THE OXFORD INSTITUTE FOR ENERGY STUDIES

May 2024

The Challenges of Incorporating Consumer Reliability Preferences into Electricity Markets with a Capacity Requirement

OIES Paper: EL54

Frank A. Felder, Senior Visiting Research Fellow, OIES

ISBN 978-1-78467-242-3

Acknowledgements

The author would like to thank the two anonymous reviewers, as well as Dr Marie Petitet and Dr Rahmat Poudineh for their comments.

Abstract

One objective of retail electricity markets is to enable end-use consumers to incorporate their reliability preferences into their purchasing decisions. This paper investigates if and how this can be done in electricity markets with a capacity requirement. It integrates the standard loss of load and cost minimization approach from the economic literature with probabilistic resource adequacy informed by the engineering literature for electricity markets with capacity requirements. For these electricity markets, a partial solution that allows retail consumers to opt out entirely or partially from capacity markets could help improve social welfare. This solution allows consumers to use their individual estimate of the cost of power outages based on the relevant outage characteristics instead of a system planner estimating a generic value of lost load. The goal of achieving optimal levels of reliability remains elusive, however, due to incomplete probabilistic reliability models, consumers' inability to opt out of capacity requirements, and transaction costs.

Contents

Tables and Figures

1. Introduction

The reliability and resiliency of electricity power systems in general and electricity markets in particular are being questioned. Recent major blackouts and close calls have occurred worldwide, such as blackouts in France and Great Britain (separately) in 2019; Texas (ERCOT grid), the US Mid-Atlantic region (PJM grid), and California in 2022; and Quebec and Ontario, Canada in 2023. Many factors cause blackouts, one being insufficient resource adequacy. Inadequate generation was a significant factor in the blackout on the ERCOT grid and the near blackout on the PJM grid.

Electricity capacity mechanisms are employed internationally to ensure resource adequacy. Resource adequacy is having sufficient electricity capacity to meet demand based on a specified resource adequacy requirement. In North America, this requirement is established by the North American Electric Reliability Corporation. Power systems are planned to have sufficient capacity such that firm load is curtailed to maintain reliability no more than once every ten years. Capacity mechanisms range from simple requirements to sophisticated markets. Typically, utilities or system operators use probabilistic resource adequacy models to determine the capacity needed to meet the resource adequacy requirement. These models calculate the capacity required, sometimes using a demand curve that reflects the diminishing returns of capacity.

A market with a capacity requirement (which may also have a capacity market with different variations) mandates that load-serving entities purchase their share of sufficient capacity to meet a system-wide loss of load probability (LOLP) criterion.[1] In contrast, an energy-only market relies on infrequent but periodically very high energy prices, combined with stipulating a value of lost load (VOLL) and an operating reserve curve that depends on the LOLP.

Historically, resource adequacy models used to set capacity requirements have made three fundamental assumptions. First, they have modelled independent generation failures; that is, the failure of one generation unit has not been caused, linked, or correlated with the failure of other generation units. Second, with some exceptions, the modelled generation units have been dispatchable; that is, their output has been able to be adjusted up or down to match changes in electricity demand. Third, for the most part, the modelled demand has not responded to prices; that is, it has been invariant to wholesale electricity prices.

As recognized by many relatively recent modelling efforts, these three assumptions no longer hold.[2] Dependent and correlated outages are substantial and significantly impact resource adequacy and, therefore, reliability. These reasons for these outages include weather conditions, operational and maintenance practices, fuel availability, output correlation with wind farms and solar panels, regulatory actions, and cybersecurity. The substantial anticipated increase in renewable resources due to decarbonization means that weather conditions, combined with the variability and non-dispatchability of wind and solar photovoltaics, make the first two resource adequacy assumptions untenable. Finally, introducing and increasing energy storage, microgrids, and load control devices means that demand and behind-the-meter resources will grow and be more price responsive. Increasing flexible demand is necessary to accommodate increasing shares of variable solar and wind reliably and economically.

This paper focuses on demand response by investigating if and to what extent different market structures and mechanisms can accommodate individual and varying consumer preferences—*reliability tailorization*—consistent with the underlying resource adequacy models.[3] Three fundamental features are essential.

[1] LOLP is the probability that demand exceeds supply. Loss of load expectation is the total duration of increments when loss of load occurs. It is sometimes expressed in hours instead of days, in which case the term 'loss of load hours' is used.
[2] EPRI (2021), 'Exploring the impacts of extreme events, natural gas fuel and other contingencies on resource adequacy', www.epri.com/research/products/000000003002019300.
[3] Demand response could be part of virtual power plants (that is, aggregations of distributed energy resources).

The first is the comprehensiveness of the underlying modelling framework of the electric power system's reliability. The current quantitative framework for estimating the LOLP is not comprehensive. It also does not account for uncertainty, so it restricts consumers' ability to tailor their reliability preferences.

The second essential feature is the inclusion of policies that allow for reliability tailorization given the market structure. Presently, capacity mechanisms do not consider individual consumer reliability preferences. Instead, all consumers are assumed to value resource adequacy equally and to have the same VOLL.[4] The third feature accounts for the size of the transaction costs for individual consumers and system operators in participating, including the enforcement of transactions that consumers undertake to tailor their reliability.

The paper is organized as follows. Section 2 compares the literature on engineering-based resource adequacy to the related economic literature on electricity markets. It finds merit in integrating these two approaches. Section 3 presents a theoretical framework integrating probabilistic resource adequacy within an economic framework. Three key factors affect the amount of demand response: the system operator's priority as to which load and when it is disconnected, the means of incentivizing or enforcing demand response participation, and the consumers' and system operator's transaction costs in participating and deploying demand response programmes. Section 4 provides a stylized numerical example. Section 5 concludes and points to additional research.

2. Literature Review of Resource Adequacy and Economic Models of Electric System Reliability

Although not wholly standardized worldwide, an accepted definition of power system reliability is that it is the ability of the electric system to supply firm load. It is divided into resource adequacy and operational reliability (historically referred to as security). Resource adequacy is the supply sufficient to meet the demand for a specified requirement, such as an LOLP of one day in ten years. Operational reliability is the ability of the system to withstand changes, such as the failure of generation or transmission, and continue to serve firm load.

Recently, the term 'resiliency' has emerged. It has a broad definition that encompasses reliability plus the ability of a power system to withstand and recover from extreme events beyond those that a reliable power system could withstand. The narrow definition of resiliency complements reliability; that is, resiliency is the ability to recover from power outages. The social and economic costs of power outages depend on their timing, frequency, magnitude, and duration, and these costs vary by type of electricity consumer, which may make it difficult for many consumers to predict their outage costs.[5] The importance and duration of an outage are illustrated by the resiliency trapezoid, which characterizes the power outage from start to recovery.[6]

This paper focuses on resource adequacy since it quantifies a power system's LOLP and is the potential mechanism, via capacity requirements, retail consumers could use to tailor at least part of the reliability and resiliency to their circumstances. Others have focused on system security (operational reliability) in the context of the market design of low-carbon grids.[7] Resource adequacy modelling is being extended to accommodate changes in the industry relating to dependent and correlated failures, non-dispatchable and variable renewables, and new technologies such as energy storage, microgrids, and load control.[8]

---

[4] Ovaere, M., Heylen, E., Proost, S., Deconinck, G., and Van Hertem, D. (2019). 'How detailed value of lost load data impact power system reliability decisions', *Energy Policy*, 132, 1064–75.

[5] Felder, F.A., and Petitet, M. (2022). 'Extending the reliability framework for electric power systems to include resiliency and adaptability', *The Electricity Journal* 35.8, 107186.

