risk aversion is highly nontrivial; we were only able to find one example for power markets [40], in the limited context of arbitrage between day-ahead and real-time markets, but not investment. Even for CVAR, which is widely adopted in the power market literature, general parameters on risk aversion are assumed in a sensitivity analysis mode rather than empirically calibrated from market outcomes, e.g., [45].*

The model outcomes are displayed in Tables 2–3 and Figures 2–3. In Table 2, the optimal capacity is displayed in rows of $\boldsymbol{K}^*$ (in MW), and capacity payment in rows of $\boldsymbol{C}^*$ (in \$/kW/yr).[7] Figures 2 and 3 respectively show investment (in MW) and total capacity payments (in \$/yr) for each technology for six scenarios, with technologies represented by different colors (and markers): green (star) for renewable, red (plus) for nuclear, blue (cross) for natural gas, purple (triangle) for coal, and black (square) for petroleum. The first four scenarios in the tables and figures correspond to the four combinations of the risk-averse levels in our model, i.e., both consumers and suppliers are risk-averse (with low/high levels) and the suppliers play a Nash game to determine the equilibrium capacity structure. For comparison, we add a fifth scenario "Competitive, Risk-Neutral", which is a baseline competitive (price-taking) and completely risk-neutral case that minimizes expected total costs (and hence maximizes total surplus under our assumption of inelastic demand), in order to show how market power of Nash players results in a very different technology mix and increases capacity payments. In addition, we add two other "Uniform Payment" scenarios, to investigate the role of heterogeneous capacity payments and market power. This subsection (Section 4.3) is devoted to the influence of risk aversion, i.e., the first four scenarios. We will discuss the last three scenarios in detail in Section 4.3.

In Table 3, we partially summarize the market outcomes resulting from the designs in Table 2, in terms of costs and revenue. The "Energy Price" and "Total Price" are cents/kWh prices that a consumer pays on average over the year, respectively being the average (quantity-weighted) energy spot price across the assessment period, and the total payment by consumers (for both energy and capacity) averaged over the demand for electricity. The "Supplier Revenue" is decomposed into "Energy Sales" (revenue from selling dispatched energy at the market price) and "Capacity Payment" (money received from the SO for building the required capacity). The "Total Cost" of the system is the sum of the "Variable Cost" for the suppliers to generate power, and the "Capacity

[7] Global optimality in the case of multiple equilibria is both a theoretical and practical concern for most realistic complex optimization programs. We tested for the possibility of multiple equilibria using random initialization, and observe that the model produces sufficiently similar outcomes (up to computation precision). This of course is not a proof that the solution is unique, but provides some reason for a uniqueness conjecture to be investigated in follow-on research.

| Scenarios | | Renewable | Nuclear | Natural Gas | Coal | Petroleum |
|---|---|---|---|---|---|---|
| $\gamma_s = 0.1$, | $\boldsymbol{K}^*$ | 1999.65 | 1881.70 | 1753.66 | 552.80 | 531.98 |
| $\gamma_c = 0.1$ | $\boldsymbol{C}^*$ | 98.49 | 56.52 | 32.60 | 82.25 | 56.38 |
| $\gamma_s = 0.1$, | $\boldsymbol{K}^*$ | 1302.68 | 2465.57 | 1925.33 | 528.95 | 498.26 |
| $\gamma_c = 0.001$ | $\boldsymbol{C}^*$ | 74.32 | 64.85 | 30.20 | 71.56 | 45.25 |
| $\gamma_s = 0.001$, | $\boldsymbol{K}^*$ | 1425.83 | 2342.50 | 1850.39 | 586.21 | 515.35 |
| $\gamma_c = 0.1$ | $\boldsymbol{C}^*$ | 79.48 | 59.73 | 31.18 | 84.43 | 58.22 |
| $\gamma_s = 0.001$, | $\boldsymbol{K}^*$ | 875.23 | 2833.64 | 1945.88 | 561.45 | 504.52 |
| $\gamma_c = 0.001$ | $\boldsymbol{C}^*$ | 55.89 | 65.24 | 28.85 | 79.50 | 51.43 |
| Competitive, | $\boldsymbol{K}^*$ | 0 | 4267.83 | 2451.95 | 0 | 0 |
| Risk-Neutral | $\boldsymbol{C}^*$ | 27.97 | 27.97 | 27.97 | 27.97 | 27.97 |
| Uniform Payment, | $\boldsymbol{K}^*$ | 988.42 | 2713.53 | 1991.45 | 503.65 | 523.47 |
| High Aversion | $\boldsymbol{C}^*$ | 66.27 | 66.27 | 66.27 | 66.27 | 66.27 |
| Uniform Payment, | $\boldsymbol{K}^*$ | 934.03 | 2643.69 | 2070.45 | 503.86 | 568.92 |
| Low Aversion | $\boldsymbol{C}^*$ | 55.83 | 55.83 | 55.83 | 55.83 | 55.83 |

Table 2: Capacity Amounts and Payments with Different Levels of Risk Aversion, Competition, and Imposition of Uniform Capacity Payments

Cost” for building the capacity; consequently, the “Supplier Profit” is the difference between the suppliers’ revenue and cost, and is a measure of the impact of market power. The values of unserved energy is not shown in the cost columns, and are negligible compared to the other costs (on the order of 0.1–1 million \$/year). All values in this table are average results from 50 independent simulation runs.

Here we focus on the first four scenarios in this subsection and leave the last three to next subsection. We note that all four risk aversion scenarios have similar levels of total capacity. In all four scenarios, the resulting LOLE (Loss of Load Expectation, the expected proportion of time that the system’s load exceeds capacity) is slightly lower than the LOLE required by the regulator (the first constraint in ($P_{SO}$)), which indicates that, given the capacity payments, the assumed energy price cap of \$2000/MWh is high enough to incentivize sufficient investment to attain the desired system reliability. However, the capacity mix and the resulting distribution of energy dispatch among technologies vary among the four scenarios.

In general, the interactions of risk aversion and market power (Nash behavior) result in some expected relationships between risk aversion and the distribution of revenues between energy and capacity markets. In particular, more risk-neutral suppliers are willing to sacrifice a constant stream of capacity payments for more volatile energy sales (compare third and fourth scenarios with the first two scenarios). In general, as the market participants become less risk-averse, we observe a decrease in total capacity payments and an increase in expected energy market revenues. Surprisingly, different risk aversion parameters lead to very different mixes of baseload capacity (i.e., high investment cost/low variable cost), with higher risk aversion favoring renewable capacity (the most extreme in terms of high investment cost/low variable cost), although the total amounts of baseload capacity are similar across the four scenarios.

But the overall effect of consumer risk aversion is less straightfoward. For instance, comparing scenarios 1 and 2, or scenarios 3 and 4 (the first of each pair having more risk-averse consumers), we see that although the relatively more risk-averse consumers experience more frequent very high and very low energy prices, the face lower prices on average. This is a result of having simultaneously more renewable capacity, which has the cheapest variable cost, as well as more petroleum, which

| Scenarios | Consumer Payment left(¢/kWh) | | Supplier Revenue left(Million \$/year) | | Total Cost left(Million \$/year) | | Supplier Profit left(Million \$/year) |
|---|---|---|---|---|---|---|---|
| | Energy Price | Total Price | Energy Sales | Capacity Payment | Variable Cost | Capacity Cost | |
| $\gamma_s = 0.1$, $\gamma_c = 0.1$ | 9.56 | 10.76 | 3496.70 | 439.80 | 627.79 | 1093.16 | 2215.56 |
| $\gamma_s = 0.1$, $\gamma_c = 0.001$ | 10.04 | 11.06 | 3673.38 | 374.72 | 739.75 | 1017.21 | 2291.25 |
| $\gamma_s = 0.001$, $\gamma_c = 0.1$ | 10.23 | 11.30 | 3742.72 | 390.44 | 733.47 | 1034.52 | 2365.18 |
| $\gamma_s = 0.001$, $\gamma_c = 0.001$ | 10.57 | 11.55 | 3867.74 | 360.50 | 809.58 | 977.29 | 2441.37 |
| Competitive, Risk-Neutral | 3.72 | 4.23 | 1360.56 | 187.95 | 668.96 | 879.54 | 0.01 |
| Uniform Payment, High Aversion | 10.64 | 11.86 | 3895.07 | 445.37 | 806.38 | 981.25 | 2552.81 |
| Uniform Payment, Low Aversion | 11.81 | 12.84 | 4322.92 | 375.23 | 901.60 | 965.99 | 2830.56 |

Table 3: Average Revenue and Cost with Different Levels of Risk Aversion, Competition, and Imposition of Uniform Capacity Payments

incurs the highest variable cost. In other words, highly risk-averse consumers are willing to pay more for capacity, and prefer lower energy prices for most time even if that comes with very high prices for a limited amount of time.

### 4.3 Nash Equilibrium Compared to Perfect Competition

We now further explore how the solutions in Section 4.2 contrast with the classical economic capacity model as analyzed by Stoft [62] by considering the row with results for the "Competitive, Risk-Neutral" case in Tables 2–3. That solution corresponds to the market equilibrium when all suppliers are risk-neutral and fully competitive (i.e., price takers rather than Nash players). In such a competitive world, under our assumptions (linear costs, continuous investment variables), competitive pressures drive profits to zero, with capacity and energy payments being just sufficient to cover the costs.

