Long-Term Resource Adequacy in Wholesale Electricity Markets with Significant Intermittent Renewables

Frank A. Wolak, *Stanford University and NBER,* United States of America

Executive Summary

Growing amounts of intermittent renewable generation capacity substantially increase the complexity of determining whether sufficient energy will be available to meet hourly demands throughout the year. As the events of August 2020 in California and February 2021 in Texas demonstrate, supply shortfalls can have large economic and public health consequences. An empirical analysis of these two events demonstrates that similar supply shortfalls are likely to occur in the future without a paradigm shift in how long-term resource adequacy is determined for an electricity supply industry with significant intermittent renewables. An alternative approach to determining long-term resource adequacy that explicitly recognizes the characteristics of different generation technologies is outlined and its properties explored relative to current approaches.

JEL Codes: Q4, L51, L94

Keywords: electricity markets, intermittent renewables, long-term resource adequacy

I. Introduction

Restructured electricity supply industries have one glaring weakness that is becoming increasingly apparent as the share of intermittent renewables in a region increases and more consumers shift to using electricity for space heating and personal transportation. There is no single entity responsible for ensuring that the supply of electricity equals demand under all possible current and future demand conditions. Generation unit

Environmental and Energy Policy and the Economy, volume 3, 2022.
© 2022 National Bureau of Economic Research. All rights reserved. Published by The University of Chicago Press for the NBER. https://doi.org/10.1086/717221

owners can only supply electricity up to the available capacity of their generation units. Transmission network operators can only dispatch the set of available generation units or curtailable demands in the geographic region under their control. Electricity retailers can only withdraw the amount of energy produced by generation unit owners less any transmission network losses.

Under the vertically integrated monopoly regime, the geographic monopoly electricity supplier was the single entity responsible for ensuring supply equals demand under all possible current and future system conditions. Consequently, politicians and regulators knew precisely what entity to penalize if supply shortfalls occurred. In the restructured regime, generation unit owners, retailers, and the system operator can all shift blame to some other entity for a supply shortfall.

Fortunately, in a restructured electricity supply industry composed of dispatchable thermal generation units and predictable peak demands, ensuring that supply will equal demand throughout the year is relatively straightforward. The system operator first multiplies the installed capacity of each generation unit by its availability factor, the fraction of hours of the year the unit is expected to be available to operate. If the sum of the availability-factor-adjusted capacities across all generation units is greater than the annual demand peak by a 10%–15% margin, the system operator can be confident that there will be sufficient supply to meet demand throughout the year.

This process becomes more complicated if a substantial fraction of energy comes from hydroelectric resources because water availability determines how much energy these resources can produce at any time during the year. There are substantial unpredictable differences across seasons and years in the amount of water that is available to produce electricity, and many examples from hydro-dominated markets around the world where unexpectedly low water conditions have led to periods with supply shortfalls and/or extremely high prices in the short-term market.[1] The first evidence that the traditional capacity-based approach to long-term resource adequacy is inappropriate for regions with significant intermittent renewable resources is that these outcomes occurred because insufficient energy was available to be produced by the hydroelectric units, and not because there was insufficient hydroelectric generation capacity in the region.

As the share of intermittent renewable energy from wind and solar generation units in a region increases, it becomes even more difficult to ensure that supply equals demand during all hours of the year. Wind and solar resources can stop producing energy with little advance

notice, produce very little energy during extreme hot and cold weather conditions, and have long durations of low energy output. These facts make it virtually impossible to determine the amount of energy wind and solar resources can reliably supply during any specific time interval during the year.

Many regions of the United States are transitioning to electricity and away from fossil fuels for space heating and personal transportation services. Charging of electric vehicles significantly increases both the level and variability of electricity demand. Electric space heating significantly increases the sensitivity of electricity demand to cold weather conditions. A significant share of space heating with electricity can change a region from one where the annual demand peak occurs during the summer to one where it occurs during the winter.

These facts imply the need for revisions to the existing approach to long-term *resource adequacy* (RA)—the process of ensuring that supply will equal demand during all hours of the year—in regions with significant amounts of wind and solar resources and goals to transition to electricity for space heating and personal transportation. The purpose of this paper is to propose a long-term RA mechanism that is more likely to achieve a reliable supply of electricity in this environment.

The first step in this process is a statement of why, unlike other product markets, all existing wholesale electricity markets require a long-term RA mechanism. This is because of what Wolak (2013) calls the *reliability externality*, caused by a regulator-mandated upper bound on the offer price a supplier can submit to the short-term market in all existing wholesale markets in the United States. This cap on offer prices creates an incentive for electricity retailers and consumers to underprocure their expected real-time demands in the forward energy market, which can result in energy shortfalls during high-demand conditions and expose customers to extremely high prices for sustained periods of time.

Empirical evidence from California during August 2020 is used to illustrate the increasing risk of relying on a capacity-based approach to address the reliability externality in a wholesale electricity market with a substantial share of intermittent renewables and policy goals to transition to electricity for space heating and personal transportation. The experience of Texas during February 2021 is used to illustrate the risk of not having a formal long-term RA mechanism in place in a wholesale electricity market with a significant share of intermittent renewables, even if there is an extremely high offer cap on the short-term market. The amount of energy supplied by renewable resources during high-demand periods in these two markets can be unexpectedly low, and

for both markets this was a major factor determining the need to curtail demand during these two time periods.

The experience of California, Texas, and several international markets demonstrates that having adequate energy available to serve demand, not adequate generation capacity, is the fundamental long-term RA challenge in renewables-dominated regions. I propose a standardized fixed-price forward contract (SFPFC) approach to long-term RA to address this challenge. This approach assigns the risk of meeting system demand throughout the year to generation unit owners. It encourages cross-hedging of energy-supply risk between dispatchable generation units and intermittent renewable resource owners. It also fosters the active participation of final consumers and storage resources in managing the real-time supply and demand balance.

The SFPFC mechanism shares a key feature with the existing long-term RA mechanisms that exist in Chile and Peru, two regions with significant hydroelectric resources as well as growing shares of intermittent wind and solar resources, particularly in Chile. As outlined in Wolak (2021), both these regions assign the risk of meeting system demand throughout the year to generation unit owners by operating a supplier-only short-term market where electricity retailers and large consumers must purchase full-requirements contracts from participants in the short-term market to meet their hourly energy demands throughout the year. These regions have successfully served system demands with average annual growth rates in Chile of more than 7% since 1992 and more than 5% since 1990 in Peru. This outcome emphasizes the necessity of high-powered financial incentives for suppliers to ensure that system demand is met every hour of the year in regions with significant intermittent renewables. The goal of the SFPFC mechanism is to subject suppliers to this high-powered financial incentive while still maximizing the opportunities for active participation of final consumers in the short-term market.

The remainder of the paper proceeds as follows. The next section defines the reliability externality and argues that it exists in all markets with finite offer caps on the short-term market. Section III describes the conventional solution to the reliability externality—a capacity-based long-term RA mechanism. This section explains why such an approach to long-term RA is likely to work as intended in a system with dispatchable thermal generation units, and why it is has led to supply shortfall periods in regions with significant renewable energy shares. Section IV uses California and Texas market outcomes during each region's supply shortfall period to illustrate, for the case of California, the inappropriateness of a capacity-based long-term RA mechanism in renewables-dominated markets and, for the case

of Texas, the need for a long-term RA mechanism even in a short-term market with an extremely high offer cap. Section V presents the SFPFC mechanism and explains why it is more likely to achieve long-term RA in a renewables-dominated market with electrification goals for space heating and transportation. This section uses the performance of the existing long-term RA mechanisms in Chile and Peru as empirical evidence in favor of the SFPFC mechanism. Section VI concludes and suggests directions for future research.

II. The Reliability Externality in Wholesale Electricity Markets

Why do wholesale electricity markets require a regulatory mandate to ensure long-term RA? Electricity is essential to modern life, but so are many other goods and services. Consumers want cars, but there is no regulatory mandate that ensures enough automobile assembly plants to produce these cars. They want point-to-point air travel, but there is no regulatory mandate to ensure enough airplanes to accomplish this. Many goods are produced using high fixed cost, low marginal cost technologies similar to electricity supply. Nevertheless, these firms recover their production costs, including a return on the capital invested, by selling their output at a market-determined price.

So, what is different about electricity that requires a long-term RA mechanism? The regulatory history of the electricity supply industry and the legacy technology for metering electricity consumption results in what Wolak (2013) calls a reliability externality.

Unlike the case of wholesale electricity, the markets for automobiles and air travel do not have a regulatory limit on the level of the short-term price. Airlines adjust the prices for seats on a flight over time in an attempt to ensure that the number of customers traveling on that flight equals the number of seats flying. This ability to use price to allocate the available seats is also what allows the airline to recover its total production costs and can result in as many different prices paid for the same flight as there are customers on the flight.

Using the short-term price to manage the real-time supply and demand balance in a wholesale electricity market is limited by a finite upper bound on a supplier's offer price and/or a price cap set by the regulator that limits the maximum market-clearing price. Although offer caps and price caps can limit the ability of suppliers to exercise unilateral market power in the short-term energy market, they also reduce the revenues suppliers can receive during scarcity conditions. This is often referred to as the *missing money problem* for generation unit owners.

However, this missing money problem is only a symptom of the existence of the reliability externality.

This externality exists because offer caps limit the cost to electricity retailers of failing to hedge their expected purchases from the short-term market. Specifically, if the retailer or large consumer knows the price cap on the short-term market is $250 per megawatt-hour (MWh), then it is unlikely to be willing to pay more than that for electricity in any earlier forward market. This creates the possibility that real-time system conditions can occur where the amount of electricity demanded at or below the offer cap is less than the amount suppliers are willing to offer at or below the offer cap.

This outcome implies that the system operator must be forced to either abandon the market mechanism or curtail firm load until the available supply offered at or below the offer cap equals the reduced level of demand, as occurred several times in California between January 2001 and April 2001, and most recently on August 14–15, 2020. A similar, but far more extreme, set of circumstances arose during February 14–18, 2021, in Texas, and this required significant demand curtailments during February 15–18.[2]

Because random curtailments of supply to different distribution grids served by the transmission network—also known as rolling blackouts—are used to make demand equal to the available supply at or below the offer cap under these system conditions, this mechanism creates a reliability externality because no retailer bears the full cost of failing to procure adequate energy to meet their demand in advance of delivery. A retailer that has purchased sufficient supply in the forward market to meet its actual demand is equally likely to be randomly curtailed as another retailer of the same size that has not procured adequate energy in the forward market. For this reason, all retailers have an incentive to underprocure their expected energy needs in the forward market. When short-term prices rise because of the supply shortfalls, retailers that do not hedge their wholesale energy purchases will go bankrupt. If they attempt to pass these short-term prices on to their retail customers, many are likely to be unable to pay their electricity bills. As we discuss in Subsection IV.B, both outcomes occurred in Texas following the events of February 14–18, 2021.

The lower the offer cap, the greater the likelihood that the retailer will delay their electricity purchases to the short-term market. Delaying more purchases to the short-term market increases the likelihood of insufficient supply in the short-term market at or below the offer cap. Because retailers do not bear the full cost of failing to procure sufficient energy in the forward market, there is a missing market for long-term contracts for energy with

long enough delivery horizons into the future to allow new generation units to be financed and constructed to serve demand under all future conditions in the short-term market. Therefore, a regulator-mandated long-term RA mechanism is necessary to replace this missing market.

Regulatory intervention is necessary to internalize the resulting reliability externality unless the regulator is willing to eliminate the offer cap and commit to allowing the short-term price to clear the real-time market under all possible system conditions. There are no short-term wholesale electricity markets in the world that make such a commitment. All of them have either explicit or implicit caps on the offer prices suppliers can submit to the short-term market. The Electricity Reliability Council of Texas (ERCOT) has a $9,000/MWh offer cap, which is the highest in the United States. The National Electricity Market (NEM) in Australia has a A$15,000/MWh offer cap, which is currently the highest in world.