[6] Petitet, M., AlHadhrami, K., and Felder, F.A. (2022). 'One year after the Texas blackout.' KAPSARC discussion paper, www.kapsarc.org/research/publications/one-year-after-the-texas-blackout-lessons-for-reliable-and-resilient-power-systems.

[7] Billimoria, F., Mancarella, P., and Poudineh, R. (2020). 'Market design for system security in low-carbon electricity grids: from the physics to the economics', OIES Paper No. 41, www.oxfordenergy.org/publications/market-design-for-system-security-in-low-carbon-electricity-grids-from-the-physics-to-the-economics.

[8] Carvallo, J.P., et al. (2023). 'A guide for improved resource adequacy assessments in evolving power systems: institutional and technical dimensions', Lawrence Berkeley National Laboratory. The term 'virtual power plants' is used to refer to the aggregation of distributed energy resources and loads that can provide services comparable to utilities.

In concert, the design basis of one-time-in-ten-years[9] and other criteria are being reconsidered, such as by limiting the expected amount of unserved energy and incorporating risk-based measures.[10] Further efforts to extend to transmission and distribution are needed in the research literature.

In practice, utilities and power system operators worldwide model generation adequacy to set resource adequacy requirements. Resource adequacy is an essential component of integrated resource planning and is used worldwide to plan power systems.[11] South Africa requires generation resource adequacy as part of its Grid Code.[12] In the United States, all regional transmission organization/independent system operator (RTO/ISO) markets except Texas ERCOT have capacity requirements based on generation resource adequacy modelling.[13] Resource adequacy is also a crucial element in non-RTO regions, such as the western United States, excluding California.[14]

The benefits of demand response in electricity markets are numerous and well recognized.[15] There are, however, limitations to demand response, resulting in it not achieving its potential.[16] The PJM Independent Market Monitor tracks cleared demand resources as a percentage of its unforced capacity market and has found that, since 2012, this percentage has fluctuated between 5.9 per cent and 9.3 per cent.[17] It has concluded that in PJM, "The demand side of wholesale electricity markets is underdeveloped."[18] A substantial increase in demand response in Texas could have significantly reduced the amount of load shedding that occurred during the 2021 outage.[19]

Many have noted the importance of incorporating individual consumer preferences into resource adequacy models. For example, a priority service has been proposed that offers a selection of contingent contracts for the distribution of scarce electricity, leading to efficiency gains.[20] Capacity subscriptions have been proposed in which consumers subscribe to their anticipated demand for capacity during peak periods, which they cannot exceed.[21] Various insurance designs have been proposed to accommodate reliability preferences or distributed resources.[22] Others argue that technological progress may enable the greater use of economic incentives to maintain reliability,

---

[9] There are multiple interpretations of the one-time-in-ten-years (or one-day-in-ten-years or one-in-ten) standard. One event in ten years is 0.1 loss of load expectation per year; one day in ten years is 2.4 loss of load hours per year. See Pfeifenberger, J.P., et al. (2013). *Resource Adequacy Requirements: Reliability and Economic Implications*, Washington, DC: Federal Energy Regulatory Commission.

[10] Carvallo, J.P., et al. (2023). 'A guide for improved resource adequacy assessments in evolving power systems: institutional and technical dimensions', Lawrence Berkeley National Laboratory.

[11] International Renewable Energy Agency (2018). *Insights on Planning for Power System Regulators*, Abu Dhabi: IRENA.

[12] Eskom (2022). 'Medium-term system adequacy outlook', www.eskom.co.za/wp-content/uploads/2022/10/Medium-Term-System-Adequacy-Outlook-2023-2027.pdf.

[13] Pfeifenberger, J.P., et al. (2013). *Resource Adequacy Requirements: Reliability and Economic Implications*, Washington, DC: Federal Energy Regulatory Commission.

[14] WECC (2022). 'Western assessment of resource adequacy', www.wecc.org/Reliability/2022%20Western%20Assessment%20of%20Resource%20Adequacy.pdf.

[15] Albadi, M.H., and El-Saadany, E.F. (2008). 'A summary of demand response in electricity markets.' *Electric Power Systems Research* 78.11, 1989–96; Baker, P. (2015). 'Resource adequacy, regionalisation, and demand response', www.raponline.org/knowledge-center/resource-adequacy-regionalisation-demand-response.

[16] Kim, J.-H., and Shcherbakova, A. (2011). 'Common failures of demand response', *Energy* 36.2, 873–80.

[17] Monitoring Analytics (2022). *State of the Market Report for PJM*, Eagleville, PA: Monitoring Analytics, p. 362.

[18] Monitoring Analytics (2022). *State of the Market Report for PJM*, Eagleville, PA: Monitoring Analytics, p. 353.

[19] Wu, D., et al. (2022). 'How much demand flexibility could have spared Texas from the 2021 outage?' *Advances in Applied Energy* 7, 100106.

[20] Chao, H.-P., and Wilson, R. (1987). 'Priority service: pricing, investment, and market organization.' *The American Economic Review*, 899–916.

[21] Doorman, G.L. (2005). 'Capacity subscription: solving the peak demand challenge in electricity markets.' *IEEE Transactions on Power Systems* 20.1, 239–45.

[22] Billimoria, F., et al. (2022). 'An insurance mechanism for electricity reliability differentiation under deep decarbonization', *Applied Energy* 321, 119356; Niromandfam, A., et al. (2020). 'Electricity consumers' financial and reliability risk protection utilizing insurance mechanism', *Sustainable Energy, Grids and Networks* 24, 100399; Fuentes, R., Blazquez, J., and Adjali, I. (2019). 'From vertical to horizontal unbundling: a downstream electricity reliability insurance business model', *Energy Policy* 129, 796–804; Billimoria, F., and Poudineh, R. (2019). 'Market design for resource adequacy: a reliability insurance overlay on energy-only electricity markets.' *Utilities Policy* 60, 100935.

specifically the ability to implement the reliability preferences of individuals or groups of consumers.[23] Explicitly linking resource adequacy models with economic characteristics in the context of resource adequacy requirements and capacity markets is necessary.

Two theoretical frameworks for the reliability of bulk power systems have been applied to power systems and markets and have significant public policy impacts.[24] These frameworks use similar terminology, although with different definitions and interpretations. The first framework is engineering based. The stochastic failure and repair of generation units and transmission components are modelled, and the likelihood of demand exceeding supply, and the likely amounts, are calculated. These calculations are then used to assess whether a bulk power system is sufficiently reliable according to an engineering-based reliability standard or design basis, such as the North American Electric Reliability Council's one-day-in-ten-years standard.[25] The most cost-effective investments are undertaken to ensure that the reliability standard is met, through the means of utility planning and regulatory processes or market design.

The second framework is economically based. The reliability of a power system has been framed economically in different ways. Many have argued that reliability is a public good.[26] Others have suggested that generation capacity is a positive externality.[27] Still others have found that capacity is not a valid construct and is unnecessary to electricity markets; instead, an energy-only market with sufficient price-responsive demand is preferable.[28] The set-up for this economic framework is a deterministic optimization problem in which the cost of generation plus the cost of unserved energy are minimized and the solution is the level and mix of capacity. For a perfectly competitive market, the benevolent social planning problem is the same as the market outcome. When the price elasticity of electricity demand is not considered, then the social welfare maximization problem is the same as the cost minimization problem.[29] Table 1 compares these two frameworks.

---

[23] Borenstein, S., Bushnell, J., and Mansur, E. (2023). 'The economics of electricity reliability.' *The Journal of Economic Perspectives* 37.4, 181–206. See also, International Renewable Energy Agency (2019). *Redesigning Capacity Markets*, Abu Dhabi: IRENA.

