Master's Thesis

Matthew Davis

Missing Markets in the Electricity Industry

Examining capacity markets, ancillary services, capture price divergence, mispricing and other challenges in the energy transition

Master of Environmental Policy and Energy Economics

Supervisor: François Salanié

Contents

Introduction

The wholesale price of electricity tends to be far more volatile than the prices of most other commodities. The main reason is that electricity cannot be stored cheaply (even with recent advances in battery technology), so supply and demand must be balanced on the timescale of seconds. This is coupled with relatively high reliability requirements compared to most goods, due to the essential nature of the service. Another reason is that consumers are usually exposed to flat tariffs. This means that in the short term demand is inelastic. In most markets an increase in price will increase supply and decrease demand. In electricity markets an increase in price will barely decrease demand. Therefore large price variations are required to balance supply and demand, because only one side of the market responds meaningfully to short term price signals.

Price volatility can be politically undesirable, as consumers prefer predictable electricity bills. Investors building generators also tend to prefer price certainty[1]. However, the large magnitude of variation in electricity spot prices reflects how the true cost and value to society of electricity consumption varies drastically over hours, minutes and even within a single second. Some units of energy (megawatt hours) are worth ten thousand times more than others, due to a difference in when they are delivered. Consumers and generators often face incentives which do not reflect this variation. The purpose of this thesis is to investigate such missing markets, and market design failures, through a series of example cases, with a focus on Europe and Australia.

This thesis is the culmination of a two-year Master of Environmental Policy and Energy Economics at the Toulouse School of Economics, in France. It was written from March to August 2025 under the supervision of Professor François Salanié.

Through this thesis I have combined the economic theory taught in this masters course with my experience in the industry to generate several interesting insights. I designed several formal algebraic models for this thesis. A key challenge for theoreticians when constructing such models is to simplify the problem enough to be tractable, whilst retaining enough complexity and nuance to capture the true nature and tradeoffs of the problem. My experience in electricity trading and dispatch inspired these models and helps ensure their relevance. In addition to theoretical models, this thesis also includes a policy review of capacity markets. This entails the collation of many facts to distil a complex issue into a research-driven story.

Most parts of this thesis include data analysis, either to motivate the problem or test a model. Regressions were performed in R. Data wrangling, simulations and exploratory analysis and graphing was done with Python. The code for all simulations and regressions is available on GitHub[2]. Due to the large size of some datasets (> 100GB) and the memory intensive nature of some optimisations and simulations, the analysis cannot be run on a normal laptop. Instead it was instead run on a larger server in the cloud.

[1] Batteries are a notable exception. Their arbitrage strategies depend on volatility.

[2] https://github.com/mdavis-xyz/masters-thesis

Part 1° explores payments for spinning reserves, also known as contingency raise services. Payments for reserves are somewhat unique to the electricity sector. Dairy farmers are not paid for the ability to increase production in case their neighbour's production drops unexpectedly. Electricity generators are. This is due to the fact that after a large shock, electricity supply and demand must be balanced on a subsecond timescale. (The reserves may also be called upon after a demand shock or transmission outage.)

A contribution of Part 1° is the clear explanation for a non-technical reader of the difference between spinning reserves and stationary reserves. This is followed by an algebraic derivation of the equilibrium cost of reserves. This model is based on Gilmore, Nolan, and Simshauser (2024), but is extended in several respects. Increasing marginal costs are used instead of a constant marginal cost. The model includes a fixed operating cost component and the possibility of a generator turning off completely, to model the important distinction between stationary and spinning reserves. The model is extended to make demand for reserves endogenous, depending on demand for energy. The nature of this endogeneity is tested empirically with autoregressive integrated moving average (ARIMA) regressions in Appendix B°. Part 1° also includes mathematical proofs of an intriguing case where aggregate marginal costs for a good decrease as the quantity demanded increases. It concludes with a discussion drawing parallels between ancillary markets for spinning reserves and markets for installed capacity.

Capacity markets for *installed* capacity are a hot topic in the industry at the moment. In theory an energy-only market (where generators are paid only per unit of energy they produce) should drive investment to the optimal level (Newbery 2016; Cramton and Stoft 2005). Proponents of capacity markets argue that the energy-only paradigm yields insufficient revenue in practice, resulting in underinvestment. This can only be true if there is a market failure, or a missing market. However, many proponents are unable to precisely identify such a cause.

Part 2° is a literature review and discussion of the potential market failures and missing markets which motivate the introduction of capacity markets. The most pertinent market failure is the presence of an inefficiently low spot price cap. The most relevant missing market is typically the absence or low liquidity of long term hedging markets. The section explores how capacity markets often fail to achieve their objectives. They are often designed without regard for the reliability targets which they are supposed to achieve, and they can exacerbate the very "missing money" problem which they are intended to solve. Part 2° also includes a discussion of how the metric for capacity is quite difficult to define in a way which does not create perverse incentives. Capacity payment metrics typically lead to operational and investment distortions, rewarding poor performers who contribute little energy during critical periods. This metric is typically defined in a way which explicitly discriminates based on fuel type, directly counteracting climate policy objectives, and is effectively a departure from a liberalised market to one of a central planner picking winners. Section 2.5 includes an analysis of data showing trends in capture price ratios, as a demonstration of how two generators with the same capacity can provide vastly different social value.

Part 3° explores these differences in the value of a megawatt within a given fuel type, from a theoretical perspective. It does so by introducing a model concerning the angle which solar panels are mounted at. Conventional wisdom says that panels should be installed at an angle equal to their latitude, and facing north/south towards the equator. This maximises the volume of energy produced. However, a panel which faces further west and more vertically, produces more energy in winter and late afternoons, when each unit of energy is more valuable. This tradeoff is explored in the model to demonstrate how the objective of investment should be to maximise the *value* of energy produced, not the *volume*. This is coupled with simulations showing that rooftop solar support schemes can distort this tradeoff such that up to half the social value of solar panels is wasted.

The model from Part 3° is then extended in Part 4° to combine solar panels with a battery. The purpose of this extension is to challenge the common misconception that combining solar panels with a battery increases the project's spot market revenue. Batteries do indeed add value by acting as a 'solar sponge', shifting surplus generation from the middle of the day when prices are low and the sun is high, to the afternoon and evening when prices are high and the sun is low or absent. However, batteries can provide this service without necessarily being located on the same premise, or even owned by the same firm or household. There are many logistical benefits from installing solar and batteries together ("colocation"). The purpose of this model is to demonstrate that from a pure spot market revenue perspective, colocation provides no increase in revenue, and may even reduce revenue, compared to solar and batteries installed separately.

Many of the challenges and quirks of electricity markets are due to the fact that most consumers pay a price for energy which is different to the marginal cost to produce that energy. Due to a strong political aversion to bill shock and the historical impracticalities of real time metering, customers face a fixed tariff instead of the wholesale price[3]. Part 5° introduces a model, based on a simplification of Borenstein and Holland (2005), to explore the inefficiency which this creates.

Electricity retailers offer these fixed tariffs to consumers, and are exposed to the variable wholesale price. This introduces a risk, which depends on the joint distribution of spot prices and consumers' daily consumption profiles. Part 6° explores the risk-return tradeoff of a retailer choosing an optimal portfolio of heterogenous customers. Parallels then are drawn to show that this is similar to to the capital asset pricing model (CAPM).

In Part 7° the discussion returns to the supply side. In electricity spot markets the price changes abruptly at the boundary between trading intervals, but the physical output of electricity generators cannot change instantaneously. Technical limits on the rate of adjustment (ramping) of generators have been well studied in the engineering and economic literature. However, some markets such as Australia's National Electricity Market (NEM) specify a minimum duration for adjustment, even if the generator is physically capable of adjusting far more quickly. Part 7° explores the impact of this limit. A unique contribution of this section is to show that even with perfect foresight, no market power, no startup costs and no *physical* ramp rate limits, a rational profit-maximising firm may still want to submit bids which differ from their marginal cost. Additionally, an algebraic model is introduced to demonstrate that this ramping constraint mutes the incentives created by price spikes. Many researchers and market participants ignore this limitation imposed by the market operator. Part 7° concludes with simulations in Australia's NEM, to show empirically that the impact of this limit on modelling results can be economically significant.

[3]Time of use (TOU) and critical-peak pricing (CPP) tariffs are also discussed in Part 5°.

Part 1 Contingency Raise Service: Payment for Spinning Reserves

1.1 Motivation and Context

"Contingency raise" is an ancillary service (i.e. reliability service) in electricity grids. First, I will define some terms. In the electricity industry, "capacity" or "reserves" may refer to two different concepts. They both refer to the amount of power a generator could potentially produce, but correspond to two different timescales. An investor may decide to construct a generator, which will take several years. The investor makes a decision about the installed capacity. Once built, the operator of the generator must make the decision to be either on (spinning, warm), or off (stationary, cold)[4]. When spinning and producing energy, the amount of *spinning reserve* equals the amount of installed capacity minus the portion currently used to produce power. For example, if a spinning generator with 100 MW of installed capacity is generating 30 MW of power, then there is 70 MW of *spinning reserve*, because the generator can quickly increase production by 70 MW. However, if a generator with 100 MW of installed capacity is stationary (producing 0 MW of power), then it has 100 MW of *stationary* reserves, and 0 MW of *spinning* reserves. This distinction is crucial because turning on a generator can take many hours (and is costly), whereas supply and demand must be balanced within minutes or seconds after a shock. Thermal generators tend to have a minimum output level. For example, a 100 MW generator might be able to produce 20 MW, but not 10 MW.

Contingency raise markets are extremely short-term capacity markets, for *spinning reserves*. Generators able to increase their output on the timescale of a few seconds (or even subsecond) are paid to be available to do so. When something goes wrong in the grid in a sudden and unpredicted way (e.g. a transmission tower falls over, or a generator has a fire and has to turn off), generators assigned to provide the contingency raise service must 'raise' their energy production level very quickly. If this is not done the grid will collapse, and all demand (more than the drop in supply) will be unmet for hours or days. Generators which are currently producing 0 MW of energy (i.e. not spinning) cannot start up quickly enough, and are thus unable to provide raise service. This is why some unused installed capacity cannot be used for the raise service.

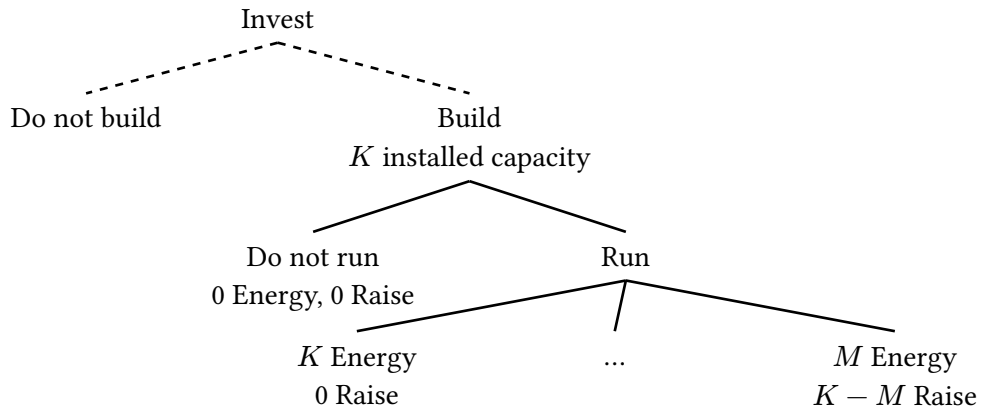


Figure 1: Sequence of decisions, to demonstrate the difference between installed capacity and spinning reserve. For this analysis, the investment decision (dashed) is neglected. I am only considering operational decisions on the timescale of hours. The "raise" service is provided by spinning reserve.

Here I focus on Australia's National Electricity Market (NEM). Most other regions have similar, but slightly different setups. Australia's frequency control and ancillary services (FCAS) market is described in detail in Gilmore, Nolan, and Simshauser (2024). Providers of contingency raise services

[4]This applies to *thermal* generators, which are are those which are combustion based: gas, coal, nuclear, biomass. Wind and hydro turbines also have a startup delay, but it is far smaller. Batteries and solar can adjust from 0 MW almost instantly. This section focuses on thermal generators.

are paid for having spinning reserve available in each 5-minute period. If called upon to actually utilise this capacity, generators will additionally be paid the energy price for their increased production. The payment for energy alone is not sufficient, because energy prices are fixed over a 5-minute period, but the raise service is used to balance supply and demand after shocks on the timescale of seconds. If the spot price varied every second and was not capped with an administrative price ceiling, then an explicit raise service would not be required. The market for the raise service was introduced because with only a standard energy market the ultra-short-term market is missing.

The market for the raise service is about increasing (raising) supply in response to a shock which decreased supply or increased demand. There is an equivalent market in the other direction. This other product, called the "lower" service, is used to reduce supply after a shock which increased supply or decreased load. For example, if region A exports to region B, and the transmission line between them is abruptly disconnected, then the raise service will be used to balance region B, and the lower service will be used to balance region A. In this analysis I focus on the raise service, which is harder to provide and tends to be more expensive.

The objective of these services is to prevent a total grid outage. Thus it is a public good, which different consumers may value differently. It is a good both for energy consumers, and for energy producers, since producers incur an opportunity cost if there is a grid outage (Billimoria, Mancarella, and Poudineh 2022). In Australia's NEM, costs for these public goods are allocated proportionally to energy volume. Generators pay for raise services (i.e. they pay for other generators to pick up the slack if they were to fail), and consumers pay for lower services. For this model I will neglect the fee that all generators pay for this service, and only consider the revenue that some generators earn by providing this service. In practice, there are a series of ancillary markets of differing timescales (1 second, 5 second, 60 second etc.). Greve et al. (2018) discuss how policy makers should choose these boundaries. For this model, I assume there is only one relevant timescale.

1.2 The Model

1.2.1 Model Setup

There are N identical firms, each with a fixed amount of installed capacity K, and cost $C(q_e)$. New assets cannot be built within the timescale considered. The implications for ex ante investment incentives are left for future research. I assume perfect competition. As a simplification, dynamic considerations are neglected and the model considers only comparative statics. Total demand for the raise service is an exogenous constant R, which is a lower bound (supply can exceed demand).

For a given energy output, making reserves available incurs no cost ($C(q_e)$ is not a function of q_r). If firms are called upon to utilise these reserves and actually increase their output, they will additionally be paid the energy price to do so. In the NEM, raise services have historically been utilised in 0.026% of trading intervals (Gilmore, Nolan, and Simshauser 2024). This is so rare that this additional revenue and cost (pre-event energy price minus fuel cost) will be neglected for now. It will be considered in Section 1.3 as an extension .

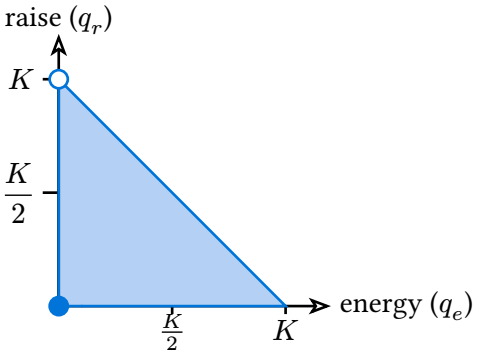
1.2.2 Simple Model: No Fixed Cost, No Minimum Generation Level

I start with a very simplified cost function $C(q_e)$ as shown in Figure 2b.

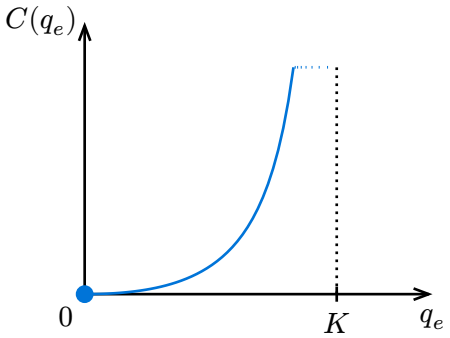
- $C(0) = 0$
- $C'(0) = 0$
- $C'(K) = \infty$
- $C''(q_e) > 0$

For this simple exposition there is no minimum energy level and no fixed cost component, as shown in Figure 2a. For now I make no distinction between spinning reserve and stationary reserve (i.e. between generators which are 'on' and 'off') Thus I am only considering the bottom-right decision node in

Figure 1. Later in Section 1.2.4 I will extend the model to consider unit commitment (startup) decisions. This model differs from Gilmore, Nolan, and Simshauser (2024), who use a constant marginal cost of energy.



(a) Possible production set. For now I assume a generator can provide the raise service without providing energy ($q_e = 0$ whilst $q_r > 0$)



(b) Cost Function

Figure 2: Simple model setup with no fixed cost component, and no minimum generation level.

1.2.2.1 Decentralised Solution

For given energy and raise service price p_e, p_r, the firm chooses energy and raise service quantities q_e, q_r.

$$\pi = \max_{q_e, q_r} p_e q_e + p_r q_r - C(q_e)$$

Subject to:

- $0 \leq q_e \leq K$
- $0 \leq q_r \leq K$
- $0 \leq q_e + q_r \leq K$

In this model, the raise service incurs no marginal cost and there are no fixed-cost components. Given the additional assumption of perfect competition, for a given q_e the optimal choice is $q_r = K - q_e$ (maximum q_r). So the problem of the firm becomes:

$$\pi = \max_{q_e} p_e q_e + (K - q_e) p_r - C(q_e) = \max_{q_e} (p_e - p_r) q_e + K p_r - C(q_e)$$

Subject to $0 \leq q_e \leq K$.

Taking first order conditions:

$$\frac{\partial \pi}{\partial q_e} = (p_e - p_r) + 0 - C'(q_e) = 0$$

Therefore the (interior) solution is q_e such that:

$$p_r = p_e - C'(q_e)$$

In equilibrium the price of the raise service will equal the opportunity cost of missing out on energy revenue, adjusted by fuel cost. This matches the findings of Gilmore, Nolan, and Simshauser (2024).

Since $C'(K) = \infty$, $q_e = K$ cannot be an exterior solution. $q_e = 0$, $q_r = K$ is an exterior solution if $p_e \leq p_r$. However, this case will be neglected because in full equilibrium and with more realistic assumptions, the energy price would rise sufficiently to prevent this situation.

After considering the firm's best response to given prices, now I consider what prices will be in equilibrium. Since all firms are identical, marginal costs are increasing and there is no fixed cost component for the decision to run, it must be that the optimal allocation is symmetric.

The demand for raise service is exogenously given by constant R. This is perfectly inelastic. Thus I treat it as an inequality constraint, not an objective. We must consider two possible cases:

When the raise service reserve constraint is binding: All capacity is used for either the raise service or energy, and the supplied raise service quantity equals the demand R. Each generator provides $q_r = \frac{R}{N}$, and the price of energy rises to reduce energy demand (D) to the remaining capacity $D_e(p_e) = Nq_e = NK - R$. Scarcity drives a wedge between marginal benefit and marginal cost of energy.

When the reserve constraint is not binding: The supply and demand curves for energy balance such that there are excess reserves for meeting the demand for the raise service. The price of raise services must be $p_r = 0$, so the allocation becomes the same as for the single energy market on its own.

$$p_e = D_e^{-1}(Nq_e) = C'(q_e)$$

The constraint for the raise service will be binding if the demand for raise service is large, demand for energy is low, or energy supply costs are low. (The precise conditions depend on the form of $C(q_e)$ and $D_e(p_e)$.)

1.2.3 Endogenous Multiplicative Demand

In practice demand for the raise service reserves is not an exogenous constant, but varies based on the total quantity of energy supplied, and other factors. Suppose it varies linearly based on energy demand. $R = \overline{R} + \beta Q_e$ where β reflects the marginal level of reserves required (e.g. $\beta = 0.1$ means that for every additional 10 MW of energy supplied there must be at least 1 MW of additional raise service reserves) and Q_e is aggregate energy demand ($Q_e = Nq_e$, $\overline{R} = N\overline{r}$ in the symmetric case). $\beta = 0$ is the case of exogenous demand described before in Section 1.2.2.