As the experience of February 14–18, 2021, in Texas demonstrated, an extremely high offer cap on the short-term market does not eliminate the reliability externality. It just shrinks the set of system conditions when random curtailments are required to balance real-time supply and demand. For the same reason, there have also been a small number of instances when the NEM of Australia has experienced supply shortfalls despite having an extremely high offer cap.

If customers do not have the ability to shift their demand away from these high-priced periods because a significant fraction of their demand for electricity is caused by space heating needs in response to the freezing outside temperatures, charging customers an extremely high wholesale price for their consumption is largely punitive. This was the case for many retail electricity customers in Texas during February 2021. They were committed to buy a substantial fraction of their wholesale electricity at the short-term price at a time when their demand for electricity for space heating is extremely price inelastic. This experience underscores the importance of a long-term RA mechanism in regions with significant intermittent renewables, growing electrification of space heating, and increasing adoption of electric vehicles.

III. Conventional Solution to Reliability Externality with Intermittent Renewables

Currently, the most popular approach to addressing the reliability externality is a capacity procurement mechanism that assigns a firm capacity

value to each generation unit based on the amount of energy it can provide under stressed system conditions. Retailers are then required to demonstrate that they have purchased sufficient firm capacity to meet their monthly or annual demand peaks. Having sufficient firm capacity typically means that the retailer has purchased firm capacity equal to between 1.10 and 1.20 times its annual demand peak. The exact multiple of peak demand chosen by a region depends on the mix of generation resources and the reliability requirements of the system operator.

Under the current long-term RA mechanism in California, firm-level capacity procurement obligations are assigned to retailers by the California Public Utilities Commission (CPUC) to ensure that monthly and annual system demand peaks can be met. Electricity retailers are free to negotiate bilateral capacity contracts with individual generation unit owners to purchase firm capacity to meet these obligations. The eastern United States wholesale electricity markets in the PJM Interconnection, Independent System Operator (ISO) New England, New York ISO, and Midcontinent ISO each have a centralized market for firm capacity. These involve periodic capacity auctions run by the wholesale market operator where all retailers purchase their capacity requirements at a market-clearing price. ERCOT does not currently have a formal long-term RA mechanism besides its $9,000/MWh offer cap and an ancillary services scarcity pricing mechanism.

All capacity-based approaches to long-term RA rely on the credibility of the firm capacity measures assigned to generation units. This is a relatively straightforward process for dispatchable thermal units. As noted earlier, the nameplate capacity of the generation unit times its annual availability factor—the fraction of hours of the year a unit is expected to be available to produce electricity—is the typical starting point for estimating the amount of energy the unit can provide under stressed system conditions. As discussed below, if all retailers have met their firm capacity requirements in a sizable market with only dispatchable thermal generation, there is a very high probability that the demand for energy will be met during peak demand periods.

A simple model helps to illustrate the logic behind this claim. Suppose that the peak demand for the market is 1,000 MW and the market is composed of equal size generation units, each with a 90% annual availability factor, meaning that it is available to produce electricity any hour of the year with a .90 probability. Suppose that the event that one generation unit fails to operate is independent of the event that any other generation unit fails to operate. This independence assumption is

reasonable for dispatchable thermal generation units because unavailability is typically due to an event specific to that generation unit. If each generation unit has a nameplate capacity of 100 MW, each has a firm capacity of 90 MW. If there are 13 generation units, then with probability .96 demand peak will be covered.[3] In this case, a firm capacity requirement of 1.17 times the demand peak would ensure that system demand is met with .96 probability. Assuming that each generation unit is one-tenth of the system demand peak is unrealistic for most electricity supply industries, but it does illustrate the important point that smaller markets require firm capacity equal to a larger multiple of peak demand to achieve a given level of reliability.

Suppose that each generation unit is now 50 MW and each still has the same availability factor, so the firm capacity of each unit is now 45 MW. In this case, the same firm capacity requirement of 1.17 times the demand peak, or 26 generation units, would ensure system demand is met with .988 probability. If each generation unit had a nameplate capacity of 20 MW with the same availability factor, each unit would have a firm capacity of 18 MW. This 1.17 times peak demand firm capacity requirement, or 65 generation units, would ensure that system demand is met with .999 probability. This example illustrates that in an electricity supply industry based on dispatchable thermal generation units, where each unit has a 10% chance of being unavailable, the system demand peak will be met with a very high probability with a firm capacity requirement of 1.17 times peak demand if all the generation units are small relative to the system demand peak.

Introducing renewables into a capacity-based long-term RA mechanism considerably complicates the problem of computing the probability of meeting system demand peaks for two major reasons. First, the ability to produce electricity depends on the availability of the underlying renewable resource. A hydroelectric resource requires water behind the turbine, a wind resource requires wind to spin the turbine, and a solar facility requires sunlight to hit the solar panels. Second, and perhaps most important, the availability of water, wind, or sunshine to renewable generation resources is highly positively correlated across locations for a given technology within a given geographic region. This fact invalidates the assumption of independence of energy availability across locations that allows a firm capacity mechanism to ensure system demand peaks can be met with a very high probability. For example, if the correlation across locations in the availability of generation units is sufficiently high, then a 0.9 availability factor at one location would imply

only a slightly higher than a 0.9 availability factor for meeting system demand, almost regardless of the amount intermittent renewable capacity that is installed.

Hydroelectric facilities have been integrated into firm capacity regimes by using percentiles of the distribution of past hydrological conditions for that generation unit to determine its firm capacity value. However, this approach only partially addresses the problem of accounting for the high degree of contemporaneous correlation across locations in water availability in hydroelectric-dominated systems. There is typically a significant amount of data available on the marginal distribution of water availability at individual hydroelectric generation units. However, the joint distribution of water availability across all hydro locations is likely to be more difficult to obtain. The weather-dependent intermittency in energy availability for hydroelectric resources is typically on an annual frequency. There are low-water and high-water years depending on global weather patterns such as the El Niño and La Niña weather events, as discussed in McRae and Wolak (2016).

Incorporating wind or solar generation units into firm capacity mechanisms is extremely challenging for several reasons, and increasingly so as the share of energy produced in a region from these resources increases. The intermittency in energy supply is much more frequent than it is for hydroelectric energy. There can be substantial differences across and within days in the output of wind and solar generation units. Moreover, if stressed system conditions occur when it is dark, the firm capacity of a solar resource is zero. Similarly, if stressed system conditions occur when the wind is not blowing, a likely outcome on extremely hot days, the firm capacity of a wind resource is zero.

The contemporaneous correlation across locations in the output of solar or wind generation resources for a given geographic area is typically extremely high. There is even a high degree of correlation across locations in the output of wind and solar resources. Again, information on the marginal distribution of wind or solar energy availability at a location is much more readily available than the joint distribution of wind and solar energy availability for all wind and solar locations in a region. For these reasons, calculating a defensible estimate of the firm capacity of a wind or solar resource that is equivalent to the firm capacity of a dispatchable thermal generation resource is extremely difficult, if not impossible.

Wolak (2016) demonstrates the extremely high degree of contemporaneous correlation between the energy produced each hour of the year by solar and wind facilities in California. For each of the 13 solar locations

and 40 wind locations in the California ISO control area studied, Wolak (2016) computes the hourly capacity factor—hourly output of the generation unit divided by its nameplate capacity—from April 1, 2011, to March 31, 2012. Let $F_h = (f_{1h}, f_{2h}, \ldots, f_{Kh})'$ equal the vector of locational capacity factors for K renewable energy locations for hour h, where f_{ih} is capacity factor for hour h at location i. Each of the locational capacity factors, f_{ih}, is a random variable that takes on values in the interval [0, 1]. Let the contemporaneous covariance matrix F_h equal Σ, a positive definite $(K \times K)$ matrix, where the (i, i) element is the variance of f_{ih} and the (i, j) element is the covariance between f_{ih} and f_{jh}. Using the singular value decomposition of a positive definite matrix implies that $\Sigma = S\Lambda S'$, where Λ is a diagonal matrix composed of the eigenvalues of Σ and S is an orthogonal matrix ($S'S = I$, a $K \times K$ identity matrix) composed of the eigenvectors of Σ. It can be shown that the sum of the eigenvalues of Σ, which is the sum of the diagonal elements of Λ, is equal to the sum of diagonal elements of Σ, which is also equal to the sum of the variances of f_{ih} for $i = 1, 2, \ldots, K$. The ratio of the largest eigenvalue to the sum of the diagonal elements of Σ (sum of the variances of the f_{ih}) is a measure of the extent of contemporaneous correlation between the elements of F_h.

For the case of the 13 solar locations in California, the largest eigenvalue is equal to 80% of the sum of the variances of the f_{ih} for all 13 solar locations, which indicates a substantial degree of contemporaneous correlation in hourly values of f_{ih} across the K locations. For the 40 wind locations in California, the largest eigenvalue is more than 50% of the sum of the variances of all 40 locations. For the 53 wind and solar locations, the first eigenvalue is equal to slightly less than 50% of sum of the variances of all these 53 locations. For comparison, if the f_{ih} for $i = 1, 2, \ldots, K$ were independently distributed random variables, all with the same variance, these percentages would equal $100 \times (1/K)$, where K is number of locations.

The high degree of contemporaneous correlation across locations in hourly capacity factors requires a methodology for computing firm capacity that accounts for the joint distribution of hourly capacity factors across locations throughout the year. Not only does this methodology need to account for the contemporaneous correlation in capacity factors across locations, but also the high degree of correlation of capacity factors over time for the same locations and other locations. California currently uses an effective load carrying capacity (ELCC) methodology for computing the firm capacity values of wind and solar generation units. The ELCC methodology was introduced by Garver (1966) and it measures

the additional load that the system can supply from a specified increase in the MW of that generation technology with no net change in reliability. The loss of load probability, which is the probability that system demand will exceed the available supply, is the measure of reliability used in the ELCC calculation.

Consistent with the results of Wolak (2016), the ELCC values for solar generation resources in California have declined as the amount of solar generation capacity in the state has increased. For example, a recent study prepared for California's three investor-owned utilities (Carden, Dombrowsky, and Winkler 2020), Southern California Edison, Pacific Gas and Electric, and San Diego Gas and Electric, recommended ELCC values for a MW of fixed-mount solar photovoltaic capacity for 2022 of approximately 5% of the nameplate capacity. Their estimates for 2026 are less than half that amount, and those for 2030 are less than one-fourth that amount. These declines in ELCC values are due to the forecast increase in the amount of solar generation capacity in California.

To understand the computational and data compilation challenge associated with calculating the ELCC of a new generation unit, suppose there are K existing generation units in an electricity supply industry and C_i is the installed capacity at location i in MW, so that the MWh produced at location i in hour h is equal to $f_{ih} \times C_i$, where f_{ih} is the capacity factor at location i during hour h. Let L_h equal system load during hour h and $\lambda(F, L)$ equal the joint density of $(F'_h, L_h)'$, where F'_h is defined above. The loss of load probability during hour h is equal to $\text{LOLP}_h = E_\lambda(I[\Sigma_{i=1}^{K} f_{ih} \times C_i < L_h])$, where $I[A]$ is equal to 1 if the event is true and equal to zero otherwise and $E_\lambda(.)$ denotes the expectation taken with respect to $\lambda(F, L)$. The annual loss of load expectation is defined as $\text{LOLE} = \Sigma_{h=1}^{8760}\text{LOLP}_h$. The "1 day in 10 years" criteria for the LOLE implies 2.4 hours per year or 0.00027397.