[24] See Chao, H.-P. (1983). 'Peak load pricing and capacity planning with demand and supply uncertainty.' *The Bell Journal of Economics*, 179–190. This paper theoretically integrates random generation outages and their economics with a continuous probability distribution that predates electricity markets.

[25] NERC (2020). 'Ensuring energy adequacy with energy constrained resources', www.nerc.com/comm/RSTC/ERATF/ERATF%20Energy%20Adequacy%20White%20Paper.pdf.

[26] Joseph, K. (2022). 'Coordinating markets for reliability: resource adequacy as a public good.' *The Electricity Journal* 35.3, 107097.

[27] Jaffe, A.B., and Felder, F.A. (1996). 'Should electricity markets have a capacity requirement? If so, how should it be priced?' *The Electricity Journal* 9.10, 52–60.

[28] Aagaard, T.S., and Kleit, A.N. (2022). *Electricity Capacity Markets*, Cambridge University Press; Hogan, W.W. (2005). 'On an ''energy only'' electricity market design for resource adequacy', https://scholar.harvard.edu/whogan/files/hogan_energy_only_092305.pdf.

[29] Biggar, D.R., and Hesamzadeh, M.R. (2014). *The Economics of Electricity Markets*, Wiley Blackwell.

Table 1: Comparison of Engineering versus Economic Resource Adequacy Frameworks

Assumption/Result	Engineering Framework	Economic Framework
Generation and transmission	Random failures and repairs of generation and transmission components. For resource adequacy purposes, the transmission system is typically assumed to be perfectly reliable or, if not, modelled in a limited fashion.	Only considers generation.
Distribution	Not considered.	Not considered.
Failures of bulk power system components	Considers failure and repair of components; historically, only independent failures.	Assumes generation is perfectly reliable.
Explicit input assumptions	Load duration curve or chronological. Generation unit failure and repair rates.	Load duration curve. VOLL. Capacity and operational costs of generation.
Implicit assumptions	Only independent failures until relatively recently. No demand price elasticity. No supply price elasticity. Generation is dispatchable (until relatively recently).	Perfectly reliable bulk power system. No demand price elasticity. No supply price elasticity. Generation is dispatchable.
Resiliency	Not considered.	Not considered.
Key result	Minimum capacity margin that meets a specified resource adequacy criterion. No closed-form solution.	LOLP = hourly capacity rental cost/(VOLL – marginal generator's marginal cost)*
Application	Sets capacity requirement for a particular bulk power system, whether market based or utility.	Informs setting of capacity requirements and energy-only market designs.

Source: *Cretì, A, and Fontini, F. (2019). *Economics of Electricity: Markets, Competition and Rules*, Cambridge University Press.

The VOLL literature is long-standing. For this paper, its essential characteristics are that consumers generally have high heterogenous VOLLs and that incorporating these variations into system planning enhances efficiency.[30] When applied to assess resource adequacy, however, a single VOLL is used for tractability and that VOLL is a proxy for the aggregate cost of power outages.

This literature review suggests the need to integrate a variety of elements to improve understanding and outcomes. The concept of reliability needs to be expanded to include resiliency in resource adequacy

[30] Ovaere, M., et al. (2019). 'How detailed value of lost load data impact power system reliability decisions', *Energy Policy* 132, 1064–75; Gorman, W. (2022). 'The quest to quantify the value of lost load: a critical review of the economics of power outages', *The Electricity Journal* 35.8, 107187; Krause, F., and Eto, J. (1988). *Least-Cost Utility Planning Handbook for Public Utility Commissioners*. Washington, DC: National Association of Regulatory Utility Commissioners.

modelling, which has several limitations that need to be addressed. Full-throttle demand response continues to be the elusive solution to improving reliability, and perhaps the ability of consumers to tailor their reliability preferences can increase demand response. Finally, when integrated, the engineering and economic reliability frameworks may provide a more useful approach to analysing and constructing consumer-based reliability policies.

3. Integrated Model of Resource Adequacy and Economics

3.1 Overarching Reliability and Economic Framework

This section describes the reliability and economic framework used to investigate the possible incorporation of reliability preferences into electricity markets with a capacity requirement. The electricity regulator is assumed to be a benevolent social welfare maximizer, which is the equivalent of minimizing total cost, including the cost of unreliability, since demand is assumed to be inelastic. Generators are profit maximizers, and the electricity market is assumed to be perfectly competitive. Electricity retail consumers are price takers and have different VOLLs. The system operator implements the regulator's policy efficiently. These assumptions are relaxed as the paper progresses.

It is assumed that due to high transaction costs, few individual electricity consumers have the necessary metering and associated equipment to be disconnected in real time based on their VOLL. Similarly, the system operator cannot selectively disconnect individual retail consumers and, therefore, does not collect the different VOLL of individual retail consumers (that is, at a retail meter).

The base case is that all consumers must purchase their prescribed share of the total capacity required to satisfy an LOLP requirement. When there is insufficient available capacity to meet the inelastic demand, the system operator implements a prescribed load-shedding procedure without knowing individual consumers' VOLLs. The major policy considered is one that allows individual consumers to opt out of the capacity requirement; in other words, it is an optional capacity requirement. The major policy is contrasted with another policy that allows consumers to sell their capacity purchase back to the system operator during power shortages.

The overall efficiency of each case and the relative costs of each are analysed. The qualitative results supported by numerical examples show that three critical factors drive the findings. The first is the accuracy of the underlying reliability model, which is vital. Its accuracy in modelling the actual reliability of the power system determines the value of the capacity requirement for society and the individual consumer. The second factor is that efficiency improves when consumers can opt out of purchasing capacity and not be provided electricity during power shortages. The third factor is the associated transaction costs, including the efficacy of the enforcement mechanism, which determine the efficiency improvement. Likewise, the compensation for selling capacity back to the system operator determines whether such an approach is efficient and what its relative efficiency is compared to the alternatives considered.

3.2 Theoretical Model

3.2.1 Social Welfare Maximization

Table 2 summarizes the notation used throughout the paper.

Table 2: Summary of Notation

Notation	Description
A	Vector of the generation units' availabilities, 1 to N
C	Vector of generation units' annualized capital and fixed cost, \$/MW-yr
c	Vector of generation units' variable costs, \$/MWh
C_O	Individual consumer's capacity cost
c_O and c_{OO}	Individual consumer's energy cost when purchasing capacity and not purchasing capacity, respectively
δ	$USE_O - USE_{OO}$, see definitions of USE_O and USE_{OO} below
G	Vector of the capacity (MW) of individual generation units, 1 to N
g	Vector of the annual amount of energy (MWh) dispatched from each generation unit
L	Average actual load
L^F	Load forecasted by the system operator
L_O	Individual consumer (load) considering not purchasing capacity
N	Total number of generation units
P_{cap}	Price of capacity, \$/MW-yr
P_e	Vector of energy prices, \$/MWh
P_{emax}	Maximum price of energy (the energy price cap), \$/MWh
P_i	The hourly price of energy
π_{emax}	Number of hours per year where the price of energy is P_{emax}
R	Reliability level set by social planner; additional superscripts are used in specific examples discussed in the text
R^A	Reliability level based on $VOLL^A$
R^I	Reliability level based on only independent generation unit failures
R_{ucap}	Reliability level based on unforced capacity requirement
R*	Optimal reliability level based on VOLL*
T	Incremental transaction costs associated with a consumer installing remote real-time disconnection by the system operator
USE_O	Individual consumer's expected unserved energy when purchasing capacity
USE_{OO}	Individual consumer's expected unserved energy when not purchasing capacity
$VOLL^A$	Assumed VOLL, which may be different from VOLL*
VOLL*	Optimal or correct VOLL
$VOLL^H$ & $VOLL^L$	High and low VOLL for two different classes of consumers
$VOLL^{avg}$	Average of high and low VOLLs
$VOLL_O$	Individual consumer's VOLL