Although our model could obtain this result by inserting a risk-neutral (linear) utility function in all expressions, and allowing the number of firms providing each generation type to be very large, it is most convenient and much faster to use the classic "screening curve" method for least-cost generation expansion planning (e.g., [5]). The screening curves in Figure 4 provide insight as to how suppliers' cost structures impact their market share in a competitive market. The suppliers are represented by total cost lines (as a function of number of hours of operation per year), with the slope being the corresponding marginal cost and the intercept being the investment cost. The minimum of these curves for a given level of operation indicates which technology is cheapest to operate at a particular operation time (hours/yr), balancing variable and construction costs. Figure 5 shows the load duration curve, cut at the reliability standard of one-day-in-10-years (equivalent to 0.027%) and showing the optimal technology mix in a competitive market. In our case, the cost-efficient

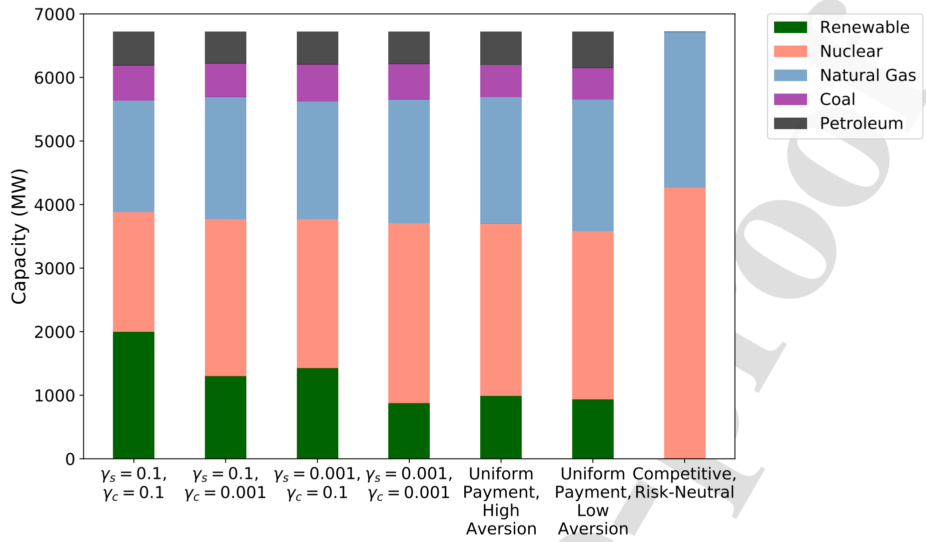


Figure 2: Model Outcome: Technology Mix

solution is that only nuclear (shown as red pluses in Figure 4) and natural gas plants (as blue crosses) get built, with the former being the baseload generator with a capacity shown on the y-axis of Figure 5 (4268 MW), and the latter supplying the remaining load with annual operation hours shown on the x-axis of Figure 5 (3399 MW). To provide the desired reliability, all suppliers require a capacity payment of \$27.97/kW/yr, which is less than that in any risk averse scenario (Table 2). In contrast, peaker plants in the Nash solutions, i.e., natural gas, coal and petroleum, get dispatched more hours per year, which greatly increases the energy spot price, as indicated by the first column in Table 3.

Such a competitive solution is very different from the four Nash solutions in Table 2, where the prices and suppliers' revenues/profits are significantly higher than the former. One of the reasons is risk aversion, which is absent in the competitive solution. Still, we may see in Figures 2 and 3 that as all participants become more risk-neutral, nuclear and natural gas account for a larger proportion in the total technology mix (thus becoming more similar to the competitive case), and get a higher capacity payment. However, the prices do not move in the same direction. In fact, as risk aversion decreases, we see consumers pay more in spite of the Stackelberg leader's best efforts to maximize benefits for consumers. Thus risk aversion is not the complete story, and even when all market participants are relatively risk-neutral ($\gamma_s = \gamma_c = 0.001$), there are still large discrepancies between that solution and the competitive solution. This is due to the effect of market power in the Nash scenarios.

In a perfectly competitive market under the linear cost assumptions, total cost is minimized and suppliers earn zero total profit — short-term gross margins just cover investment costs. However, oligopolies exist in real markets; in our Nash case, the suppliers recognize that reducing their investment will increase spot prices by placing more expensive capacity on the margin more often in the energy market. The SO counters this incentive to reduce investment by increasing capacity payments, but the result is still higher costs for consumers and positive profits for suppliers.

As a result, both risk aversion and Nash behavior tend to distort capacity mixes relative to the fully competitive case. In particular, the cost-inferior technologies (renewables, coal, and

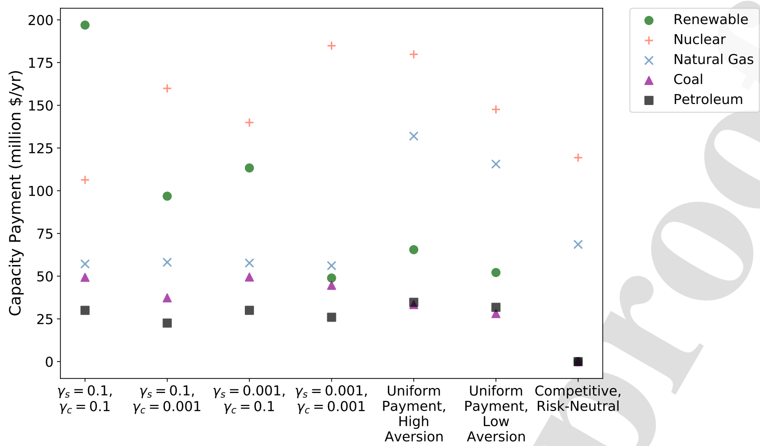


Figure 3: Model Outcome: Capacity Payment per Technology

petroleum) always enter the market in the Nash case, which is possible because prices are roughly twice as high as the competitive case and they get higher capacity payment. For instance, renewable capacity, which gets no dispatch or payment in the competitive solution, may get even more payment than cost-efficient nuclear/gas capacity, especially when risk aversion is high. It may appear surprising that the Stackelberg leader (SO) encourages the entry of inefficient technologies by requiring consumers to pay for their capacity. The explanation of this phenomenon is that from a consumer cost perspective, the cost inefficiency from building the inefficient technologies must be more than compensated for by the lower prices that result from more intensive competition if there are five sizable suppliers rather than a duopoly of just nuclear and natural gas. But these second-best prices are only partially effective in promoting consumer welfare, because consumer costs are still more than double what they are in the competitive solution.

Another noteworthy difference between the oligopolistic and competitive solutions is the discriminatory capacity payments to different technologies. In a competitive, risk-neutral market, the optimal capacity payment is uniform among all suppliers who build capacity in order to provide just enough incentives to make up for the "missing money". However, our model suggests in the second-best world, it's optimal to provide higher capacity payments to some technologies to increase their market share. To shed some light on why this might be true, we consider the "uniform payment" solution, where we require that all technologies receive equal capacity payments in addition to the existing constraints in ($P_{SO}$). The results are summarized in Tables 2–3 and Figures 4–5 as two additional scenarios denoted by "Uniform Payment". We consider two different riskiness levels: the high risk-aversion case $\gamma_s = \gamma_c = 0.1$ (second-to-last row in Tables 2–3) and the low risk-aversion case $\gamma_s = \gamma_c = 0.001$ (last row). Comparing uniform and discriminatory payment solutions under conditions of market power (comparing the second-to-last row to the first row, or the last row to the fourth row in the table), we see a rise in production costs and prices due to loss of baseload capacities, and an increase in suppliers' profits as well as total capacity payments. This testifies to the efficiency benefits arising from allowing the SO to discriminate among technologies when paying for capacity.

As a final remark, we compare the uniform capacity payments without and with market power (comparing the last row to the second- or third-to-last row in the table), and observe that the efficiency improvement does not only come from discriminatory capacity payments, but also from

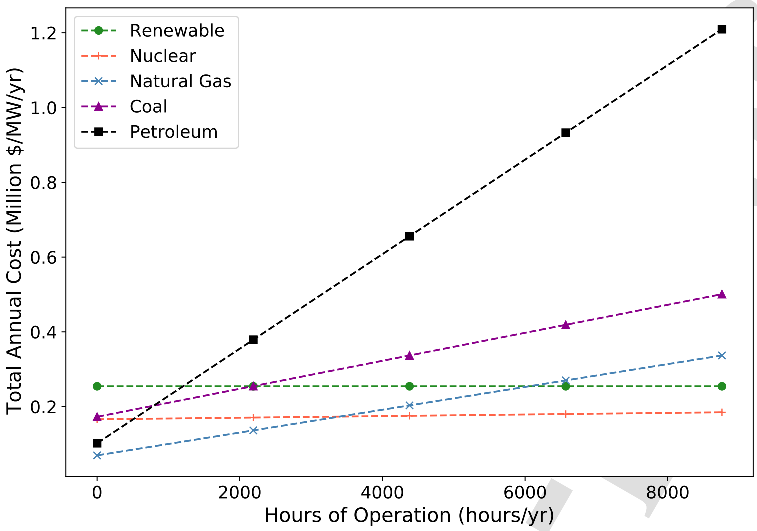


Figure 4: Screening Curves

elimination of market power: The capacity payments with market power (the second- or third-to-last row in the table) can be twice as high as that in the competitive solution (the last row) and accompanied by substantial suppliers' profits, which we attribute to the market power that suppresses capacity in order to increase profits, whether they come from capacity payments or raised energy prices.

### 4.4 Effect of Alternative Price Caps

Recall that in the examples of Sections 4.2–4.3, we imposed a price cap of \$2000/MWh with VOLL fixed at \$6000/MWh. As pointed out in Section 1, a regulator might be interested in seeking an optimal price cap that promotes system efficiency. Theorems 3.10–3.11 indicate that a high reliability requirement is more likely to be fulfilled with high price caps, and this implies that it could be interesting to ask the question what price cap would maximize consumers' utility while guaranteeing target reliability. Such an optimization requires an extra layer (a federal/political level) on top of the existing model, which we do not attempt to model in the present paper. In this subsection, we test our model with different levels of price caps (gradually increased from the previous \$2000/MWh to the value of VOLL), in order to provide some initial results on how price caps impact technology mix, consumer costs, and profits. We maintain a relatively high level of risk aversion $\gamma_s = \gamma_c = 0.1$. The corresponding model outcomes are summarized in Tables 4–5. The first rows (price cap=\$2000/MWh) in the tables are identical to the first rows ($\gamma_s = \gamma_c = 0.1$) in Tables 2–3.