Let the current value of the LOLE equal LOLE(Current). To define the ELCC of a Q MW of intermittent renewable capacity at a new location, let Δ equal the increase in load that can be served by this additional Q MW and still maintain LOLE(Current) and $\Lambda(F, L, f(\text{new}))$ the joint density of $(F', L, f(\text{new}))'$, where $f(\text{new})$ is the hourly capacity factor at the new generation unit location. The system load increment associated with the Q MW of new investment is the solution to the following equation in Δ,

$$\text{LOLE(Current)} = E_\Lambda\left(I\left[\sum_{i=1}^{K} f_{ih} \times C_i + Q \times f(\text{new})_h < L_h + \Delta\right]\right),$$

where $E_\Lambda(.)$ is the expectation with respect to $\Lambda(F, L, f(\text{new}))$. The ELCC of this Q MW unit is defined as $\text{ELCC} = \Delta/Q$. This expression illustrates

the massive data and computational requirements associated with computing the ELCC. First, an estimate of the joint density of $\Lambda(F, L, f(\text{new}))$ is required for every hour of the year. Second, a $K + 2$ dimensional integration must be performed 8,760 times to compute the right-hand-side expression. Third, this equation must be solved numerically for Δ. Consequently, all implementations of this process make simplifying assumptions that can have a substantial impact on the resulting ELCC value as shown in Kahn (2004).

An additional problem with computing the firm capacity of a solar or wind generation resource using the ELCC methodology is that the same Q MW investment is likely to be able to serve different increments to system demand depending on the location of the investment, the location of the increment to demand, and the size and location of other renewable resources in the region. This leaves the system operator with two difficult choices for setting the value of firm capacity for solar and wind resources. The first would be to set different values of firm capacity for resources based on their location in the transmission network. This would likely be a very politically contentious process because of the many assumptions that go into computing the ELCC of a resource. The second approach would set the same firm capacity value for all resources employing the same generation technology. This means that two resources with very different ELCC values could sell the same product, to the potential detriment of overall system reliability.

These facts, and the fact that what is predicted to be the major source of electricity in the future in California has been estimated to have a little firm capacity value, imply that it would be prudent for California to consider alternatives to its capacity-based long-term resource mechanism if it intends to meet its goals of obtaining 50% of the state's energy from renewable sources by 2025 and 60% by 2030 and increase the use of electricity in space heating and personal transportation.

IV. Experience with Long-Term RA Mechanisms

This section presents an analysis of the performance of the California and Texas markets during stressed system conditions. These states are also the two regions of the continental United States with the largest shares of intermittent renewables in their energy mix. The experience of the California market during August 2020 provides an example of the shortcomings of the existing capacity-based long-term RA mechanism described in the previous section. The experience of Texas demonstrates that

even a wholesale market with an extremely high offer cap still suffers from the reliability externality discussed in Section II.

A. California

Figure 1*A* plots the time series of in-state generation capacity in MW by technology in California between 2001 and 2020. Figure 1*B* plots the time series of in-state electricity generation in gigawatt-hours (GWh) for the same time period. California's renewables portfolio standard (RPS) was established in 2002 with the requirement that California obtain 20% retail electricity sales from renewable resources by 2017. Figure 1*A* shows that the major increase in renewable generation capacity did not begin until later in the decade, and most of that came in the form of wind generation units. The RPS requirement was accelerated to 33% by 2020, starting in 2013. This was followed by a significant increase in investments in solar photovoltaic (PV) capacity.

Between 2013 and 2019, California retired 2,254 MW of nuclear capacity at the San Onofre Nuclear Generating Station (SONGS). Over the same time period, natural gas generation capacity in California fell by 8,529 MW. Solar PV and solar thermal capacity increased by 8,471 MW and wind generation capacity increased by 188 MW over this same period. It is important to bear in mind that the SONGS facility typically ran at annual average capacity factor of more than 90%, whereas solar facilities in California had an annual average capacity factor in 2020 of 24.67% and wind facilities had a 24.09% annual average capacity factor in 2020.[4] Natural gas facilities typically have annual availability factors greater than 85% but currently run at a significantly lower annual average capacity factors because of the large amount of renewable generation capacity in the state. Consequently, replacing the 10,750 MW reduction in thermal generation capacity with 8,712 MW of intermittent wind and solar capacity significantly reduces the amount of firm capacity available to the California ISO.

An important factor in allowing the California ISO to meet demand with significantly less firm capacity is that California has more than 18,000 MW of transmission capacity between it and the rest of the Western Interconnection, also called the Western Electricity Coordinating Council (WECC), which contains all western US states and Canadian provinces.[5] Historically, California obtains between 25% and 33% of its annual consumption from electricity imports from hydroelectric units in the Pacific Northwest and coal-fired and natural gas-fired generation units in the desert Southwest.

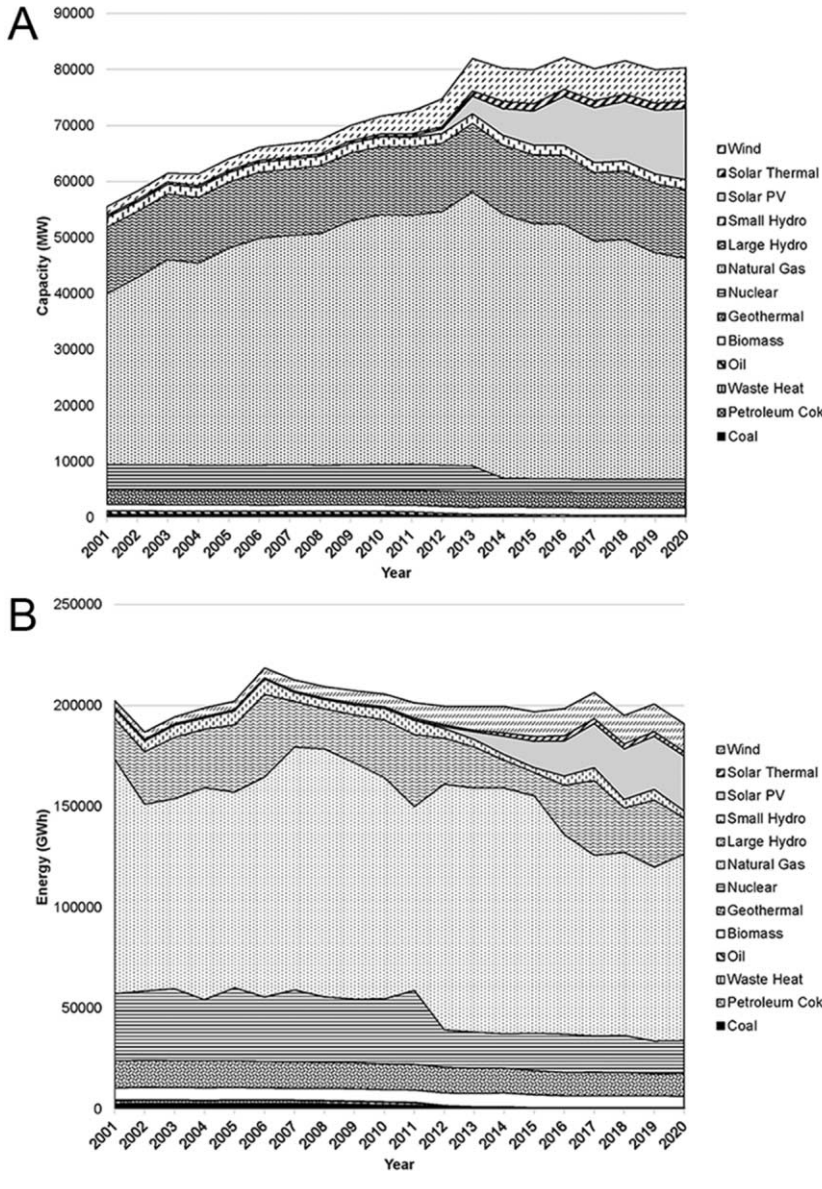

Fig. 1. *A*, In-state generation capacity in megawatts (MW) by fuel type 2001–2020. *B*, In-state generation in gigawatt-hours (GWh) by fuel type 2001–2020. Color version available as an online enhancement.

California's substantial import dependence is another strong argument against a capacity-based long-term RA mechanism. Kirchoff's laws governing the flow of electricity in transmission and distribution networks imply that electricity imports from neighboring states occur because these regions produce more electricity than they consume and California consumes more electricity that it is producing.[6] This requires system operators in neighboring states to ensure that the agreed-upon amount of excess generation in their states is produced, so that the agreed-upon imports will flow into California. Consequently, as a rule, California cannot purchase firm capacity from neighboring states. At best, California can purchase commitments from suppliers located outside the state that they will schedule specified quantities of energy imports into the state. Exactly which generation units located outside of California will provide this energy is largely unknown until real-time operation. It depends on many factors including the real-time output of all generation units in California and the rest of the WECC, the configuration of the transmission network in the WECC, and location and level of demand at all locations throughout the WECC.

Rolling Blackouts on August 14–15, 2020

In mid-August of 2020, California and neighboring states in the rest of the Western Interconnection experienced a sustained period of extremely hot weather. This led the California ISO to curtail firm load by declaring rolling blackouts during the late evening on August 14 and 15. The California ISO also came very close to having to curtail firm load during August 16–18. This section documents the failure of the state's firm capacity-based long-term RA mechanism to ensure sufficient energy to meet system demand during the portions of August 14 and 15 when the rolling blackouts occurred.

Figure 2*A* presents the 5-minute demand, 5-minute net demand (the difference between demand and wind and solar energy production), and the hour-ahead demand forecast for August 14, 2020. This net demand must be met by dispatchable generation resources in California or electricity imports. As shown in figure 1*A*, the vast majority of these dispatchable in-state resources are powered by natural gas. The rectangle between 18:00 and 19:00 denotes the time interval when the rolling blackout occurred. Figure 2*B* presents the same information for August 15, 2020, along with a rectangle denoting when rolling blackouts occurred. Figure 3*A* compares these demands to those on August 16–18. The dashed

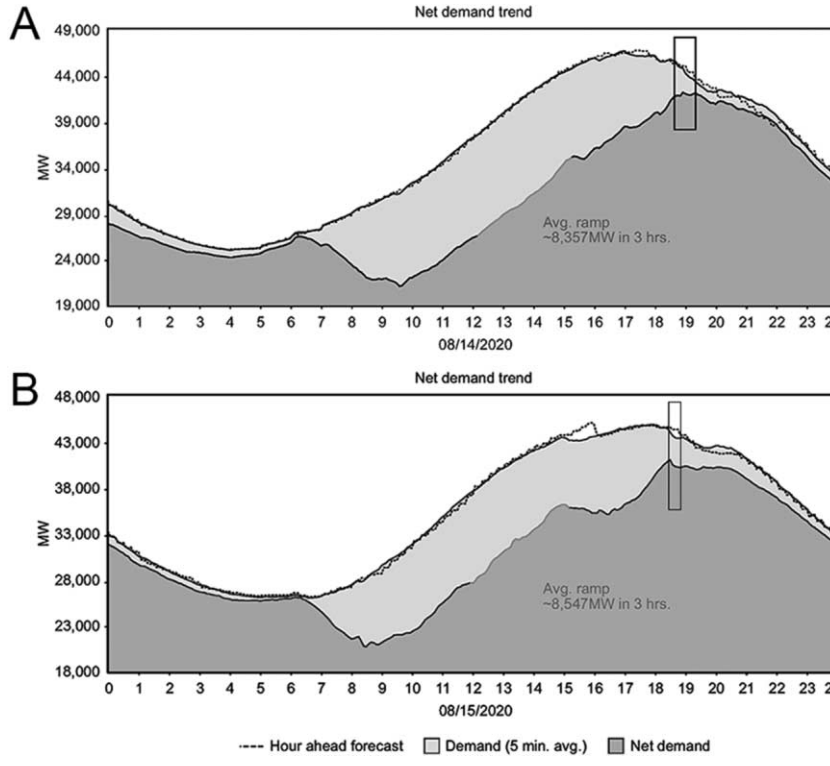

Fig. 2. *A*, System demand, net demand, and hour-ahead forecast demand in megawatts (MW) on August 14, 2020. *B*, System demand, net demand, and hour-ahead forecast demand on August 15, 2020. Color version available as an online enhancement.

line on the bottom of the graph plots the hourly demands on June 29, 2020, which is an ideal day for solar energy production, as shown in figure 3*B*. The hourly demands on August 18 were uniformly higher than the demands on August 14 and 15 and the demand on August 17 was higher than the demand on August 14, even though blackouts occurred on August 14 and 15.