The social planner minimizes the total capital, fixed and operating costs of serving electricity consumers, and the cost of power outages by selecting a reliability level, R, and an optimal mix of generation, G. This corresponds to the following objective function (1):

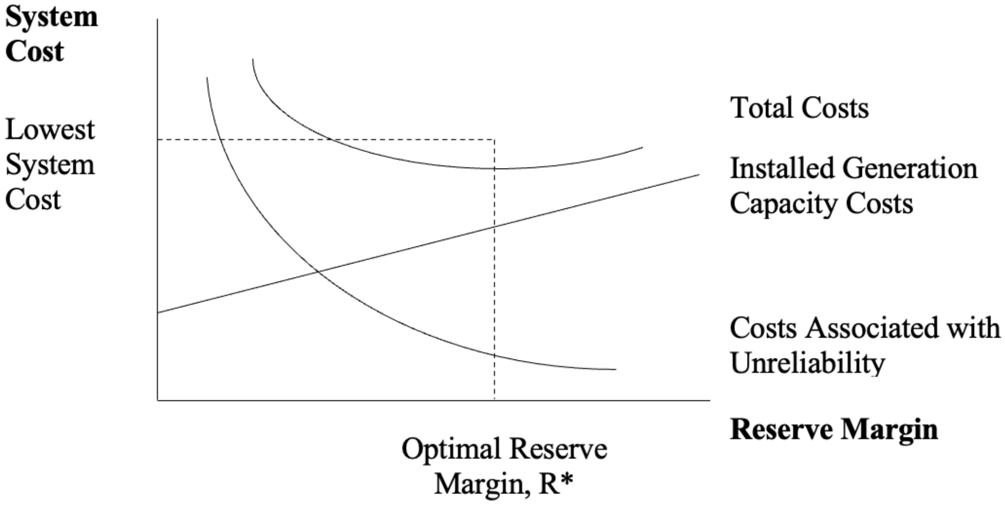
$$\min_{R,G} \text{ generation cost + cost of unserved energy} \quad (1)$$

As noted above, the optimal solution is the following:

$$\text{LOLP = hourly capacity rental cost / (VOLL – marginal generator's marginal cost)} \quad (2)$$ [31]

The LOLP is translated into R using power system resource adequacy techniques. Figure 1 summarizes this conceptual framework.[32]

Figure 1: Illustration of Standard Resource Adequacy Optimization Framework



This framework assumes that (1) the VOLL is known and correct, (2) the VOLL does not vary by consumer, (3) the system operator, whether a utility, an RTO/ISO, or a transmission system operator, is not behaving strategically, and (4) there are no causes of power outages other than inadequate generation due to independent random generation failures.

Let $VOLL^A$ be the assumed VOLL and let $VOLL^*$ be the correct VOLL, with associated reliability levels of R^A and R^*. If $VOLL^A > VOLL^*$, then $R^A > R^*$, and the social planner will inefficiently spend too much money on generation costs, thereby increasing total costs above the optimal level. There is evidence that this has been occurring in the United States. In North America, R^A is explicitly set based on the one-time-in-ten-years standard. Some have argued that, based on economics, this requirement is inefficiently high; in other words, that $R^A > R^*$ because of the implicit $VOLL^A > VOLL^*$.[33]

Complementing this view is a principal–agent analysis applied to RTOs/ISOs that finds that these system operators overstate the load forecast, resulting in over-procurement of capacity.[34] If the actual

---

[31] This equation can also be expressed on an annual basis: LOLP approximately = (annual capital rental cost) / VOLL. See Cretì, A., and Fontini, F. (2019). *Economics of Electricity: Markets, Competition and Rules*, Cambridge University Press.

[32] Based on Pfeifenberger, J.P., et al. (2013). *Resource Adequacy Requirements: Reliability and Economic Implications*, Washington, DC: Federal Energy Regulatory Commission. See also Billinton, R., and Allan, R.N. (2012). *Reliability Assessment of Large Electric Power Systems*, Springer Science & Business Media.

[33] Cramton, P, and Stoft, S. (2006). 'The convergence of market designs for adequate generating capacity with special attention to the CAISO's resource adequacy problem', white paper, Electricity Oversight Board. There are recent concerns, however, that the US transition to clean electricity is resulting in resource adequacy and other reliability problems. See NERC (2023). '2023–2024 winter reliability assessment', www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/NERC_WRA_2023.pdf; MISO (2024). 'MISO's response to the reliability imperative', https://cdn.misoenergy.org/2024%20Reliability%20Imperative%20report%20Feb.%2021%20Final504018.pdf.

[34] Aagaard, T., and Kleit, A.N. (2022). 'Too much is never enough: constructing electricity capacity market demand', *Energy Law Journal* 43, 79; Aagaard, T., and Kleit, A. (2022). 'Why capacity market prices are too high.' *Utilities Policy* 75, 101335. For a

load is on average less than the load forecast ($L < L^F$), the system operator purchases more capacity than needed, resulting in $R^A > R^*$ and, implicitly therefore, $VOLL^A > VOLL^*$.

As previously noted, the VOLL varies among consumers. To illustrate the importance of accounting for different VOLLs for different consumers, Equation (1) can be reformulated for two classes of consumers, one with a high VOLL and the other with a low VOLL,[35] and becomes Equation (3):

$\min_{R,G}$ generation cost + cost of unserved energy to $VOLL^H$ + cost of unserved energy to $VOLL^L$ (3)

The solution to this problem depends on whether the system operator can selectively disconnect $VOLL^L$ consumers before $VOLL^H$ ones. This can depend on the technologies in place but also on the event that occurred. For example, even if the technology to selectively disconnect consumers is available, the location of the event on the gird could prevent the system operator from disconnecting consumers in the most cost-effective order. If the system operator cannot distinguish between these two types of consumers, it has to use some aggregate value, perhaps an average, $VOLL^{avg}$, which, depending on the characteristics of the power outages, may not be accurate.

The final relaxed assumption is that there are other causes of power outages besides inadequate generation due to independent failures. Only considering independent failures implies that the reliability of independent generation failures, R^I, is lower than the actual reliability, R, for a given level of resource adequacy (that is, only including independent failures overestimates reliability). Thus, an optimal resource adequacy standard alone results in too little reliability.[36] Furthermore, to achieve the optimal level of reliability, using other solutions combined with increasing the amount of generation are as cost-effective or more cost-effective than using generation alone.[37]

3.2.2 Generation Model

Electric generators and consumers produce and consume wholesale electricity in a competitive market with a capacity requirement. Multiple types of generation are built and dispatched to serve demand, represented by a load duration curve. Generators are profit maximizers:

$\max_{G,g}$ capacity revenue + energy and ancillary services revenue – capital costs – operating costs (4)

where G is the vector of the capacity of individual generation units (in megawatts). For simplicity, G can be thought of as the amount of baseload, intermediate, and peaking generation capacity. The variable g is the vector of the annual amount of energy dispatched from each generation unit. There are N generation units, indexed by i.

The capacity requirement is:

$$\sum_{i=1}^{n} A_i G_i \geq R_{ucap} \quad (5)$$

where A is the vector of average annual availabilities between 0 and 1 for each generation unit, and R_{ucap} is the unforced capacity requirement.

$$\sum_{i=1}^{n} g = \min\left(L, \sum_{i=1}^{n} A_i G_i\right) \quad (6)$$

since generation must always equal demand, which cannot exceed the available supply.

$$\text{total capacity revenue} = P_{cap} \sum_{i=1}^{n} A_i G_i \quad (7)$$

---

general discussion of the RTO/ISO principal–agent problem, see Felder, F.A. (2012). 'Watching the ISO watchman.' *The Electricity Journal* 25.10, 24–37.