Suppose that $\beta > 0$. This yields solutions equivalent to the exogenous case ($\beta = 0$), except the scarcity pricing is less extreme (for the same $\overline{R}$). This is shown in Figure 3. As a baseline, if the reserve constraint is not binding then the equilibrium would be at the intersection of supply and demand curves: p_e^0, q_e^0, q_r^0. Suppose now that the installed capacity is not sufficient ($K < q_e^0 + q_r^0$), so the energy market cannot clear as usual. A binding constraint means that $q_r + q_e = K$, and $q_r = \overline{r} + \beta q_e$, so $q_e = \frac{K - \overline{r}}{1+\beta}$. Scarcity drives a wedge between the value of marginal demand for energy and the cost of marginal supply of energy, yielding a constrained allocation q_e^1, q_r^1. (Even though the cost function is designed to be infinite when supplying energy at full capacity, the costs here are finite, because *energy* is supplied at less than full capacity. The remainder is allocated as reserves, which incur no cost.) Consumers are charged high energy price p_e^1, and generators earn scarcity rents above their marginal cost c^1. The price of the raise service is still set by the indifference of firms between supplying energy or reserves. $p_r = p_e - C'(q_e)$

If there was an exogenous demand for the raise service capacity ($\beta = 0$), this would be the end of the story. However, if demand is endogenous ($\beta > 0$) then the reduction in *energy* demand from q_e^0 to q_e^1 results in a reduction in *raise* reserve demand of $\Delta q_r = \beta(q_e^0 - q_e^1) = q_r^0 - q_r^1$. This frees up capacity to be used for energy, so the energy allocation can increase from q_e^1 to q_e^2. This continues until a convergence strictly above q_e^1 (not drawn). Note that this iterative behaviour is not a dynamic outcome, but rather a pedagogical explanation of the static equilibrium. The scarcity pricing in the endogenous

case is less extreme ($p_e^2 < p_e^1$) because every unit of energy demand reduction from a higher energy price results in $1 + \beta$ total demand reduction across both energy and raise service markets.

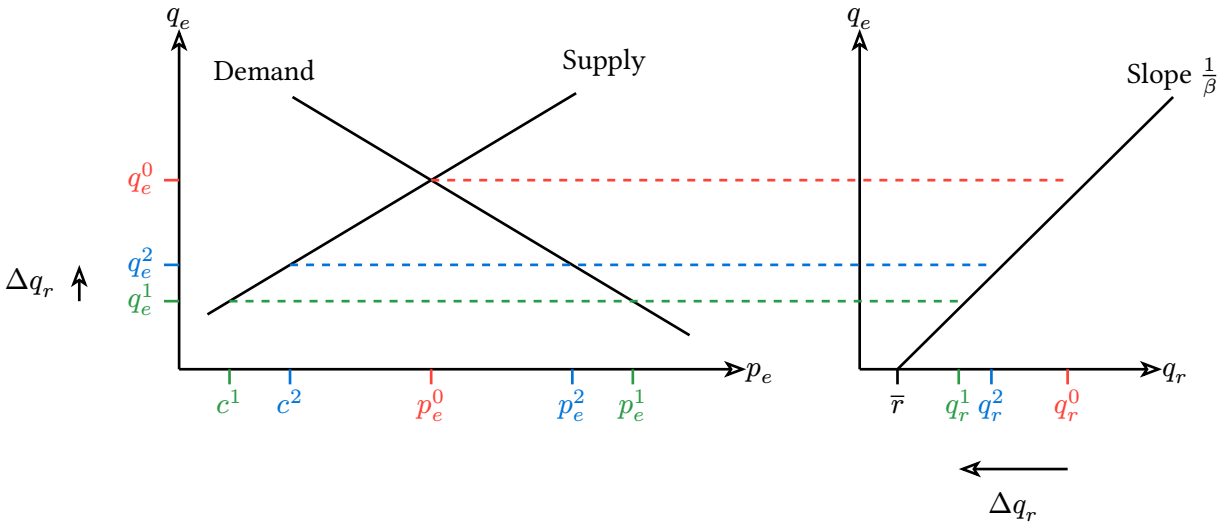


Figure 3: Endogeneity of raise service reserve requirements: Suppose that scarcity pushes the allocation away from the baseline of p_e^0, q_e^0, q_r^0 to p_e^1, q_e^1, q_r^1. Since $q_e^0 + q_r^0 < K$ cannot be physically provided, the energy price must rise to scarcity level p_e^1 to reduce energy demand to q_e^1. This loosens the requirements for raise service reserves by Δq_r, which frees up that capacity to be used to serve more energy.

In Appendix B° on page 51 the assumption that β is positive is empirically tested with ARIMA regressions, using data for every generator in Australia's NEM. After controlling for a confounding variable and serial correlation, the estimate of β is positive and statistically significant, except for rooftop solar. The algebra above does not require the assumption that $\beta > 0$. If $\beta < 0$, the finding about scarcity price extremity is merely reversed. Negative feedback is turned into positive feedback, but the pedagogic iteration will still converge for $|\beta| < 1$.

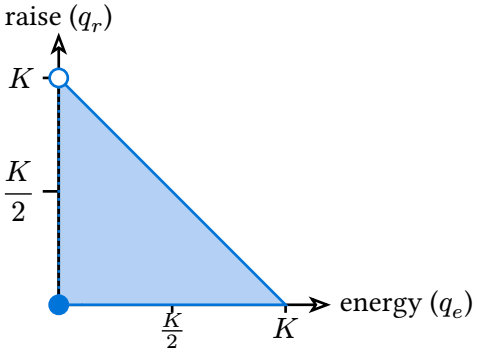
Section 1.2.4 and subsequent sections revert to a fixed exogenous demand R. The combination of fixed costs and endogenous demand is discussed in Appendix B.3 on page 53.

1.2.4 Fixed Cost Component

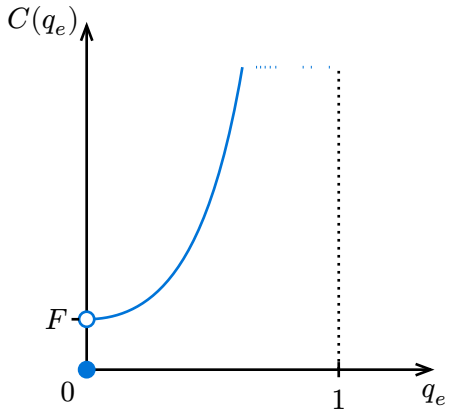
Next I extend the model by making a crucial distinction between:

- generators which are providing some energy ($q_e > 0$), which are 'warm' or 'spinning', and can thus provide raise service; and
- generators which are providing no energy ($q_e = 0$), which are 'cold' or 'stationary', and cannot provide any raise service.

This is shown in Figure 4. The consequence of this is that now firms face a non-trivial decision about how much raise service to provide. Previously the optimal choice for a given energy output was always the maximum amount $q_r = K - q_e$. Now, from the perspective of a given generator, it may be optimal to provide neither energy nor raise service: $q_e = 0 = q_r$, to avoid paying the fixed cost component F.



(a) Possible production set. A generator cannot provide any raise service without providing energy (If $q_e = 0$ then $q_r = 0$). The difference between this and Figure 2a is that now the left side of the triangle is open (dotted line), not part of the set.



(b) Cost Function, with fixed component F

Figure 4: Model setup with fixed cost component

Due to the fixed cost component the socially optimal allocation may be asymmetric, despite symmetric firms. A subset of generators N will be running (supplying energy at positive cost) and the remainder will be off (zero cost). The possibility of a generator turning off is a key difference between this model and that of Gilmore, Nolan, and Simshauser. As a simplification, I do not explicitly model the minimum generation level of each generator. The production set includes arbitrarily small output levels, which is unrealistic. However the fixed cost component ensures that such outputs are never chosen in equilibrium, thus capturing the effect and tradeoffs of minimum output levels.

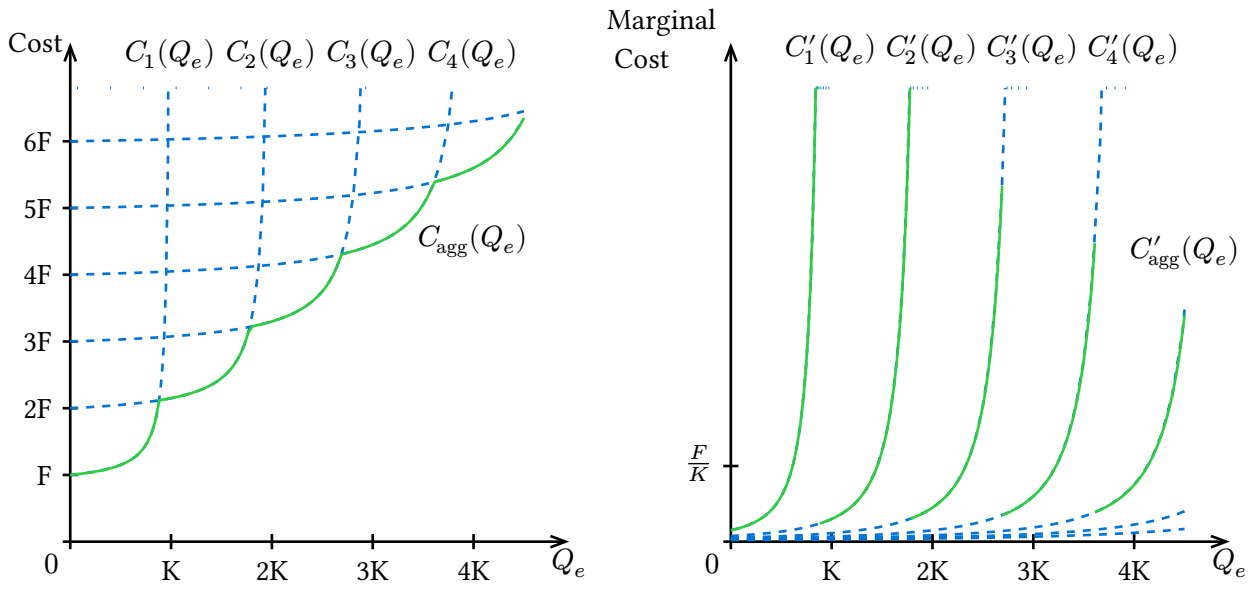
Let $C_N(Q_e)$ be the cost of supplying aggregate energy Q_e with N generators running. Let $C_{\text{agg}}(Q_e) = \min_N C_N(Q_e)$ be the minimum of these, as shown in Figure 5a. The derivative of this function is shown in Figure 5b. To supply a marginal increase in energy demand, a social planner may either:

- increase the output of generators currently running, even though they all have a relatively high marginal cost; or
- turn on an additional generator, which has a low marginal cost, but incurs an additional fixed cost (and potentially reduce the output of the generators currently running, which are more expensive).

For most output levels, the cheapest approach to supply a marginal increase in demand is to increase the output of the generators which are currently on, even though they have a higher marginal cost than generators which are currently off ($C'(q_e) > C'(0^+)$). Turning on an additional generator is only optimal when the additional fixed cost is outweighed by the savings from reducing the output of the currently-running generators which are operating in regions of relatively high marginal costs. Let these breakpoints be called $\widehat{Q_N}$, which satisfy $C_N\left(\widehat{Q_N}\right) = C_{N+1}\left(\widehat{Q_N}\right)$. This leads to a marginal aggregate cost curve which is piece-wise continuous. It is locally increasing, yet globally decreasing.

To supply a marginal increase in demand for the raise service, a social planner may either:

- decrease the output of generators currently running (and decrease energy demand); or
- turn on an additional generator (and potentially reduce the energy output of existing generators).



(a) Aggregate cost function

(b) Derivative of cost function, i.e. marginal cost

Figure 5: Optimal allocation to supply finite demand from pool of many generators.

1.2.4.1 Propositions

Locally increasing: The marginal energy cost curve has a positive slope everywhere (except at the breakpoints where marginal cost is undefined): $C''_{\text{agg}}(Q) > 0 \forall Q \notin \{\widehat{Q_N}\}$

Globally decreasing: $\forall \varepsilon > 0, \exists Q : \forall Q' > Q, C'_{\text{agg}}(Q') < \underline{C'} + \varepsilon$ where $\underline{C'} > 0$. Equivalently, the marginal energy cost at the breakpoints are a decreasing sequence converging to a strictly positive value

Finite allocation: Even if an infinite number of generators are available, the optimal generation output is a finite number of generators producing a finite amount of energy. This per-generator output does not approach zero as aggregate supply increases: $\lim_{Q\to\infty} \frac{Q}{N_Q} > 0$

A contribution of this thesis is that this is an example of a situation in the electricity industry where marginal costs are decreasing (in a sense) with respect to quantity. The proofs are in Appendix A° on page 49.

1.2.4.2 Paying for Installed Capacity or Spinning Capacity Instead of Spinning Reserves

Generators in the NEM are paid for spinning reserves, not installed capacity. e.g. a 100 MW generator providing 60 MW of energy can be paid for up to 40 MW of raise service, but a 100 MW generator providing 0 MW of energy is paid for 0 MW of raise service.

Next I consider two other hypothetical forms of payment structure, and explain why they are less desirable than paying for spinning reserves.

Installed capacity: Firms could be paid for the amount of capacity installed (K), regardless of whether it is spinning. (100 MW in the example above.) Ex ante this will provide incentives for firms to enter the market (at the top node of Figure 1) by building more infrastructure, which helps ensure that the demand for the raise service is met. However, ex post, for a given asset the decision to run at all, or reduce energy output to provide more spinning reserves (the 2nd and 3rd decisions in Figure 1) would not change the payment to a generator under this scheme. In practice the value of R varies intraday, thus the market would not clear on operational timescales (hours), only on investment timescales (decades).

Spinning capacity (spinning reserves plus utilisation): Instead of paying for just unused spinning reserves (q_r), firms could be paid based on total spinning capacity, including the capacity already used to provide energy. ($q_e + q_r = 60 \text{ MW} + 40 \text{ MW} = 100 \text{ MW}$ in the example above.) This creates an incentive to run when a generator would not otherwise run (the second decision in Figure 1). However, conditional on running at all, it does not create any incentive to reduce energy output to provide more raise service (the third decision in Figure 1). This creates two problems:

- The market for raise services will not clear if energy prices are high enough that all generators would run anyway. This would result in energy demand being met, but reserves being unmet, which means the grid would be unacceptably vulnerable to shocks. It is preferable to raise energy prices to suppress demand, to allow some generation capacity to be held in reserve in case of a shock.
- Even if the market clears, it might do so by paying generators to turn on to provide the additional raise services, incurring high fixed costs, when it would have been socially cheaper to increase the raise service output (decrease the energy output) of the generators which are already running (and decrease energy demand through higher energy prices).

In summary, structuring the payments based on a different definition of capacity leads to inefficiencies or the market failing to clear.

In practice generators are heterogenous. Only some can provide raise services, and marginal costs differ. The optimal allocation may be to decrease the energy output of a generator which can provide raise services, and increase the energy output of a generator with higher marginal costs which cannot provide raise services, rather than start new generators. Such co-optimisation of energy and ancillary services is a strength of market design in Australia, compared to Europe.

1.3 Accounting for Usage Payments

The *Contingency* raise service is used to respond to large, unexpected emergencies. These are so rare that until now, I neglected the payment for energy and the increased fuel cost when the reserve provider is called upon to increase output. A similar service called "regulation raise" in Australia, or "frequency containment reserve" (FCR) in Europe, is used to respond to small, frequent shocks (e.g. noisy demand). Regulation raise services are called upon in most trading intervals. Therefore the energy revenue and fuel costs for utilisation are non-negligible in expectation. In this section I extend the foundational model from Section 1.2.2 to account for these utilisation costs (without the fixed cost extension introduced in Section 1.2.4).

After offering reserves, generators may be called upon to utilise all, none, or only some of those reserves. As a simplifying assumption I choose a binomial distribution. Suppose that there is a probability α that the full quantity of reserves offered (q_r) will be called upon, and $1 - \alpha$ probability that no reserves will be utilised.

For now, assume that the utilisation of reserves is uncorrelated to other parameters, such as demand and supply of energy (i.e. uncorrelated with energy price). As a justification for this assumption, consider the recent example of the Iberian blackout on 28 April 2025. Immediately prior to the incident demand, prices, weather, supply availability and fuel mix were typical (ENTSO-E 2025; OMIE 2025). As another example, consider the unexpected fire at the Callide C coal turbine in Queensland in 2021. Immediately prior to the incident demand, prices, weather, supply availability and fuel mix were normal (AEMO 2021). Energy prices rise after such events. However, the purpose of contingency raise service reserves is to respond within an energy trading interval. Energy prices cannot change within such an interval, so when reserves are utilised, the generator is paid ordinary pre-incident energy prices, not elevated post-incident energy prices.

For given prices p_e, p_r, the profit of a firm is:

$$\pi = \begin{cases} p_e(q_e + q_r) + p_r q_r - C(q_e + q_r) & \text{if called upon: } \Pr = \alpha \\ p_e q_e + p_r q_r - C(q_e) & \text{if not called upon: } \Pr = (1-\alpha) \end{cases}$$

I assume risk neutrality. Expected profit is:

$$\mathbb{E}\pi = p_e \times (q_e + \alpha q_r) + p_r q_r - \alpha C(q_e + q_r) - (1-\alpha)C(q_e) \tag{1}$$

Suppose a firm offers all their capacity as either energy or reserves ($q_e + q_r = K$), and is called upon to provide energy with all those reserves. Since $C(q_e + q_r) = C(K) = \infty$, the firm will make an infinite loss. So instead (for $\alpha > 0$) firms will offer $q_e + q_r < K$. (This may be why Gilmore, Nolan, and Simshauser assume a constant marginal cost in their model.) Therefore we can assume an internal solution. Combining first order conditions when maximising Equation 1 yields:

$$p_r = (1-\alpha)(p_e - C'(q_e)) \tag{2}$$

$$p_r = \alpha(C'(q_e + q_r) - p_e) \tag{3}$$

In words, Equation 2 says that firms should choose energy quantity q_e such that the expected opportunity cost of not generating more energy equals the price earned for providing reserves. Equation 3 says that firms should choose the reserve quantity q_r such that the expected marginal cost of being called upon to provide more reserves equals the price earned from providing those reserves. These two equations are symmetric, because the problem could instead be considered as one of providing $q_e + q_r$ of energy, with a $(1-\alpha)$ probability of being called upon to provide a *lower* service (i.e. being available to reduce production level quickly).

$\alpha = 0$ is the case where utilisation is so rare that it can be neglected, which yields the same results as Section 1.2.2. For $0 < \alpha < 1$, a higher α leads to a lower p_r. This is because the more likely the utilisation of the reserves is, the lower the opportunity cost is in expectation.

Extending this to a full equilibrium model is outside the scope of this thesis. Doing so is complex because in reality the demand for raise services is correlated with energy demand (Billimoria, Mays, and Poudineh 2025), and energy prices are *positively* correlated with supply availability (Gilmore, Nolan, and Simshauser 2024).

1.4 Link Between Ancillary Services and Capacity Markets

Over the past few years "capacity markets" have become quite topical in the industry. These are for *installed* capacity, as an alternative to the energy-only paradigm. There are parallels between installed capacity markets (discussed in Part 2°) and ancillary markets for spinning reserves (discussed in this section).

Ancillary markets exist because wholesale energy prices vary on the timescale of minutes, whereas electricity supply and demand must be balanced on the timescale of seconds and hundreds of milliseconds. Hypothetically, if energy spot prices varied on the timescale of milliseconds and had no administrative price cap, then ancillary markets would not be needed (Newbery 2016). After a shock (such as a transmission line failure) the spot price would rise and fall within seconds, rewarding the generators which left aside spinning reserve to fill the gap with sufficient responsiveness. This is not technically feasible due to the time required for independent system operators (ISOs) to run constrained linear optimisers to clear the market, and liquidity concerns. Additionally, due to the short and infrequent nature of such events the energy-only price required to incentivise spinning reserves or installing new generators would be extremely high, even compared to today's energy spot market caps. Thus the slow time scales and price caps on energy are market failures, which is why many regions have markets for *raise* services, which are markets for spinning capacity on the timescale of seconds.

Part 2 Capacity Markets

Capacity markets are used to pay for the *capability* to produce energy (installed capacity) in addition to the payment for energy itself. This is typically justified by the existence of an exogenous, regulatory price cap, which creates the "missing money problem". (Other justifications exist, such as market power, which are not examined here.)

The definition of capacity markets is that producers are paid regardless of whether they produce energy. Thus it should not be surprising that a large drawback of capacity markets is that payments are made to generators who do not produce energy when it is most needed (McRae and Wolak 2019). The purpose of this section is to review the justifications for capacity markets, and discuss the extent to which they distort firms' decisions to provide energy when it is most needed.

In this section the term "capacity" refers to installed capacity, not the spinning reserves for the *raise* service, which were discussed in Part 1°.