The rolling blackouts on August 14 and 15 were necessitated by the fact that the net demand in California exceeded the amount of available dispatchable generation capacity in California and amount of electricity imports available during the evening hours. This outcome occurred for a variety of reasons. First, the demand for electricity in California was high because of the intensive use of air conditioning due to high in-state temperatures. Second, a reduced supply of intermittent renewable generation increased the net demand that must be met by dispatchable generation

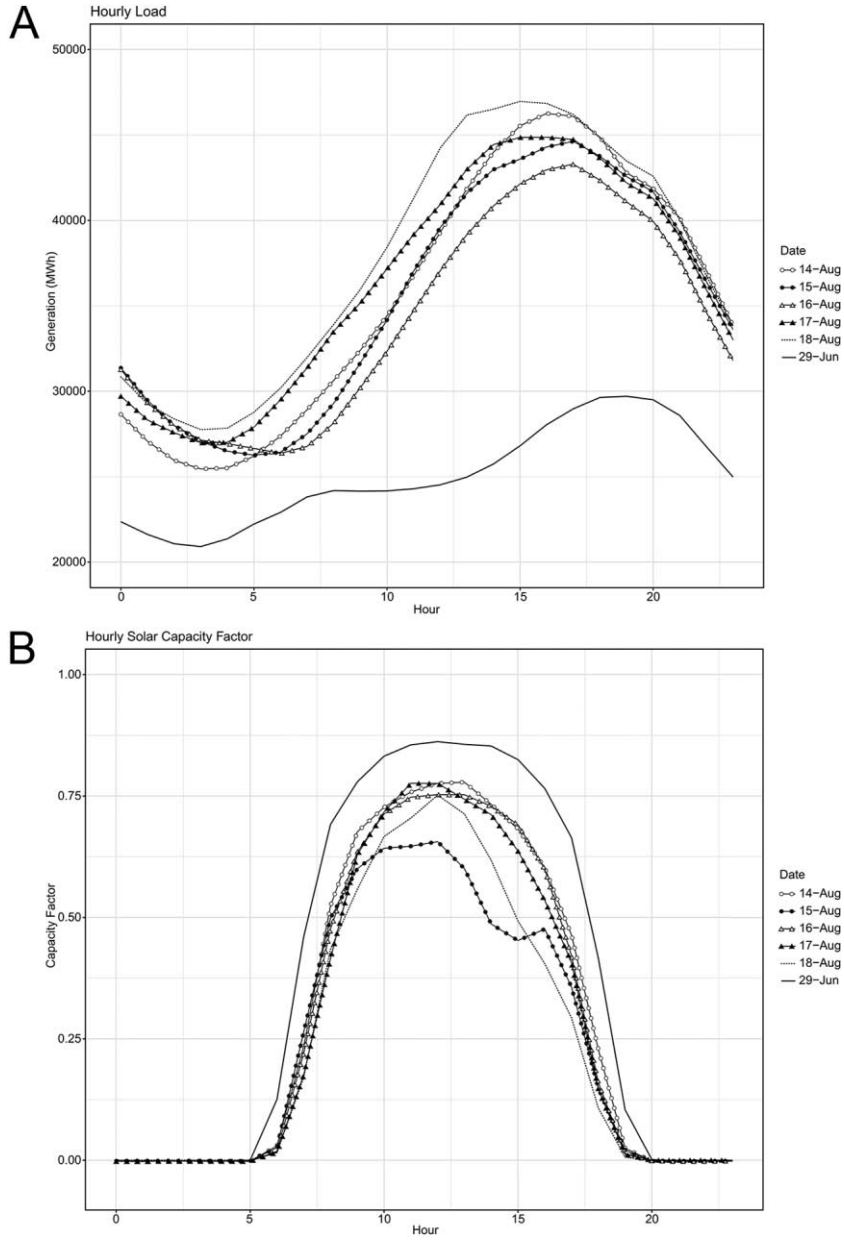

Fig. 3. *A*, Hourly system demands in megawatt-hours (MWh) August 14–18 and June 29, 2020. *B*, Hourly capacity factor of solar generation units August 14–18 and June 29, 2020. Color version available as an online enhancement.

units in California or electricity imports. Third, extreme temperatures in the rest of the western United States and Canada reduced the amount of electricity produced in the rest of the WECC that could be imported into California.

Figure 3*B* presents the first factor contributing to the events of August 14–18 by plotting the hourly capacity factor of solar generation units in California on these days. The dashed line is the hourly capacity factor for solar generation units in California for June 29, 2020. During much of the day on August 14 and 16–18, the hourly capacity factor of the solar generation units in California is lower than it was on June 29, 2020. This is particularly true for the evening hours when the rolling blackouts were declared.

Figure 4*A* provides one explanation for this outcome. It plots the hourly temperatures within the day in Barstow, California, for August 14–18 and for June 29, 2020. Barstow is located near a significant fraction of the solar generation capacity in California. Hourly temperatures during the day on August 14–18 were much higher than they were on June 29, which is close to an ideal day for electricity to be produced from a solar PV facility. Solar panels convert light into electricity, and this occurs with maximum efficiency at a panel temperature of 77°F. The efficiency of a solar panel declines linearly with every degree its temperature is above 77°F. The extremely high temperatures during the day on August 14–18 significantly reduced the efficiency at which the solar panels converted light into electricity. The solar panels were also likely to be significantly hotter later in the day than earlier in the day given the pattern of daily temperatures shown in figure 4*A*. Another contributing factor to the lower injections of electricity to the transmission grid from solar generation facilities during August 14–18 is the larger demand for electricity for on-site cooling on these days relative to June 29, 2020.

As shown in the next subsection, the firm capacity numbers assigned to solar generation units in California only vary by month, and do not depend on the outside temperature. However, as the share of solar energy increases in California, even a 5% reduction in solar output on high temperature days coupled with the likely increase in the demand for electricity for space cooling can lead to more days like August 14 and 15 in California.

Figure 4*B* plots the hourly capacity factors within the day for California's wind generation units for the same days as figure 4*A*. Consistent with the high temperatures throughout the state, the amount of wind energy produced was extremely low, particularly during the middle

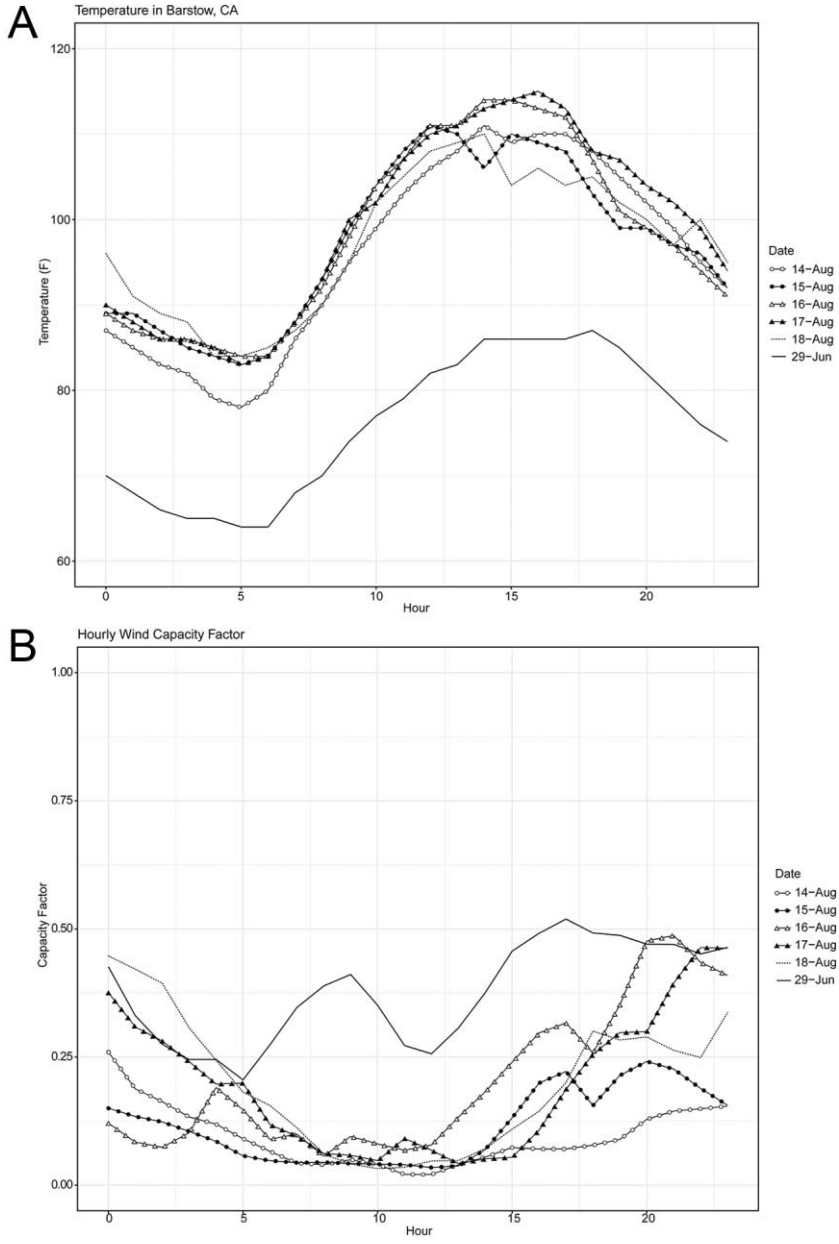

Fig. 4. *A*, Hourly temperature in Barstow, California, on August 14–18 and June 29, 2020. *B*, Hourly capacity factor of wind generation units on August 14–18 and June 29, 2020. Color version available as an online enhancement.

of the day, as well during the period of the rolling blackouts. This is consistent with the fact that wind blows because of temperature differentials between locations, and on extremely hot days in California and neighboring states temperatures are similar across locations. The hourly capacity factors on June 29, 2020, are significantly higher throughout the day, consistent with the milder temperatures throughout that day. The hourly capacity factors are significantly below the firm capacity values for August 2020 for wind generation capacity assigned by CPUC of 21% for the entire day on August 14 and for virtually all daylight hours for August 15–18.

To investigate the extent to which the various technologies used to produce electricity in California had statistically distinguishable lower or higher mean capacity factors during the extreme weather period of August 14–18 of 2020 than the remainder of the month of August, I ran the following regression for the hourly capacity factors for each technology for January 1, 2020, to December 31, 2020:

$$CF_{hdm} = \alpha_{hm} + \delta_d + I_{hdm}\beta + \varepsilon_{hdm}$$

where CF_{hdm} is the capacity factor in hour h of day d of month m, α_{hm} is an hour-of-day h for month m fixed effect, δ_d is the fixed effect for weekend days (Saturday and Sunday), I_{hdm} is an indicator variable that is equal to 1 if hour h of day d of month m is during the August 14–18 of 2020 period, and ε_{hdm} is a zero-mean disturbance.