[35] This formulation can be extended to many classes of consumer with different VOLLs. Furthermore, it could be extended to account for the characteristics of power outages that affect an individual consumer's outage costs, such as an outage's timing, frequency, duration, and magnitude and the ability of the consumer to respond.

[36] It may be that, in practice, system operators know that the assumption of independent failures overstates resource adequacy, so they set the resource adequacy requirement above that which would be based solely on the VOLL.

[37] Felder, F.A., and Petitet, M. (2022). 'Extending the reliability framework for electric power systems to include resiliency and adaptability.' *The Electricity Journal* 35.8, 107186.

where P_{cap} is the exogenous price of capacity and is a function of the load forecast, L^F, and P_{cap} is a function that decreases in G and A and increases in L^F and R.

$$\text{total energy revenue} = \sum_{j=1}^{8760} \mathrm{P_j} \sum_{i=1}^{N} \mathrm{g_{i,j}} \qquad (8)$$ [38]

where P_j is the hourly price of energy and $g_{i,j}$ is the output of generation unit i in hour j. In some hours, P_j is P_{emax}, the energy price cap when demand exceeds supply. In markets with a price cap, P_{emax} is set below the VOLL for market power reasons.[39]

$$\text{annual cost} = \sum_{i=1}^{n} \mathrm{C_i G_i} + \sum_{i=1}^{n} \mathrm{c_i g_i} \qquad (9)$$

where C is the vector of generation units' annual fixed costs ($/MW-yr) of generators' G, and c is the vector of generators' variable cost ($/MWh).

From the first-order conditions for a competitive electricity market (not considering any network representation), the marginal revenue (MR) equals the marginal cost (MC):

$$\text{MRenergy} + \text{MRcapacity} = \text{MCenergy} + \text{MCcapacity} \qquad (10)$$

Equation (9) drives the market to equilibrium. In equilibrium, the optimal energy prices are P^*_e for each hour in the year, and the capacity price is P^*_{cap}. For the mandatory capacity requirement, since L is fixed, consumers are price takers in the energy and capacity markets.

In practice, generation units' availabilities are estimates with error bars based on historical data and are (until recently) assumed to be independent.[40] The amount of capacity that generators are permitted to sell is typically the product of a unit's rated capacity by season times its availability, providing a strong incentive for generators to be certified by the system operator at high values for these parameters. Errors in estimating availabilities can either over- or underestimate a system's resource adequacy. If availabilities are not independent, then the system's resource adequacy is overestimated.[41]

3.2.3 Retail Consumer Model

This section presents the retail consumer model in the context of allowing consumers to opt out of the capacity requirement. Retail consumers purchase electricity based on their load profile and average retail electricity prices. During shortages, these consumers may not be able to purchase electricity due to rolling power outages or cascading blackouts.[42]

There are several ways to implement an optional capacity requirement (Table 3). One is to allow load to opt out of purchasing capacity.[43] Such load, L_O, would not be allowed to purchase energy during shortages. This provision could be enforced via smart metering that disconnects this load during shortages or through penalty payments enforced through a credit requirement.[44] Another way is to allow load to sell the capacity it bought back to the market during shortage events via demand response programmes, which is the approach that US RTOs/ISOs have taken.[45]

---

[38] For simplicity, revenues from ancillary services are included in energy revenues.

[39] Markets may cap energy offers by generators to effectuate a price cap.

[40] System operators collect data from generators to estimate availabilities and associated capacities (A and G).

[41] See Felder, F.A. (2004). 'Incorporating resource dynamics to determine generation adequacy levels in restructured bulk power systems.' *KIEE International Transactions on Power Engineering* 4.2, 100–5; Murphy, S., Sowell, F., and Apt, J. (2019). 'A time-dependent model of generator failures and recoveries captures correlated events and quantifies temperature dependence', *Applied Energy* 253, 113513.

[42] Although this section is written in the context of retail and wholesale markets, the policies considered could also be applied to traditionally regulated utility consumers.

[43] Monitoring Analytics (2022). *State of the Market Report for PJM*, Eagleville, PA: Monitoring Analytics, Section 6.

[44] Joskow, P., and Tirole, J. (2007). 'Reliability and competitive electricity markets.' *The RAND Journal of Economics* 38.1, 60–84, discusses the implication of not being able to ration individual consumers, among other factors that affect achieving efficient allocation of resources in retail electricity markets.

[45] Monitoring Analytics (2022). *State of the Market Report for PJM*, Eagleville, PA: Monitoring Analytics, Section 6. 'Since the implementation of the RPM [Reliability Pricing Model, PJM's capacity market] on June 1, 2007, the capacity market (demand response) has been the primary source of demand response revenue' (p. 359). See Rodilla, P., Mastropietro, P., and Brito-Pereira, P. (2023). 'The challenge of integrating demand response: providing a comprehensive theoretical framework', *IEEE Power and Energy Magazine*, 21.4, 64–71, for a more general discussion.

Table 3: Possible Capacity Opt-Out versus Capacity Resale Policies

Possible Policies	Policy	Enforcement	Transaction Costs
Opting out of capacity requirement	Load voluntarily opts out of capacity requirement and associated costs and is required to not consume electricity during shortages	Physical disconnection	Metering, communication, and disconnection costs
		Penalties for non-compliance and credit policy	Penalties, credit policy costs, administration, and possible non-compliance
Demand response	Load selects from a menu of demand response programmes administered by the system operator with associated requirements and payments	Measurement and verification protocols	Development and administration of demand response programmes, assessment, enforcement, and possible non-compliance

The first policy for consideration is allowing retail consumers to opt out of the capacity requirement by permitting the system operator to physically disconnect these consumers during capacity shortages. They would not have to purchase the associated capacity resources for their demand. The system operator would also require the necessary information, metering, and control technologies to disconnect and reconnect them in near real time based on system conditions. Who pays for these metering-related technologies would affect whether consumers opted out. If individual consumers had to pay, this would reduce their incentive to opt out; these costs would be a substantial portion of their savings from not having to purchase capacity. Allowing consumers to opt out of the capacity market is efficient so long as they cannot consume electricity during shortages because it those consumers that opt out avoid paying something that they value less than the payment.

Table 4 compares the costs of capacity to consumers' electricity costs for PJM (an RTO). Historically, capacity costs range from 7 per cent to 22 per cent of a consumer's pre-tax wholesale electricity bill including transmiss.

Table 4: Capacity Costs as Part of Wholesale Electricity Costs for PJM, 2019–2022[46]

Year	Capacity ($/MWh)	Total ($/MWh)	Percentage of Total
2022	8.03	105.34	7.6
2021	10.95	66.78	16.4
2020	9.45	44.57	21.2
2019	11.27	50.33	22.4

Consider an optional capacity requirement in which load does not have to purchase capacity. For the optional capacity purchase, an individual consumer, denoted with a "O" subscript, chooses between purchasing capacity or not.

[46] Monitoring Analytics (2022). *State of the Market Report for PJM*, Eagleville, PA: Monitoring Analytics, Table 10; Monitoring Analytics (2020). *State of the Market Report for PJM*, Eagleville, PA: Monitoring Analytics, Table 8.

If the consumer purchases capacity, then it has the following costs:

$$c_O + C_O + (VOLL_O USE_O) \tag{11}$$

where c_O is the consumer's energy cost, C_O is the consumer's capacity cost, $VOLL_O$ is the consumer's VOLL, and USE_O is the consumer's unserved energy during power outages.

If the consumer declines to purchase capacity, then it pays the following:

$$c_{OO} + VOLL_O USE_{OO} + T \tag{12}$$

where c_{OO} is the consumer's energy cost, USE_{OO} is the consumers unserved energy during power outages, and T is the incremental transaction costs associated with opting out of the capacity market, such as paying for additional metering and controls to disconnect L_O or posting sufficient credit to pay penalties for not complying.