As one may expect, Table 5 shows that as the price cap increases, the regulator demands a smaller amount of total capacity payment in order to incentivize capacity investment. This is consistent with established results in analyses of competitive markets in which capacity payments are proposed as a remedy to the "missing money" problem incurred by price caps (e.g., [23][34]). Seemingly counterintuitive, however, is the result that the energy prices as well as supplier profits decrease as the price cap rises. Concerning the technology mix, we observe that a higher price cap results in a slightly larger baseload capacity (renewable together with nuclear), but a significantly

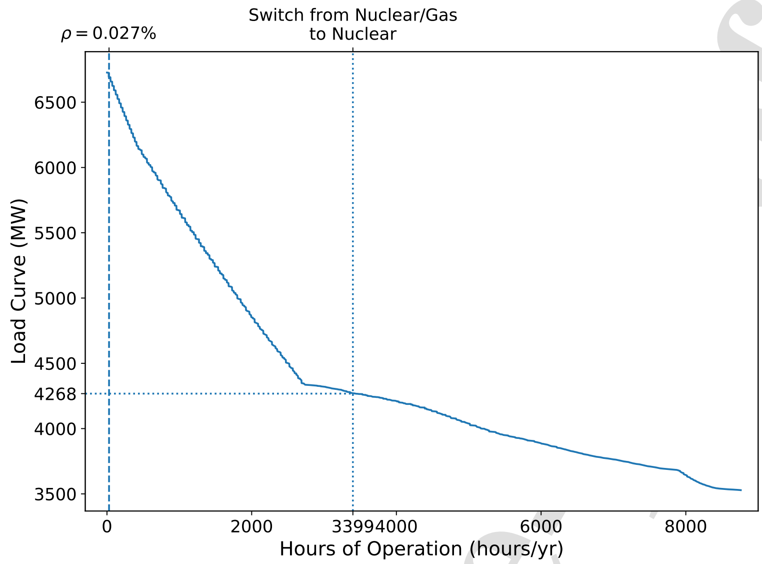


Figure 5: Load Duration Curve

larger capacity from cost-efficient technologies (nuclear together with natural gas). This is because as the price cap increases and prices decrease, there is less margin for the inefficient producers to stay in the market and they gradually shrink their market share. In particular, despite a drop in capacity for the marginally cheapest power (renewable), the increase in nuclear capacity is such that the overall baseload capacity increases.

The interactions of market power, price caps, and risk aversion are complex, and we leave further analysis of their effects to future research. This example serves to illustrate that these interactions can lead to counterintuitive outcomes, and that raising price caps can, in at least some cases, increase market efficiency and consumer welfare.

### 4.5 Remarks on Optimal Reliability

In our model we impose an upper-bound constraint on the SO's optimization (the first constraint in ($P_{SO}$)): it wants to at least attain a targeted grid reliability $\rho$ and therefore tries to incentivize energy providers to comply with it. The exogenously specified reliability standard, LOLE of 2.4 hours/year, is one interpretation of the common US industry standard of one day's outage in 10 years,[8] and is imposed in our base numerical simulations. A relevant question is whether a different (potentially less stringent) reliability can in fact lead to higher consumer welfare, despite the higher risk of load shedding. We try to shed some light by removing the reliability constraint in ($P_{SO}$), and investigate how consumers' riskiness affects the reliability they prefer.

The results in Tables 6–7 correspond to the scenario where the SO pays capacity payments to maximize consumer utility, without imposing a separate constraint upon reliability. We suppose both the price cap and VOLL to be \$6000/MWh, and suppliers are almost risk-neutral ($\gamma_s = 0.001$). We consider two cases: when consumers are almost risk neutral with $\gamma_c = 0.001$, and when they are more risk-averse with $\gamma_c = 0.1$.

[8]See, e.g., `https://pubs.naruc.org/pub/FA865D94-FA0B-F4BA-67B3-436C4216F135`

| Price Cap | | Renewable | Nuclear | Natural Gas | Coal | Petroleum |
|---|---|---|---|---|---|---|
| 2000 | $\boldsymbol{K}^*$ | 1999.65 | 1881.70 | 1753.66 | 552.80 | 531.98 |
| | $\boldsymbol{C}^*$ | 98.49 | 56.52 | 32.60 | 82.25 | 56.38 |
| 3000 | $\boldsymbol{K}^*$ | 1778.46 | 2226.95 | 1800.52 | 461.72 | 452.36 |
| | $\boldsymbol{C}^*$ | 96.95 | 52.48 | 30.37 | 80.02 | 52.42 |
| 4000 | $\boldsymbol{K}^*$ | 1631.94 | 2432.00 | 1842.38 | 389.15 | 425.18 |
| | $\boldsymbol{C}^*$ | 92.14 | 50.03 | 29.32 | 75.67 | 47.92 |
| 5000 | $\boldsymbol{K}^*$ | 1522.83 | 2587.3 | 1897.92 | 331.03 | 381.85 |
| | $\boldsymbol{C}^*$ | 89.03 | 48.29 | 27.51 | 69.22 | 44.85 |
| 6000 | $\boldsymbol{K}^*$ | 1465.56 | 2703.22 | 1928.6 | 294.92 | 330.16 |
| | $\boldsymbol{C}^*$ | 87.21 | 45.43 | 26.64 | 61.8 | 38.44 |

Table 4: Market Power/Discriminatory Payment Model Outcomes with Different Price Caps

| Price Cap | Consumer Payment left(¢/kWh) | | Supplier Revenue left(Million \$/year) | | Total Cost left(Million \$/year) | | Supplier Profit left(Million \$/year) |
|---|---|---|---|---|---|---|---|
| | Energy Price | Total Price | Energy Sales | Capacity Payment | Variable Cost | Capacity Cost | |
| 2000 | 9.56 | 10.76 | 3496.70 | 439.80 | 627.79 | 1093.16 | 2215.56 |
| 3000 | 8.37 | 9.47 | 3061.25 | 404.63 | 573.04 | 1073.62 | 1819.22 |
| 4000 | 7.64 | 8.66 | 2793.97 | 375.88 | 555.28 | 1058.01 | 1556.56 |
| 5000 | 7.00 | 7.96 | 2561.00 | 352.77 | 541.83 | 1045.44 | 1326.50 |
| 6000 | 5.95 | 6.86 | 2177.11 | 332.91 | 523.63 | 1040.75 | 945.65 |

Table 5: Average Revenue and Cost for Market Power/Discriminatory Payments Model with Different Price Caps

We observe a substantially higher LOLE in both cases than the LOLE (2.4 hr/year) imposed in Sections 4.2–4.4. That is, utility-maximizing consumers prefer to keep the (average) price low at the risk of insufficient capacity and more frequent load shedding. As consumers become more risk-averse, they are willing to pay more for capacity in order to avoid possible inadequacy, and achieve a LOLE similar to the cost-efficient value in the corresponding competitive risk-neutral case, which equals 11.66 hr/year.

### 4.6 Incorporation of Supplier Utility in the SO Objective

Before concluding our discussion of results, we would like to remark on how the SO may also incorporate the suppliers' utility into its optimization. In Remark 2, we stated our assumption that the SO only focuses on consumer welfare and hence uses a weighted sum of consumers' utility as its objective. In this subsection we propose a weighting method for the SO to evaluate the total utility of both consumers and suppliers, by using a weight $\beta$ that quantifies the importance of suppliers' welfare relative to consumers'. Given $\beta \in [0, 1)$, the total-utility function is defined as

$$u^{(total)}(\boldsymbol{C}, \boldsymbol{K}; M) := u^{(SO)}(\boldsymbol{C}, \boldsymbol{K}; M) + \frac{\beta}{1-\beta} \frac{\sum_{i=1}^{N_c} N^{(i)}}{N_s} \sum_{j=1}^{N_s} u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M),$$

| $\gamma_c$ | | Renewable | Nuclear | Natural Gas | Coal | Petroleum |
|---|---|---|---|---|---|---|
| 0.001 | $\boldsymbol{K}^*$ | 1433.50 | 2881.32 | 1765.43 | 204.82 | 302.25 |
| | $\boldsymbol{C}^*$ | 92.40 | 58.23 | 22.05 | 42.36 | 30.96 |
| 0.1 | $\boldsymbol{K}^*$ | 1698.82 | 2815.32 | 1524.65 | 252.60 | 318.22 |
| | $\boldsymbol{C}^*$ | 104.82 | 68.58 | 20.70 | 47.64 | 32.75 |

Table 6: Market Power/Discriminatory Payments Model Outcome with No Reliability Requirement

| $\gamma_c$ | Consumer Payment left(¢/kWh) | | Supplier Revenue left(Million \$/year) | | Total Cost left(Million \$/year) | | Supplier Profit left(Million \$/year) | LOLE left(hr/year) |
|---|---|---|---|---|---|---|---|---|
| | Energy Price | Total Price | Energy Sales | Capacity Payment | Variable Cost | Capacity Cost | | |
| 0.001 | 6.16 | 7.13 | 2252.72 | 357.20 | 488.75 | 1032.40 | 1088.77 | 16.56 |
| 0.1 | 5.72 | 6.88 | 2093.85 | 425.16 | 416.16 | 1082.06 | 1020.79 | 12.32 |

Table 7: Market Power/Discriminatory Payments Model: Average Revenue and Cost with No Reliability Requirement

and the total-utility-maximizing SO solves the following problem similar to $(P_{SO})$:

$$\sup_{(\boldsymbol{C},\boldsymbol{K})\in\Omega^{\epsilon}_{\boldsymbol{C}}\times\mathbb{R}^{N_s}_{+}} u^{(total)}(\boldsymbol{C},\boldsymbol{K};M) \quad \text{s.t. } \mathbb{E}\left[\int_0^T \frac{1}{T}\mathbb{I}_{\{D_t>\mathbf{1}_{N_s}(\boldsymbol{K}_0+\boldsymbol{K})\}}dt\right] \le \rho,\ \boldsymbol{K}\in\boldsymbol{K}^*(\boldsymbol{C};M). \qquad (P^{(total)}_{SO})$$

Due to the monotonicity of the individual utility functions, the total-utility function provides Pareto-type optimality among all consumers and suppliers, but extra caution is needed to choose the value of $\beta$ as illustrated in the following example.