Table 1 presents the 2020 annual mean hourly capacity factors for wind, solar, and natural gas generation units in California and the estimate of the change in the mean hourly capacity factor during the August 14–18 period for each technology. For the case of natural gas generation units, there was no statistically discernable change in the mean capacity factor during August 14–18, 2020. For the case of both solar and wind, the mean capacity factor was lower during the August 14–18 period relative to the remainder of the month of August. For solar it was 0.033 lower, which,

Table 1
Change in Mean Capacity Factor (CF) by Technology for August 14–18, 2020

	Solar_CF	Wind_CF	Natural Gas_CF
Sample Mean of Dependent Variable	.2467	.2409	.5946
β	–.0330	–.161	.000000649
Standard error	(.0140)	(.0336)	(.00000673)

Note: All regressions include hour-of-day fixed effects for each month of the year. Standard errors are clustered by day of sample.

when applied to an installed capacity of solar of close to 14,000 MW, implies an average hourly reduction in output of close to 500 MWh. For wind it was 0.161 lower, which, when applied to an installed capacity of close to 6,000 MW, implies an average hourly reduction in output of more than 900 MWh. This average shortfall in intermittent renewable output of 1,400 MWh (= 500 MWh + 900 MWh) is significantly larger than the amount of load that was curtailed during each of the rolling blackout events on August 14 and 15.

Given the similarities between hourly system demands on August 14–18 and the output of renewables on these days, an obvious question is why rolling blackouts occurred on August 14 and 15 but not on August 16–18. Figure 5 provides an answer to this question. Figure 5*A* plots the hourly net imports (imports minus exports) scheduled in the day-ahead market. These are commitments that market participants make to import energy into California the day before the energy actually flows. The day-ahead imports during the late afternoon and early evening are very low during August 14–17 relative to the day-ahead imports on June 29, 2020. This outcome is consistent with temperatures in neighboring states in the Western Interconnection being extremely high on these 5 days in August and relatively mild on June 29, 2020. This means the opportunity cost of scheduling an import into California, typically the highest priced region in the WECC, was extremely low on June 29, 2020. However, there were lucrative opportunities for selling electricity outside of California on August 14–18, because of the extremely high temperatures and high demand outside the state.

The high net imports scheduled in the day-ahead market on August 18 shown in figure 5*A* hint at what ultimately led to rolling blackouts on August 14 and 15, but not during the period August 16–18. Figure 5*B* presents the hourly real-time net imports into California for the same set of days as figure 5*A*. The real-time net imports on August 16–18 are uniformly higher by substantial margins during the late afternoon and early evening than the same magnitudes on August 14 and 15. The real-time net imports on August 14 are also significantly lower than those on August 15. After the events of August 14 and 15, the California ISO operators and entities throughout the Western Interconnection significantly increased the supply of imports willing to sell into the California market in real-time.

A final point about this 5-day period in August is particularly important to emphasize. That is the impossibility of preventing sellers of electricity from finding the highest possible price for their electricity. There

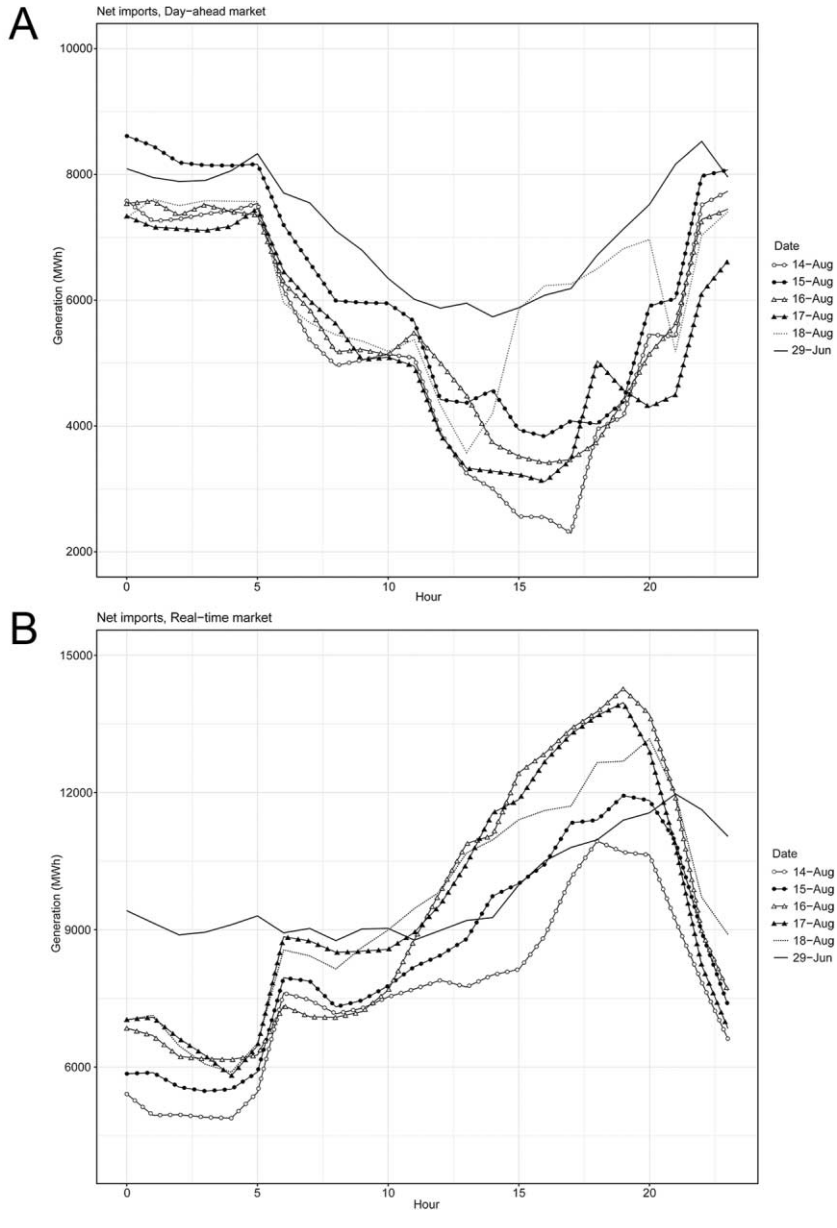

Fig. 5. *A*, Hourly day-ahead imports in megawatt-hours (MWh) on August 14–18 and June 29, 2020. *B*, Hourly real-time imports on August 14–18 and June 29, 2020. Color version available as an online enhancement.

is evidence that during the August 14–18 period, suppliers committed to sell energy to California in the day-ahead market under the state's long-term RA mechanism did so, but other market participants found more attractive options for this energy and bought it for export and sold it in neighboring states at a higher price. California had a $1,000/MWh cap on offer prices at this time, whereas there was a higher cap on prices outside of the state. This fact illustrates another shortcoming of a capacity-based long-term RA mechanism for California. If California purchases a commitment for sellers outside the state to supply imports to California and prices outside the state are higher than California's offer cap, market participants can purchase this energy at or below the state's offer cap and sell it outside the state at a higher price.

One response of California to this set of circumstances would be to suspend exports of electricity from the state. This market intervention would discourage suppliers from selling energy into California in the day-ahead market, because they know they are foregoing the option to sell at a higher price outside of the state if they do. This fact illustrates what I like to call the "tyranny of electricity imports," because if California wants to attract imports to the state it must be willing to pay a higher price than neighboring control areas or violate the integrity of its market mechanisms. For this reason, suspending exports is likely to have adverse long-term energy-supply consequences for an import-dependent region like California.

The Performance of California's Capacity-Based Long-Term RA Mechanism

This section evaluates the performance of California's capacity-based long-term RA mechanism based on the experience of August 14–18, 2020. Figure 6*A* plots the monthly average wind capacity factors (CF) for 2020 and the monthly values of the firm capacity (ELCC) for wind units set by the CPUC. Figure 6*B* plots these same to magnitudes for solar generation units.

Except for May for wind and July for solar, the monthly values of firm capacity are slightly below the average capacity factors for the month. However, it is important to bear in mind that the firm capacity of a generation unit is supposed to measure what the facility can reliably produce under extreme system conditions, not what it produces on average. Consequently, a monthly average capacity factor less than the firm capacity value assigned to wind or solar generation resources provides

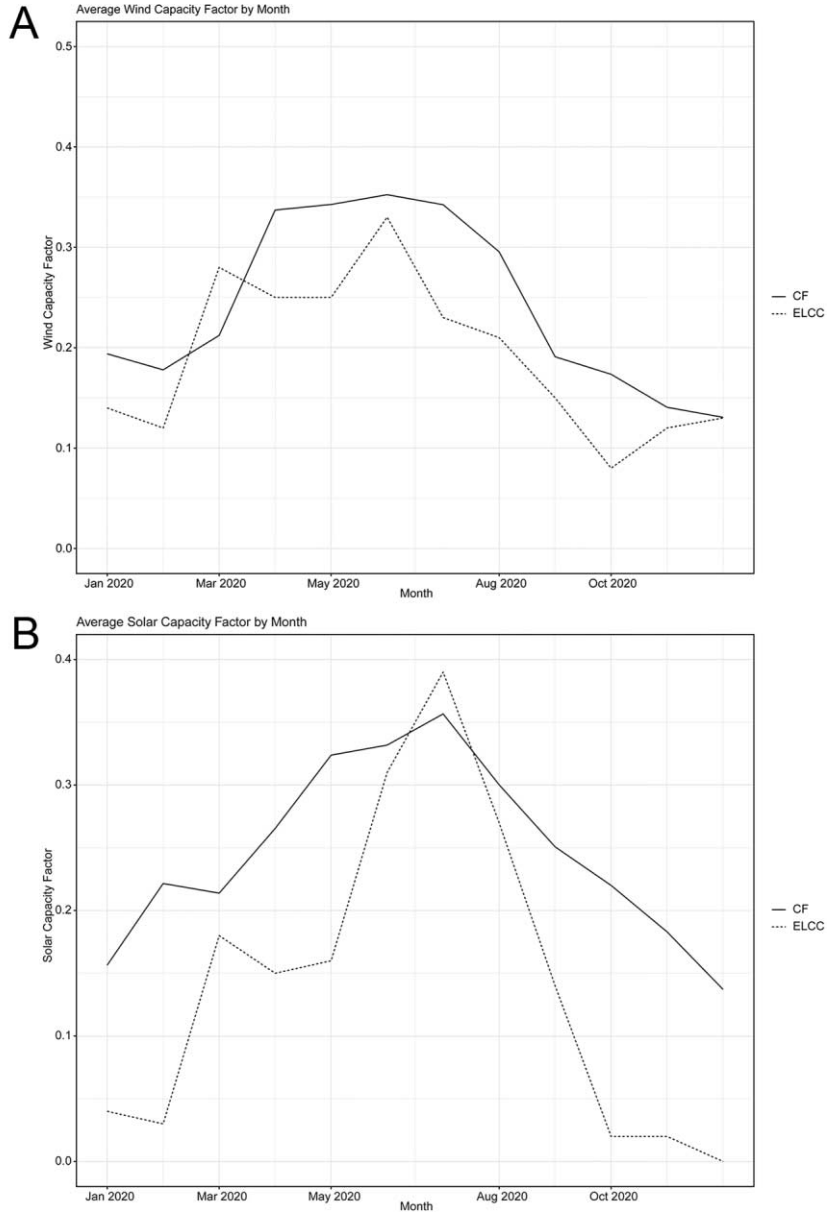

Fig. 6. *A*, Monthly average wind capacity factor (CF) and effective load carrying capacity (ELCC) value for firm capacity for 2020. *B*, Monthly average solar capacity factor and ELCC value for firm capacity for 2020. Color version available as an online enhancement.

further evidence against the viability of a capacity-based long-term RA mechanism with a large share of intermittent renewables. This outcome implies there are many hours in the month when the intermittent wind or solar resource is producing less than its firm capacity. Given the unpredictable intermittent nature of these resources, there is a nonzero probability this outcome will occur during a time with stressed system conditions, similar to those in August 2020.