2.1 Justifications for Capacity Markets

The main justification for capacity markets is that the energy-only paradigm is allegedly insufficient, covering short run costs but not upfront capital costs, thus creating the "missing money" problem. For example this missing money problem is the official justification given for the United Kingdom's capacity market (Department of Energy and Climate Change 2014). However, they are not precise about *why* there is a missing money problem in an energy-only paradigm. It can be shown that with an energy-only market, infra-marginal profits during periods of scarcity pricing will drive entry into the market to the efficient level (Newbery 2016; Cramton and Stoft 2005). Therefore energy-only markets do not necessarily have a missing money problem, and capacity markets can only be justified in the presence of market failures. The main relevant market failures are:

Price Caps: There is generally no upper limit on the price a supermarket can charge for milk, but most liberalised electricity markets have legal limits on the price at which a generator can sell electricity. Some examples are 4,000 €/MWh in Europe, 20,300 AUD/MWh in Australia and 120 ₹/MWh (120 €/MWh) in India[5] (EPEX SPOT 2020; AEMC 2025a; CERC 2023). If this limit is equal to value of lost load (VOLL)[6] there will be enough scarcity rent to drive investment to the efficient level (Borenstein and Holland 2005; Newbery 2016). However, for political reasons and to mitigate market power, the price cap is typically set far below VOLL (Cramton and Stoft 2005; EC 2016). This leads to inefficiently low investment in generation.

Difficulty Hedging Long Term: Electricity futures contracts are typically only available up to a one year maturity (Newbery 2016), compared to the decade or multi-decade lifespan of generation assets. Retailers may not fully hedge their consumption (reducing liquidity for generators) because they face *quantity* risk (customers may leave) as well as price risk.

Paternalistic Intervention: Consumers and their retailers are not incentivised to hedge fully, because they anticipate that if prices rise to extreme levels the government will paternalistically intervene (Keppler, Quemin, and Saguan 2022; Batlle et al. 2023). A notable example is France's response during the gas crisis. The government's objective was to reduce electricity bills for consumers. If investors anticipate this style of intervention, they will underinvest, which will exacerbate the crisis. Subsequent french legislation now *guarantees* intervention (Batlle et al. 2023). Thus such interventions are counterproductive.

[5] India's electricity commission goes so far as to use supply scarcity as the *justification* for their extremely low cap, which deters investment (CERC 2023).

[6] In situations of extreme scarcity the last resort to balance supply and demand to avoid a grid-wide outage is load shedding. i.e. a random subset of consumers are forcefully disconnected. This forgone consumption is a mix of low value and high value use cases. The average value per unit of energy of this mix is called the VOLL.

Public Good Nature of Adequacy: One of the unique aspects of electricity compared to other goods is that a slight shortfall in supply can quickly lead to all demand being unmet[7]. Investors do not internalise the reduction in outage probability from additional investment. Consumers do not internalise the increase in outage probability from their consumption in times of scarcity.

Regulatory Volatility: Over the last two decades there have been frequent policy changes in the electricity sector. Many regions have swapped between multiple different tools to tackle climate change, and adopted or changed capacity markets. Regulatory volatility has been particularly common in Europe. Several countries are forcing the premature closure of nuclear generators, often followed by extensions and reversals of these plans (e.g. Belgium, Germany). Newbery (2016) provides an excellent example of how the United Kingdom's carbon price floor was to be escalated in a predictable schedule, but that schedule was discarded just three years later, "subject to the whim of chancellors". Policy volatility significantly impairs investor confidence in a sector dependent on infrastructure with high upfront capital costs and multi-decade lifespans.

These market failures are linked. Price caps reduce the cost of insufficient hedging for retailers, so they under-hedge. Regulatory volatility can still yield efficient investment, as long as firms can hedge against that risk with Arrow-Debreu securities (Newbery 2016), however futures markets are illiquid.

2.2 Not All Megawatts Are Created Equal

In Section 1.2.4.2 I demonstrated that in capacity markets for *spinning* reserves, using the wrong metric for capacity may yield inefficiencies, or even result in a market which is unable to clear. In the context of capacity markets for *installed* capacity, the same thorny questions arise. What is the definition of capacity? Capacity markets are intended to increase the number of megawatts of installed capacity, but how are those megawatts measured? Are some megawatts more useful than others? The purpose of this section is to discuss how the value of a megawatt of capacity varies greatly (even for a given fuel type) and how capacity markets try, and fail, to capture this difference.

The legislation for the Belgian capacity market describes it as "technology neutral". However, by royal decree 1 MW of gas capacity is treated as equal to 100 MW of solar capacity (Elia 2024). Spain uses a similar approach (Wynn and Julve 2016). Differences in capacity factor are one reason for this derating based on fuel type. (e.g. 1 MW of solar capacity produces 0 MW at night time, so 1 MW of nameplate capacity corresponds to less than 0.5 MW production on average.) However these ratios are too extreme to be explained by capacity factor alone. The main motivation for this discrimination based on fuel type is that it is a crude heuristic to pay more to generators who provide energy in times of need, than those who do not. Whilst that objective is sound, there are two problems with this approach.

This first problem is that these "derating coefficients" are somewhat arbitrary. Elia based their 1:100 ratio on modelling about how much solar may contribute in critical periods in the future. This is circular, since the amount of solar available will be impacted by the extent to which investors are rewarded by the capacity market, which depends on which derating coefficient is chosen. Once some projects are receiving money for capacity payments, the likelihood that others will build generators without it will be reduced. Thus the job of a capacity market designer becomes that of a central planner, picking winners and deciding the optimal fuel mix (Newbery 2016). This is a concerning reversion from efficient free-market designs, both from a cost and emissions perspective.

The other issue with this approach is that it suppresses market signals to make improvements to generators within a fuel type. Conditional on choosing a fuel type, there are many investment and

[7] If supply is less than demand, the electrical frequency will start dropping below 50Hz. As the shortfall continues, the frequency continues to drop. Eventually generators and loads will disconnect automatically to protect themselves from damage, cascading into a grid-wide outage.

operational decisions which can improve the probability that a generator will supply energy during critical periods, at a cost.

Location: It may be socially optimal to build new wind farms not where the wind is strongest, but where it is uncorrelated with existing wind farms. (Analogous to the tradeoff in Part 3°.)

Grid Constraints: It is increasingly common that generators (particularly new renewables) have their output restricted due to finite grid transmission capacity. Location choice is one way to mitigate this. For example, Germany should build new wind farms in the south, where there is high demand and low supply. They continue to build them in the north, where there is already much wind generation, and limited transmission capacity to the south. The same applies on a smaller geographical scale. Another approach to mitigate grid constraints is to increase utilisation of finite transmission capacity by over-provisioning solar (McArdle 2022), or adding a battery onsite to spread generation across the day[8].

Design: As discussed in Part 3°, the tilt of solar panels can increase their output in the late afternoon and early evenings, which is the precisely the period in the day when capacity concerns are most relevant. The same is true of double-axis and single-axis tracking (where solar panels move to follow the sun). Wind turbines can be optimised for low wind speeds (when available capacity from other wind farms are most lacking).

Dynamic Performance: Some generators can start up, and ramp (adjust output level) quicker than others of the same fuel type. This allows them to better take over when sunshine and wind drops quickly, or faster than expected.

Outage Scheduling: In countries with cold winters and high penetration levels of solar, capacity will likely be most scarce in winter. Thus solar farms should schedule maintenance outages during summer. However, most capacity markets and renewables subsidies incentivise choosing winter for outages, when the forgone quantity (MWh) is lower.

Maintenance and Outage Prevention: Generators can choose to keep more spare parts in stock, and invest more in maintenance. Coal can be kept unfrozen (Cramton and Stoft 2005).

These measures are generally costly but of value to society. Thus a capacity market should reward them. Derating factors which cannot be improved through design and operational improvements fail to do so. Cramton and Stoft (2005) give the example of a hypothetical "dog" generator which is out for maintenance most of the time, and has a startup time too slow to respond to a crisis. In the New England capacity market this generator would be rewarded, despite being unable to contribute capacity when needed.

Early capacity market designs aimed to replace the forgone scarcity rents during the few hours per year when price caps bind, and pay them back as capacity payments, spread over the whole year (Cramton and Stoft 2005). If capacity payments are not based on the performance during those hours or there are insufficiently strong performance penalties, then they will distort investment. This leads to overinvestment in poorly performing generators (such as legacy coal generators which are unreliable and inflexible), and perhaps underinvestment in more performant assets (EC 2016).

If capacity markets are to solve the missing money problem, and the missing money problem is caused by administrative price caps which are lower than VOLL, then a first-best capacity market design would pay per unit of energy (MWh) delivered during periods of scarcity, at a price equal to the difference between VOLL and the price cap. This would be algebraically equivalent to raising the energy price cap, albeit through two payment channels instead of one. However, this is not feasible, for the same reasons that the price cap is too low to begin with. (Susceptibility to market power, political

[8]The findings in Part 4° are based on the assumption that the grid connection is not restricted nor costly.

undesirability of price volatility, and difficulty in estimating VOLL.) Thus, such a design is merely useful as a baseline for a second-best capacity market.

The New England capacity market involves a subtraction of infra-marginal rents during scarcity from the capacity payments, to mitigate the market power concern (Cramton and Stoft 2005). However, estimating infra-marginal rents is difficult. Would a hypothetical reference generator already be on before the start of a price spike, or start turning on at the start of the price spike? Inserting estimations of gas generators' marginal costs into market regulations is fraught with danger. Australia's NEM includes provisions to mitigate market power based on whether the average spot price remains above gas generator prices for extended periods (300 AUD/MWh). During the gas crisis of 2022, higher gas prices doubled marginal costs to 600 AUD/MWh, whilst the threshold in regulation remained at 300 AUD/MWh. The ensuing mismatch (in addition to other factors such as weather-based outages[9] of coal generators) resulted in the first ever suspension of the market (AEMO 2022). This same kind of problem could occur if legislators design capacity market performance criteria using marginal cost figures.

McRae and Wolak (2019) propose a subtraction in the other direction. If the cause of the missing money problem is a lack of long term hedging options for generation investors, the government can offer one-sided contracts for difference (CfDs) with a very high strike price. In this "Colombian model" the generators receive a constant (low-risk) payment each time period, and pay back the government (or consumers) during scarcity periods. The price of the CfD payment is the difference between the spot price (p_{spot}) and strike price (p_{strike}). The quantity is the difference between sent out energy and contracted capacity. Thus generators are penalised for underperforming, and rewarded for overperforming, but less so than in the absence of the CfD (since they are penalised/rewarded at $p_{\text{spot}} - p_{\text{strike}}$ instead of the more efficient p_{spot})[10]. Defining the quantity of the CfD as a pre-agreed amount would retain the efficient short-term incentives, whilst providing risk-reduction and additional cash to generators[11]. This is the same issue as for CfDs to support renewables, which I discussed in IEA (2025).

As a closing example of the importance of performance penalties and firm choices affecting reliability, consider the case of the outage in Texas in winter 2021. Many customers were disconnected due to load shedding. However, the installed gas capacity was sufficient to meet demand. The issue was that many gas generators were unavailable due to frozen pipes. Texas has no capacity market. If there had been a capacity market in place, then it is unlikely that the load shedding would have been prevented by a capacity market (Borenstein, Bushnell, and Mansur 2023). If a hypothetical capacity market withheld payments for those unavailable generators, this would improve incentives. However, many capacity market contracting periods contain no scarcity periods, making performance assessment impossible, resulting in overpayment to underperformers in periods prior to their unavailability.

2.3 How Much Capacity To Procure

After deciding how to measure capacity (and penalise under-delivery) a capacity market designer must choose the quantity to be procured (or the price). Despite the objective of capacity markets being an assurance of system reliability, several European countries have introduced capacity markets without defining a reliability standard (EC 2016). Several more have done so without using their existing reliability standard to guide decisions about quantity.

[9] Whilst coal generators are less weather dependent than renewables, coal mines are susceptible to floods and fire.

[10] Coal and gas generators can produce slightly more than their designed capacity for short periods of time, at the cost of greater wear and tear.

[11] As efficient as can be in the presence of price caps.

The quantity decision is often left to policy makers who are overly cautious, and tend to neglect import capacity (because it is outside their control) . This leads to the procurement of excessive capacity, which suppresses energy prices, which exacerbates the missing money problem that capacity markets were supposed to solve (Newbery 2016; EC 2016). This vicious cycle is illustrated in Figure 6. As an example, Spain has a low price cap and capacity markets. (This was the case even before renewables were common.) Their gas generators have exceptionally low capacity factors (utilisation) of 10% (Wynn and Julve 2016; IEA 2015), which is indicative of overcapacity from overinvestment.

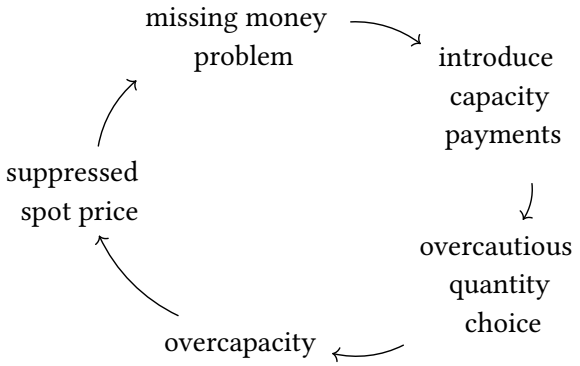


Figure 6: Flow chart illustrating how capacity markets can exacerbate the problem they are intended to solve

2.4 Capacity Markets and Emissions

Capacity markets are not *directly* motivated by climate policies or variable renewable energy (VRE). The market failures which motivate capacity markets (described in Section 2.1) can be present without renewables. However, recent increases in penetration levels of renewables, and the shutting down of aging coal generators have been a catalyst for introducing capacity markets.

Emissions reduction and capacity adequacy are two different objectives. As per the Tinbergen Rule, these two objectives should be achieved through two policy instruments. However, a more cynical view would be that capacity markets are a way to deliberately subsidise uncompetitive fossil fuels[12].

Economic theory clearly shows that the most effective climate policy is to charge a Pigouvian tax on emissions, equal to the marginal damage of those emissions (or equivalently, to implement a cap-and-trade quota). Despite this, carbon prices tend to be politically infeasible. As a second-best option, policy makers in most regions resort to subsidies for low-emissions renewables.

It is worth making explicit what the mechanism is through which subsidies for renewables reduce emissions. Solar and wind generators do not *remove* carbon from the atmosphere. In fact, the manufacturing of those generators results in some emissions (such as CO_2 from concrete setting in wind turbine foundations). These generators are subsidised to help them out-compete more emissions-intense fossil fuel generators. The market failure is that fossil fuel firms do not face the true cost of their emissions. Subsidies to low-emissions renewables artificially reduce the apparent cost of renewables, as a second-best attempt to level the playing field from the other side.

Whilst subsidies to renewables may increase uptake of renewables, that alone would not reduce emissions if the amount of fossil fuel burning remained the same. Rather, the suppression of spot price by the subsidies is intended to make some fossil fuel plants unviable, so that they produce less, and many shut down permanently. Without this, renewables subsidies could not possibly reduce emissions

[12]Wynn and Julve (2016) claim that this was even stated explicitly by the Spanish energy minister in 2015, who allegedly proposed a levy on renewables to fund capacity payments for fossil fuels. However, I cannot find any sources to corroborate this claim.

in the electricity sector[13]. Thus **the mechanism through which renewables subsidies achieve their policy objective is to deliberately create a missing money problem for fossil fuels**. If a capacity market is introduced to reintroduce the fossil-fuel profit which renewables subsidies deliberately took away, the capacity market would be directly negating the climate policy. If many coal and gas generators find it economically unviable to enter or remain in the market, that is not necessarily a sign of market failure. It may be the recovery from one.

2.5 Not All Megawatt Hours Are Created Equal

When estimating how much generation capacity is needed, capacity market designers and ISOs tend to size the capacity based on maximum demand. However, this may not be appropriate going forward. As a case study, consider California's grid on August 14 and 15, 2020. Solar power dropped quickly in the late afternoon, demand rose quickly, and due to a confluence of factors other generation could not pick up the slack. Customers were forcefully disconnected (load shedding) *after* the maximum demand for that day (CAISO 2021). This example shows that when considering how much generation capacity is adequate, one must consider *when* the energy can be produced.

In grids with a mix of VRE and fossil fuels, VRE provides cheap electricity when it can, and fossil fuels fill in the gaps with expensive electricity when VRE cannot. As penetration levels of renewables increase, those gaps shrink. The fixed operating costs of existing gas peaker plants and the entry costs of new ones must be recovered through a shrinking window of hours. Therefore the price in those hours will tend to increase. The purpose of this section is to look at volume-weighted average prices of different fuel types. This is important from a climate perspective, as well as another perspective to view the derating coefficients used by most capacity markets to discriminate based on fuel type.

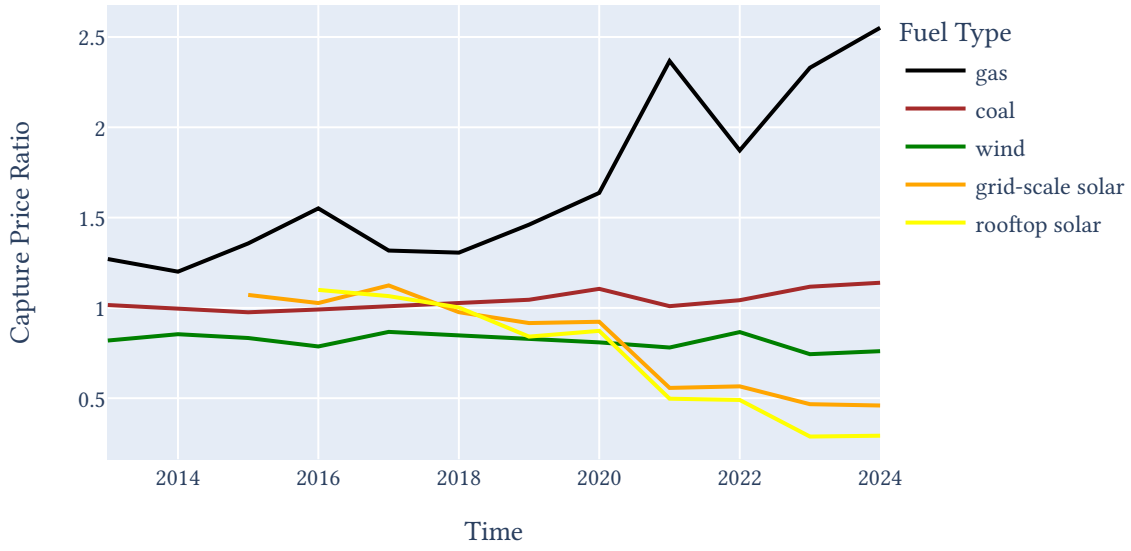
The volume-weighted average price of a generator's energy is called the *capture price*[14]. The ratio of this to the unweighted average price grid is called the capture price ratio (Cabot and Villavicencio 2024)[15]. A generator which provides a constant power output at all times has a capture price ratio of exactly 1. A generator which turns on only for high price spikes (such as a gas peaking plant) would have a capture price ratio higher than 1.

The capture price ratio over time for generators in Australia (aggregated across all regions) is shown in Figure 7a. There are some clear trends. Coal plants have a capture price ratio slightly above 1, which reflects the fact that they are almost always on. The capture price ratios of gas and (non-hydro) renewables are diverging starkly. In 2024 grid-scale solar power was sold for a volume-weighted average of 53 AUD/MWh, compared to 266 AUD/MWh for gas. If we ignore climate externalities, these numbers reflect the difference in value provided by each fuel type. Gas generators can start, stop and adjust output level more quickly and cheaply than coal generators, so gas generators can provide electricity which is more than twice as valuable (per MWh) than electricity from coal. These trends are expected to be similar around the world. Despite this, The royal decree for Belgium's capacity market values gas and coal almost equally per MWh (Elia 2024).