Table 8–9 shows outcomes of $(P^{(total)}_{SO})$ in a relatively risk-averse setting $\gamma_s = \gamma_c = 0.1$ with a price cap of \$2000/MWh. The first case $\beta = 0$ corresponds to the consumer-utility-maximizing model in previous sections, and the numbers are identical to the first rows ($\gamma_s = \gamma_c = 0.1$) in Tables 2–3. The second case $\beta = 2\times10^{-6}$ corresponds to the naive case where the total-utility is a simple sum of the utility of all suppliers and consumers, or equivalently, $\beta = \frac{N_s}{N_s+\sum_{i=1}^{N_c} N^{(i)}}$. The last case shows an equal weight $\beta = 0.5$ for consumers and suppliers.

| $\beta$ | | Renewable | Nuclear | Natural Gas | Coal | Petroleum |
|---|---|---|---|---|---|---|
| 0 | $\boldsymbol{K}^*$ | 1999.65 | 1881.70 | 1753.66 | 552.80 | 531.98 |
| | $\boldsymbol{C}^*$ | 98.49 | 56.52 | 32.60 | 82.25 | 56.38 |
| $2\times10^{-6}$ | $\boldsymbol{K}^*$ | 1999.86 | 1879.71 | 1756.86 | 553.25 | 530.28 |
| | $\boldsymbol{C}^*$ | 98.43 | 55.32 | 33.41 | 85.29 | 56.12 |
| 0.5 | $\boldsymbol{K}^*$ | 1612.25 | 1767.03 | 1941.42 | 775.28 | 623.84 |
| | $\boldsymbol{C}^*$ | 56.68 | 36.85 | 21.82 | 91.21 | 60.92 |

Table 8: Market Power/Discriminatory Payments Model Outcomes When Including Weighted Supplier Utility in the SO Objective

As one may expect, the results for the first two cases are similar, because the few suppliers' utility values are diluted when simply added to the large sum of consumers' utility. Comparing the $\beta = 0.5$ case with the first case, we observe a drastic increase in the suppliers' profit as a result

| $\beta$ | Consumer Payment left(¢/kWh) | | Supplier Revenue left(Million \$/year) | | Total Cost left(Million \$/year) | | Supplier Profit left(Million \$/year) |
|---|---|---|---|---|---|---|---|
| | Energy Price | Total Price | Energy Sales | Capacity Payment | Variable Cost | Capacity Cost | |
| 0 | 9.56 | 10.76 | 3496.70 | 439.80 | 627.79 | 1093.16 | 2215.56 |
| $2 \times 10^{-6}$ | 9.57 | 10.77 | 3503.20 | 436.47 | 628.53 | 1093.00 | 2218.14 |
| 0.5 | 13.93 | 14.77 | 5098.83 | 307.58 | 1032.25 | 1036.45 | 3337.71 |

Table 9: Average Revenue and Cost for Market Power/Discriminatory Payments Model When Including Weighted Supplier Utility in the SO Objective

of the SO considering the suppliers' profitability and allowing energy prices to be more frequently set by technologies that produce expensive electricity. In that situation, the suppliers need lower capacity payments to recover their investment costs, yet the consumers suffer from higher prices, which becomes a major revenue increase for the suppliers. We note that when the SO weights suppliers' and consumers' equally, the total cost increases, and the system is less efficient than the scheme where it only considers consumers' welfare. We interpret this loss of system efficiency as a result of the regulator overweighting suppliers in total utility evaluation, and the SO must appropriately choose the value of $\beta$ while fully recognizing the suppliers' advantage over consumers granted by their market power.

## 5 Conclusions

We model the system operator's and suppliers' decisions on capacity payments and investments, respectively, with a bi-level model that is solved as a mathematical program with equilibrium constraints (MPEC). In the lower (Stackelberg follower) level of the model, suppliers play a Nash game in capacity investments, considering how capacity investments affect spot prices. In the upper (Stackelberg leader) level, the system operator (SO) chooses capacity payments such that the suppliers voluntarily build enough capacity to guarantee network reliability, and the consumers are, to the extent achievable by the SO, protected from power shortage and high prices. Unlike the classical (implicit) fixed-point approach (e.g., [50]), we propose an explicit open-loop approach (via variational inequalities) to compute the suppliers' Nash equilibrium. We also show the SO's problem can be solved via standard constrained optimization methods, and how values for two market constraint parameters, reliability requirement and scarcity price, can be selected to guarantee model solvability.

The model yields several results that differ from the literature. One is that a stochastic bi-level model can be formulated and solved to recommend capacity payments when suppliers are risk-averse, play a Nash game in capacity investment, and participate in a spot energy market that is subject to effective short-run market power mitigation. Another result is that differentiated capacity payments can be second-best; standard analyses, such as [62], instead define a uniform payment (perhaps adjusted for plant reliability) that is made to all types of capacity. Differentiation can be optimal if different suppliers have different amounts of market power. However, as the numerical example illustrates, such second-best payments are in general insufficient to fully mitigate long-run market power resulting from the ability to refrain from investing in capacity. Our hypothetical example illustrates that the level of those payments can represent a trade-off between reducing market concentration through encouraging entry of inefficient additional suppliers, and the resulting increase in electricity production costs.

An interesting topic for future work is to endogenize part of the market parameters by adding one more layer of optimization. As we pointed out earlier in this paper, the input parameters to our model (such as capital costs, risk aversion, the price cap, and target reliability) can be endogenized, in order to get more applicable results and to address market design questions such as "What is the optimal price cap in terms of system welfare?" Additional analysis of the effect of alternative objective functions for the Stackelberg system operator is also of interest. We provide initial investigation of these questions in Sections 4.4–4.6 , and leave more detailed work for future research. Numerically, a broader list of questions includes more realistic parameterizations of the model and inclusion of additional uncertainties such as fuel price and renewable output fluctuations, and policy uncertainties.

# Appendices

## Appendix A Proofs of Lemmas and Propositions

*Proof (Lemma 3.1).* For continuity and smoothness see the closed-form formula that follows. The second statement is a result from existence and closeness of continuous VI problems on compact sets (e.g., [25],[44]).

For simplicity we will omit the subscript $s$ in $\gamma_s^{(j)}, r_s^{(j)}$ when there is no obscurity. We have

$$\begin{aligned}
&u^{(j)}(K^{(j)};\boldsymbol{K}^{(-j)},\boldsymbol{C},M)\\
=&-\int_0^T \mathbb{E}\left[e^{-r^{(j)}t-\gamma^{(j)}(C^{(j)}-F^{(j)})K^{(j)}}\mathbb{I}_{\{D_t\le \mathbf{1}_j^T(\boldsymbol{K}_0+\boldsymbol{K})\}}\right]\\
&-\sum_{l=j+1}^{N_s}\mathbb{E}\left[e^{-r^{(j)}t-\gamma^{(j)}\left((C^{(j)}-F^{(j)})K^{(j)}+(G^{(l)}-G^{(j)})(K_0^{(j)}+K^{(j)})\right)}\mathbb{I}_{\{\mathbf{1}_{l-1}^T(\boldsymbol{K}_0+\boldsymbol{K})<D_t\le \mathbf{1}_l^T(\boldsymbol{K}_0+\boldsymbol{K})\}}\right]\\
&-\mathbb{E}\left[e^{-r^{(j)}t-\gamma^{(j)}\left((C^{(j)}-F^{(j)})K^{(j)}+(M-G^{(j)})(K_0^{(j)}+K^{(j)})\right)}\mathbb{I}_{\{D_t>\mathbf{1}_{N_s}^T(\boldsymbol{K}_0+\boldsymbol{K})\}}\right]dt\\
=&-e^{-\gamma^{(j)}(C^{(j)}-F^{(j)})K^{(j)}}\int_0^T e^{-r^{(j)}t}Q_t^{(j)}(\boldsymbol{K})dt\\
&-\sum_{l=j+1}^{N_s}e^{-\gamma^{(j)}(G^{(l)}-G^{(j)})K_0^{(j)}-\gamma^{(j)}(C^{(j)}-F^{(j)}+G^{(l)}-G^{(j)})K^{(j)}}\int_0^T e^{-r^{(j)}t}\left(Q_t^{(l)}(\boldsymbol{K})-Q_t^{(l-1)}(\boldsymbol{K})\right)dt\\
&-e^{-\gamma^{(j)}(M-G^{(j)})K_0^{(j)}-\gamma^{(j)}(C^{(j)}-F^{(j)}+M-G^{(j)})K^{(j)}}\int_0^T e^{-\gamma^{(j)}t}\left(1-Q_t^{(N_s)}(\boldsymbol{K})\right)dt,
\end{aligned}\tag{7}$$

with $Q_t^{(l)}(\boldsymbol{K}) := \Phi\left(\frac{\mathbf{1}_l^T(\boldsymbol{K}_0+\boldsymbol{K})-\mu_t}{\sigma_t}\right)$.

Note we can switch differentiation and integration, and get

$$\begin{aligned}
&\frac{\partial}{\partial K^{(j)}}u^{(j)}(K^{(j)};\boldsymbol{K}^{(-j)},\boldsymbol{C},M)\\
=&e^{-\gamma^{(j)}(C^{(j)}-F^{(j)})K^{(j)}}\int_0^T e^{-r^{(j)}t}\left(\gamma^{(j)}(C^{(j)}-F^{(j)})Q_t^{(j)}(\boldsymbol{K})-\frac{1}{\sigma_t}q_t^{(j)}(\boldsymbol{K})\right)dt\\
&+\sum_{l=j+1}^{N_s}e^{-\gamma^{(j)}(G^{(l)}-G^{(j)})K_0^{(j)}-\gamma^{(j)}(C^{(j)}-F^{(j)}+G^{(l)}-G^{(j)})K^{(j)}}\\
&\times\int_0^T e^{-r^{(j)}t}\Big(\gamma^{(j)}(C^{(j)}-F^{(j)}+G^{(l)}-G^{(j)})\big(Q_t^{(l)}(\boldsymbol{K})-Q_t^{(l-1)}(\boldsymbol{K})\big)\\
&-\frac{1}{\sigma_t}\big(q_t^{(l)}(K^{(j)};\boldsymbol{K}^{(-j)})-q_t^{(l-1)}(K^{(j)};\boldsymbol{K}^{(-j)})\big)\Big)dt\\
&+e^{-\gamma^{(j)}(M-G^{(j)})K_0^{(j)}-\gamma^{(j)}(C^{(j)}-F^{(j)}+M-G^{(j)})K^{(j)}}\\
&\times\int_0^T e^{-r^{(j)}t}\Big(\gamma^{(j)}(C^{(j)}-F^{(j)}+M-G^{(j)})\big(1-Q_t^{(N_s)}(K^{(j)};\boldsymbol{K}^{(-j)})\big)+\frac{1}{\sigma_t}q_t^{(N_s)}(K^{(j)};\boldsymbol{K}^{(-j)})\Big)dt,
\end{aligned}\tag{8}$$

with $q_t^{(l)}(K^{(j)};\boldsymbol{K}^{(-j)}) := \phi\left(\frac{\mathbf{1}_l^T(\boldsymbol{K}_0+\boldsymbol{K})-\mu_t}{\sigma_t}\right)$. $\square$

*Proof (Lemma 3.2).* It suffices to compute $\bar{\boldsymbol{K}}_\epsilon$ such that for each $j \in \{1,\dots,N_s\}$, we have that $\bar{K}_\epsilon^{(j)} \geq \boldsymbol{0}$, such that for any $\boldsymbol{K}^{(-j)} \geq 0, 0 \leq \boldsymbol{C} \leq \boldsymbol{F} - \epsilon$, it holds that $\frac{\partial}{\partial K} u^{(j)}(K; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M) < 0$ for all $K \geq \bar{K}_\epsilon^{(j)}$.