With the optional capacity requirement, the probability that L_O is not served increases if the consumer opts out of the capacity requirement and bears the cost of VOLL times the amount of L_O's unserved energy because the opt-out consumer is disconnected first. Also, with less capacity, the likelihood that L_O pays higher prices than it would if it had purchased capacity increases. The trade-off is between avoiding the capacity payment and paying higher energy and unserved energy costs and transaction costs. As more and more loads opt out of the capacity market, those that purchase capacity pay higher energy costs as well since the non-capacity load still buys energy in the energy market when sufficient supplies exist but there is less capacity available.[47]

If (11) exceeds (12), L_O does not purchase capacity. Subtracting the two equations and rearranging terms, if

$$(c_O - c_{OO}) + C_O + VOLL_O(USE_O - USE_{OO}) - T > 0 \tag{13}$$

then the consumer does not purchase capacity.[48]

$$\text{Let } \delta = USE_O - USE_{OO} \tag{14}$$

For ease of explanation, assume that the change in energy costs between the two choices is small. This comparison, then, depends on C_O, the product of δ and $VOLL_O$, and T. A small $VOLL_O\delta$ product and low T favours not buying capacity; a low C_O favours purchasing capacity. Given that the product of $VOLL_O$ and δ is what matters, consumers with high VOLL, such as data centres, may opt out of the capacity market if they have extremely reliable backup power because even if they are disconnected from the grid, they still have power. When making this evaluation, consumers may change their behaviour to reduce their outage costs (for example, install energy storage, distributed generation, or backup generation) when forgoing the purchase of capacity, accounting for how the expected timing, duration, and magnitude of power outages affect them (in effect reducing their $VOLL_O$).[49]

3.2.4 Policy Discussion

Policymakers determine, in part, δ and T. When a generation shortage occurs, typically (at least according to the modelling results based on independent failures) only a small amount of load is not served. Thus, if the policy for load that does not purchase capacity is that it is disconnected first, then δ increases, whereas, if the load purchased capacity, it is disconnected based on the system operator's procedure. Furthermore, if the policy of disconnecting load is based on the type of load, then δ could be small for some load. For instance, critical facilities such as hospitals and police stations would be

[47] See Cramton, P., Ockenfels, A. and Stoft, S. (20130. 'Capacity market fundamentals.' *Economics of Energy & Environmental Policy* 2.2, 27–46 for a discussion of the 'price suppression effect'.

[48] For simplicity, this analysis assumes that the consumer's VOLL does not vary between the two cases, although a variation could easily be accommodated. Furthermore, the characteristics of power outages and their associated costs in both cases could be considered.

[49] Consumers who continue to purchase capacity may also take similar actions if, even with the purchase of capacity, the costs associated with these actions help avoid greater costs expected from the power outages.

expected to be disconnected last. Thus, a rule would be necessary that requires critical facilities to purchase capacity; otherwise, such load would not purchase capacity, knowing that they would likely not be disconnected.

The conditions under which L_O is disconnected are also essential. System operators' responses to potential and actual shortages follow procedures that include many steps and discretion.[50] These responses may include calling for voluntary load reductions, operating generation and transmission at emergency limits, reducing voltages, purchasing power from neighbouring systems, reducing operating reserves, or disconnecting load. Where in this list L_O is disconnected is important because it affects δ. This priority list raises the question of whether resource adequacy measures the probability of disconnecting firm load or the probability of taking emergency actions to avoid and limit doing so. Presently, load is disconnected by disconnecting portions of a distribution system, which disconnects multiple loads instead of individual loads, so a load's location on the distribution system affects its likelihood of being disconnected by the system operator.

Multiple causes and contributing factors also require the disconnection of load. A combination of generation, transmission, and distribution outages may require load reductions. If the capacity requirement is only based on independent failures of generation units, load that does not purchase capacity but pays via the transmission tariff for reliable service should not be disconnected prior to load that purchases capacity during transmission failures. Determining this in real time during a system emergency may be challenging. Furthermore, many power outages, including some of the most severe ones, are not due to independent failures of generation.[51] The system operator must be able to distinguish between resource adequacy events and other causes of power outage since consumers should only be disconnected when there are supply shortages. This ensures that consumers' compensation aligns with the benefit they are providing to the system operator. This requires the resource adequacy model to explicitly articulate which causes of capacity shortage are included and not included in the model, such as dependent failures and correlated failures.

The incremental transaction costs, T, associated with load that does not purchase capacity are also important. If smart meters and disconnection technology are required for all loads, then the T is 0. If individual consumers are required to pay for such equipment as a requirement of not purchasing capacity, then the amount of T becomes important. The appropriate technologies must be available, installed, and highly reliable for these improvements to be implemented effectively.

Note that L_O is very small compared to the size of shortages. Even large industrial peak loads are only in the order of tens of megawatts. This represents a small percentage of power outages, which run into the hundreds if not thousands of megawatts. Thus, for the system operator to invest time and effort into integrating this type of load into its planning, systems, operations, and training, the benefits must be substantial. One hundred megawatts of peak demand may consist of tens of thousands of consumer meters. As the amount of demand that opts out of purchasing capacity increases, the number of consumers per megawatt likely increases as additional smaller consumers are needed to obtain the next increment of megawatt reduction.[52]

The second opt-out policy for consideration is self-enforced disconnection. Under this policy, consumers can also opt out of the capacity market and must reduce their electricity consumption to zero when told to do so by the system operator during capacity shortage s. If they do not, they pay a penalty that is well above the energy price during the shortages, and they must post sufficient credit to cover possible penalties. This credit avoids having to install disconnection equipment (although a real-time meter is required) but does not completely ensure that the consumer will stop consuming electricity when instructed to do so. Both the physical opt-out and the financial opt-out options could be applied partially,

---

[50] Joskow, P., and Tirole, J. (2007). 'Reliability and competitive electricity markets.' *The RAND Journal of Economics* 38.1, 60–84. The implication from this observation is that the VOLL should be replaced with the cost of the system operator's action, a cost function that depends on power outage timing, frequency, magnitude, and duration and on consumers' ability to adapt.

[51] Felder, F.A. (2001). '"An island of technicality in a sea of discretion": a critique of existing electric power systems reliability analysis and policy', *The Electricity Journal*, 14.3, 21–31.

[52] Zarnikau, J.W. (2010). 'Demand participation in the restructured Electric Reliability Council of Texas market', *Energy*, 35.4, 1536–43, notes that declining technological costs may increase the ability of small consumers to be active demand-response participants.

meaning that such consumers could opt out of some of their capacity requirement, for example based on 30 per cent of their peak load.

The third possible improvement is consumers selecting from a menu of demand response alternatives provided by the system operator. The capacity value of these demand reduction profiles could be assessed similarly to how the capacity value of energy storage and energy-limited resources are in resource adequacy models.

For these types of demand response programme, the baseline demand consumption for each participating consumer needs to be estimated. One way is to require consumers to bid in the day-ahead market (assuming this market exists). This prevents consumers gaming the system by creating artificially high baselines to reduce their demand. Given that these proposals depend on estimated and but-for calculations, it is necessary to identify possible gaming opportunities and responses. There have been many examples of such gaming tactics.[53]

Of course, these policy proposals occur within a political context that may affect their acceptability. The implementation of such proposals would depend, in part, on the willingness of policymakers to be accountable setting policies that may lead to some consumers being disconnected and the willingness of consumers to accept the consequences of their decisions.