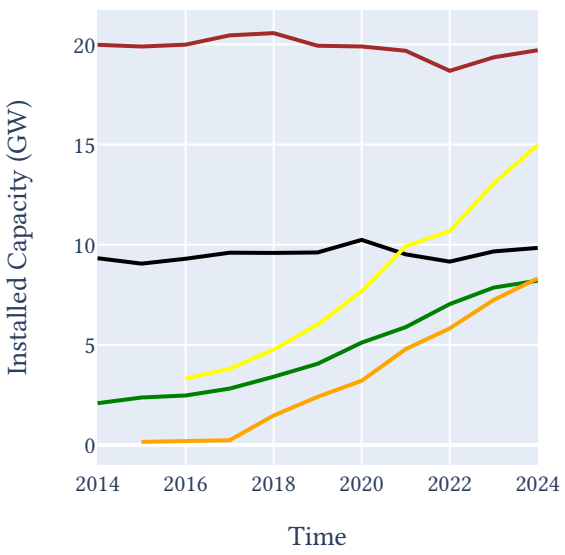
[13]Renewables subsidies could reduce emissions through suppressing electricity prices to incentivise changes in end-uses from other energy sources, such as incentivising a swap from petrol cars to electric.

[14]Revenue from ancillary services is excluded here. Including it does not change the story.

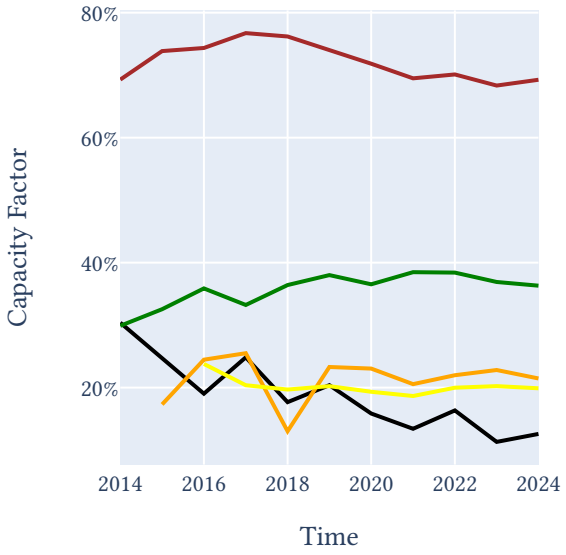
[15]The ratio of capture price to the average of price weighted by the total market volume is called *participation factor*. A representative generator which adjusts its output proportional to the overall supply would have a participation factor of 1. Both metrics yield similar insights.



(a) Capture Price Ratio



(b) Installed Capacity



(c) Capacity Factor

Figure 7: Divergence of capture price ratio in Australia's NEM, for selected fuel types. Installed capacity of wind and solar is increasing over time (b). This new VRE capacity is crowding out gas, leading to lower utilisation (c). However, since gas is now turning on for only the highest-priced times, it is earning far more per unit of energy than wind and solar (a). In 2024 gas earned 5-9 times more per unit of energy than solar (2.5 vs 0.5 and 0.3).

The concerning drop in capture price ratio for solar and wind is endogenous. Wind or solar farm owners want strong wind or strong sunshine, so they can produce large quantities of energy. However, when that happens other wind and solar farms are also producing more. Thus there is a strong, negative correlation between wind or solar output and price, leading to low capture price ratios. This *correlation penalty* is increasing over time, due to increasing penetration levels of renewables. It is stronger for solar than for wind, because wind speeds vary over large distances, whilst sunshine is strongly correlated across longitudes (especially in the NEM's narrow north-south grid). The capture

price ratio for rooftop solar is lower than for grid-scale solar, because rooftop solar is not economically curtailed for very negative prices, and many grid-scale solar plants use tracking systems to angle panels to follow the sun[16].

2.6 Capacity Market: Are They Worth It?

The main argument in favour of capacity markets is that low spot price caps create a missing money problem. This argument can be reversed. The distortions created by capacity markets are a strong justification for increasing price caps closer to the efficient value of VOLL, to avoid needing capacity markets. The European Commission described the tradeoff succinctly (prior to the 2022 gas crisis), by saying that "no capacity mechanism should be a substitute for market reforms" (EC 2016).

If there is underinvestment due to an inability to hedge long term, it is not clear that introducing capacity payments with a tenor far less than a generator's lifespan will resolve the issue. Capacity markets typically discriminate based on technology type. In addition to hindering climate goals, derating coefficients which are not based on actual per-generator performance disincentivise within-technology choices which could increase supply availability when it is most scarce. If capacity payments are to be made, performance penalties should be strong, and based on a metric similar to capture price ratio, calculated per generator, in a way that can be influenced by their design and operational decisions. The quantity of capacity to procure tends to be too much, set independently of reliability goals. This leads to overinvestment, which exacerbates the missing money problem which capacity markets were supposed to solve. This may lead to higher electricity bills for consumers, with worse supply availability.

[16]Whilst households are not exposed to the spot price, this metric is still useful to capture the *social* value of their generation. Furthermore, the way rooftop solar investors are shielded from spot prices with a fixed tariff is not unique. It is equivalent to the way grid-scale solar investors are shielded from spot prices with power purchase agreements.

Part 3 Optimal Solar Panel Slope

This section introduces a model and extensions to explore a tradeoff between maximising solar energy *volume* and *value*, because solar power output is negatively correlated with price.

3.1 Motivation and Context

When installing fixed-tilt solar panels, conventional wisdom is that you should point them north/south towards the equator, and tilt them at an angle equal to the latitude of the site. This maximises the total sunshine captured, which maximises the volume of energy produced. However, due to the volatility of the spot price of electricity, some megawatt-hours are 10,000 times more valuable than others. Whilst there is some unpredictability in the spot market, on average there are clear trends in most regions of the world. Solar power tends to be more valuable (per unit of energy produced) during mornings and evenings, and during winter.

The causal link is straightforward. When there is a lot of sunlight (such as a summer day at midday), there is an abundant supply of solar power, so there is an abundant supply of electricity from all sources, so wholesale electricity prices are low. When there is less sunlight (late afternoon or winter), the supply of solar power is more scarce, so supply of all power is more scarce, so prices are higher. Thus a solar investor, if exposed to the wholesale spot price, should point their panels more west, and more vertically than conventional wisdom suggests, to shift the times when they produce large quantities of energy to when prices tend to be higher. This research assumes that the wholesale spot price reflects the marginal value of energy to society. Of course a lack of carbon pricing undermines this assumption. Differences in marginal pollution abatement throughout each day are left for future research. It is expected that they would only strengthen the findings here.

For this paper the foundational model will be focused on large scale solar farms with fixed-tilt panels. I will abstract from from the distinction between investors, operators and offtakers, using the single term "firm"[17]. As an extension, the same model will be applied to household solar, examining how investor decisions are distorted by fixed time-of-use tariffs which shield consumers from spot prices.

Katzen and Leslie (2024) contribute to the literature of zonal vs nodal pricing by demonstrating that paying renewable generators the same price across a geographical region yields inefficient investment outcomes. This model is similar, but for the case when generators face the same price across time.

Badran and Dhimish (2024) conducted an engineering experiment with panels which are completely vertical. They demonstrate that "bifacial" (two-sided) vertical panels can produce even larger quantities of power than standard diagonal, fixed-tilt, monofacial panels. This surprising finding is mostly because bifacial panels can consume sunlight on both sides of the panel, including ambient reflected light. They compare returns on investment based on higher volumes of generated energy, and lower engineering costs. However, they assume that each unit of energy produced by a solar panel is of equal value (because rooftop solar tariffs tend to pay a flat rate). I will strengthen their findings by investigating how the energy from vertical panels has higher *social* value per unit, due to the shift in time of day when it is produced, for otherwise identical panels. Single-axis and double-axis panels which move within each day to follow the sun are out of scope of this analysis.

3.2 Simple Algebraic Model

The stylised model consists of 3 time periods, as shown in Table 1. Prices increase as the sun sets (due to scarcity of solar power, and typical demand patterns), so $0 < p_1 < p_2 < p_3$. Prices are taken as given. A full equilibrium model is left for future research. I neglect constraint management, balancing, solar forecast errors and negative prices. Therefore ex post there are no *operational* decisions for a

[17] Offtakers and investors both wish to maximise the size of the pie. Relative negotiating power should merely change how the pie is divided. This is discussed further in Section 3.5.

solar farm to make. In this model, the only decision is an ex ante investment decision about the angle at which to install the panels at. I simplify the decision to installing the panels either:

Horizontally: to maximise output at midday, with zero output in the afternoon; or
Vertically: to maximise output in the afternoon (when prices are higher), but generating less volume.

Quantities are normalised such that a horizontal panel generates 1 at midday, and a vertical panel generates a smaller amount $0 < \delta < 1$ in the afternoon.

Time Period		Sunlight	Price		Volume	
					Horizontal	Vertical
1	midday	strong	p_1	low	1	0
2	afternoon	weak	p_2	medium	0	δ
3	evening	none	p_3	high	0	0

Table 1: Setup of time, prices and quantities. $0 < p_1 < p_2 < p_3$ and $0 < \delta < 1$. The firm chooses whether to install panels horizontally or vertically.

The profit from a horizontal panel is $\pi_h = 1 \times p_1 + 0 \times p_2 + 0 \times p_3 = p_1$, and the profit from a vertical panel is $0 \times p_1 + \delta \times p_2 + 0 \times p_3 = \delta p_2$.

"Capture price" is the volume-weighted price of the energy produced by a generator. In this example with only one period with non-zero production, the horizontal panels obtain a capture price of p_1 and the vertical panels have a higher capture price of p_2.

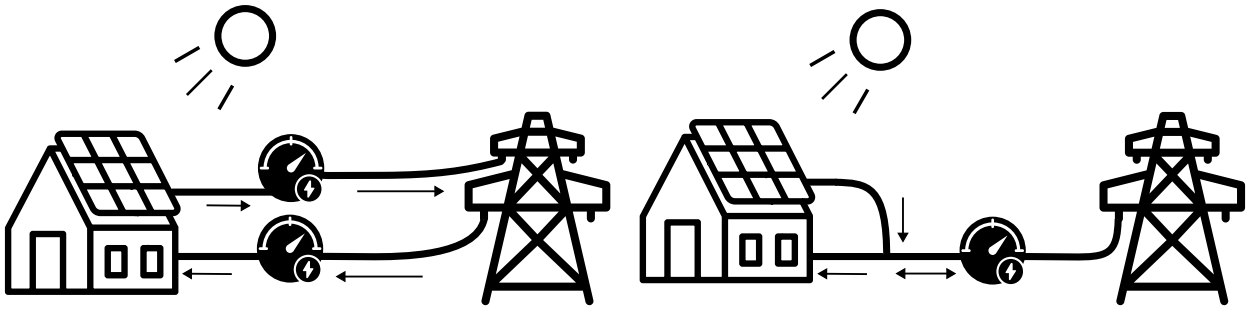
The firm should choose to orient the panel vertically if $\delta p_2 > p_1$. i.e. only if the increase in capture price in the afternoon outweighs the loss of production volume from a geometrically inferior tilt. This result is unsurprising, and is merely presented as a baseline.

3.3 Potentially Distortive Fixed Tariffs

Next I consider how different types of feed-in tariffs can distort investor decisions. Here I focus on household rooftop solar. In practice, rooftop solar panels are often installed at whatever slope the roof happens to be, so the choice is between roof faces spread about the compass (e.g. north facing vs west facing). Such a choice involves the same tradeoff between maximising quantity and capture price, so I will retain the horizontal vs vertical stylised model.

Many rooftop solar installations are paid a fixed feed-in tariff, which only changes on the timescale of years. These are often for the purpose of subsidising renewables, although some are merely an unsubsidised risk-hedging retail offering, similar to residential consumption contracts.

Suppose a rooftop solar installation receives a fixed feed in tariff, with "gross" metering, as shown in Figure 8a. This means that all generated power is sent into the grid. If any solar energy happens to immediately flow into that same house for consumption, it will be through the 'front door', charged the same as power consumed from other sources. This means that decisions about solar panels are additively separable from the household's electricity consumption, so I can neglect the household's consumption.



(a) Gross metering: there are two different measurements[18], and potentially two different prices

(b) Net metering: power from the rooftop solar panel can be consumed "behind the meter". The retailer sees, and charges, only one flow

Figure 8: Illustration of the difference between net and gross metering for a house with rooftop solar[19]

Let the price paid to the household for its solar output be ρ_{out}, which is independent of wholesale prices p_t. If the panel is installed horizontally or vertically, the owner receives $\pi_h = 1 \times \rho_{\text{out}}$ or $\pi_v = \delta\rho_{\text{out}} < \pi_h$. The flat tariff removes the tradeoff between large volumes at a low price and small volumes at a high price, because the price is the same either way. In the case where a vertical panel is socially optimal ($\delta p_2 > p_1$), such a tariff distorts the incentives, costing society $\delta p_2 - p_1$. Increasing ρ_{out} through a subsidy does not change this incentive. It only impacts total investment quantity (outside the scope of this model), not the decision about the installed angle.

Next I consider the case of net metering, also known as "behind the meter" solar, shown in Figure 8b. Let s_t be solar output in time t, and d_t be exogenous electricity consumption inside the house (e.g. television, stove). Suppose the household faces tariff ρ for both consumption and generation, such that their electricity bill is $C = \rho \sum_{t=1}^{3} d_t - s_t$. This minimisation problem is separable, thus the decision about which angle at which to install the panels at remains the same as for the gross metering case, which yields distorted outcomes.

3.4 Impact of Subsidies

In practice, the spot price of electricity in many regions is lower than the social cost, due to the externality of untaxed carbon emissions. Instead renewables may receive an additive subsidy s per unit of energy they produce. In this case, a rooftop investor would choose to install panels vertically only if $\delta(p_2 + s) > p_1 + s$. When $p_1 + s(1 - \delta) > \delta p_2 > p_1$, investors will make the socially sub-optimal decision of installing panels horizontally. Larger subsidies lead to larger distortions.

Macklin (forthcoming) proposes an improved subsidy where projects earn $p_t \times (1 + s')$ instead of $p_t + s$. Under such a scheme, investors would choose to install solar panels vertically if $\delta p_2(1 + s') > p_1(1 + s') \Rightarrow \delta p_2 > p_1$. Thus a multiplicative subsidy does not create this particular distortion.

These inefficiencies are in terms of the direct non-climate costs of electricity production. As the sun sets each afternoon, solar power output tends to reduce and eventually reach zero. Fossil fuel generators generally increase their output to pick up the slack. Thus emission intensities tend to be lowest when solar power is producing the most. This means that shifting solar output through installation angle to be later in the afternoon may produce larger marginal emissions abatement per unit of energy. I hypothesise that once marginal emissions abatement is accounted for, the inefficiencies produced by paying residential solar installations a flat price will be inflated even more than my model suggests. However, investigating this is left for future research.

[18] In practice there is one meter, one box, but it contains two logical "circuits".

[19] Image components were sourced from Rutmer Zijlstra, yode and Vectors Point via The Noun Project (CC BY 3.0)

3.5 Large Scale Project: Power Purchase Agreements

So far these fixed tariffs have been discussed in the context of residential rooftop solar. Most large solar farms are contracted under power purchase agreements, such that the investor receives a fixed price ρ for their power, independent from the spot price (ARENA 2021). This setup is equivalent to that of the household, except the price ρ offered to investors by offtakers is endogenous. Since offtakers receive the spot price and pay a fixed price, they will be willing to offer a higher fixed price if the panels are oriented such that power is generated at times of higher spot prices (yielding a higher capture price). The offtaker and investor have conflicting interests. The investor wants to maximise volume (for a given ρ). The offtaker wants to maximise capture price (for a given volume). Assuming perfect information and a spot price which matches social value (e.g. a carbon price is applied), vertical panels are a Pareto improvement over horizontal ones if $\delta p_2 > p_1$, which matches the social optimum. Therefore by the Coase theorem the two parties should decide to reduce quantity to increase the capture price only if it is socially desirable to do so.

Many grid-scale solar projects use single-axis or double-axis tracking, which move the panels throughout the day to follow the sun. This increases both volume and capture price. As the cost of panels continues to decrease compared to the motors which move them, we may see a trend towards cheaper fixed-tilt panels at large sites because they yield a lower levelised cost of energy (Dedenbach 2025). The decision to install fixed-tilt or single-axis panels is comparable to the decision to install vertical fixed tilt panels or horizontal fixed tilt panels. Additive subsidies could distort decisions away from single-axis (high value) to fixed-tilt (low cost) installations when not socially optimal. The same is true of power purchase agreements if the offtaker does not give adequate consideration to how design decisions influence capture price. Thus the findings in this paper will have increasing policy relevance over the coming decade.

3.6 Data Analysis

I have analysed historical pricing data for Australia's NEM in 2024. For each possible azimuth (compass) and zenith (slope) angle, the volume of energy (MWh) and value ($) of a hypothetical 1 MW fixed-tilt monofacial solar panel was calculated. Atmospheric and weather effects were neglected, because accounting for them typically requires expensive commercial software. The results are shown in Figure 9.

For each region, the orientation which yields the maximum revenue is somewhere between that which yields the maximum solar energy volume, and that which yields the maximum price per unit of energy. In Queensland the revenue-maximising orientation earns approximately double the revenue of the volume-maximising position. Fixed feed-in tariffs for rooftop solar therefore incentivise household investors to angle panels such that up to **half the total economic value is lost**, even without considering differences in emissions abatement. These results align with Brown et al. (2024).

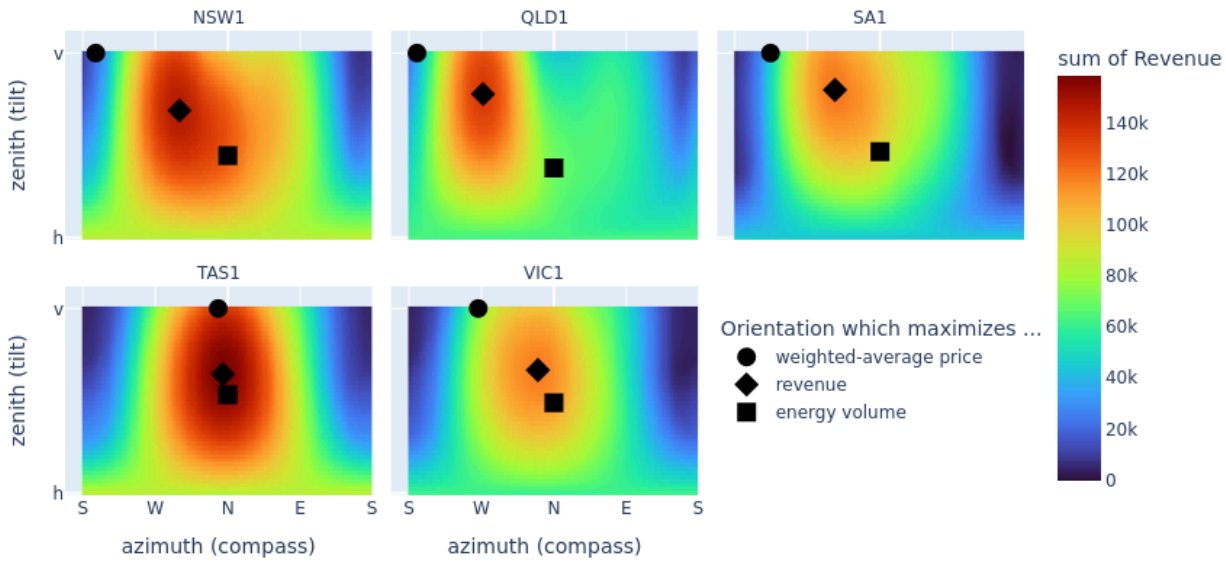


Figure 9: Simulations of optimal zenith (horizontal to vertical tilt) and azimuth (compass orientation) of 1 MW fixed-tilt solar panels, when earning the spot price, and trying to maximise energy volume (MWh), revenue ($) or capture price (MWh / $). Colour represents revenue. Prices for Australia's NEM, 2024.

3.7 Model Implications

In conclusion, configuring solar panels to maximise the volume of energy they produce may not be socially optimal. If the prices are sufficiently volatile ($p_3 \gg p_2 \gg p_1$), then vertical panels yield higher spot revenue, despite lower energy volume. Feed-in tariffs for rooftop solar can distort investment decisions so much that up to half the social value of the solar panels is lost.

The decision about the angle to install solar panels at is just one stylised example of the tradeoff between value and volume. The findings here are relevant to other decisions. For example, maintenance outages for a solar farm can be scheduled for summer, when the forgone volume is large, or winter, when solar is more scarce so the forgone capture price is large. This discussion is therefore generalisable.