Recall by definition $0 \leq G^{(j)} \leq G^{(l)} < M$ for all $j < l \leq N_s$. Following (8),we have

$$
\begin{aligned}
& e^{\gamma^{(j)}(C^{(j)}-F^{(j)})K^{(j)}} \cdot \frac{\partial}{\partial K^{(j)}} u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M) \\
=& \gamma^{(j)}(C^{(j)} - F^{(j)}) \int_0^T e^{-r^{(j)}t} Q_t^{(j)}(\boldsymbol{K})dt - \int_0^T \frac{1}{\sigma_t} e^{-r^{(j)}t} q_t^{(j)}(\boldsymbol{K})dt + \sum_{l=j+1}^{N_s} e^{-\gamma^{(j)}(G^{(l)}-G^{(j)})(K_0^{(j)}+K^{(j)})} \\
& \times \int_0^T e^{-r^{(j)}t} \gamma^{(j)}(C^{(j)} - F^{(j)} + G^{(l)} - G^{(j)})(Q_t^{(l)}(\boldsymbol{K}) - Q_t^{(l-1)}(\boldsymbol{K}))dt - \frac{1}{\sigma_t}(q_t^{(l)}(\boldsymbol{K}) - q_t^{(l-1)}(\boldsymbol{K}))dt \\
& + e^{-\gamma^{(j)}(M-G^{(j)})(K_0^{(j)}+K^{(j)})} \int_0^T e^{-r^{(j)}t} \gamma^{(j)}(C^{(j)} - F^{(j)} + M - G^{(j)})(1 - Q_t^{(N_s)}(\boldsymbol{K}))dt \\
& + \frac{1}{\sigma_t} q_t^{(N_s)}(\boldsymbol{K})dt \\
\leq & - \sum_{l=j+1}^{N_s} \int_0^T \frac{e^{-r^{(j)}t}}{\sigma_t} \left( \frac{1}{N_s} q_t^{(j)}(\boldsymbol{K}) + e^{-\gamma^{(j)}(G^{(l)}-G^{(j)})(K_0^{(j)}+K^{(j)})} \left( q_t^{(l)}(\boldsymbol{K}) - q_t^{(l-1)}(\boldsymbol{K}) \right) \right) dt \\
& - \int_0^T \frac{e^{-r^{(j)}t}}{\sigma_t} \left( \frac{1}{N_s} q_t^{(j)}(\boldsymbol{K}) - e^{-\gamma^{(j)}(M-G^{(j)})(K_0^{(j)}+K^{(j)})} q_t^{(N_s)}(\boldsymbol{K}) \right) dt + \sum_{l=j+1}^{N_s} \gamma^{(j)} \int_0^T e^{-r^{(j)}t} \\
& \times \Big( e^{-\gamma^{(j)}(G^{(l)}-G^{(j)})(K_0^{(j)}+K^{(j)})} (C^{(j)} - F^{(j)} + G^{(l)} - G^{(j)})(Q_t^{(l)}(\boldsymbol{K}) - Q_t^{(l-1)}(\boldsymbol{K})) \\
& + \frac{C^{(j)} - F^{(j)}}{N_s} Q_t^{(j)}(\boldsymbol{K}) \Big) dt - \gamma^{(j)} \int_0^T e^{-r^{(j)}t} \Big( \frac{C^{(j)} - F^{(j)}}{N_s} Q_t^{(j)}(\boldsymbol{K}) \\
& + e^{-\gamma^{(j)}(M-G^{(j)})(K_0^{(j)}+K^{(j)})} (C^{(j)} - F^{(j)} + M - G^{(j)})(1 - Q_t^{(N_s)}(\boldsymbol{K})) \Big) dt.
\end{aligned} \tag{9}
$$

Now take $K^{(j)} \geq \max_{t \geq 0}\{\mu_t\}$, and observe that at all $t$, we have $0 \leq Q_t^{(l)} - Q_t^{(l-1)} < Q_t^{(j)}$, $0 < 1 - Q_t^{(N_s)} \leq Q_t^{(j)}$, $-q_t^{(j)} < q_t^{(l)} - q_t^{(l-1)} < 0$, $0 < q_t^{(N_s)} \leq q_t^{(j)}$ for all $1 \leq j < l \leq N_s$. Denote $\mathcal{I}_1^{(j)} := \{l \in \{j+1,\dots,N_s\} : G^{(l)} > G^{(j)}\}$, $\mathcal{I}_2^{(j)} := \{l \in \{j+1,\dots,N_s\} : G^{(l)} \leq G^{(j)} - (C^{(j)} - F^{(j)})\}$. Continuing with (9), we have

$$
\begin{aligned}
& e^{\gamma^{(j)}(C^{(j)}-F^{(j)})K^{(j)}} \cdot \frac{\partial}{\partial K^{(j)}} u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}) \\
<& - \sum_{l \notin \mathcal{I}_1} \int_0^T e^{-r^{(j)}t} \Big( \big((C^{(j)} - F^{(j)}) Q_t^{(l)}(\boldsymbol{K}) - \frac{1}{\sigma_t} q_t^{(l)}\big) - \big((C^{(j)} - F^{(j)}) Q_t^{(l-1)}(\boldsymbol{K}) - \frac{1}{\sigma_t} q_t^{(l-1)}(\boldsymbol{K})\big) \Big) dt \\
& - \int_0^T \frac{e^{-r^{(j)}t}}{\sigma_t} \Big( \sum_{l \in \mathcal{I}_1} \big( \frac{1}{N_s} - e^{-\gamma^{(j)}(G^{(l)}-G^{(j)})(K_0^{(j)}+K^{(j)})} \big) + \frac{1}{N_s} - e^{-\gamma^{(j)}(M-G^{(j)})(K_0^{(j)}+K^{(j)})} \Big) q_t^{(j)}(\boldsymbol{K})dt \\
& + \int_0^T e^{-r^{(j)}t} \Big( \sum_{l \in \mathcal{I}_2} \Big( \frac{C^{(j)} - F^{(j)}}{N_s} + e^{-\gamma^{(j)}(G^{(l)}-G^{(j)})(K_0^{(j)}+K^{(j)})} (C^{(j)} - F^{(j)} + G^{(l)} - G^{(j)}) \Big) Q_t^{(j)}(\boldsymbol{K}) \\
& - \Big( \frac{C^{(j)} - F^{(j)}}{N_s} + e^{-\gamma^{(j)}(M-G^{(j)})(K_0^{(j)}+K^{(j)})} (C^{(j)} - F^{(j)} + M - G^{(j)}) \Big) Q_t^{(j)}(\boldsymbol{K}) \Big) \\
& \times \mathbb{I}_{\{M \geq G^{(j)} - (C^{(j)} - F^{(j)})\}} dt \cdot \gamma^{(j)}.
\end{aligned} \tag{10}
$$

A sufficient condition such that (10) is nonpositive for all $\boldsymbol{K}^{(-j)} \geq \boldsymbol{0}, \boldsymbol{0} \leq \boldsymbol{C} < \boldsymbol{F} - \epsilon$, is

$$
\begin{aligned}
K^{(j)} \geq \bar{K}_\epsilon^{(j)} & \\
:= \max \Big\{ & \max_{t\geq 0}\{\mu_t\}, F^{(j)} \max_{t\geq 0}\{\sigma_t\}, \frac{\log N_s}{\gamma^{(j)}(G^{(l_j)} - G^{(j)})}, \frac{\log N_s + \log\left(1 + \frac{M - G^{(j)}}{\epsilon}\right)}{\gamma^{(j)}\left(M - G^{(j)}\right)} \mathbb{I}_{\{M \geq G^{(j)} + \epsilon\}}, \\
& \max_{j < l \leq N_s} \Big\{ \frac{\log N_s + \log\left(1 + \frac{G^{(l)} - G^{(j)}}{\epsilon}\right)}{\gamma^{(j)}\left(G^{(l)} - G^{(j)}\right)} : G^{(l)} \geq G^{(j)} + \epsilon \Big\} \Big\},
\end{aligned} \tag{11}
$$

with $l_j := \min\{l : l \in \mathcal{I}_1^{(j)}\}$.

□

*Proof (Proposition 3.3).* The statements follow from computation based on (11). □

*Proof (Lemma 3.4).* The goal is to show for for any given $j$ and $\boldsymbol{K}^{(-j)} \geq \boldsymbol{0}$, $u^{(j)}(\cdot; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M)$ is pseudo-concave for all $\boldsymbol{C} \in \Omega^\epsilon_{\boldsymbol{C}}$. It suffices to show $\frac{\partial^2}{(\partial K^{(j)})^2} u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M) < 0$ whenever $\frac{\partial}{\partial K^{(j)}} u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M) = 0$. From Lemma 3.2, we only need to consider $\boldsymbol{K} \in \operatorname{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)$, because any point outside would have a negative gradient.

Fix $j, \boldsymbol{K}^{(-j)}$. For convenience, we will omit the index $j$ in following quantities. Denote for $l \geq j$ that $\Delta G^{(l)} := G^{(l)} - G^{(j)} \geq 0$. By nature of the merit order dispatch, we note $\Delta G^{(l)} \leq \Delta G^{(l-1)}$. Without loss of generality we further assume $\Delta G^{(l)} < \Delta G^{(l-1)}$, otherwise the $l$- and $(l+1)$-th suppliers can be merged and the computation can be processed in an identical way.