4. Numerical Example

4.1 Overall System Description and Outcomes

The following stylized numerical calculations integrate resource adequacy and costs. They can be applied to electricity markets with capacity requirements but also to traditionally regulated systems that would allow consumers to avoid charges related to generation capacity. These calculations illustrate that resource adequacy results can vary with relatively small changes in assumptions, making it challenging for consumers to know exactly what their capacity purchases provide, and that if a consumer decides to opt out of purchasing capacity, that consumer's outage probability may substantially increase.

All generation units are assumed to be the same size and have the same availability, which allows for the use of a binomial formula to calculate available generation and associated probabilities. Table 5 lists the generation and load characteristics of the model.

Table 5: Generation and Load Assumptions

Generation					Load	
Generation Type	**Generation Size (MW)**	**Capital Cost ($/kW)**	**Variable Cost ($/MWh)**	**Generation Availability**	**Load (MW)**	**Load (% Time)**
Baseload	100	1,000	20	0.9	3,000	40.0
Intermediate	100	600	40	0.9	5,000	50.0
Peaking	100	250	100	0.9	8,500	10.0
						100

Table 6 is the size in megawatts and availabilities of the different types of generation unit that satisfy the one-time-in-ten-years resource adequacy standard at the least cost.

[53] See, for example, Howland, E. (2024). 'FERC enforcement office seeks $27M from Ketchup Caddy for MISO demand response fraud', www.utilitydive.com/news/ferc-enforcement-ketchup-caddy-miso-market-manipulating/708183; Howland, E. (2024). 'NIPSCO, Linde to pay $66.7M to settle charges for gaming MISO demand response program', www.utilitydive.com/news/nipsco-linde-ferc-miso-demand-response-settlement/703888; Multer, M. (2023). 'FERC's enforcement in demand response case a lesson for utilities', www.powermag.com/fercs-enforcement-in-demand-response-case-a-lesson-for-utilities. The PJM Market Monitor has also identified other problems with demand response. See Monitoring Analytics (2022). *State of the Market Report for PJM*, Eagleville, PA: Monitoring Analytics, Section 6.

Table 6: Generation Capacity and Expected Availability

Unit Type	Number	Unit Size (MW)	Total (MW)	Availability	Expected Availability (MW)
Baseload	33	100	3,300	0.9	2,970
Intermediate	22	100	2,200	0.9	1,980
Peaking	49	100	4,900	0.9	4,410
	104		10,400		9,360

Capacity Margin	18%
Reserve Margin*	22%

Note: *The difference between capacity margin and reserve margin is that in the denominator, the former uses capacity and the latter uses peak load.

Table 7 calculates the expected annual costs of this system, assuming an annual capital carrying charge of 15 per cent.

Table 7: Expected Annual Capital and Dispatch Costs ($, thousands)

Annual Capital Costs ($000)	Expected Dispatch Cost ($000)	Annual Capital + Dispatch Costs ($000)
$ 495,000	$ 525,600	$ 1,020,600
$ 198,000	$ 1,051,200	$ 1,249,200
$ 183,750	$ 6,745,200	$ 6,928,950
$ 876,750	$ 8,322,000	$ 9,198,750

Interpreting 'one time in ten years' as having insufficient generation to meet demand for 24 hours in ten years (or 2.4 hours in one year), the LOLP (2.4 / 8,760 = 0.0003, or equivalent of 0.9997) and expected unserved energy are presented in Table 8.[54] The available megawatts are calculated using the binomial distribution, with all units having an assumed availability of 0.9. The boxed row indicates the probability associated with this LOLP criterion.

[54] There are 8,760 hours in a non-leap year. Table 8 only considers peak hours, which is 10% of the time. During non-peak hours, it is assumed that there is always sufficient generation to meet demand. Thus, if during peak hours the probability of demand exceeding supply is 0.997 during all hours, it is 0.9997.

Table 8: Loss of Load Probability and Expected Unserved Energy during Peak Hours

# Units Available (no.)	MW Available	Probability during Peak Hours	Cumulative Probability during Peak Hours	Expected Unserved Energy (MWh)	Unserved Energy during Power Outage (MWh)
104	10,400	0.0000	0.0000		
103	10,300	0.0002	0.0002		
102	10,200	0.0012	0.0014		
101	10,100	0.0044	0.0057		
100	10,000	0.0122	0.0179		
99	9,900	0.0271	0.0451		
98	9,800	0.0498	0.0948		
97	9,700	0.0774	0.1722		
96	9,600	0.1043	0.2765		
95	9,500	0.1236	0.4001		
94	9,400	0.1305	0.5306		
93	9,300	0.1239	0.6544		
92	9,200	0.1067	0.7611		
91	9,100	0.0839	0.8449		
90	9,000	0.0606	0.9055		
89	8,900	0.0404	0.9459		
88	8,800	0.0250	0.9709		
87	8,700	0.0144	0.9852		
86	8,600	0.0077	0.9929		
85	8,500	0.0039	0.9968		
84	8,400	0.0018	0.9986	160	87,600
83	8,300	0.0008	0.9994	143	175,200
82	8,200	0.0003	0.9998	90	262,800
81	8,100	0.0001	0.9999	47	350,400
80	8,000	0.0001	1.0000	22	438,000
79	7,900	0.0000	1.0000	9	525,600
78	7,800	0.0000	1.0000	4	613,200
77	7,700	0.0000	1.0000	1	700,800
76	7,600	0.0000	1.0000	0	788,400
75	7,500	0.0000	1.0000	0	876,000
74	7,400	0.0000	1.0000	0	963,600
73	7,300	0.0000	1.0000	0	1,051,200
72	7,200	0.0000	1.0000	0	1,138,800
				477	7,971,600

Note: Probability values rounded to 4 decimal places. Generation availabilities of less than 72 units are ignored since the probabilities are negligible. The values in the row immediately below the boxed row are calculated as follows: 84 is the number of generation units that are available, for a total of 8,400 MW; 0.0018 is the probability that out of 104 generation units, each with an availability of 0.9, 84 are available; 0.9986 is the cumulative probability that 84 or more generation units are available during peak hours; 160 MWh is (8,500 – 8,400) × 0.0018 × 8,760 × 10% (with rounding); and 87,600 MWh is (8,500 – 8,400) × 8,760 × 10%.

Recalculating Table 8 with one less peaking unit, the increase in expected unserved energy is 502 MWh at an annual cost saving of $3,750,000, resulting in an implicit VOLL of $7,470/MWh. Table 8 also indicates that only a small percentage, both in megawatt-hours and number of consumers, need to reduce their consumption during capacity shortages. The amount of load during the peak hours is 7,446,000 MWh (8,500 MW × 8,760 hours × 0.1).

As discussed in the literature review section, experience has shown that generation unit failures are not independent. Thus, having a capacity requirement, even if formally based on the assumption of independent generation failures, when combined with a one-time-in-ten-year requirement is informally capturing dependent and correlated failures. If this is the case, then the expected unserved energy values presented in the above tables would be too low, and an individual consumer would have to have access to more accurate values to make an informed decision of whether to opt out of the capacity requirement. This makes it challenging for market designers, system planners, and consumers to forecast power outages with certainty.

Assume that the actual availability is lower than the assumed 90 per cent. Table 9 recalculates the expected unserved energy and implicit VOLL. Even a small change in the assumed availabilities results in a substantial change in the reported values.