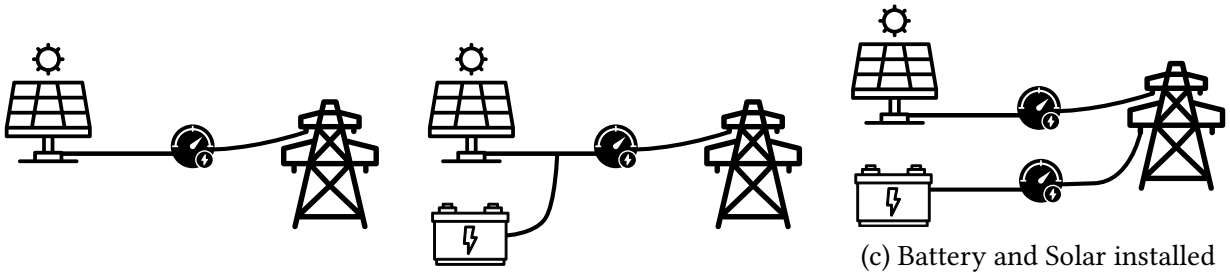
Part 4 Adding Batteries: The Lack of Spot Revenue Benefit from Colocation

4.1 Motivation

"Colocation" refers to installing batteries and solar (or wind) together at the same site, as shown in Figure 10b. When examining the benefits of colocation, many investors, grid operators and policy makers choose a solar project without batteries as the counterfactual (Figure 10a). Given the high upfront cost of batteries, a more suitable counterfactual is that of solar and batteries installed at separate locations, possibly by separate firms (Figure 10c). From this perspective, there are many practical benefits of battery colocation. For example:

- Only one network connection is needed. This matters because grid connections are often scarce, and costly to acquire (Zhao et al. 2015).
- Network charges, charges for ancillary services, and dispatch target deviations due to weather forecast errors can be improved through co-optimising (Yang et al. 2021; Ma et al. 2019).
- There are efficiencies of scale for project management. e.g. Electricians only need to visit one site to maintain both panels and batteries.

Many researchers and industry participants conflate these benefits with the direct spot-revenue impact of charging a battery from solar. Naemi, Davis, and Brear (2022), Zhao et al. (2015) and Wong et al. (2019) suggest that batteries help solar (or wind) generators store power when the sun is shining and prices are cheap, and discharge it later when prices are higher. Whilst this 'solar sponge' approach does yield higher revenue than a solar farm without a battery, it also costs more to build. Conditional on buying a battery, it is not obvious whether colocation has any impact on spot revenue, compared to a solar farm and battery installed separately.



(a) Solar Only (b) Battery and Solar Colocated (c) Battery and Solar installed separately

Figure 10: Visual explanation of colocation[20]. It is often claimed that colocation (b) yields higher revenue than (a). Since batteries are expensive, a more suitable comparison is (b) and (c)

As an example, a common argument is that when spot prices go negative (or below zero minus the subsidy price) a colocated project can charge the battery for free using on-site solar generation, without wasting that sunlight or paying to export. However, in such a situation it would be more profitable to withhold the solar generation (waste sunlight) and get paid to charge from the grid at the negative price. This is exactly how separate batteries and solar would behave (either if owned by the same firm, or separate firms). Therefore I argue that time-shifting renewable power to increase revenue (or decrease emissions) is not a valid argument in favour of colocation.

The purpose of this section is to extend the model in Part 3° to investigate the impact of colocation on spot market revenue. The angle at which solar panels are installed at is an investment decision, and the charging schedule of the battery is an operational decision. I will investigate whether colocating batteries with solar changes the optimal angle which panels should be installed at, the optimal battery

[20]Image components from Larea°, Vectors Point° and yode° via The Noun Project° (CC BY 3.0)

charging revenue, or total spot market revenue. The logistical considerations mentioned above will be neglected for this model. Only spot revenue will be considered. Upfront investment costs will be neglected, since they would be the same with both colocation and separate sites. It is assumed that the grid connection is not constrained. This extension applies to large scale projects, because I assume revenue is based on spot price. For households with solar and a battery, the same principles would apply, were they not distorted by fixed tariffs.

4.2 Solar With a Battery Which Can Charge From the Grid

I now extend the model shown in Table 1, Section 3.2 on page 25. Suppose that a battery is installed with the solar panels. Now the operator faces the choice of whether to immediately export solar power when it is generated, or store it for later.

The battery is sized such that its depth (duration) is one time interval, and maximum charge and discharge is normalised to 1, which is the same as the size of the solar panels. i.e. when the solar panels are producing at their maximum, the battery will charge from completely empty to completely full in exactly one interval. I assume there are no subsidies. For now, I assume that the battery could also be charged from the grid. If the solar panel is producing $0 \leq s_t \leq 1$ the battery may take s_t from the solar panel, and up to $1 - s_t$ from the grid, to charge up to its maximum of 1.

Now the firm has two sets of decisions to makes:

Ex ante: whether to position the solar panel horizontally or vertically; and
Ex post: when to charge and discharge the battery. Equivalently, whether to immediately export solar, or store it for later.

The sequence of events is:

- Prices are known in advance.
- The owner chooses whether to install the solar panel horizontally or vertically.
- The battery starts completely empty.
- T_1:
  - the solar produces 1 if horizontal or 0 if vertical
  - the battery can be charged from the grid and/or from the solar (if any)
- T_2:
  - the solar produces 0 if horizontal or δ if vertical
  - the battery can be charged from the grid and/or from the solar (if any), or can discharge (if not empty)
- T_3: Solar produces 0. The battery may be discharged into (or charged from) the grid

Assume the battery has a round-trip efficiency of $0 < \gamma < 1$. i.e. if charged until full (consuming 1 unit of energy) it will later provide γ units when discharged until empty.

I proceed with backwards induction. Since $p_3 > 0$, the optimal plan is to discharge whatever power remains at the start of the final period. Proceeding backwards to T_2, suppose the firm chose to install the panels horizontally. The operational decision regarding when to charge and discharge becomes either:

Store solar: Use all the solar power in T_1 to charge the battery to full, then discharge in the last (highest priced) period, earning revenue of γp_3
Export solar immediately then stop: Export solar to the grid as soon as it is generated (T_1), to earn p_1 of revenue. No solar is able to be generated in subsequent periods. Choose to not charge from the grid at all. Do not use the battery.

Export solar immediately then arbitrage: Export solar to the grid as soon as it is generated (T_1), to earn p_1 of revenue. In T_2 and T_3, do battery-only arbitrage. i.e. Charge from the grid in T_2 and discharge in T_3, yielding $\gamma p_3 - p_2$ in additional revenue (only worthwhile if $\gamma p_3 > p_2$). Thus the total revenue is $p_1 - p_2 + \gamma p_3$.

The optimal charge schedule is to store solar power and discharge later during high prices, if the price increase is large enough to outweigh the round-trip storage losses. Otherwise, the best choice is to immediately export solar, and then do nothing. (Battery-only arbitrage in periods 2 and 3 is not profitable unless it is even more profitable to store from the solar panel.) Note that this matches whether and when a standalone battery without solar would be charged to do arbitrage from the grid.

If instead the solar panels were installed vertically, the charge decision is between:

Charge from solar, without additional arbitrage: In T_2, store the δ of solar energy that is generated. Discharge in T_3 to export this, yielding $\delta\gamma p_3$ in revenue.

Charge from solar, with additional arbitrage: Since $\delta < 1$, the battery is not filled up by the solar output. So we can fill up the remaining capacity by charging from the grid in the prior period. (We could instead charge from the grid in T_2, but that charge would cost more because $p_2 > p_1$.) Thus we charge $1 - \delta$ from the grid in T_1 at a cost of $(1 - \delta)p_1$, charge δ from solar in T_2, and discharge γ (i.e. 1 minus round trip storage losses) in T_3, giving γp_3 in discharge revenue. Total profit is $\gamma p_3 - (1 - \delta)p_1$

Immediately export solar, without additional arbitrage: Export all the solar produced in T_2, giving δp_2 in revenue. Do not charge or discharge the battery.

Immediately export solar, with additional arbitrage: Export all the solar produced in T_2, when we produce it, earning δp_2 in revenue. Charge the battery fully from the grid in T_1, at a cost of $1 \times p_1$. Store it through T_2 until T_3. Discharge fully in T_3 to receive γp_3 in revenue, giving $-p_1 + \delta p_2 + \gamma p_3$ in profit.

The revenue is:

$$\pi = \begin{cases} \delta\gamma p_3 & \text{(charge from solar, without additional arbitrage) Never Optimal} \\ -p_1 + \delta p_1 + \gamma p_3 & \text{(charge from solar, with additional arbitrage) Never Optimal} \\ \delta p_2 & \text{(immediately export solar, without additional arbitrage)} \\ -p_1 + \delta p_2 + \gamma p_3 & \text{(immediately export solar, with additional arbitrage)} \end{cases}$$

It is not worth charging the battery with solar power if the prices vary too little for the arbitrage to cover the storage losses ($\gamma p_3 < p_2$). When that is not the case ($\gamma p_3 > p_2$) it is *still* not optimal to charge the battery with solar power, because even greater profit can be obtained by immediately exporting the solar in T_2, and using the battery to capture a larger price difference ($p_3 - p_1 > p_3 - p_2$) from T_1 to T_3 with energy from the grid[21]. This is a **key takeaway** of this model. Figure 11 shows the difference between the two strategies. There is value in a 'solar sponge' battery which stores solar power when the sun is bright, and discharges it later when prices are high and the sunlight is low, when compared to not being able to store the power, and ignoring the sunk cost of the battery. However, this is not strictly better than charging from the grid using even cheaper power (which may be solar power from a neighbour, instead of one's own solar). In simpler terms, moving a battery and solar farm together, such that charging from solar is no longer an explicit expense, does not eliminate the opportunities and opportunity costs of supplying and consuming from the spot market.

[21] This is true even if the battery is smaller in size, to match the maximum vertical solar output δ.

A vertical panel gives higher profit than a horizontal panel if and only if $\delta p_2 \geq p_1$. This is the same condition as when deciding the optimal tilt for solar without a battery. Thus, the solar tilt decision, and battery charge decisions are separable. **Colocation of batteries and solar does not provide any new opportunities in the spot market, compared to separate battery and solar projects.** (Distortive household tariffs or subsidies may change this.)

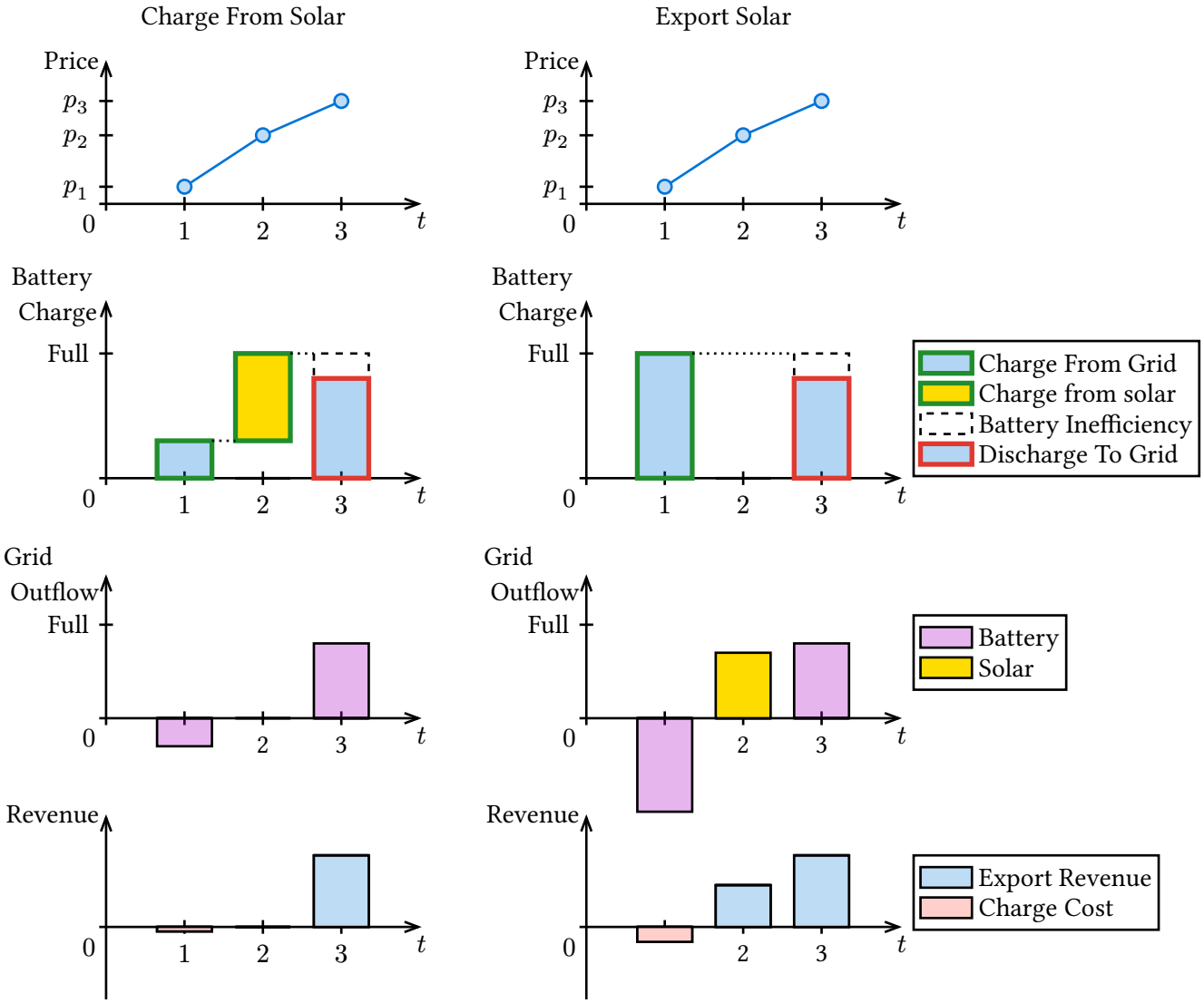


Figure 11: Charging strategies for a battery and vertical solar panel. On the left is what happens when the battery is charged from the solar power in T_2 (with the remaining battery capacity charged from the grid in T_1). On the right is what happens when the solar power is immediately exported in T_2, and the battery arbitrages power from the grid by charging fully from the grid in T_1 and discharging in T_2. Charging from the grid (on the right) yields higher revenue than charging from solar (left).

4.3 Solar With a Battery Which Cannot Charge From the Grid

In practice, many colocated 'solar with battery' projects are configured such that the battery can only be charged from the solar panels, never from the grid. This may be due to engineering reasons, or regulatory reasons. Now the choice of the firm becomes:

Ex ante: Choose whether to install the panels vertically or horizontally; and

Ex post: Choose whether to immediately export solar when it is generated, or use all of it to charge the battery. (Interior solutions cannot be optimal.)

Conditional on charging, the optimal discharge decision is to wait until the last period, because that is when the price is the highest.

For a horizontally positioned solar panel, the choice is:

- Use the 1 unit of solar production in T_1 to charge the battery. Discharge the battery (less the losses) to earn γp_3 in T_3; or
- Immediately export all solar generation in T_1, to earn p_1. The battery is not able to charge from the grid in subsequent periods.

Both cases yield equivalent outcomes to having a solar panel and battery installed and operated separately. The battery should be charged if and only if $\gamma p_3 > p_1$. (i.e. if price increase is enough to overcome the storage round trip loss.)

For a vertically positioned solar panel, the choice is:

- Immediately export δ units of power in T_2, earning δp_2. Do not use the battery.
- Charge the battery with δ of solar power in T_3. Even though it is not full, the remainder cannot be filled from the grid. Discharge it in T_3 to give $\delta\gamma p_3$ in revenue.

Note that in the case where $\gamma p_3 > p_1$, the outcome is strictly worse than if the battery could be charged from the grid (or equivalently, if the battery was installed at a separate site, which could be charged from the grid). This is similar to Figure 11, except the inability to charge from the grid means the more profitable situation on the right is not possible, and on the left the small arbitrage of grid power from T_1 to T_3 is also not possible. Another example, beyond this 3 period model, is that if a battery cannot charge from the grid, then it cannot earn revenue by arbitraging power within the time between sunset and sunrise. If firms are able to pay more in engineering or regulatory costs to allow the battery to be charged from the grid as well as from solar, then for a sharp enough daily price curve it may be worth it to do so. This would require an explicit cost-benefit analysis. Anecdotally, recognition of the spot revenue benefit of being able to charge from the grid is lacking in the industry.

4.4 Model Implications

- When the battery can charge from the grid, the optimal charge/discharge plan is identical to if a firm had a battery and solar project separately.
- If the batteries can also charge from the grid, then it is optimal for the solar to be vertical if and only if it is optimal for solar on its own to be vertical.
- If the battery cannot also charge from the grid, then the optimal tilt of the solar panels may be horizontal in cases when the optimal tilt of solar on its own is vertical. The intuition of this coupling is that the constraint preventing the battery from charging from the grid limits arbitrage opportunities (making the firm worse off). Keeping the solar panels horizontal instead of vertical shifts the potential charging time earlier. This yields a larger duration, and thus larger price difference for temporal arbitrage with the battery.

Contrary to popular opinion, colocating batteries and solar on the same site compared to different sites provides no improvement or even a reduction in spot market revenue. For colocated projects, preventing the battery from charging from the grid reduces the set of possible arbitrage strategies.

Part 5 Inelastic Demand and Flat Tariffs

5.1 Motivation and Context

Electricity consumers typically sign contracts which offer energy at a fixed price which is independent of the volatile wholesale price (in the short term). Retailers provide these contracts as a hedging service.

In a typical market an increase in price should increase supply and decrease demand. However, with fixed retail prices for consumers, an increase in wholesale price can only increase supply, not decrease demand (in the short term). Therefore, after a given shock, wholesale prices must change by more, in order to balance supply and demand, compared to if consumers were exposed to the spot price. This makes electricity prices quite volatile, and leads to inefficiencies because the value of marginal consumption differs from the marginal cost of supply.

5.2 Model

I will model time-invariant pricing using a simplified version of the model by Borenstein and Holland (2005). The problem is static (only one time period). Suppose that the future state of the world is random: $s \in \mathbb{S}$ (e.g. weather) with known probability α_s where $\sum_{s\in\mathbb{S}} \alpha_s = 1$. These states affect demand (e.g. through heating/cooling) and supply (e.g. through wind power availability). Let the wholesale price be p_s. Let the retail price be ρ per unit of energy, plus a fixed fee A. Assume the supply capacity will be sufficient to satisfy demand for all states of the world. (No outages, no load shedding, no price caps.) Retailers must choose the retail price ρ prior to knowing the state s.

Let:

- $D_s(p)$ represent the aggregate demand curve (quantity as a function of price)
- $S_s(p)$ represent the aggregate supply curve (quantity as a function of price)
- $D_s^{-1}(q)$ represent the inverse demand curve (price as a function of quantity)
- $S_s^{-1}(q)$ represent the inverse supply curve (price as a function of quantity)
- $C(q) = \int_{q'=0}^{q} S^{-1}(q')dq'$ be the (cumulative) supply cost curve.
- $B(q) = \int_{q'=0}^{q} D^{-1}(q')dq'$ be the consumer surplus curve.

In the first best case, where consumers are exposed to the wholesale price, the price would be p_s such that $D_s(p) = S_s(p_s)$. In expectation, the consumer would face price $\sum_{s\in\mathbb{S}} \alpha_s p_s$.

If consumers are exposed to a fixed retail price ρ, then the total quantity demanded is given by $q_s = D_s(\rho)$. To supply that quantity, generators produce with marginal cost $p_s = S_s^{-1}(q_s) = S_s^{-1}(D_s(\rho))$.

A social planner's problem is to maximise expected social welfare W by choosing ρ (and transfer A).

$$W = \arg\max_{\rho} \sum_{s\in\mathbb{S}} \alpha_s[B_s(q_s) - C_s(q_s)] = \arg\max_{\rho} \sum_{s\in\mathbb{S}} \alpha_s \int_{q'=0}^{D_s(\rho)} D_s^{-1}(q') - S_s^{-1}(q')dq'$$

Taking the derivative with respect to ρ (assuming an internal solution):

$$\frac{\partial W}{\partial \rho} = \sum_{s\in\mathbb{S}} \alpha_s[B_s'(D_s(\rho)) - C_s'(D_s(p_s)))]D_s'(\rho) = 0$$

Consumers and suppliers will choose a quantity for a given price so that their marginal benefit/cost matches the price. Therefore $B_s'(D_s(\rho)) = \rho$ and $C_s'(D_s(p_s)) = p_s$.

$$0 = \sum_{s\in\mathbb{S}} \alpha_s(\rho - p_s)D_s'(\rho)$$

Rearranging:

$$\rho = \frac{\sum_{s \in \mathbb{S}} \alpha_s D'_s(\rho) p_s}{\sum_{s \in \mathbb{S}} \alpha_s D'_s(\rho)}$$

This shows that the (second-best) optimal retail price is an average of spot prices, weighted by the probability of each outcome, and also demand elasticity, which impacts how much the equilibrium differs from the first best case. In the special case where $D_s(\rho) = A\rho + B$, then the optimal tariff is simply the expectation of the spot price.