Given $M$, consider $K^{(j)}$ such that $\frac{\partial}{\partial K^{(j)}} u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M) = 0$. If $K_0^{(j)} + K^{(j)} = 0$, we have

$$
\begin{aligned}
& \frac{\partial^2}{(\partial K^{(j)})^2} u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}) \\
= & -\left(\gamma^{(j)}(C^{(j)} - F^{(j)} + M - G^{(j)})\right)^2 J_1 + 2\gamma^{(j)}(C^{(j)} - F^{(j)} + M - G^{(j)}) J_2 + I_0,
\end{aligned}
$$

where $J_1 := \int_0^T e^{-r^{(j)}t}\Big(1 - \Phi\big(\frac{\boldsymbol{1}_{N_s}^T(\boldsymbol{K}+\boldsymbol{K}_0)}{\sigma_t}\big)\Big)dt$, $J_2 := \int_0^T e^{-r^{(j)}t} \frac{1}{\sigma_t} \phi\big(\frac{\boldsymbol{1}_{N_s}^T(\boldsymbol{K}+\boldsymbol{K}_0)}{\sigma_t}\big)\Big)dt$, and $I_0$ is a continuous function of $\boldsymbol{K}, \boldsymbol{C}$ and independent of $M$. Note we restrict $\boldsymbol{K} \in \operatorname{cl}\left(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\right)$, so $J_1, J_2, I_0$ are bounded and $J_1$ is bounded away from 0. That allows us to claim that if $K_0^{(j)} = 0$ and $\frac{\partial}{\partial K^{(j)}} u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M) = 0$ at $K^{(j)} = 0$, then $u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}) < 0$ will eventually be concave at $K^{(j)} = 0$ for sufficiently large $M$. In the following computation, we fix such a large $M$ and consider $K^{(j)}$ such that $\frac{\partial}{\partial K^{(j)}} u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M) = 0$ and $K_0^{(j)} + K^{(j)} > 0$.

Denote $R^l := \int_0^T e^{-r^{(j)}t} \Phi\big(\frac{\boldsymbol{1}_{N_s}^T(\boldsymbol{K}_0+\boldsymbol{K}) - \mu_t}{\sigma_t}\big)dt > 0$, $R_1^l := \frac{\partial}{\partial K^{(j)}} R^l > 0$, $R_2^l := \frac{\partial^2}{(\partial K^{(j)})^2} R_l$, and note they are all bounded. For arbitrary $\lambda > 0$, at $K^{(j)} > 0$ such that $\frac{\partial}{\partial K^{(j)}} u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M) = 0$, we have

$$
\begin{aligned}
& e^{\gamma^{(j)}(C^{(j)} - F^{(j)})K^{(j)}} \frac{\partial^2}{(\partial K^{(j)})^2} u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M) \\
= & e^{\gamma^{(j)}(C^{(j)} - F^{(j)})K^{(j)}} \left( \frac{\partial^2}{(\partial K^{(j)})^2} u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M) + \lambda \frac{\partial}{\partial K^{(j)}} u^{(j)}(K^{(j)}; \boldsymbol{K}^{(-j)}, \boldsymbol{C}, M_0) \right) \\
= & \sum_{l=j}^{N_s - 1} I_1^{(l)} + \sum_{l=j}^{N_s - 1} I_2^{(l)} + \sum_{l=j}^{N_s} I_3^{(l)} + I_4,
\end{aligned} \tag{12}
$$

(The last step is a result of rearranging terms in $\frac{\partial}{\partial K^{(j)}}u^{(j)}(K^{(j)};\boldsymbol{K}^{(-j)},\boldsymbol{C},M)=0.$) with

$$
\begin{aligned}
I_1^{(l)} &:= 2\big(-g_1(K_0^{(j)}+K^{(j)};\gamma^{(j)}\Delta G^{(l)})+g_1(K_0^{(j)}+K^{(j)};\gamma^{(j)}\Delta G^{(l+1)})\big)R_1^l,\\
I_2^{(l)} &:= \big(g_2^{\lambda}(K_0^{(j)}+K^{(j)};\gamma^{(j)}\Delta G^{(l)})-g_2^{\lambda}(K_0^{(j)}+K^{(j)};\gamma^{(j)}\Delta G^{(l+1)})\big)R^l,\\
I_3^{(l)} &:= \big(-e^{-\gamma^{(j)}\Delta G^{(l)}(K_0^{(j)}+K^{(j)})}+e^{-\gamma^{(j)}\Delta G^{(l+1)}(K_0^{(j)}+K^{(j)})}\big)(\lambda R_1^l-R_2^l),\\
I_4 &:= \gamma^{(j)}(C^{(j)}-F^{(j)}+M_0-G)e^{-\gamma^{(j)}(M-G)(K_0^{(j)}+K^{(j)})},
\end{aligned}
$$

where $g_1(x;a):=(\gamma^{(j)}(C^{(j)}-F^{(j)})+a)e^{-ax}$, $g_2^{\lambda}(x;a):=\big(\lambda\big(\gamma^{(j)}(C^{(j)}-F^{(j)})+a\big)+\big(\gamma^{(j)}(C^{(j)}-F^{(j)})+a\big)^2\big)e^{-\gamma^{(j)}ax}$. Since $\frac{\partial}{\partial K^{(j)}}u^{(j)}(K^{(j)};\boldsymbol{K}^{(-j)},\boldsymbol{C},M)=0$, altering $\lambda$ in (12) does not change its value. We show that there is $\lambda$ such that (12) is negative.

Note $I_1^{(l)}, I_4$, as continuous functions on $(\boldsymbol{K},\boldsymbol{C})\in \mathrm{cl}\,\big(\mathcal{S}(\bar{\boldsymbol{K}}_\epsilon)\big)\times\boldsymbol{\Omega}_{\boldsymbol{C}}$, are bounded for all $\boldsymbol{C}$.

For $I_2^{(l)}$, recall by assumption $K_0^{(j)}+K^{(j)}>0$. We have $\frac{\partial}{\partial a}g_2^{\lambda}(K_0^{(j)}+K^{(j)};a)<0$ for $a\in\big(\gamma^{(j)}\Delta G^{(l)},\gamma^{(j)}\Delta G^{(l+1)}\big)$ and hence $g_2^{\lambda_2}(K_0^{(j)}+K^{(j)};\gamma^{(j)}\Delta G^{(l)})-g_2^{\lambda_2}(K_0^{(j)}+K^{(j)};\gamma^{(j)}\Delta G^{(l+1)})<0$, as soon as $\lambda\geq\frac{2}{K_0^{(j)}+K^{(j)}}$. It takes no further effort to claim $I_2^{(l)}<0$ for sufficiently large $\lambda$.

For $I_3^{(l)}$, note $R_1^l, R_2^l$ are continuous in $\boldsymbol{K}$ with $R_1^l>0$, and achieve their extreme values on the compact region. So $R_2^l$ is bounded and $R_1^l$ is bounded away from 0. Again recall $K_0^{(j)}+K^{(j)}>0$, and hence $e^{-\gamma^{(j)}\Delta G^{(l)}(K_0^{(j)}+K^{(j)})}>e^{-\gamma^{(j)}\Delta G^{(l+1)}(K_0^{(j)}+K^{(j)})}$. So $I_3^{(l)}$ can be arbitrarily large and negative for sufficiently large $\lambda$.

Putting the four pieces together, we have $\frac{\partial^2}{(\partial K^{(j)})^2}u^{(j)}(K^{(j)};\boldsymbol{K}^{(-j)},\boldsymbol{C},M)<0$ if $K_0^{(j)}+K^{(j)}>0$ and $\frac{\partial}{\partial K^{(j)}}u^{(j)}(K^{(j)};\boldsymbol{K}^{(-j)},\boldsymbol{C},M)=0$, which holds for any $M$. Recall that we have chosen $M$ large enough such that any stationary point with $K_0^{(j)}+K^{(j)}=0$ is concave. That completes the proof. □

*Proof (Lemma 3.7).* Let $L(\mathbb{R}^{N_s},\mathbb{R})$ be the space of linear mappings from $\mathbb{R}^{N_s}$ to $\mathbb{R}$. For fixed $\boldsymbol{C},M$, the mapping $\mathcal{T}:\mathbb{R}_+^{N_s}\to L(\mathbb{R}^{N_s},\mathbb{R})$ with $\mathcal{T}(\boldsymbol{K}):=\langle\nabla\boldsymbol{u}(\boldsymbol{K};\boldsymbol{C},M),\cdot\rangle$ is continuous. With $\mathbb{R}_+,\mathbb{R}_+^{N_s}$ being closed, Lemma 3.7) is a direct consequence of Proposition 1.1,Theorem 2.2, [37]. □

*Proof (Lemma 3.8).* For arbitrary $\boldsymbol{0}\leq\boldsymbol{C}\leq\boldsymbol{F}-\epsilon,\boldsymbol{K}\geq\boldsymbol{0}$, we give the explicit formula of $u^{(SO)}(\boldsymbol{C},\boldsymbol{K};M)$ and the continuity follows. Define

$$
\begin{aligned}
\mu_t^{(-i)} &:= \mathbb{E}\big[\sum_{l\neq i}d_t^{(l)}\big]=\sum_{l\neq i}\mu_t^{(l)},\ (\sigma_t^{(-i)})^2:=\mathrm{Var}\big(\sum_{l\neq i}d_t^{(l)}\big)=\sum_{l_1,l_2\neq i}\rho_t^{(l_1l_2)}\sigma_t^{(l_1)}\sigma_t^{(l_1)},\\
\rho_t^{(i)} &:= \mathrm{Corr}\big(d^{(i)},\sum_{l\neq i}d_t^{(l)}\big)=\frac{\sum_{l\neq i}\rho_t^{(il)}\sigma_t^{(l)}}{\sigma_t^{(-i)}}.
\end{aligned}
$$