Table 9: Expected Unserved Energy When Actual Availability is Different Than Assumed

Actual Availability	Expected Unserved Energy (MWh)	Change in Unserved Energy (MWh)	Multiplier Change
0.90	477		
0.89	1,439	961	2 = 961/477
0.85	31,419	30,942	65 = 30,942/477
0.80	224,723	224,246	470 = 224,246/477

4.2 Individual Consumer Opt-Out Analysis Assuming Independent Generation Failures

Now consider a 1 MW consumer that has the option of opting out of the capacity market. Table 10 compares what a 1 MW consumer is expected to experience when purchasing capacity versus not purchasing capacity. If the consumer purchases capacity, it is assumed that the consumer is randomly disconnected. For example, if the system operator is short 100 MW in a peak hour, the probability that the consumer who purchased capacity is disconnected in that hour is 100 / 8,500. If the consumer opts out of purchasing capacity, the probability of being disconnected in that hour is assumed to be 1. The change in the expected unserved energy is 2.7 MWh is (2.8 – 0.06).

Table 10: Comparison of Disconnection of 1 MW Consumer with and without Capacity Purchases

Capacity Shortage (MW)	Probability	Unserved Energy during Power Outage (MWh)	Expected Unserved Energy 1 MW Customer Disconnected with Capacity (MWh)	Expected Unserved Energy 1 MW Customer Disconnected without Capacity (MWH)
100	0.002	87,600	0.02	1.6
200	0.001	175,200	0.02	0.7
300	0.000	262,800	0.01	0.3
400	0.000	350,400	0.01	0.1
500	0.000	438,000	0.00	0.0
600	0.000	525,600	0.00	0.0
700	0.000	613,200	0.00	0.0
800	0.000	700,800	0.00	0.0
900	0.000	788,400	0.00	0.0
1000	0.000	876,000	0.00	0.0
1100	0.000	963,600	0.00	0.0
1200	0.000	1,051,200	0.00	0.0
			0.06	2.8

The annual capacity cost for a 1 MW consumer with a 18 per cent required capacity margin is $44,250. For a change in expected unserved energy of 2.7 MWh, the breakeven VOLL in which a consumer is indifferent to purchasing capacity is $16,389/MWh assuming no transaction costs. In a capacity structure that allows for opting out of purchasing capacity, those opting out dramatically increase their probability of being called on to curtail their electricity consumption as they would be curtailed first.

Table 11 presents the results of repeating the above consumer-specific analysis but using actual availabilities less than the assumed availabilities. If actual generation availabilities are below the assumed availabilities, the decision to opt out of the capacity market becomes less attractive. In this case, the breakeven VOLL decreases to $250/MWh, again assuming no transaction costs.

Table 11: Comparison of Disconnection of 1 MW Consumer with and without Capacity Purchases—Actual Generation Availability of 85% but Assumed to Be 90%

Capacity Shortage (MW)	Probability	Unserved Energy during Power Outage (MWh)	Expected Unserved Energy 1 MW Customer Disconnected with Capacity (MWh)	Expected Unserved Energy 1 MW Customer Disconnected without Capacity (MWh)
100	0.067	-	0.69	58.4
200	0.050	86,000	1.03	43.8
300	0.035	172,000	1.09	30.9
400	0.024	258,000	0.97	20.6
500	0.015	344,000	0.76	13.0
600	0.009	430,000	0.54	7.7
700	0.005	516,000	0.36	4.4
800	0.003	602,000	0.22	2.3
900	0.001	688,000	0.13	1.2
1000	0.001	774,000	0.07	0.6
1100	0.000	860,000	0.03	0.3
1200	0.000	946,000	0.02	0.1
			5.9	183.4

5. Overall Conclusions and Further Research

To analyse if and how retail electricity consumers can tailor their welfare-enhancing reliability needs in an electricity market with a capacity requirement requires assessing whether the current resource adequacy framework is sufficient. If it is, the analysis can proceed by considering various approaches and under what conditions they are welfare enhancing. Unfortunately, several essential factors suggest that the resource adequacy reliability framework is insufficient, and therefore, if consumers can opt out of their capacity requirement, the making of general statements is challenging.

The ability of retail electricity consumers to tailor their reliability preferences in electricity markets is limited. Electricity markets with capacity requirements are optimal only under restrictive assumptions of a single VOLL for all consumers; power outages caused by independent generation failures; an accurate estimation of generation availabilities; and non-strategic behaviour by generators, system operators, and electricity consumers. The barriers to allowing electricity consumers to tailor their reliability purchases are the high transaction costs associated with metering, telecommunication, and disconnection and enforcement, as well as the high system operator costs.

Permitting electricity consumers to opt out of purchasing capacity, subject to sufficient monitoring and enforcement, is welfare enhancing. It would enable low VOLL consumers or those who can adapt, for example by having sufficient backup power, to avoid paying for capacity they do not need. This opt-out option allows consumers to use their private information on outage costs, based on the outage characteristics that are costly to them (rather than having a social planner using a generic VOLL), to decide whether to purchase capacity. Under what conditions opt-out consumers are required to disconnect is a crucial determinant of the expected cost of opting out. Consumers, however, must have sufficiently accurate information regarding the probability of experiencing a power outage with and without purchasing capacity.

Table 12 summarizes the overall findings.

Table 12: Summary of Overall Findings

Finding	Generation Resource Adequacy as an Accurate Reliability Probability Measure (can be interpreted physically)	Generation Resource Adequacy as an Index (a relative measure of reliability, not to be interpreted physically)
1. Engineering resource adequacy requirement is not economically optimal.	Does not explicitly account for VOLL. There is some evidence that it implies too high of a VOLL. Does not account for other causes of power outages. Does not account for other means of reducing power outages besides generation.	An explicit model to account for all causes of power outage is required to evaluate the optimality of a resource adequacy requirement. The numerical values of generation resource adequacy metrics (such as loss of load expectation) are very sensitive to generation availability assumptions and correlated failures. Additional generation may provide reliability benefits beyond resource adequacy in such a way that allowing for opting out of capacity requirements may reduce welfare.
2. An economically based resource adequacy requirement is suboptimal.	The use of a single VOLL value is suboptimal. Allowing for opting out of capacity requirements may improve social welfare. Capacity-related demand response programmes may be cost-effective depending on their details.	
3. Transaction costs associated with opting out and capacity-related demand response programmes are critical.	Capacity opt-out options require sufficient metering and either disconnection capability or credit requirements. Capacity-related demand response programmes require administrative and monitoring costs.	
4. Levels of capacity requirements need more research.	Further research is required on whether there can be levels of capacity purchases, such as one-time purchases in five or 15 years.	

There are numerous avenues for further research. One is considering whether each of the above approaches could be modified by allowing consumers to select their own level of resource adequacy that is different from the one stipulated by the capacity mechanism. For instance, could some consumers choose a one-in-five-year or a one-in-20-year standard? The resource adequacy model could be run with these types of consumers individually to determine the corresponding amount of capacity that must be bought. There may be a resiliency offer option that allows consumers to propose their own demand curtailment profile instead of selecting from a menu of provided options. Under this approach, the resource adequacy model and analytical process must be able to examine individual consumers' submissions to determine the associated capacity value. As the number of submissions increases to potentially millions, this may be a computational challenge. One possibility is that the aggregation of consumers' submissions is analysed based on load aggregators' submissions, which would vastly reduce the number of submissions to be considered. It would then be up to each load aggregator to further disaggregate the capacity value to each consumer.

Implementing tailored resource adequacy policies in a real power system or market needs further investigation to develop specific rules, including market monitoring and manipulation. Broadening the notion of reliability to include resiliency—particularly in the context of increasing the amount of variable and intermittent generation resources such as wind and solar photovoltaics—is essential.

A final avenue of future work is revisiting this discussion in the context of energy-only markets such as Texas (ERCOT), Australia, Alberta, and New Zealand. At first glance, energy-only markets seem to be the ultimate example of consumers being able to select their desired level of reliability through their willingness to pay for energy. During power outages, however, if consumers are not disconnected based on their willingness to pay, then policies that enable this should be considered due to their potential to improve social welfare.