A is a transfer which does not matter for efficiency. Under perfect competition amongst retailers it will be set such that retailers break even in expectation. There are no externalities or public goods in this model, so by the first fundamental theorem of welfare economics, competitive retailers will offer the optimal ρ.

5.3 Discussion

Electricity retail tariffs are typically flat. More complex prices are commercially or politically infeasible, because they are seen as complex, risky or confusing. Borenstein and Holland (2005) estimate that this inefficiency is at least 5% of total wholesale costs, which is economically significant. Imelda, Fripp, and Roberts (2024) find that this inefficiency is larger in a high-renewables grid, which gives this topic renewed policy relevance.

Common approaches to find a middle ground include time of use (TOU) tariffs, where consumers face 2 or 3 flat prices depending on time of day, day of week or season. Critical-peak pricing (CPP) is another approach where consumers face higher rates for a handful of hours or days per year, typically with one day of notice. Cabot and Villavicencio (2024) estimate that CPP reduces the deadweight loss from flat prices by 25% to 50%. In contrast, Hinchberger et al. (2024) estimate that TOU and CPP each reduce the inefficiency by only 10%. They find that far greater efficiency can be obtained by exposing consumers to real time pricing, with aggressive price caps to mitigate risk. However, they do not consider the missing money problem created by such caps, which distort investment decisions as discussed in Part 2°.

5.4 Conclusion

Electricity consumers generally face a flat tariff, so they base their consumption decisions on a price signal which does not match the marginal cost to producers. This leads to inefficiencies. The optimal retail tariff is an average of spot prices across all potential states of the world, weighted not just by probability but also by demand elasticity. Higher penetration levels of renewables increases the deadweight losses caused by flat tariffs. This effect gives renewed policy relevance to middle ground solutions such TOU, CPP and potentially other more dynamic tariffs.

Part 6 The Optimal Retail Customer Mix

6.1 Motivation

Energy retailers face two main risks:

Price Risk: Retailers offer customers a fixed price, and are exposed to a variable spot price

Quantity Risk: Customers may consume more or less quantity than expected

The interaction between these two risks is crucial. For example, a retailer may invest in a solar farm to hedge long term trends in electricity spot prices. However, even if the daily volume of energy (MWh) consumed by customers equals the daily volume of energy generated by the solar farm, the retailer is still exposed to *shape risk*. If customers mostly consume in the evenings, after the sun has set, the solar investment cannot hedge the consumption, so the retailer is left with a short position for those hours. (This shape risk exists for other generator types, such as coal, which typically produces an output that varies within a day by less than a typical customer.)

In Section 5.2, and papers such as Borenstein and Holland (2005), it is assumed that all customers are identical. Of course this is not true. Many businesses consume most of their power during daylight hours, when the spot price of electricity is low. A typical household may consume much of their power around dinnertime (when spot prices are high), and almost none during midday when cheap solar power is abundant, because the occupants are not home. However, a household with a remote worker may consume a greater portion of their consumption during the middle of the day, when spot prices tend to be low, thus making them more profitable for retailers (for the same retail price).

Suppose that retailers are able to discriminate based on these different types of customers (e.g. through targeted advertising)[22]. Some customers have consumption profiles which happen to be strongly and positively correlated with price, others are weakly correlated, and some may even be negatively correlated. The purpose of this section is to examine which mix of customers is optimal. The key contribution of this model is to show that the problem of an electricity retailer targeting different customer types is similar to the problem of an investor choosing an optimal portfolio in the capital asset pricing model (CAPM).

6.2 Model Setup

The model is static (one time period). Suppose a retailer expects the spot price to be $p = \overline{p} + \varepsilon$ and customers of type i have consumption: $q_i = \overline{q} + \eta_i$, where $\mathbb{E}(\varepsilon) = 0$, $\mathbb{E}(\eta_i) = 0$. Intraday load shape is captured by this model through the joint distribution of quantity and price. The retailer's cost, in expectation is given by:

$$\mathbb{E}(c_i) = \mathbb{E}(p \cdot q_i) = \overline{p} \cdot \overline{q} + \mathbb{E}(\varepsilon \cdot \eta_i) = \overline{p} \cdot \overline{q} + \mathrm{Cov}(\varepsilon, \eta_i)$$

Suppose that retailers are risk averse. Thus the retail price in equilibrium for risky customers will be higher than the cost in expectation. Let ρ_i (per-unit price) and A_i (flat subscription cost) be the equilibrium retail price for customer type i.

Suppose there are two types of customers, s (safe) and r (risky), with low and high values of $\mathrm{Cov}(\varepsilon, \eta_i)$ respectively. Each retailer chooses to have a portion α of type s, and $1 - \alpha$ of type r. The risky customers generally command a higher risk premium, so an optimal customer portfolio may still include some risky customers.

6.3 Insights

The profit of the retailer is given by:

[22] Other potential methods include a front book/back book spread, and bundling with other products.

$$\pi = \underbrace{\alpha\rho_s q_s + (1-\alpha)\rho_r q_r}_{\text{usage revenue}} + \underbrace{\alpha A_s + (1-\alpha)A_r}_{\text{subscription revenue}} - \underbrace{(\overline{p}+\varepsilon)\cdot(\alpha(\overline{q}+\eta_s)+(1-\alpha)(\overline{q}+\eta_r))}_{\text{spot cost}}$$

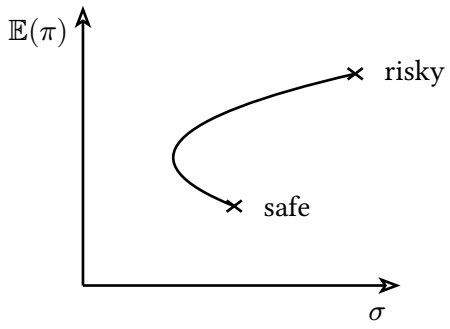
The expected value and variance of profit are derived in Appendix C° on page 54. The results are:

$$\mathbb{E}(\pi) = \alpha\rho_s\overline{q} + (1-\alpha)\rho_r\overline{q} + \alpha A_s + (1-\alpha)A_r - \overline{p}\cdot\overline{q} - \alpha\mathbb{E}(\varepsilon\eta_s) - (1-\alpha)\mathbb{E}(\varepsilon\eta_r)$$

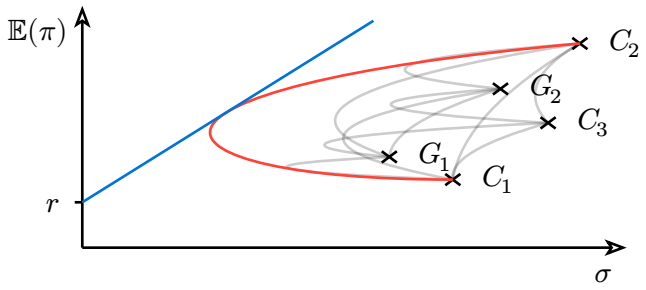
$$\begin{aligned}\operatorname{Var}(\pi) = {} & \alpha^2\rho_s^2 \operatorname{Var}[\eta_s] + (1-\alpha)^2\rho_r^2 \operatorname{Var}[\eta_r] + \alpha^2 \operatorname{Var}[pq_s] + (1-\alpha)\operatorname{Var}[pq_r] \\ & +2\alpha(1-\alpha)\rho_s\rho_r \operatorname{Cov}(q_s, q_s) + 2\alpha(1-\alpha)\operatorname{Cov}(pq_s, pq_r) \\ & +2\alpha^2\rho_s \operatorname{Cov}(q_s, pq_s) + 2(1-\alpha)^2\rho_r \operatorname{Cov}(q_r, pq_r) \\ & +2\alpha(1-\alpha)\rho_s \operatorname{Cov}(q_s, pq_r) + 2\alpha(1-\alpha)\rho_r \operatorname{Cov}(q_r, pq_s)\end{aligned}$$

The key finding here is that $\mathbb{E}(\pi)$ is linear in terms of α, and $\operatorname{Var}(\pi)$ is quadratic. Combining them parametrically yields the efficient frontier shown in Figure 12a. This is qualitatively the same tradeoff as for an investor in the standard CAPM. However, the algebra differs due to the multiplicative nature of the spot price shocks (whereas the standard equity investor model entails additive risks).

Retailers cannot easily hedge against customer *quantity* risk, but they can try to hedge price risk by investing in generation (which has its own quantity risk). The short and long positions (with respect to spot price) of consumers and generators complement each other more so than between customers of different types. This leads to more strongly bent pairwise frontiers in Figure 12b between generator and customer than between two customer types. Generators with daily output profiles most closely matching a given customer type are the strongest hedging pairs (most bent pairwise frontiers).



(a) Efficient frontier for a retailer choosing an optimal mix of risky and safe customers.



(b) Efficient frontier (red) for a vertically integrated 'gentailer' choosing an optimal mix of customers (C_i) and generation (G_i). When including the the risk-free rate of return r, the efficient frontier becomes the tangent in blue.

Figure 12: Efficient frontiers for electricity retailers. σ is the volatility (standard deviation of realised returns), and $\mathbb{E}(\pi)$ is the expected return. Intuitively the tradeoff between expected returns and correlated volatility is the same as the CAPM, although the algebra differs.

The contribution of this model is to highlight the similarities between an investor choosing an optimal portfolio of assets, and a retailer targeting customers with different daily load shapes.

Part 7 Diagonal Dispatch Targets

7.1 Context and Motivation

In a typical electricity spot market time is divided up into discrete periods or intervals. Within these intervals prices are constant. There are discontinuous steps at the boundaries between trading intervals. However, to maintain the stability of the electrical grid, the physical power output of generators (and storage) based on these discontinuous prices are not themselves discontinuous. Rather, the quantity obtained for each generator from the intersection of supply and demand curves is used as a target that the generator must move towards, over time. In Australia's NEM generators must take the entire period to adjust their output, even if they could physically adjust more quickly (AEMC 2025b). The California Independent System Operator (CAISO) requires generators to adjust linearly within each 5 minute dispatch interval, with prices changing every 15 minute *trading* interval (CAISO 2018). This means that prices are piecewise-*constant*, or 'step' functions, but quantities are 'diagonal' piecewise-*linear* 'dot-to-dot' functions. Papers such as Xia and Elaiw (2010) and Wei et al. (2020) consider physical limits on ramp rates (production level adjustment rates) arising from technical limitations of each generator, and how this impacts bidding strategies and social optimums. In contrast, this section considers limits imposed by the market operator (a.k.a ISO) which may restrict adjustment speed to rates slower than what the asset can physically do.

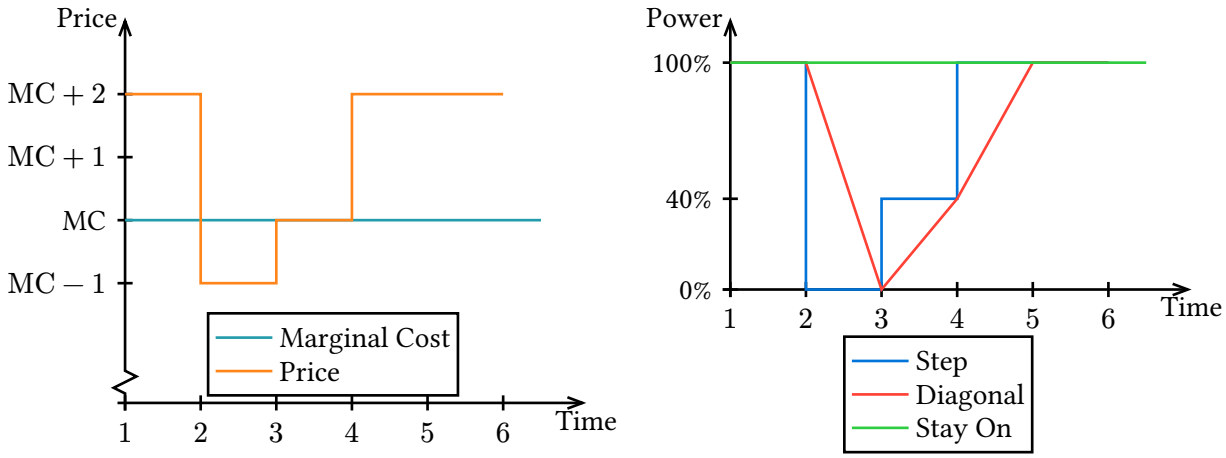


Figure 13: Example price and power series

Figure 13 shows this constraint graphically. In this example the price starts above a generator's marginal cost (assumed to be bid truthfully), drops below, and then rises back up. Under standard economic assumptions the generator and ISO want the generator to instantaneously reduce power output to 0% at the start of period 2 (shown as the blue curve). However, to ensure grid stability, ISOs will instead instruct the generator to ramp diagonally from its starting power level (100% in this case) to 0, over the interval (shown as the red curve). The same is true of scheduled loads and storage. Note that this is true even for assets which are physically able to adjust production level almost instantly (e.g. solar, batteries).

Consequently, the red curve in Figure 13 has an average power of 50% over interval 2, so it is 'half-on', despite having a bid higher than the cleared price. This means that the generator is losing money in interval 2, and missing out on money in interval 4. Table 2 shows that this reduces producer profit in this stylised example by 20%. In general this piecewise-linear production schedule will reduce social surplus compared to a hypothetical piecewise-constant schedule, as each firm's output becomes an average between an optimally efficient solution based on present physical parameters, and what was optimal based on parameters one interval prior.

The generator in this example would make more profit by strategically lowering their bid in interval 2 below marginal cost, to ensure they are fully on in interval 4 (the green curve). This strategy is detailed in Table 2. A contribution of this section is to show that even in the absence of market power, startup costs, *physical* ramp rate limits, and with perfect foresight of prices, rational profit-maximising firms should submit bids that do not match their marginal costs. This novel finding illustrates how simplified analysis using stepped power levels yields qualitatively incorrect conclusions about optimal strategies. Despite these quantitative and qualitative differences, many researchers use the step model for their theory and empirics (Lamp and Samano 2022; Giulietti et al. 2018). The remainder of this paper explores when and to what extent this simplification is an *over*-simplification.

		Step					Diagonal Off-On					Stay On				
Time	Period Start	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	Period End	2	3	4	5	6	2	3	4	5	6	2	3	4	5	6
Power (MW)	Start	1	0	0.4	1	1	1	1	0	0.4	1	1	1	1	1	1
	End	1	0	0.4	1	1	1	0	0.4	1	1	1	1	1	1	1
	Average	1	0	0.4	1	1	1	0.5	0.2	0.7	1	1	1	1	1	1
Price - MC ($/MWh)		+2	−1	0	+2	+2	+2	−1	0	+2	+2	+2	−1	0	+2	+2
Bid - MC ($/MWh)		0	0	0	0	0	0	0	0	0	0	0	<−1	<0	0	0
Profit	Period	2	0	0	2	2	2	−0.5	0	1.4	2	2	−1	0	2	2
	Total			6					4.9					5		

Table 2: Example revenue for different strategies based on prices in Figure 13, for a 1 MW load and 1 hour intervals. Prices and bids are relative to marginal cost (MC). Whilst 'Step' yields maximum profit, it is not allowed, even if physically possible for the firm. 'Diagonal off-on' is when the generator bids truthfully based on their constant marginal cost. A higher profit can be obtained by bidding below marginal cost.

7.2 Model

Assume there are two time periods. Let T refer to the discrete time intervals, and t refer to any instant in continuous time. So $T = 1$ covers period $t \in (0, 1]$, and $T = 2$ covers period $t \in (1, 2]$ Let prices P_T be exogenously given (for $T \in \{1, 2\}$). Assume no market power.

Let Q_T be the nominal power level given by the ISO. The ISO takes all generators' bids (supply curves), and demand curves, and intersects them, then tells each generator to supply a corresponding quantity. The novel consideration of this paper is that generators cannot instantaneously change production output to the new level. A generator assigned power level Q_T (based on its bids) should smoothly adjust its production level towards Q_T, to reach Q_T exactly at $t = T$ (not earlier), from whatever it started at (Q_{T-1}). Even if the firm could physically adjust output more quickly, they are not allowed to. The power in continuous time $q(t)$ is a 'join the dots' sequence of diagonal lines.

$$q(t) = \begin{cases} Q_T \text{ if } t = T \\ Q_{T-1} \text{ if } t = T - 1 \\ Q_{T-1} \times (T - t) + Q_T \times (t - (T - 1)) \text{ if } T - 1 < t < T \end{cases}$$

Define $\overline{Q_T}$ as the *average* power output over period T.

$$\overline{Q_T} = \int_{t=T-1}^{T} q(t)dt = \frac{Q_{t-1} + Q_T}{2}$$

We additionally assume that there are no startup/shutdown costs nor delays, and that firms always

comply exactly with the piecewise-linear production schedule[23]. Without loss of generality, the maximum production level (nameplate capacity) is normalised to 1 MW, with a minimum of 0. (i.e. it is a generator, not storage. The results would be similar for storage.).

Assume starting power $Q_0 = q(0)$ is exogenously given, which reflects the production level at the end of the prior period. Without loss of generality, define prices relative to marginal cost (assumed constant), so $P_T = 0$ is the break even point. Profit in interval T is given by:

$$\pi_T = \int_{t=T-1}^{T} q(t) P_T dt = \overline{Q_T} \times P_T$$

Profit across both intervals is given by:

$$\begin{aligned}
\Pi &= \sum_{T=1}^{2} \overline{Q_T} \times P_T \\
&= \frac{Q_0 + Q_1}{2} \times P_1 + \frac{Q_1 + Q_2}{2} \times P_2 \\
&= \left(\frac{P_1}{2}\right) \times Q_0 + \frac{P_1 + P_2}{2} \times Q_1 + \left(\frac{P_2}{2}\right) \times Q_2
\end{aligned}$$

Assuming perfect foresight of prices, a generator can submit bids to control power Q_1, Q_2 to any level. The maximisation problem becomes:

$$\arg\max_{Q_1, Q_2} \left(\frac{P_1}{2}\right) \times Q_0 + \frac{P_1 + P_2}{2} \times Q_1 + \left(\frac{P_2}{2}\right) \times Q_2$$

The solution for Q_1 is:

$$Q_1 = \begin{cases} 1 \text{ (maximum) if } \frac{P_1+P_2}{2} > 0 \\ [0,1] \text{ (anything) if } \frac{P_1+P_2}{2} = 0 \\ 0 \text{ (minimum) if } \frac{P_1+P_2}{2} < 0 \end{cases}$$

In a more typical model where production levels can be stepped discontinuously at the start of each period, the solution would be to maximise production ($Q_1 = 1$) if $P_1 > 0$ (above marginal cost), else minimise it ($Q_1 = 0$). In this model the produced quantity is an average of each consecutive pair of nominal power targets, so firms should optimise based on the average of consecutive pairs of prices. Therefore the impact of this linear production smoothing (compared to a hypothetical discontinuous stepping) is that the incentives for firms to respond to volatile price spikes are somewhat muted, and smoothed. If a firm submits a bid to produce quantity Q_T and price P_T, they are implicitly consenting to produce at least $\frac{Q_T}{2}$ in the next interval, even if $P_{T+1} \ll P_T$.

The solution for the last period is trivially:

$$Q_2 = \begin{cases} 1 \text{ (maximum) if } P_2 > 0 \\ \in [0,1] \text{ (anything) if } P_2 = 0 \\ 0 \text{ (minimum) if } P_2 < 0 \end{cases}$$

This solution is the same as in the standard model where generators may change their production level discontinuously at the start of each interval. This is merely an artefact of the finite-time model, because there is no subsequent $T = 3$ period that is impacted by Q_2.

[23] In practice generators deviate slightly from the plan due to technical lags, data lags, intra-period weather variability etc.

7.3 Simulations

As shown earlier, the requirement that firms must adjust their output linearly instead of instantaneously introduces a hysteresis. This results in incentives to bid differently to marginal cost, and reduces profit, even for fast assets. Despite these impacts, many researchers model electricity markets using stepped production levels. The aim of this section is to determine whether this simplification introduces material errors. To do so, I construct a simple battery operation optimisation problem, using real price data from several regions, and compare the energy, revenue and cycles both of diagonal and stepped power levels.