Denote $\mathcal{I}_t:=\{i\in\{1,\ldots,N_c\}:|\rho_t^{(i)}|=1\}$. With $(d_t^{(i)},\sum_{l\neq i}d_t^{(l)})$ bi-variate normally distributed, we have

$$
\begin{aligned}
& u^{(SO)}(\boldsymbol{C},\boldsymbol{K};M) \\
=& -\int_t^T \sum_{i\notin\mathcal{I}_t} N^{(i)} \Bigg( \sum_{j=1}^{N_s} \mathbb{E}\Big[ e^{-r^{(i)}t-\gamma_c{}^{(i)}\left(G^{(j)}d^{(i)}+\alpha^{(i)}\boldsymbol{C}^T\boldsymbol{K}\right)} \mathbb{I}_{\{\mathbf{1}_{j-1}^T(\boldsymbol{K}+\boldsymbol{K}_0)<D_t\le\mathbf{1}_j^T(\boldsymbol{K}+\boldsymbol{K}_0)\}} \Big] \\
& + \mathbb{E}\Big[ e^{-r^{(i)}t-\gamma_c^{(i)}\left(Md_t^{(i)}+\alpha^{(i)}(V^{(i)}-M)(D_t-\mathbf{1}_{N_s}^T(\boldsymbol{K}+\boldsymbol{K}_0))+\alpha^{(i)}\boldsymbol{C}^T\boldsymbol{K}\right)} \mathbb{I}_{D_t>\mathbf{1}_{N_s}^T(\boldsymbol{K}+\boldsymbol{K}_0)\}} \Big] \Bigg) dt \\
& - \int_t^T \sum_{i\in\mathcal{I}_t} N^{(i)} \Bigg( \sum_{j=1}^{N_s} \mathbb{E}\Big[ e^{-r^{(i)}t-\gamma_c{}^{(i)}\left(G^{(j)}d^{(i)}+\alpha^{(i)}\boldsymbol{C}^T\boldsymbol{K}\right)} \mathbb{I}_{\{\mathbf{1}_{j-1}^T(\boldsymbol{K}+\boldsymbol{K}_0)<c_t^{(i)}d_t^{(i)}+m_t^{(i)}\le\mathbf{1}_j^T(\boldsymbol{K}+\boldsymbol{K}_0)\}} \Big] \\
& + \mathbb{E}\Big[ e^{-r^{(i)}t-\gamma_c^{(i)}\left((M+\alpha^{(i)}(V^{(i)}-M)c_t^{(i)})d_t^{(i)}+\alpha^{(i)}(V^{(i)}-M)(m_t^{(i)}-\mathbf{1}_{N_s}^T(\boldsymbol{K}+\boldsymbol{K}_0))+\alpha^{(i)}\boldsymbol{C}^T\boldsymbol{K}\right)} \\
& \times \mathbb{I}_{c_t^{(i)}d_t^{(i)}+m_t^{(i)}>\mathbf{1}_{N_s}^T(\boldsymbol{K}+\boldsymbol{K}_0)\}} \Big] \Bigg) dt \\
=& -\int_0^T \sum_{i\notin\mathcal{I}_t} N^{(i)} \Bigg( \sum_{j=1}^{N_s} e^{-\gamma_c{}^{(i)}\alpha^{(i)}\boldsymbol{C}^T\boldsymbol{K}+2(\gamma_c^{(i)}G^{(j)}\sigma_t^{(i)})^2-2\gamma_c^{(i)}G^{(j)}\mu_t^{(i)}-r^{(i)}t} \big(\Phi(a_t^{(i,j)}(\boldsymbol{K}))-\Phi(a_t^{(i,j-1)}(\boldsymbol{K}))\big)dt \\
& + e^{\gamma_c^{(i)}\alpha^{(i)}(V^{(i)}-M)\mathbf{1}_{N_s}^T(\boldsymbol{K}+\boldsymbol{K}_0)-\gamma_c{}^{(i)}\alpha^{(i)}\boldsymbol{C}^T\boldsymbol{K}+\frac{1}{2}\left(\gamma_c^{(i)}\alpha^{(i)}(V^{(i)}-M)\right)^2} (1-\Phi(b_t^{(i)}(\boldsymbol{K}))) \Bigg) dt \\
& \times e^{(\gamma_c^{(i)})^2(M+\alpha^{(i)}(V^{(i)}-M))\alpha^{(i)}(V^{(i)}-M)\rho_t^{(i)}\sigma_t^{(i)}+\frac{1}{2}\left(\gamma_c^{(i)}(M+\alpha^{(i)}(V^{(i)}-M))\sigma_t^{(i)}\right)^2-\gamma_c^{(i)}\alpha^{(i)}(V^{(i)}-M)\frac{\mu_t^{(-i)}}{\sigma_t^{(-i)}}-r^{(i)}t} \\
& - \int_0^T \sum_{i\in\mathcal{I}_t} N^{(i)} \Bigg( \sum_{j=1}^{N_s} e^{-\gamma_c{}^{(i)}\alpha^{(i)}\boldsymbol{C}^T\boldsymbol{K}+\frac{1}{2}(\gamma_c^{(i)}G^{(j)}\sigma_t^{(i)})^2+\gamma_c^{(i)}G^{(j)}\mu_t^{(i)}\sigma_t^{(i)}-r^{(i)}t} \big(\Phi(\tilde{a}_t^{(i,j)}(\boldsymbol{K}))-\Phi(\tilde{a}_t^{(i,j-1)}(\boldsymbol{K}))\big)dt \\
& + e^{\gamma_c^{(i)}\alpha^{(i)}(V^{(i)}-M)(\mathbf{1}_{N_s}^T(\boldsymbol{K}+\boldsymbol{K}_0)-m_t^{(i)})-\gamma_c{}^{(i)}\alpha^{(i)}\boldsymbol{C}^T\boldsymbol{K}+\frac{1}{2}(\gamma_c^{(i)}(M+\alpha^{(i)}(V^{(i)}-M)c_t^{(i)})\sigma_t^{(i)})^2} \\
& \times e^{\gamma_c^{(i)}(M+\alpha^{(i)}(V^{(i)}-M)c_t^{(i)})\mu_t^{(i)}\sigma_t^{(i)}-r^{(i)}t} \big(1-\Phi(\tilde{b}_t^{(i)}(\boldsymbol{K}))\big) \Bigg) dt,
\end{aligned}
$$

where

$$
\begin{aligned}
a_t^{(i,j)}(\boldsymbol{K}) &:= \frac{\mathbf{1}_j^T(\boldsymbol{K}+\boldsymbol{K}_0) - \mu_t - 2\gamma_c^{(i)} G^{(l)}(\rho_t^{(i)}\sigma_t^{(-i)} - \sigma_t^{(i)})}{\sqrt{\sigma_t^2 - 2\rho_t^{(i)}\sigma_t^{(i)}\sigma_t^{(-i)}}}, \\
b_t^{(i)}(\boldsymbol{K}) &:= \frac{\mathbf{1}_j^T(\boldsymbol{K}+\boldsymbol{K}_0) - \mu_t + \gamma_c^{(i)}\big(\alpha^{(i)}(V^{(i)} - M)\rho_t^{(i)} - (M + \alpha^{(i)}(V^{(i)} - M))\sigma_t^{(i)}\big)(\rho_t^{(i)}\sigma_t^{(-i)} - 1)}{\sqrt{\sigma_t^2 - 2\rho_t^{(i)}\sigma_t^{(i)}\sigma_t^{(-i)}}} \\
&\quad + \frac{\gamma_c^{(i)}\alpha^{(i)}(V^{(i)} - M)(1 - (\rho_t^{(i)})^2)\sigma^{(-i)}}{\sqrt{\sigma_t^2 - 2\rho_t^{(i)}\sigma_t^{(i)}\sigma_t^{(-i)}}}, \\
c_t^{(i)} &= \frac{\rho_t^{(i)}\sigma_t^{(-i)} + \sigma_t^{(i)}}{\sigma_t^{(i)}},\ m_t^{(i)} = \mu_t^{(-i)} - \frac{\rho_t^{(i)}\sigma_t^{(-i)}\mu_t^{(i)}}{\sigma_t^{(i)}} \\
\tilde{a}_t^{(i,j)}(\boldsymbol{K}) &:= \frac{\frac{1}{c_t^{(i)}}(\mathbf{1}_j^T(\boldsymbol{K}+\boldsymbol{K}_0) - m_t^{(i)}) - \mu_t^{(i)} - \gamma_c^{(i)} G^{(j)}\sigma_t^{(i)}}{\sigma_t^{(i)}}, \\
\tilde{b}_t^{(i)}(\boldsymbol{K}) &:= \frac{\frac{1}{c_t^{(i)}}(\mathbf{1}_{N_s}^T(\boldsymbol{K}+\boldsymbol{K}_0) - m_t^{(i)}) - \mu_t^{(i)} - \gamma_c^{(i)}(M + \alpha^{(i)}(V^{(i)} - M)c_t^{(i)})\sigma_t^{(i)}}{\sigma_t^{(i)}}.
\end{aligned}
$$

(Note $a_t^{(i,j)}, b_t^{(i)}$ are well-defined for $i \notin \mathcal{I}_t$, since $\sigma_t^2 - 2\rho_t^{(i)}\sigma_t^{(i)}\sigma_t^{(-i)} = (\sigma_t^{(i)} - \rho_t^{(i)}\sigma_t^{(-i)})^2 + (1 - \rho_t^{(i)})^2(\sigma_t^{(-i)})^2 > 0$.) $\square$

## References

[1] 2020 annual report on market issues and performance. California Independent System Operator, Department of Market Monitoring, 2021.

[2] Todd S Aagaard and Andrew N Kleit. *Electricity Capacity Markets*. Cambridge University Press, 2022.

[3] I. Abada, G. de Maere dAertrycke, and Y. Smeers. On the multiplicity of solutions in generation capacity investment models with incomplete markets: A risk-averse stochastic equilibrium approach. *Mathematical Programming*, 165(1):5–69, 2017.

[4] I. Abada and A. Ehrenmann. A new game theoretical approach for modeling export energy markets equilibria. *The Energy Journal*, 39(5):54–84, 2018.

[5] International Atomic Energy Agency. Expansion planning for electricity generating systems. `https://www-pub.iaea.org/MTCD/Publications/PDF/TRS1/TRS241_Web.pdf`, 1984.

[6] R. Aïd, L. Li, and M. Ludkovski. Capacity expansion games with application to competition in power generation investments. *Journal of Economic Dynamics and Control*, 84:1–31, 2017.

[7] R. Aïd, D. Possamaï, and N. Touzi. Optimal electricity demand response contracting with responsiveness incentives. *arXiv preprint arXiv:1810.09063*, 2018.

[8] M. Assili, M.H.J. DB, and R. Ghazi. An improved mechanism for capacity payment based on system dynamics modeling for investment planning in competitive electricity environment. *Energy Policy*, 36(10):3703–3713, 2008.

[9] H. Bahar and J. Sauvage. Cross-border trade in electricity and the development of renewables-based electric power: lessons from Europe. OECD Trade and Environment Working Papers, 2013.

[10] C. Benavides, R. Álvarez, R. Torres, R. Moreno, M. Matus, D.F. Muñoz, J. González, G. Jiménez-Estévez, and R. Palma-Behnke. Capacity payment allocation in hydrothermal power systems with high shares of renewable energies. In *E3S Web of Conferences*, volume 140, page 11008. EDP Sciences, 2019.

[11] D.P. Bertsekas. Convexification procedures and decomposition methods for nonconvex optimization problems. *Journal of Optimization Theory and Applications*, 29(2):169–197, 1979.

[12] P.C. Bhagwat, L.J. De Vries, and B.F. Hobbs. Expert survey on capacity markets in the US: Lessons for the EU. *Utilities Policy*, 38:11–17, 2016.