To understand better the shortcomings of a capacity-based approach to long-term RA, figures 7 and 8 break down the information behind figure 6 into hourly within-day histograms of capacity factors of wind and solar generation units by month for 2020. Each monthly graph provides box and whiskers plots of the daily distribution of capacity factors for that hour of the day. The black bar for each box is the median capacity factor, the top and bottom of the box are the 75th and 25th percentiles, and top/bottom lines are 1.5 times the interquartile range from the 75th/25th percentile. Dots are values that are more than 1.5 times the interquartile range from the 75th/25th percentiles. The horizontal line on each graph is the monthly value of the firm capacity value for that month of 2020 from figure 6.

For all months of 2020, there are days when the firm capacity value for the month for solar generation units exceeds an hourly capacity factor. This outcome is particularly likely during the March to September time period. During the early daylight hours and late evening hours of these months, there are many days when there are hourly capacity factor realizations that are less than the firm capacity value assigned to solar units for that month. As shown in figure 3*B*, on all the days during August 14–18, 2020, the early morning hours and early evening hours had solar capacity factors less than the firm capacity value for solar units for August 2020 of 0.27. As shown in figure 2, rolling blackouts were declared during the early evening hours of August 14 and 15.

The situation for wind units is even worse. There are many months when the median capacity factor for an hour of the day is below the firm capacity value for the month for a substantial number of hours of the day. During August 2020, it was not unusual to have hourly capacity factors during the early evening that were below the monthly value of firm capacity of 0.21.

It is important to emphasize that the capacity factors plotted in figures 6–8 are on a fleet-wide basis. The hourly capacity factor values for specific generation sites are likely to be even more volatile. Moreover, for the reasons discussed in Subsection IV.A, there are likely to be significant

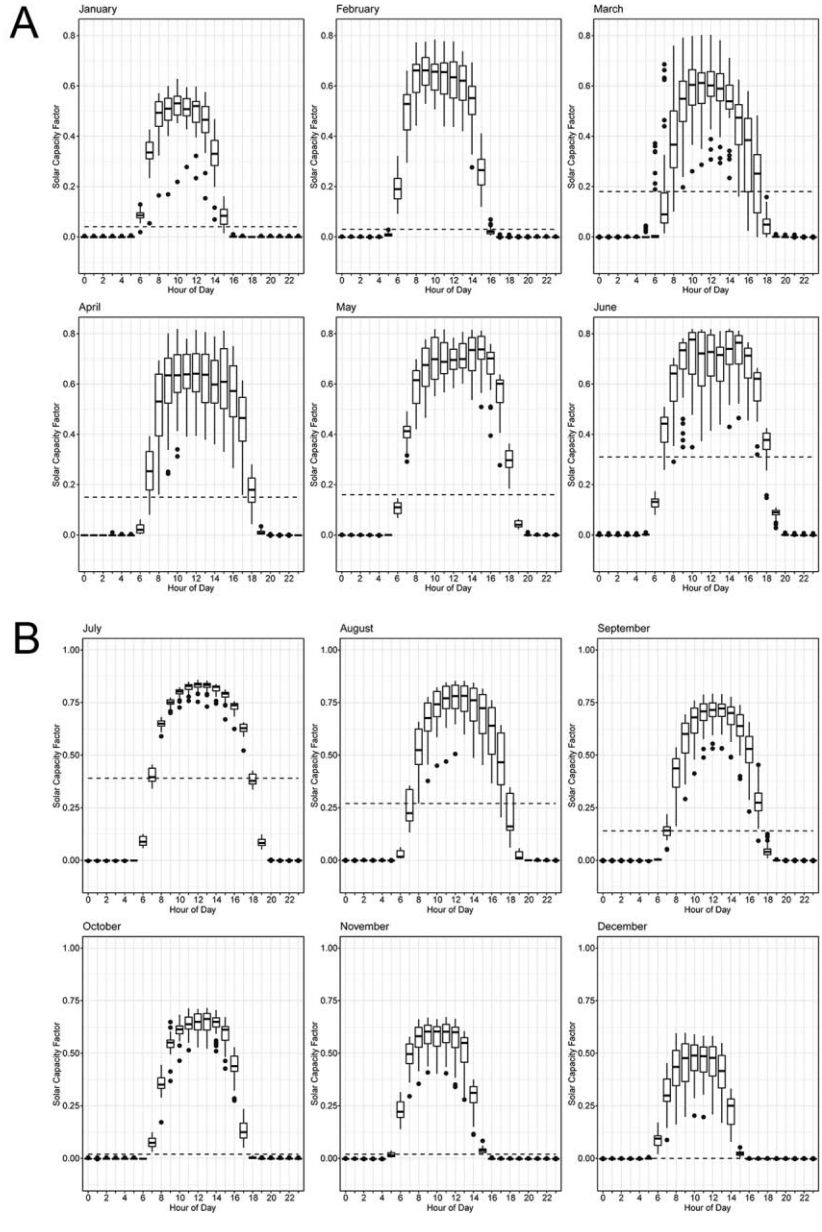

Fig. 7. A. Histograms of hourly solar capacity factors and monthly effective load carrying capacity (ELCC) value for firm capacity for 2020 (January–June). *B*, Histograms of hourly solar capacity factors and monthly ELCC value for firm capacity for 2020 (July–December).

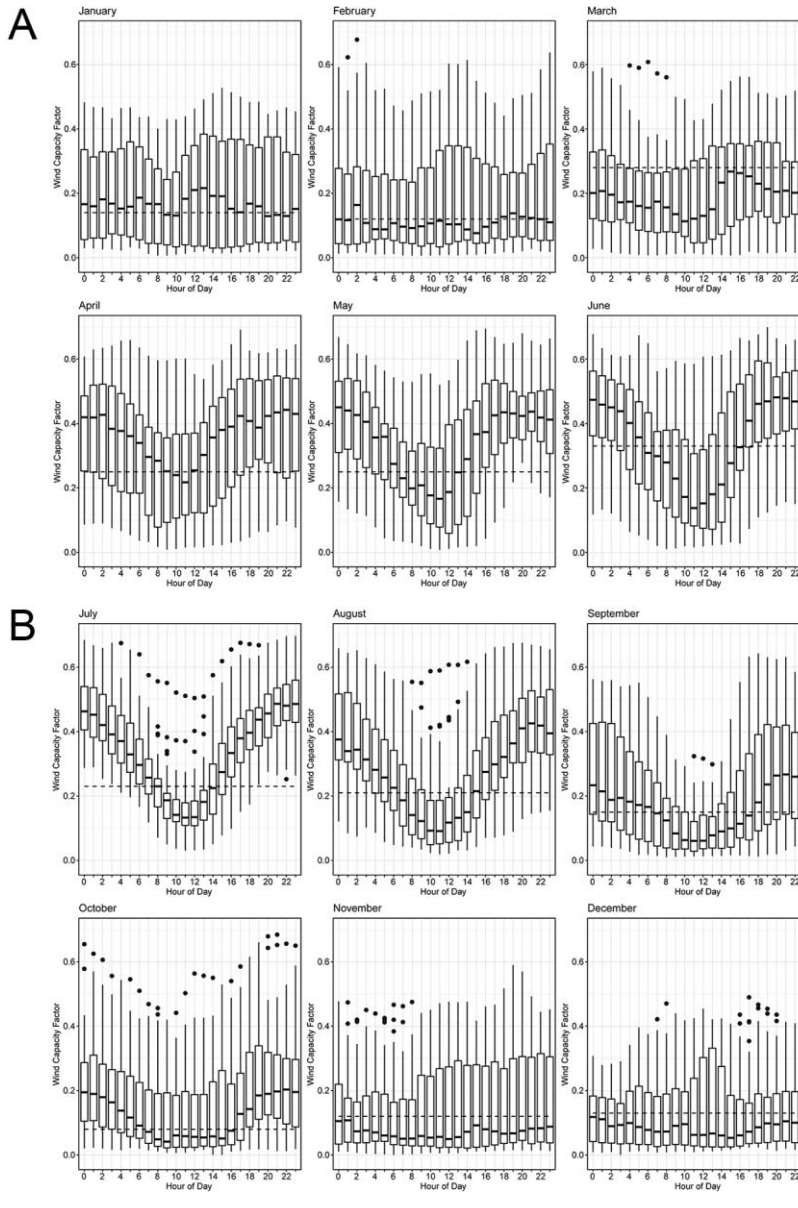

Fig. 8. *A*, Histograms of hourly wind capacity factors and monthly effective load carrying capacity (ELCC) value for firm capacity for 2020 (January–June). *B*, Histograms of hourly wind capacity factors and monthly ELCC value for firm capacity for 2020 (July–December).

differences in the distributions of hourly capacity factors across locations, even though the CPUC assigns all facilities of the same generation technology the same firm capacity factor value that changes each month of the year.[7]

These results suggest that events like August 14–15 are increasingly likely to occur under a capacity-based long-term RA mechanism in California with an increasing amount of intermittent wind and solar generation capacity. The state will increasingly need to rely on imports from neighboring states from dispatchable thermal generation resources when the net demand for electricity in California is high. Unless California builds additional controllable generation resources or makes substantial investments in energy storage, the state will be increasingly reliant on imported energy under these system conditions. These imports are also likely to be significantly more carbon intensive than electricity produced inside the state.

Figure 9 plots the mix of generation capacity in the WECC excluding California. Any available energy from the hydroelectric capacity shown in the figure will be used each year regardless of California's demand for electricity, because of its very low variable cost of production. Consequently, any marginal increase in electricity imports to California is likely to come from either natural gas-fired or coal-fired generation. This means that incremental imports will typically be more carbon intensive than electricity produced from natural gas-fired units in California, because California does not have any significant coal-fired generation capacity.

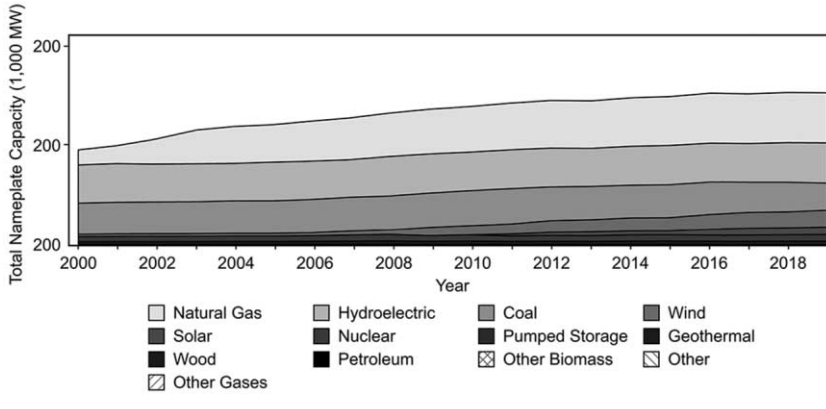

Fig. 9. Installed capacity in megawatts (MW) by technology in Western Electricity Coordinating Council (WECC) excluding California 2000–2019. Color version available as an online enhancement.

B. Texas

To illustrate the existence and consequences of the reliability externality even in a market with an extremely high offer cap, this section analyzes the performance of the ERCOT market during two periods with extreme cold weather in the state. A major difference between these two weather events was the mix of generation capacity available to meet demand during these two periods. The first period is February 1–5, 2011, and the second is February 14–18, 2021.

The significant increase in the share of energy supplied by intermittent wind and solar resources in February 2021 relative to February 2011 appears to be a major factor in explaining the difference in the performance of the ERCOT market across these two time periods. However, the more extreme weather during 2021 versus 2011 and the larger share of home heating supplied by electricity in Texas in 2021 versus 2011 cannot be ruled out as factors. As Doss-Gollin et al. (2021) note, the weather during the 2011 period was not nearly as severe at the weather during 2021 period. However, these authors also argue that the 2021 period was not as severe as a weather period of a similar length that occurred in December 1989. At that time Texas had very little wind generation and a significantly smaller fraction of households heated with electricity.