5-minute spot prices were obtained from the Australian Energy Market Operator (AEMO) for the NEM and the Midcontinent Independent System Operator (MISO) for the United States (averaged across nodes). (Note that this linear constraint is only present in a minority of ISOs, such as these two.) The analysis period is 2024. The battery is sized at 1 MW, 2 hours depth, with 80% round trip efficiency. Perfect foresight is assumed. Pyomo (a linear optimiser by Bynum et al. (2021)) is used to identify the optimal charge and discharge schedule to maximise energy arbitrage revenue. Ancillary service revenue and other 'revenue stacking' is neglected, despite making up approximately half of battery revenue in practice (Gilmore, Nolan, and Simshauser 2024), because the linear constraint in question only applies to energy.

A subset of the time series results is shown visually in Figure 14, and the aggregate results are shown in Table 3. The absolute level of each metric is unimportant for our purposes. The key question is merely whether there is an economically significant difference between the stepped and diagonal models. For the total amount of energy stored, and the profit made by batteries, the impact of neglecting the diagonal constraint is that the numbers are inflated by up to 4.5%. This is large enough to be consequential for some research questions. Note that these simulations ran faster with the diagonal ramping than without, so computational feasibility is probably not a valid justification for ignoring diagonal ramping. Therefore researchers should account for the piecewise-linear constraint unless they have a good reason not to.

For practitioners, this reduction in profitability is so economically significant that it may make or break the business case for a particular investment. The total amount of energy stored and discharged is lower in the diagonal case. As a silver lining, this corresponds to fewer full cycles, which is good from a warranty perspective.

Region	Sub-Region	Energy (GWh)			Profit ($k)		
		Step	Diagonal	Difference	Step	Diagonal	Difference
Australia (NEM)	New South Wales	2.17	2.06	4.8 %	482.3	465.4	3.5 %
	Queensland	2.27	2.17	4.3 %	394.8	381.4	3.4 %
	South Australia	2.51	2.46	2.1 %	445.9	425.8	4.5 %
	Tasmania	2.03	2.01	1.0 %	205.3	196.2	4.4 %
	Victoria	2.43	2.38	2.1 %	295.9	284.1	4.0 %
MISO	Node Average	1.46	1.40	4.0 %	46.8	45.1	3.7 %

Table 3: Simulation results for a battery's optimal charge-discharge strategy, in different regions, both with and without the assumption of continuous, piecewise-linear (diagonal) production levels. A positive difference means that the stepped model overestimates the true diagonal value.

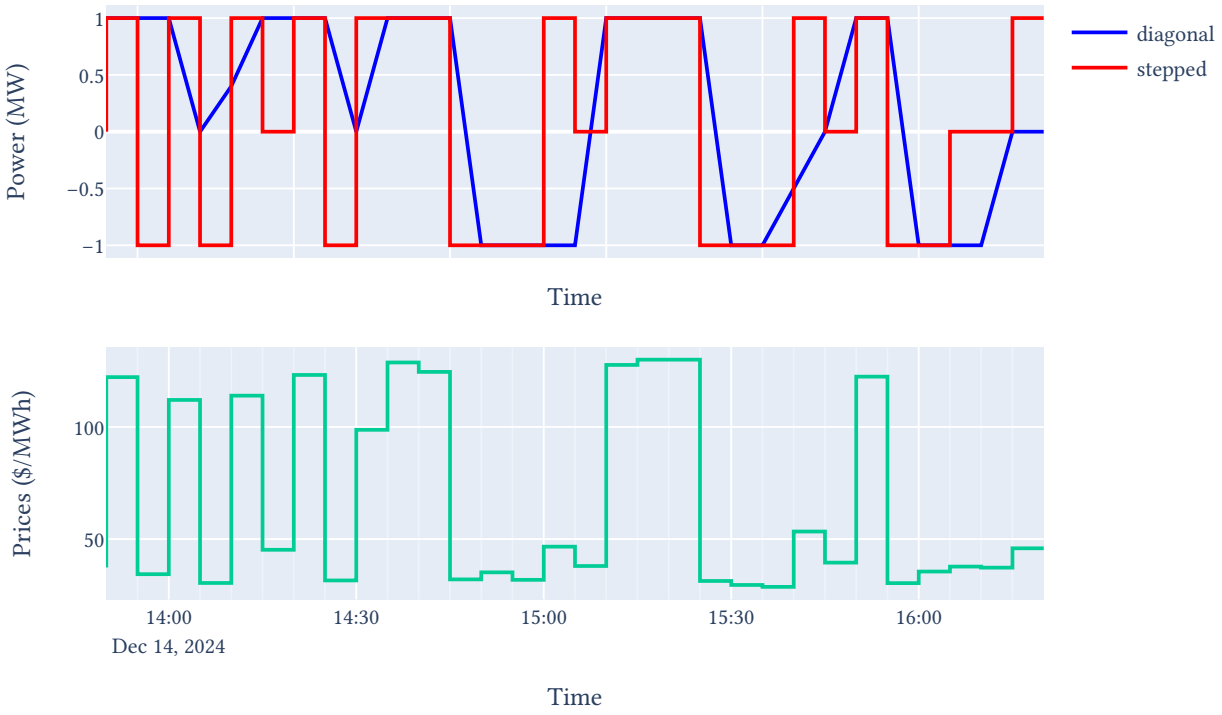


Figure 14: Simulation results for MISO, for a subset of the time period. When accounting for the diagonal piecewise-linear constraint, the optimal charge/discharge strategy is more smooth, and has fewer single-interval dips/spikes.

7.4 Comparison to Europe

In Europe, prior to 2011 there were no rules specifying how generators were to adjust output within each trading period, as long as the average physical output each period matched the commercial target. This resulted in deterministic imbalances at the start of each hour, which increased costs and emissions (ENTSO-E 2011). Since then the rules around balancing have changed. For example, the transmission lines connecting the nordic regions to the rest of Europe are required to adjust their output over 10 minute blocks, to help mitigate these deterministic imbalances (ENTSO-E 2023). Europe's balancing rules are relatively complicated compared to other markets, and vary from country to country. The findings of this model and simulation do not apply to Europe.

Conclusion

Wholesale electricity prices are extremely volatile, however this is not necessarily a problem. The true social value of each unit of energy does vary drastically, even on a second and subsecond timescale. One of the unique aspects of the electricity industry is that if some demand is unmet, this can escalate into all demand being unmet. Therefore after a large shock (such as an unexpected generator outage) it is crucial that supply be adjusted within seconds, to balance supply and demand[24]. However, energy prices are locked in on longer timescales (5 minutes to 1 hour). Therefore there is a missing market for very short term power. This justifies the creation of ancillary markets, such as the market for the Contingency raise service.

Even in the absence of surprises such as equipment failures, the marginal value of electricity varies greatly. However, most liberalised electricity markets have price caps below VOLL, which prevents generators from being paid a price equal to the marginal value of energy which they create. This artificial market failure (and others) have been used to justify payments for installed capacity, to address the missing money problem, so that reliability targets will be met. However, capacity markets are often designed without regard for those targets, if reliability targets are defined at all. Capacity markets tend to exacerbate the missing money problem they aim to solve, whilst distorting decisions which maximise supply availability during periods of scarcity. Defining the metric to use for capacity markets is a difficult task, which often involves a central planner discriminating based on a fuel type, to the detriment of climate goals and economic efficiency.

Even within a fuel type, there are decisions which investors and operators can make to maximise social value for a given installed capacity. This often involves a tradeoff between maximising volume and maximising value. For example, tilting solar panels more vertically and westward will decrease volume but increase value. Many renewables support schemes distort this tradeoff, resulting in up to half of the value of solar power being lost.

On the consumer side, loads are generally exposed to fixed retail prices, not spot prices. This creates inefficiencies because the price which consumers face differs from the cost to supply the good to them. This has been true since the beginning of the electricity industry. However, these inefficiencies are set to grow as renewable penetration increases, giving this topic renewed policy relevance. Fixed tariffs also expose retailers to risk. Heterogenous consumers present retailers with different shape risk and capture prices. This gives rise to a customer selection problem which is similar to the CAPM problem of selecting an optimal investment portfolio.

This thesis has used a series of models to explore missing markets and market failures in the electricity industry. The common theme is that not all megawatts of power capacity are created equal, even within a fuel type. Not all megawatt hours of energy are equally valuable. Inefficiencies arise when markets are designed in a what such that they do not fully reflect these substantial differences in value.

[24]Demand response is also useful, but generally not as substantial.

Bibliography

AEMC, Australian Energy Market Commission. 2025a. 'AEMC Updates Market Price Cap for 2025-26'. 27 February 2025. https://www.aemc.gov.au/news-centre/media-releases/aemc-updates-market-price-cap-2025-26°.

AEMC, Australian Energy Market Commission. 2025b. 'National Electricity Rules'. 2025. https://energy-rules.aemc.gov.au/ner/621/526154#clause_4.9.2°.

AEMO, Australian Energy Market Operator. 2021. 'Trip of Multiple Generators and Lines in Central Queensland and Associated under-Frequency Load Shedding on 25 May 2021'. https://www.aemo.com.au/-/media/files/electricity/nem/market_notices_and_events/power_system_incident_reports/2021/trip-of-multiple-generators-and-lines-in-qld-and-associated-under-frequency-load-shedding.pdf°.

AEMO, Australian Energy Market Operator. 2022. 'NEM Market Suspension and Operational Challenges in June 2022'. https://www.aemo.com.au/-/media/files/electricity/nem/market_notices_and_events/market_event_reports/2022/nem-market-suspension-and-operational-challenges-in-june-2022.pdf°.

ARENA, Australian Renewable Energy Agency. 2021. 'The Generator Operations Series Report One: Large-Scale Solar Operations'. https://arena.gov.au/assets/2021/05/report-one-large-scale-solar-operations.pdf°.

Badran, Ghadeer, and Mahmoud Dhimish. 2024. 'Comprehensive Study on the Efficiency of Vertical Bifacial Photovoltaic Systems: A UK Case Study'. *Scientific Reports* 14 (1): 18380. https://doi.org/10.1038/s41598-024-68018-1°.

Batlle, Carlos, Tim Schittekatte, Paolo Mastropietro, and Pablo Rodilla. 2023. 'The EU Commission's Proposal for Improving the Electricity Market Design: Treading Water, But Not Drowning'. *Current Sustainable/renewable Energy Reports* 10 (4): 197–205.

Billimoria, Farhad, Pierluigi Mancarella, and Rahmat Poudineh. 2022. 'Market and Regulatory Frameworks for Operational Security in Decarbonizing Electricity Systems: From Physics to Economics'. *Oxford Open Energy* 1:oiac7. https://doi.org/10.1093/ooenergy/oiac007°.

Billimoria, Farhad, Jacob Mays, and Rahmat Poudineh. 2025. 'Hedging and Tail Risk in Electricity Markets'. *Energy Economics* 141:108132.

Borenstein, Severin, and Stephen Holland. 2005. 'On the Efficiency of Competitive Electricity Markets with Time-Invariant Retail Prices'. *The RAND Journal of Economics* 36 (3): 469–93. http://www.jstor.org/stable/4135226°.

Borenstein, Severin, James Bushnell, and Erin Mansur. 2023. 'The Economics of Electricity Reliability'. *Journal of Economic Perspectives* 37 (4): 181–206. https://doi.org/10.1257/jep.37.4.181°.

Brown, Claudia, Serguey Maximov, James Price, and Michael Grubb. 2024. 'Generating Surplus: The Challenges and Opportunities of Large-Scale Renewables Deployment'.

Bynum, Michael L., Gabriel A. Hackebeil, William E. Hart, Carl D. Laird, Bethany L. Nicholson, John D. Siirola, Jean-Paul Watson, and David L. Woodruff. 2021. *Pyomo - Optimization Modeling in Python*. 3rd ed. Springer.

Cabot, Clément, and Manuel Villavicencio. 2024. 'Second-Best Electricity Pricing in France: Effectiveness of Existing Rates in Evolving Power Markets'. *Energy Economics* 136:107673. https://doi.org/10.1016/j.eneco.2024.107673°.

CAISO, California Independent System Operator. 2018. 'Eligible Intermittent Resource Dispatch Operating Target EIR Dot'. https://www.caiso.com/Documents/Presentation-DispatchOperatingTargetTariffClarificationTraining.pdf°.

CAISO, California Independent System Operator. 2021. 'Root Cause Analysis - Mid-August 2020 Extreme Heat Wave'. https://www.caiso.com/Documents/Final-Root-Cause-Analysis-Mid-August-2020-Extreme-Heat-Wave.pdf°.

CERC, Central Electricity Regulatory Commission. 2023. 'Directions by the Commission to the Power Exchanges Registered under the Power Market Regulations, 2021'. https://cercind.gov.in/2023/orders/04-SM-2023.pdf°.

Cramton, Peter, and Steven Stoft. 2005. 'A Capacity Market That Makes Sense'. *The Electricity Journal* 18 (7): 43–54. https://doi.org/10.1016/j.tej.2005.07.003°.

Dedenbach, Xavier. 2025. 'The Future of Solar Doesn't Track the Sun'. 20 April 2025. https://terraformindustries.wordpress.com/2025/04/29/the-future-of-solar-doesnt-track-the-sun/°.

Department of Energy and Climate Change. 2014. 'Implementing Electricity Market Reform (EMR): Finalised Policy Positions for Implementation of EMR'. https://www.gov.uk/government/publications/implementing-electricity-market-reform-emr°.

EC, European Commission. 2016. 'Final Report of the Sector Inquiry on Capacity Mechanisms'. Brussels. https://energy.ec.europa.eu/system/files/2016-11/com2016752.en__0.pdf°.

Elia. 2024. 'Calibration Report: Report of the Transmission System Operator Containing the Information to Determine the Volume to Be Contracted and Proposals for Other Parameters'. https://www.elia.be/en/electricity-market-and-system/adequacy/capacity-remuneration-mechanism°.

ENTSO-E. 2011. 'Deterministic Frequency Deviations – Root Causes and Proposals for Potential Solutions'. https://eepublicdownloads.entsoe.eu/clean-documents/pre2015/publications/entsoe/120222_Deterministic_Frequency_Deviations_joint_ENTSOE_Eurelectric_Report__Final_.pdf°.

ENTSO-E. 2023. 'Explanatory Document for the Amended Nordic LFC Block Methodology for Ramping Restrictions for Active Power Output'. https://consultations.entsoe.eu/system-operations/nordic-tsos-methodology-for-ramping-restrictions-f/supporting_documents/230130%20Explanatory%20Document%20for%20Ramping%20restrictions%20for%20active%20power%20output%20amended%20for%20public%20consultation.pdf°.

ENTSO-E. 2025. 'Actual Generation Per Production Type'. 28 April 2025. https://transparency.entsoe.eu/generation/r2/actualGenerationPerProductionType/show°.

EPEX SPOT. 2020. 'Single Day-Ahead Coupling (SDAC)'. https://www.epexspot.com/sites/default/files/download_center_files/Day-Ahead%20MRC%20Processes%20%2802.07.2019%29.pdf°.

Gilmore, Joel, Tahlia Nolan, and Paul Simshauser. 2024. 'The Levelised Cost of Frequency Control Ancillary Services in Australia's National Electricity Market'. *The Energy Journal* 45 (1): 201–29. https://doi.org/10.5547/01956574.45.1.jgil°.

Giulietti, Monica, Luigi Grossi, Elisa Trujillo Baute, and Michael Waterson. 2018. 'Analyzing the Potential Economic Value of Energy Storage'. *The Energy Journal* 39 (1_suppl): 101–22.

Greve, Thomas, Fei Teng, Michael G. Pollitt, and Goran Strbac. 2018. 'A System Operator's Utility Function for the Frequency Response Market'. *Applied Energy* 231:562–69. https://doi.org/10.1016/j.apenergy.2018.09.088°.

Hinchberger, Andrew J, Mark R Jacobsen, Christopher R Knittel, James M Sallee, and Arthur A van Benthem. 2024. 'The Efficiency of Dynamic Electricity Prices'. https://doi.org/10.3386/w32995°.

IEA, International Energy Agency. 2015. 'Energy Policies of IEA Countries: Spain 2015 Review'. https://www.iea.org/reports/energy-policies-of-iea-countries-spain-2015-review°.

IEA, International Energy Agency. 2025. 'Electricity 2025'. https://www.iea.org/reports/electricity-2025°.

Imelda, Matthias Fripp, and Michael J Roberts. 2024. 'Real-Time Pricing and the Cost of Clean Power'. *American Economic Journal: Economic Policy* 16 (4): 100–141.

Katzen, Matthew, and Gordon W Leslie. 2024. 'Siting and Operating Incentives in Electrical Networks: A Study of Mispricing in Zonal Markets'. *International Journal of Industrial Organization* 94:103069.

Keppler, Jan Horst, Simon Quemin, and Marcelo Saguan. 2022. 'Why the Sustainable Provision of Low-Carbon Electricity Needs Hybrid Markets'. *Energy Policy* 171:113273. https://doi.org/10.1016/j.enpol.2022.113273°.

Lamp, Stefan, and Mario Samano. 2022. 'Large-Scale Battery Storage, Short-Term Market Outcomes, And Arbitrage'. *Energy Economics* 107:105786.

Ma, Wei, Wei Wang, Xuezhi Wu, Ruonan Hu, Fen Tang, Weige Zhang, Xiaoyan Han, and Lijie Ding. 2019. 'Optimal Allocation of Hybrid Energy Storage Systems for Smoothing Photovoltaic Power Fluctuations Considering the Active Power Curtailment of Photovoltaic'. *IEEE Access* 7:74787–99.

Macklin, Graham. 2025. 'Optimal Subsidies for Intermittent Renewable Energy [Forthcoming]'.

McArdle, Paul. 2022. 'Analytical Challenge – Choosing What Measure to Use, For 'Installed Capacity''. 27 September 2022. https://wattclarity.com.au/articles/2022/09/analyticalchallenge-installedcapacity/°.

McRae, Shaun D, and Frank A Wolak. 2019. 'Market Power and Incentive-Based Capacity Payment Mechanisms'. *Unpublished Manuscript, Stanford University*.

Naemi, Mostafa, Dominic Davis, and Michael J. Brear. 2022. 'Optimisation and Analysis of Battery Storage Integrated into a Wind Power Plant Participating in a Wholesale Electricity Market with Energy and Ancillary Services'. *Journal of Cleaner Production* 373:133909. https://doi.org/10.1016/j.jclepro.2022.133909°.

Newbery, David. 2016. 'Missing Money and Missing Markets: Reliability, Capacity Auctions and Interconnectors'. *Energy Policy* 94:401–10.

OMIE. 2025. 'Day-Ahead Minimum, Average and Maximum Price'. 2025. https://www.omie.es/en/market-results/monthly/daily-market/daily-market-price?scope=monthly&year=2025&month=4°.

Wei, Mengyao, Zijiang Yang, Jiandong Wang, Song Gao, and Daning You. 2020. 'Optimal Dispatching Method Based on Actual Ramp Rates of Power Generation Units for Minimising Load Demand Response Time'. *IET Generation, Transmission & Distribution* 14 (26): 6562–68.

Wong, Ling Ai, Vigna K. Ramachandaramurthy, Phil Taylor, J.B. Ekanayake, Sara L. Walker, and Sanjeevikumar Padmanaban. 2019. 'Review on the Optimal Placement, Sizing and Control of an Energy Storage System in the Distribution Network'. *Journal of Energy Storage* 21:489–504. https://doi.org/10.1016/j.est.2018.12.015°.

Wynn, Gerard, and Javier Julve. 2016. 'Spain's Capacity Market: Energy Security or Subsidy?'. https://ieefa.org/wp-content/uploads/2017/11/Spains-Capacity-Market-Energy-Security-or-Subsidy_December-2016.pdf°.

Xia, X., and A.M. Elaiw. 2010. 'Optimal Dynamic Economic Dispatch of Generation: A Review'. *Electric Power Systems Research* 80 (8): 975–86. https://doi.org/10.1016/j.epsr.2009.12.012°.

Yang, Yuqing, Stephen Bremner, Chris Menictas, and Merlinde Kay. 2021. 'Impact of Forecasting Error Characteristics on Battery Sizing in Hybrid Power Systems'. *Journal of Energy Storage* 39:102567. https://doi.org/10.1016/j.est.2021.102567°.