[13] P.C. Bhagwat, Jörn C. Richstein, E.J.L. Chappin, K.K. Iychettira, and L.J. De Vries. Cross-border effects of capacity mechanisms in interconnected power systems. *Utilities Policy*, 46:33–47, 2017.

[14] M. Bichuch, B.F. Hobbs, X. Song, and Y. Wang. Optimal electricity distribution pricing under risk and high photovoltaics penetration. *Journal of Energy Markets*, 14(1):61–97, 2021.

[15] R.E. Bohn, M.C. Caramanis, and F.C. Schweppe. Optimal pricing in electrical networks over space and time. *The Rand Journal of Economics*, 15(3):360–376, 1984.

[16] D.P. Brown. Capacity payment mechanisms and investment incentives in restructured electricity markets. *Energy Economics*, 74:131–142, 2018.

[17] C. Byers and A. Botterud. Additional capacity value from synergy of variable renewable energy and energy storage. *IEEE Transactions on Sustainable Energy*, 11(2):1106–1109, 2019.

[18] E. Cavazzuti, M. Pappalardo, and M. Passacantando. Nash equilibria, variational inequalities, and dynamical systems. *Journal of Optimization Theory and Applications*, 114(3):491–506, 2002.

[19] M.C. Ceparano and F. Quartieri. Nash equilibrium uniqueness in nice games with isotone best replies. *Journal of Mathematical Economics*, 70:154–165, 2017.

[20] P. Cramton and A. Ockenfels. Economics and design of capacity markets for the power sector. *Interdisziplinäre Aspekte der Energiewirtschaft*, 53:191–212, 2016.

[21] P. Cramton, A. Ockenfels, and S. Stoft. Capacity market fundamentals. *Economics of Energy & Environmental Policy*, 2(2):27–46, 2013.

[22] P. Cramton and S. Stoft. A capacity market that makes sense. *The Electricity Journal*, 18(7):43–54, 2005.

[23] P.C. Cramton and S. Stoft. The convergence of market designs for adequate generating capacity with special attention to the CAISO's resource adequacy problem. MIT Center for Energy and Environmental Policy Research, 2006.

[24] S. Dempe, B.S. Mordukhovich, and A.B. Zemkoho. Necessary optimality conditions in pessimistic bilevel programming. *Optimization*, 63(4):505–533, 2014.

[25] B.C. Eaves. On the basic theorem of complementarity. *Mathematical Programming*, 1(1):68–75, 1971.

[26] J. Fortuny-Amat and B. McCarl. A representation and economic interpretation of a two-level programming problem. *Journal of the Operational Research Society*, 32(9):783–792, 1981.

[27] S.A. Gabriel, A.J. Conejo, J.D. Fuller, B.F. Hobbs, and C. Ruiz. *Complementarity modeling in energy markets*, volume 180. Springer Science & Business Media, 2012.

[28] D. Hach and S. Spinler. Capacity payment impact on gas-fired generation investments under rising renewable feed-in—A real options analysis. *Energy Economics*, 53:270–280, 2016.

[29] P. Hartman and G. Stampacchia. On some non-linear elliptic differential-functional equations. *Acta Mathematica*, 115(1):271–310, 1966.

[30] S. Hesmondhalgh, J. Pfeifenberger, and D. Robinson. Resource adequacy and renewable energy in competitive wholesale electricity markets. The Brattle Group, 2010.

[31] B.F. Hobbs, M.C. Hu, J.G. Iñón, S. Stoft, and M.P. Bhavaraju. A dynamic analysis of a demand curve-based capacity market proposal: The PJM reliability pricing model. *IEEE Transactions on Power Systems*, 22(1):3–14, 2007.

[32] B.F. Hobbs, J. Inon, and S. Stoft. Installed capacity requirements and price caps: Oil on the water, or fuel on the fire? *The Electricity Journal*, 14(6):23–34, 2001.

[33] Michael Hogan. Follow the missing money: Ensuring reliability at least cost to consumers in the transition to a low-carbon power system. *The Electricity Journal*, 30(1):55–61, 2017.

[34] P. Holmberg and R.A Ritz. Capacity mechanisms and the technology mix in competitive electricity markets. IFN Working Paper, 2019.

[35] P. Holmberg and R.A. Ritz. Optimal capacity mechanisms for competitive electricity markets. *The Energy Journal*, 41(Special Issue):33–66, 2020.

[36] A.S.M. Khan, R.A. Verzijlbergh, Oz.C. Sakinci, and L.J. De Vries. How do demand response and electrical energy storage affect (the need for) a capacity market? *Applied Energy*, 214:39–62, 2018.

[37] P.Q. Khanh and L.M. Luu. Upper semicontinuity of the solution set to parametric vector quasivariational inequalities. *Journal of Global Optimization*, 32(4):569–580, 2005.

[38] S. Komlósi. On the Stampacchia and Minty variational inequalities. In G. Giorgi and F.A. Rossi, editors, *Generalized Convexity and Optimization for Economic and Financial Decisions*, pages 231–260. Pitagora Editrice, Bologna, 1999.

[39] D.M. Kreps and J.A. Scheinkman. Quantity precommitment and Bertrand competition yield Cournot outcomes. *The Bell Journal of Economics*, 14(2):326–337, 1983.

[40] R. Li, A. Svoboda, and S. Oren. Efficiency impact of convergence bidding in the California electricity market. *Journal of Regulatory Economics*, 48(3):245–284, 2015.

[41] P. Loridan and J. Morgan. Weak via strong Stackelberg problem: new results. *Journal of Global Optimization*, 8(3):263–287, 1996.

[42] P. Luh, W. Blankson, Y. Chen, J. Yan, G. Stern, S. Chang, and F. Zhao. Payment cost minimization auction for deregulated electricity markets using surrogate optimization. *IEEE Transactions on Power Systems*, 21(2):568–578, 2006.

[43] Z.Q. Luo, J.S. Pang, and D. Ralph. *Mathematical programs with equilibrium constraints.* Cambridge University Press, 1996.

[44] A. Maugeri and F. Raciti. On existence theorems for monotone and nonmonotone variational inequalities. *Journal of Convex Analysis*, 16(3-4):899–911, 2009.

[45] J. Mays and J. Jenkins. Electricity markets under deep decarbonization. USAEE Working Paper No. 22-550, 2022.

[46] G. Meunier. Capacity choice, technology mix and market power. *Energy Economics*, 32(6):1306–1315, 2010.

[47] G. Meunier. Risk aversion and technology mix in an electricity market. *Energy Economics*, 40:866–874, 2013.

[48] F. Meyabadi and M.H. Deihimi. A review of demand-side management: Reconsidering theoretical framework. *Renewable and Sustainable Energy Reviews*, 80:367–379, 2017.

[49] R. Moita and D. Monte. Competitive equilibrium and optimal capacity matrix in electricity markets with renewables. IDB Working Paper Series, 2019.

[50] W.R. Morrow and S.J. Skerlos. Fixed-point approaches to computing Bertrand-Nash equilibrium prices under mixed-logit demand. *Operations Research*, 59(2):328–345, 2011.

[51] F.D. Munoz, S. Wogrin, S.S. Oren, and B.F. Hobbs. Economic inefficiencies of cost-based electricity market designs. *The Energy Journal*, 39(3):51–68, 2018.

[52] P.A. Neto, T.L. Friesz, and K. Han. Electric power network oligopoly as a dynamic Stackelberg game. *Networks and Spatial Economics*, 16(4):1211–1241, 2016.

[53] K. Neuhoff, J. Diekmann, F. Kunz, S. Rüster, W.P. Schill, and S. Schwenen. A coordinated strategic reserve to safeguard the European energy transition. *Utilities Policy*, 41:252–263, 2016.

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

[55] S.S. Oren. Ensuring generation adequacy in competitive electricity markets. University of California Energy Institute's Energy Policy and Economics Working Paper Series, 2003.

[56] J.S. Pang, S. Sen, and U.V. Shanbhag. Two-stage non-cooperative games with risk-averse players. *Mathematical Programming*, 165(1):235–290, 2017.

[57] J. Pfeifenberger, S. Newell, K. Spees, A. Hajos, and K. Madjarov. Second performance assessment of PJMs reliability pricing model. The Brattle Group, 2011.

[58] R.T. Rockafellar and R.J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[59] N.H. Santibáñez, A. Jofré, and D. Possamaï. Pollution regulation for electricity generators in a transmission network. *arXiv preprint arXiv:2111.13505*, 2021.

[60] D.A. Schiro, B.F Hobbs, and J.S. Pang. Perfectly competitive capacity expansion games with risk-averse participants. *Computational Optimization and Applications*, 65(2):511–539, 2016.

[61] G. Stern, J.H. Yan, P.B. Luh, and W.E. Blankson. What objective function should be used in the ISO and RTO unit-commitment models: A market participant pointview. `https://www.researchgate.net/publication/224651745_What_objective_function_should_be_used_for_optimal_auctions_in_the_ISORTO_electricity_market`, 2006.

[62] S. Stoft. *Power system economics: Designing markets for electricity*, volume 468. IEEE Press Piscataway, 2002.

[63] B. Willems and J. Morbee. Market completeness: How options affect hedging and investments in the electricity sector. *Energy Economics*, 32(4):786–795, 2010.

[64] S. Wogrin, B.F. Hobbs, D. Ralph, E. Centeno, and J. Barquin. Open versus closed loop capacity equilibria in electricity markets under perfect and oligopolistic competition. *Mathematical Programming*, 140(2):295–322, 2013.

[65] S. Wogrin, S. Pineda, and D.A. Tejada-Arango. Applications of bilevel optimization in energy and electricity markets. In S. Dempe and A. Zemkoho, editors, *Bilevel Optimization: Advances and Next Challenges*, pages 139–168. Springer, 2020.

[66] G. Zöttl. A framework of peak load pricing with strategic firms. *Operations Research*, 58(6):1637–1649, 2010.

[67] M. Zugno, J.M. Morales, P. Pinson, and H. Madsen. A bilevel model for electricity retailers' participation in a demand response market environment. *Energy Economics*, 36:182–197, 2013.

Credit Author Statement

Maxim Bichuch: Conceptualization; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Roles/Writing - original draft; Writing - review & editing

Benjamin F. Hobbs: Conceptualization; Funding acquisition; Investigation; Methodology; Validation; Roles/Writing - original draft; Writing - review & editing

Xinyue Song: Conceptualization; Formal analysis; Investigation; Methodology; Validation; Roles/Writing - original draft; Writing - review & editing