Figure 10*A* plots generation capacity in MW by fuel type in ERCOT from 2010 to 2020. Figure 10*B* plots the annual generation in terawatt-hours (TWh) by fuel type in ERCOT over this same time period. Three trends are immediately apparent. First, the installed capacity of wind generation units increased by 15,477 MW and the amount of solar generation capacity increased by 2,478 MW. Second, coal-fired generation capacity has declined by 4,619 MW and the production of coal-fired electricity declined even faster. Finally, the amount of natural gas-fired generation capacity increased by 3,356 MW and amount of natural gas-fired generation increased at a slightly lower rate over this period.

Two other facts about the Texas market help explain the severity of these two supply shortfall periods. First, legally speaking ERCOT is not electrically interconnected with the rest of the United States.[8] This means that unlike California, it is unable to rely on significant amounts of electricity imports from neighboring states when there are supply shortfalls or demand spikes in ERCOT. Second, according to the US Census Bureau, currently 61% of Texas housing units rely on electricity for heating, compared with 39.5% nationally.[9] This makes the electricity demand in Texas extremely sensitive to extreme cold-weather events.

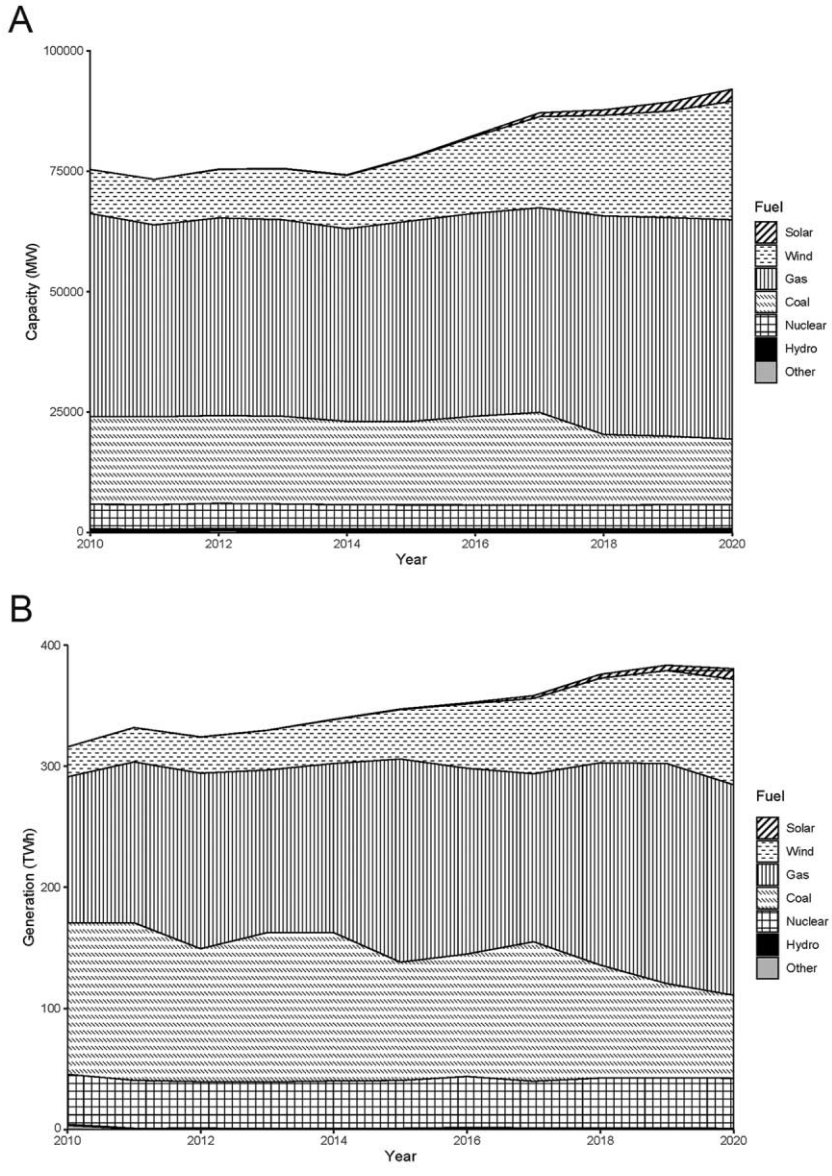

Fig. 10. *A*, Installed capacity in megawatts (MW) by technology in Electric Reliability Council of Texas (ERCOT) 2010–2020. *B*, Annual generation in terawatt-hours (TWh) by technology in ERCOT 2010–2020. Color version available as an online enhancement.

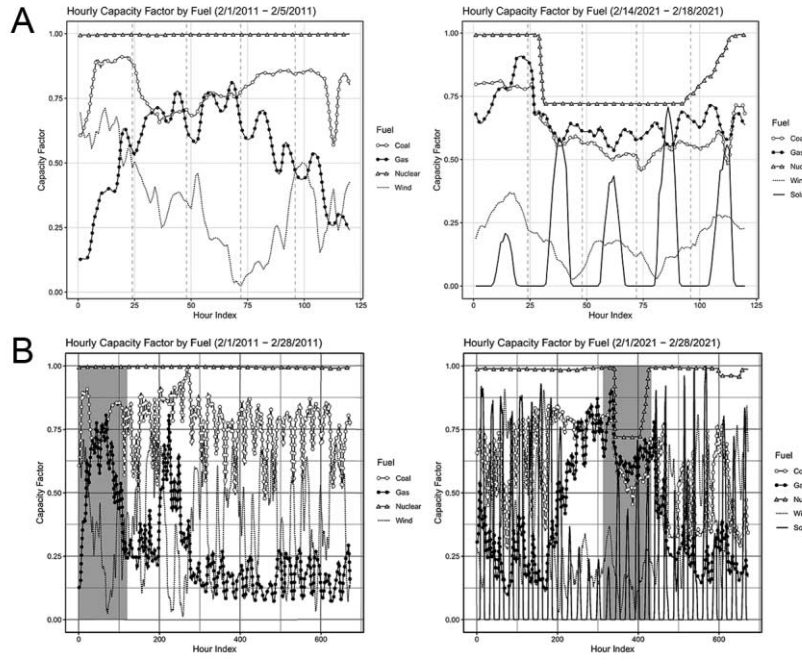

Fig. 11. *A*, Hourly capacity factors by technology in Electric Reliability Council of Texas (ERCOT) for selected 5-day periods in February, 2011–2020. *B*, Hourly capacity factors by technology in ERCOT for February, 2011–2020. Color version available as an online enhancement.

February 1–4, 2011, versus February 14–18, 2021

Figure 11*A* plots the hourly capacity factors of coal-fired, natural gas-fired, nuclear, wind, and solar generation units for the two extreme weather periods—February 1–4, 2011, and February 14–18, 2021. Because there were no solar generation units in 2011, this technology is omitted from the February 1–4 graph. Although there is significant variation in the hourly capacity factors during these two periods, two differences between them immediately stand out. First, the average capacity factor of wind generation units is significantly less during the February 2021 period relative to the February 2011 period. Recognizing that wind generation capacity increased by 15,472 MW between 2011 and 2021 implies a significant shortfall in renewable energy production throughout February 14–18, 2021. Second, there is a significant drop in the nuclear capacity factor during the second day of the 2021 period, whereas the nuclear capacity factor remains constant during the 2011 time period.

Figure 11*B* plots hourly capacity factors for the same technologies for the entire month of February 2011 and 2021 with the two extreme weather periods are shaded in the figure. These graphs demonstrate that the low capacity factors for wind generation units during February 14–18, 2021, were significantly lower than other hours during February 2021, whereas the average capacity factors of wind units during February 1–4 of 2011 was not different from that for remainder of February 2011. For the case of nuclear power, the average capacity factor during the period February 14–18 is significantly less than the mean capacity factor for remaining hours of the month. Finally, for solar units the average capacity is lower during February 14–18 of 2021 relative to the remainder of the month.

To investigate which technologies had statistically distinguishable lower or higher mean capacity factors during the extreme weather periods of February 2011 and February 2021, relative to the remainder of the month of February, I ran the following regression for the hourly capacity factor for each technology for the periods March 1, 2010, to February 28, 2011, and March 1, 2020, to February 28, 2021:

$$CF_{hdm} = \alpha_{hm} + \delta_d + I_{hdm}\beta + \varepsilon_{hdm}$$

where CF_{hdm} is the capacity factor in hour h of day d of month m, α_{hm} is an hour-of-day h for month m fixed effect, δ_d is the fixed effect for weekend days (Saturday and Sunday), I_{hdm} is an indicator variable that is equal to 1 if hour h of day d of month m is during the February time period of 2011 or 2021, and ε_{hdm} is a zero-mean disturbance.

Table 2 presents the estimates of the coefficient associated with I_{hdm} for each technology and each February period. The annual mean capacity factor for each sample period for each technology is also included in

Table 2
Estimated Change in Mean Hourly Capacity Factor (CF) by Technology during February 1–4, 2011, and February 14–18, 2021, Weather Events

Technology	2011 Mean CF	2011 Coefficient	2011 Std. Error	2021 Mean CF	2021 Coefficient	2021 Std. Error
Coal	.7793	.0189	.0258	.5993	.0167	.0572
Natural gas	.3155	.3159	.0638	.4056	.3061	.0521
Wind	.3198	–.0454	.0734	.3996	–.2236	.0443
Nuclear	.9214	.0005	.0026	.9152	–.1641	.0476
Solar	N/A	N/A	N/A	.2117	–.0763	.0283

Note: All regressions include hour-of-day fixed effects for each month of the year. Standard errors are clustered by day of sample.

the table. For coal-fired generation units there is a slight, but not statistically different from zero, increase in the mean hourly capacity factor during the extreme weather periods in February 2011 and 2021 versus other hours in the month. For natural gas units, there is a precisely estimated substantial increase in the mean hourly capacity factor during both February extreme weather periods. In both 2011 and 2021, the mean hourly capacity factor of natural gas units increased by more than 0.30 during these extreme weather periods.

For the wind generation units, the capacity factor is 0.2236 less during the extreme weather period in 2021 than in other hours of February. Because there is 24,593 MW of wind generation in ERCOT in 2020, this reduction in the average capacity factor implies an average MWh reduction of wind energy during the February 2021 extreme weather period of 5,410 MWh. The nuclear capacity factor fell by 0.1641 during the extreme weather period, which for an installed capacity of 4,973 MW implies an average hourly reduction in nuclear generation of 795 MWh. The solar energy capacity factor fell by 0.0763, which for an installed capacity of 2,478 MWs implies an average hourly reduction in solar energy of 173 MWh during the extreme weather period. The total of these average hourly supply shortfalls during February 14–18, 2021, was 6,400 MWh, with the vast majority coming from intermittent renewable resources.

These results emphasize the substantial risk of relying of intermittent renewable energy units to produce during extreme cold weather periods, even relative to system conditions that typically exist during the winter months. As the graphs for 2021 in figures 11*A* and 11*B* demonstrate, the average hourly reduction in wind energy production of 5,410 MWh implies significantly larger reductions for a number of hours during February 14–18. As shown in both figures, hourly capacity factors very close to zero occurred at least twice during this time period.

ERCOT and the Reliability Externality

Although the historical peak demand of 72,820 MWh in ERCOT occurred on August 12, 2019, during the 4 p.m. to 5 p.m. hour, demand during the February 14–18 period was expected to exceed that demand but did not because rolling blackouts were implemented.[10] From analysis of the previous subsection, it seems reasonable to expect that a similar supply shortfall could occur during future extreme weather events as Texas increases the share of wind and solar resources in the state.