Zhao, Haoran, Qiuwei Wu, Shuju Hu, Honghua Xu, and Claus Nygaard Rasmussen. 2015. 'Review of Energy Storage System for Wind Power Integration Support'. *Applied Energy* 137:545–53. https://doi.org/10.1016/j.apenergy.2014.04.103°.

Glossary

***ACF* – autocorrelation function**: A measure of the correlation between a sequence of data points and itself several time periods prior

***AEMO* – the Australian Energy Market Operator**: The electricity (and gas) market operator for Australia

***AIC* – Akaike information criterion**: A number used to select the best model amongst several specifications

***ARIMA* – autoregressive integrated moving average**: A time series model which accounts for serial correlation between time periods, including differencing

***CAISO* – the California Independent System Operator**: The electricity market operator for California

***CAPM* – capital asset pricing model**: An import model in market finance, relating the tradeoff between risk and reward, by breaking down risk into systematic and unsystematic risks

***CPP* – critical-peak pricing**: An electricity tariff structure where consumers mostly face a flat price, and a second, higher price for a few hours or days per year when the market is tight

***CfD* – contract for difference**: One sided financial derivative

***FCAS* – frequency control and ancillary services**: Reliability services used to respond to shocks within a trading interval

***FCR* – frequency containment reserve**: A reserve of spinning generation capacity, ready to increase output in response to a shock

***ISO* – independent system operator**: The operator of the wholesale electricity market. For example, CAISO in California, AEMO in Australia

***MISO* – the Midcontinent Independent System Operator**: The electricity market operator for the central United States

***MW* – megawatt**: A measure of power intensity. It is a flow, not a stock.

***MWh* – megawatt hour**: A measure of energy. It is a stock, not a flow. $1\ \text{MWh} = 1\ \text{MW} \times 1\text{h} = 2\ \text{MW} \times 0.5\text{h}$

***NEM* – National Electricity Market**: Australia's Electricity Market, excluding Western Australia and the Northern Territory

***TOU* – time of use**: An electricity tariff structure where consumers face a pre-agreed flat price , which varies by hour of the day, day of week or season

***VOLL* – value of lost load**: Marginal benefit of a randomly-selected unit of electricity

***VRE* – variable renewable energy**: Intermittent renewable power, i.e. wind and solar, not hydro nor geothermal

Appendix A Proof of Aggregate Marginal Cost Properties

This section contains the proofs for the propositions in Section 1.2.4.1.

A.1 Locally Increasing

Let N_Q be the optimal number of generators to provide quantity Q. i.e. $\forall Q \notin \left\{\widehat{Q_N}\right\}, \exists N s.t. C_{\text{agg}}(N_Q) = C_N(N_Q)$, which exists by definition of C_{agg}. The aggregate cost function is piece-wise continuous, with finite-length double-differentiable pieces. Therefore:

- $C_{\text{agg}}(Q) = C_{N_Q}(Q) = N_Q C_1\left(\frac{Q}{N_Q}\right)$
- $C'_{\text{agg}}(Q) = C'_{N_Q}(Q) = C'_1\left(\frac{Q}{N_Q}\right)$ (where $C' = \frac{\partial C}{\partial Q}$)
- $C''_{\text{agg}}(Q) = C''_{N_Q}(Q) = \frac{1}{N_Q} \cdot C''_1\left(\frac{Q}{N_Q}\right) > 0$

Therefore I have proved that the aggregate marginal cost is locally increasing ($C''_{\text{agg}}(Q) > 0$), except at the breakpoints

A.2 Finite Allocation

Next I prove that as total demand grows arbitrarily large, the optimal quantity per generator (for those which are running) converges to a strictly positive limit.

Let $q = \frac{Q}{N_Q}$ be the optimal per-generator production level for a given aggregate quantity Q. Suppose $\underline{q} = \min_Q q$ is the lower bound. The objective is to prove that $\underline{q} > 0$. To do this, I use a proof by contradiction.

Suppose that $\underline{p} = 0$. Therefore, for any arbitrarily small $\varepsilon > 0$, there exists a Q such that $\frac{Q}{N_Q} < \varepsilon$. Therefore, reducing the number of generators would not reduce total costs. (At a breakpoint, there is no change. Elsewhere total costs would increase.) The change in total costs is given by the decrease in costs for one generator (turning off), and increase in costs for the remaining generators increasing production to pick up the slack: This cost increase is given by

$$\big(N_Q - 1\big)C'_{\text{agg}}(q)q - C_1(q) = \big(N_Q - 1\big)C'_1(q)q - C_1(q) < \big(N_Q - 1\big)C'_1(q)\varepsilon - C_1(q)$$

For a choice of ε smaller than $C_1 \frac{q}{(N_Q - 1)C'_1(q)}$, this cost increase becomes negative. That means that deviating from the lowest-cost number of generator N_Q yields a cost decrease. Thus we have a contradiction. Therefore the original assumption that $\underline{p} = 0$ is not true. Therefore $\underline{p} > 0$, and the optimal allocation per generator is strictly positive for arbitrarily large demand.

A.3 Globally Decreasing

I aim to show that the height of the breakpoints in the marginal cost curve is a descending sequence, converging to a positive value. i.e. $\lim_{N\to\infty} C'_N\left(\widehat{Q_N}\right) = \underline{C'} > 0$, where $C'_N(Q) = \frac{\partial C_N(Q)}{\partial Q}$

$$C_N(Q) = N C_1\left(\frac{Q}{N}\right)$$

$$C'_N(Q) = N C'_1\left(\frac{Q}{N}\right) \cdot \frac{1}{N} = C'_1\left(\frac{Q}{N}\right)$$

$$C'_N\left(\widehat{Q_N}\right) = C'_1\left(\frac{\widehat{Q_N}}{N}\right) > C'_1\left(\underline{q}\right) > 0$$

$$\lim_{N\to\infty} C'_N\left(\widehat{Q_N}\right) = C'_1\left(\underline{q}\right) > 0$$

Therefore the marginal costs to the left of each breakpoint are a descending sequence, converting to a finite value. Since the marginal costs within each piece-wise segment are increasing (maximum to the left of each breakpoint), it follows that

$\forall \varepsilon > 0, \exists Q : \forall Q' > Q, C'_{\text{agg}}(Q') < \underline{C'} + \varepsilon$ where $\underline{C'} > 0$

Appendix B Sign of the Relationship Between Raise Service Reserve Quantity and Energy Quantity

B.1 Motivation

Section 1.2.3 on page 10 extends the model of raise services so that the demand for raise service reserves depends linearly on energy demand: $R = \overline{R} + \beta Q_e$. It was assumed that $\beta > 0$. This reflects a fractional reserve approach. For example, $\beta = 0.1$ means that for every additional 10 MW of energy demand to be met, 1 MW of additional raise service reserves must be provided. In practice the determination of raise service reserves is more complicated. The purpose of this section is to discuss those complications, and test whether $\beta > 0$.

In Australia's ancillary markets there are four different contingency raise services, of varying timescales (technical response speeds). They differ mostly in terms of the quantities demanded by the grid operator, and the set of technologies which are eligible to provide them. The 1 second market is dominated by batteries, not gas nor thermal. Batteries have a unique stack of costs which are not captured by my model. Amongst the remainder, the analysis yields similar results. The 60 second product is shown in this section.

Figure 15 shows the relationship between the energy supplied in Australia's NEM, and the raise service reserves procured, based on a random subset of historical observed quantities. In the electricity industry, aggregate supply quantities are typically reported excluding rooftop solar generation, even when rooftop solar is substantial in volume, because of its different data provenance. When excluding rooftop solar, a naive fit using ordinary least squares yields a positive slope. However, amongst the cloud of points there are some visible linear clusters with negative slopes, suggesting that this is a case of Simpson's paradox. When including rooftop solar, the overall fit has a negative slope (and negatively sloped clusters remain).

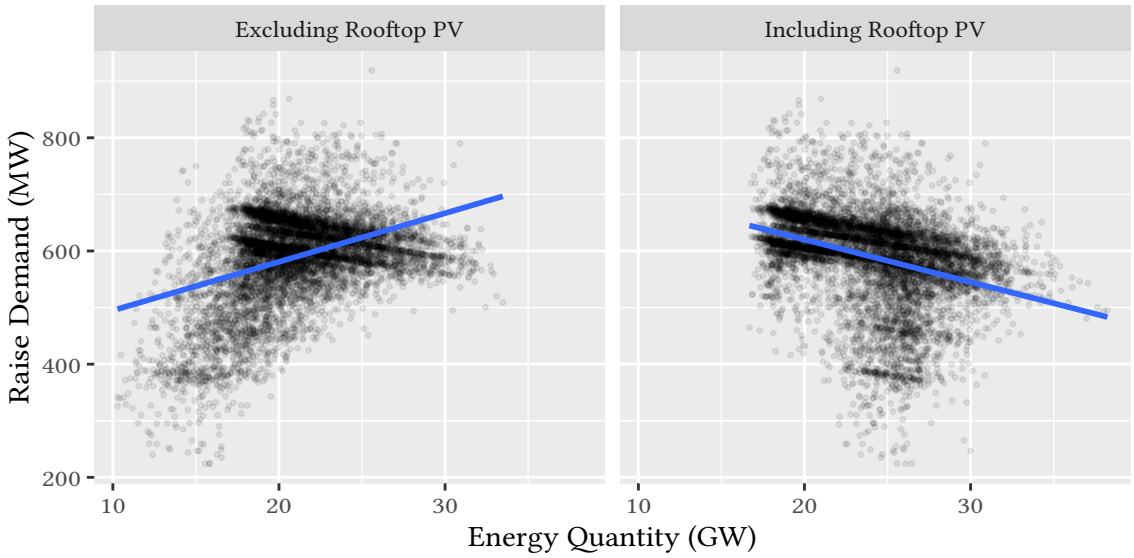


Figure 15: Relationship between observed energy quantity and raise service quantity in Australia's NEM, 2024 for the 60 second raise product. When excluding rooftop solar (as is standard in the industry) a naive linear fit has a positive slope, yet negatively sloped clusters are visible (suggestive of Simpson's paradox).

As energy demand increases, the total number of spinning generators increases. This increase in physical inertia (by having more spinning metal connected to the grid) increases the resilience of the

grid, decreasing the demand for raise services. Thus the operational decisions of generators impact the market for raise services in two ways:

Supply: A generator can choose to make q_r of their spinning *reserves* available to respond after a shock

Demand: A generator can choose to start or stop their generator ($q_e = 0$ or $q_e > 0$). Starting a generator will connect more *inertia* to the grid, which will reduce the *demand* for raise services, in a binary way which does not depend on q_e or q_r.

Thus a more realistic demand function is $R = \overline{R} + \beta Q_e + \gamma K \sum_{i=0}^{N} \mathbb{I}(q_e^i > 0)$ with $\beta > 0$ and $\gamma < 0$, where $\sum_{i=0}^{N} \mathbb{I}(q_e^i > 0)$ is the number of spinning generators. The relationship between energy quantity and raise service quantity is confounded by spinning capacity, which explains the occurrence of Simpson's paradox. This motivates a more sophisticated regression than the slope of Figure 15.

B.2 Multivariate Time series ARIMA Regression

B.2.1 Regression Setup

In this section I attempt to estimate β empirically, in Australia's market for 60 second raise service. The relationship is modelled as:

$$R_t = \overline{R} + \beta Q_{e,t} + \gamma_1 C_t + \gamma_2 X_t + \varepsilon_t$$

Where:

- R_t is the demand for raise services (MW)
- β_1 is the parameter of interest (GW)
- C_t is the confounder, which is the amount of spinning capacity, including unused spinning reserves, for thermal generators only (e.g. excluding VRE)
- X_t are controls relating to the size of the largest potential supply shock that the raise service must respond to.

Data was obtained from AEMO at nemweb.com.au° using Nemosis° (a Python package). Regressions were run in R. Values were aggregated across all regions in the NEM, because FCAS demand is typically not determined on a per-region basis. Calendar year 2024 was used as the period of analysis.

Like most electricity market data, these values are serially correlated, so a naive regression would drastically underestimate standard errors. Instead the quantities are modelled as seasonal ARIMA processes. Data was aggregated from 5-minutes to an hourly level, so that it is computationally feasible to model a daily seasonal component with enough lags (P, Q) to also capture weekly seasonality.

B.2.2 Descriptive Statistics

Visual inspection of the raw time series data for R_t suggests that it is neither a unit root process (because it remains predictably bounded), nor a stationary process (because the expectation varies). The autocorrelation function (ACF) has significant values for extremely high lags, and a Dickey Fuller test rejects the null hypothesis of a unit root. Thus we take the first difference of R_t (and of $Q_{e,t}$ and C_t, which are similar).

The ACF of ΔR_t with a 24 hour lag of itself has significant spikes around lag 1 and 24, and not elsewhere. This suggests an ARIMA process with `d = 1`, `D=1`, `Q = 1`, `P = 1`. This was used to inform the bounds for a search based on Akaike information criterion (AIC), which selected an `ARIMA(23,1,0)(2,1,0)[24]` process.

B.2.3 Regression Results and Discussion

The results of the regression for this model are shown in Table 4. For all specifications, the results are that $\hat{\beta}$ is positive (and statistically significant), and $\hat{\gamma_1}$ is negative (and statistically significant), as expected. The interpretation of $\hat{\beta} = 5.179$ is that for each additional 1 GW of energy demand, raise service demand increases by approximately 5 MW.

		(1)	(2)	(3)	(4)	(5)	(6)†
Generation Excl. Rooftop	β	8.527***	9.03***	1.096**	7.365***	7.474***	5.179***
		(0.632)	(0.803)	(0.532)	(0.873)	(0.871)	(0.87)
Rooftop PV				−12.925***	−5.995***	−6.074***	−5.449***
				(0.57)	(1.289)	(1.287)	(1.27)
Spinning Inertial Capacity	γ_1					−4.234***	−4.213***
						(1.261)	(1.242)
Electrical Controls							✓
ARIMA	(p,d,q)	(0, 0, 0)	(23, 1, 0)	(0, 0, 0)	(23, 1, 0)	(23, 1, 0)	(23, 1, 0)
	(P, D, Q)	(0, 0, 0)	(2, 1, 0)	(0, 0, 0)	(2, 1, 0)	(2, 1, 0)	(2, 1, 0)
Observations		8784	8759	8784	8759	8759	8759

† Preferred specification; * , ** , *** Significant at 10%, 5% and 1%levels

Table 4: Regression results. The dependent variable is raise service demand (MW). Independent variables are in GW. Models 1 and 3 use HC3 errors not ARIMA processes, and are shown only for reference.

The sign of the coefficient for rooftop solar is always negative and statistically significant. The interpretation of this is that for each additional unit of rooftop solar power (keeping all other energy supply constant), the demand for raise services *decreases*. This is an intriguing finding which does not match the theory. Exploring this is left for future research. Note that the derivations in Section 1.2.3 did not depend on the assumption that $\beta > 0$. That was introduced for intuition. So a negative slope for solar power does not invalidate the algebraic findings, but merely inverts them, for that type of generation.

B.3 Combining Endogenous Raise Service Demand and Fixed Costs

In Section 1.2.3 on page 10 when considering endogenous demand for the raise service, I assumed there were no fixed running cost for generators, so all generators run:$\sum_{i=0}^{N} \mathbb{I}(q_e^i > 0) = N$. In Section 1.2.4 on page 11 I separately assumed that there are fixed costs, so it may be optimal that some generators do not run. Suppose we consider both matters at once (fixed costs, yielding asymmetric optimal allocations, and endogenous demand, which depends on spinning capacity). I proceed with backwards induction. Conditional on a generator starting up ($q_e > 0$, at the decision after "Run" in Figure 1), the impact of the demand reduction from spinning capacity is already sunk. Thus γ has no impact on the allocation between q_e and q_r, conditional on running. Thus the findings in Section 1.2.4 and Section 1.2.3 still hold. Next we proceed back a step to the decision whether to run a generator (immediately after "Build" in Figure 1). Since I assume perfect competition, the reduction in demand for the raise service due to each generator's decision to run is negligible. Therefore there is a positive externality. When deciding whether to run a generator, each firm does not consider the impact on all other grid users of the reduction in *demand* for raise service reserves. This is the well known missing market for *inertia*.

Appendix C Derivations for Optimal Retail Portfolio

This section contains the derivations for the results in Section 6.3 on page 36. A retailer's expected profit is:

$$\begin{aligned}
\mathbb{E}(\pi) &= \alpha\rho_s\overline{q} + (1-\alpha)\rho_r\overline{q} + \alpha A_s + (1-\alpha)A_r \\
&\quad -\mathbb{E}((\overline{p}+\varepsilon)\cdot(\alpha(\overline{q}+\eta_s) + (1-\alpha)(\overline{q}+\eta_r))) \\
&= \alpha\rho_s\overline{q} + (1-\alpha)\rho_r\overline{q} + \alpha A_s + (1-\alpha)A_r - \overline{p}\cdot\overline{q} \\
&\quad -\mathbb{E}(\overline{p}(\alpha\eta_s + (1-\alpha)\eta_r) + \alpha\varepsilon\eta_s + \varepsilon(1-\alpha)\eta_r) \\
&= \alpha\rho_s\overline{q} + (1-\alpha)\rho_r\overline{q} + \alpha A_s + (1-\alpha)A_r - \overline{p}\cdot\overline{q} \\
&\quad -\mathbb{E}(\overline{p}\alpha\eta_s) - \mathbb{E}(\overline{p}(1-\alpha)\eta_r) - \mathbb{E}(\alpha\varepsilon\eta_s) - \mathbb{E}(\varepsilon(1-\alpha)\eta_r) \\
&= \alpha\rho_s\overline{q} + (1-\alpha)\rho_r\overline{q} + \alpha A_s + (1-\alpha)A_r - \overline{p}\cdot\overline{q} \\
&\quad -\alpha\mathbb{E}(\varepsilon\eta_s) - (1-\alpha)\mathbb{E}(\varepsilon\eta_r)
\end{aligned}$$

The variance is given by:

$$\begin{aligned}
\mathrm{Var}(\pi) &= \mathrm{Var}[\alpha\rho_s q_s + (1-\alpha)\rho_r q_r + \alpha A_s + (1-\alpha)A_r - p\cdot(\alpha q_s + (1-\alpha)q_r)] \\
&= \mathrm{Var}[\alpha\rho_s q_s + (1-\alpha)\rho_r q_r - p\cdot(\alpha q_s + (1-\alpha)q_r)] \\
&= \mathrm{Var}[\alpha\rho_s q_s] + \mathrm{Var}[(1-\alpha)\rho_r q_r] + \mathrm{Var}[p\alpha q_s] + \mathrm{Var}[p(1-\alpha)q_r)] \\
&\quad +2\ \mathrm{Cov}(\alpha\rho_s q_s, (1-\alpha)\rho_r q_r) \\
&\quad +2\ \mathrm{Cov}(\alpha\rho_s q_s, p\alpha q_s) \\
&\quad +2\ \mathrm{Cov}(\alpha\rho_s q_s, p(1-\alpha)q_r) \\
&\quad +2\ \mathrm{Cov}((1-\alpha)\rho_r q_r, p\alpha q_s) \\
&\quad +2\ \mathrm{Cov}((1-\alpha)\rho_r q_r, p(1-\alpha)q_r) \\
&\quad +2\ \mathrm{Cov}(p\alpha q_s, p(1-\alpha)q_r) \\
&= \alpha^2\rho_s^2\ \mathrm{Var}[\eta_s] + (1-\alpha)^2\rho_r^2\ \mathrm{Var}[\eta_r] + \alpha^2\ \mathrm{Var}[pq_s] + (1-\alpha)\ \mathrm{Var}[pq_r] \\
&\quad +2\alpha(1-\alpha)\rho_s\rho_r\ \mathrm{Cov}(q_s, q_s) \\
&\quad +2\alpha^2\rho_s\ \mathrm{Cov}(q_s, pq_s) \\
&\quad +2\alpha(1-\alpha)\rho_s\ \mathrm{Cov}(q_s, pq_r) \\
&\quad +2\alpha(1-\alpha)\rho_r\ \mathrm{Cov}(q_r, pq_s) \\
&\quad +2(1-\alpha)^2\rho_r\ \mathrm{Cov}(q_r, pq_r) \\
&\quad +2\alpha(1-\alpha)\ \mathrm{Cov}(pq_s, pq_r)
\end{aligned}$$