These events demonstrate that having a $9,000/MWh offer cap on the short-term market does not eliminate the reliability externality; it only reduces the frequency of supply shortfall events. The implicit assumption of the ERCOT market that the supply of energy would always exceed demand at a price at or below $9,000/MWh turned out to be false for the weather conditions experienced during February 14–18, 2021. The large share of housing units heated with electricity makes the demand for electricity in Texas extremely sensitive to extremely cold temperatures because these households must increase their demand for electricity to keep warm.

Consistent with the logic of the reliability externality, there were many households that paid for their wholesale electricity according to the hourly short-term price. This decision clearly makes economic sense in vast majority of hours of the year because short-term wholesale prices typical reflect substantial amounts of wind and solar energy production. One company, Griddy, was well known for selling retail electricity in this manner. Early during the extreme weather event, Griddy told all its customers to switch retail suppliers.[11] Of those that did not switch, many were unable to pay their bills as a result of purchasing much of their wholesale electricity during this time period at $9,000/MWh ($9/KWh). Consequently, ERCOT removed Griddy's right to operate effective February 21, 2021.

There were also several retailers that failed to fully hedge the partial or fully fixed-price retail contracts they sold to customers. These retailers had to purchase energy at $9,000/MWh and sell it at a fixed price to these retail customers. There was at least one retailer offering customers a $100 credit off their final bill and waiving all early termination fees if they switched providers before February 15, 2021.[12] This would enable the retailer to avoid having to purchase wholesale energy at a loss and sell it to its customers or avoid the likelihood that their customers would be unable to pay their bills, two outcomes with adverse financial consequences for the retailer.

Given the substantial volatility in wind and solar energy production in ERCOT, the state's dependence on electricity for space heating, and the fact that Texas cannot rely on large amounts of net imports from neighboring states when renewable energy shortfalls occur, the events of February 2021 are not unexpected. Figures 12 and 13 repeat figures 7 and 8 for the case of ERCOT for the period March 2020 through February 2021. Each monthly graph gives the same box and whiskers graph of the histogram of hourly capacity factors within that month. Because ERCOT

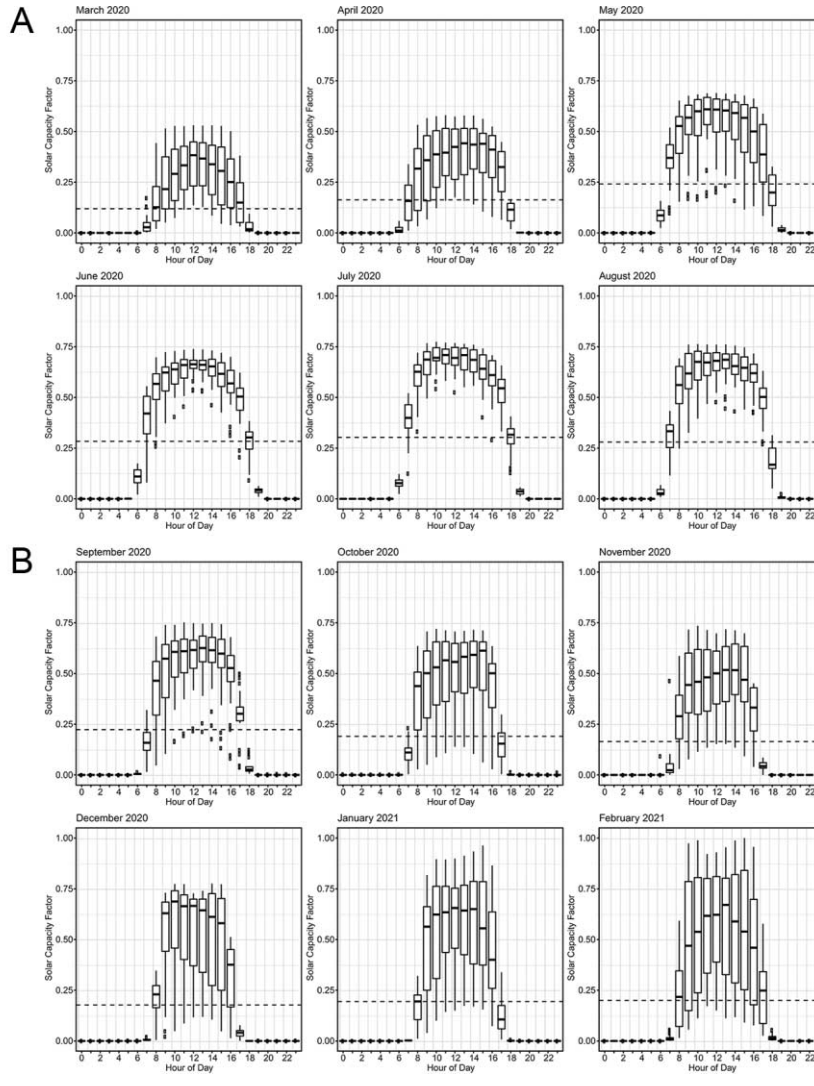

Fig. 12. *A*, Histograms of hourly solar capacity factors and monthly mean capacity factor in Electric Reliability Council of Texas (ERCOT) for March 2020–August 2020. *B*, Histograms of hourly solar capacity factors and monthly mean capacity factor in ERCOT for September 2020–February 2021.

does not have a firm capacity construct, the horizontal line on each graph is the monthly mean capacity factor for that technology. Figure 12 shows that the monthly mean capacity factor for solar resources is generally larger than the median hourly capacity factor for the early and late daylight hours of the day for many months of the year.

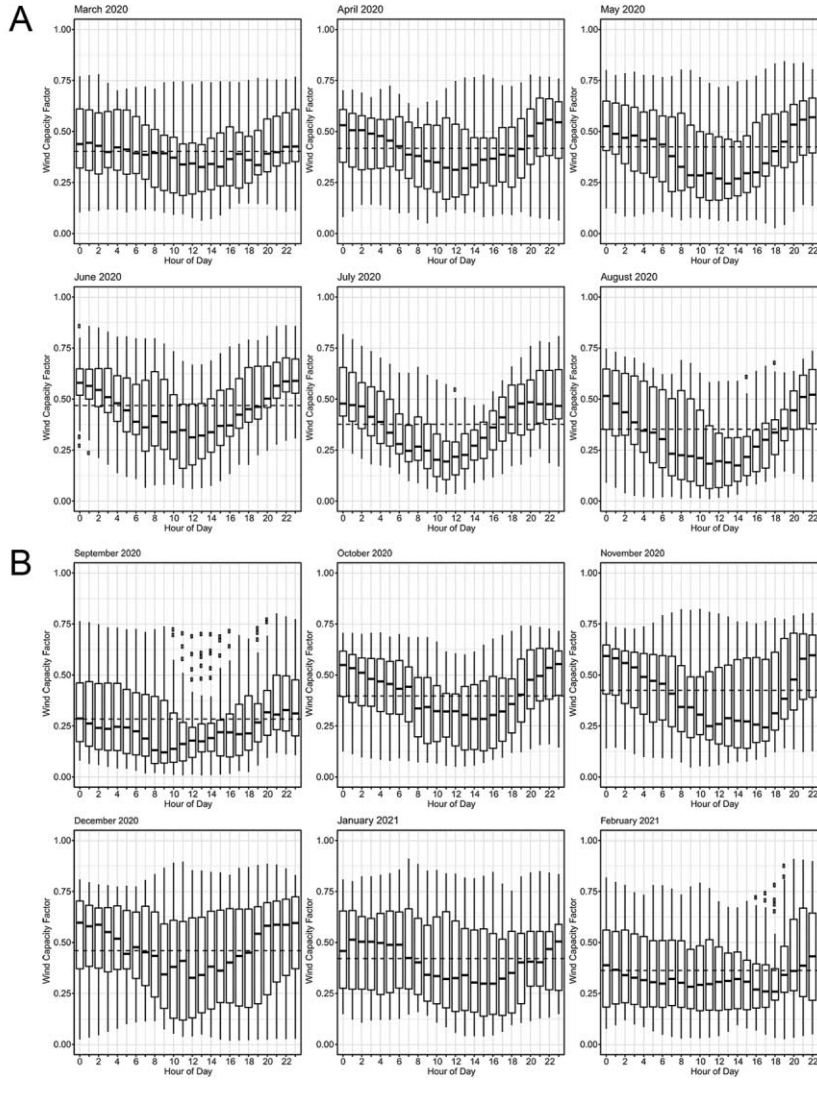

Fig. 13. *A*, Histograms of hourly wind capacity factors and monthly mean capacity factor in Electric Reliability Council of Texas (ERCOT) for March 2020–August 2020. *B*, Histograms of hourly wind capacity factors and monthly mean capacity factor in ERCOT for September 2020–February 2021.

From April to November 2020, the monthly mean capacity factor for wind units in ERCOT is above the median capacity factor for most hours in the middle of the day. During the December to March time period, the median hourly capacity factor is relatively constant across hours of the

day. All of these graphs show that there are many hours of the day during all months when extremely low capacity factors for wind occur. These extremely low capacity factors can occur during the summer months as well as the winter months.

Because ERCOT is not interconnected with the rest of the United States grid, this implies that the region will need to invest in significant storage capacity or increase the amount of natural gas-fired generation capacity to meet the demand for energy during these time periods. These thermal generation units will run at smaller capacity factors as the share of wind and solar energy increases. It is unclear whether the necessary storage units or thermal generation units will be built and remain financially viable without a long-term RA mechanism in ERCOT.

The Natural Gas Market in Texas

The severity of the electricity supply shortfall during February 15–18, 2021, was exacerbated by events in natural gas markets caused by the extremely cold weather and curtailment of electricity to natural gas processing facilities, natural gas pumping stations, and compressors on natural gas pipelines and at power plants.[13] Because much of the natural gas currently produced in Texas is associated with the production of oil, it comes to the surface at low pressure along with other liquids, including water. This has two implications for the level of natural gas production during cold weather conditions. First, low temperatures can lead to well freeze-offs, where the raw natural gas stream freezes at the wellhead or in the natural gas gathering lines, which ultimately stops production of natural gas. Second, to pump the oil and associated natural gas to the surface, process it, and compress it for transportation to natural gas generation units requires electricity. Many of the natural gas processing facilities in Texas failed to apply for outage exemptions as essential facilities before the February 2021 extreme weather event, and as a result were curtailed in the rolling blackouts.[14]

It is difficult to determine precisely what fraction of the almost 50% decline in natural gas production during the February 2021 cold snap was due to freeze-offs versus electricity outages to natural gas processing facilities.[15] A significant amount of natural gas-fired generation capacity was unable to operate during the cold snap because natural gas could not be delivered to these units.

Requiring all natural gas processing and transportation facilities to be classified as critical infrastructure that is protected from power outages

due to supply shortfalls should eliminate this cause of natural gas shortfalls. One natural gas company found that all its natural gas facilities that had been registered as critical infrastructure did not have their power turned off during the February 2021 cold snap.[16]

A more controversial recommendation made by many observers is to winterize natural gas wells to prevent freeze-offs during cold weather conditions. This recommendation was made following the 2011 cold snap but was not implemented for the same reason it is unlikely to be implemented after 2021 cold snap without an explicit regulatory mandate. Only around 0.06% of total annual natural gas production in Texas freezes off on average, and winterizing can cost as much as $100,000 per well.[17] These two figures suggest that it was and continues to be expected profit-maximizing for Texas natural gas producers not to winterize. In addition, if the political process decides to mandate winterizing all wells, this would likely cause producers to abandon many wells rather than pay the cost to winterize them, which would ultimately reduce the supply of natural gas in Texas.

C. The Need for Long-Term Storage with Significant Renewables

This section identifies an important characteristic of electricity supply industries with significant intermittent renewable generation capacity that provides further evidence against a capacity-based long-term RA mechanism. This is the potential for long durations of low levels of renewable output, particularly in regions where a significant amount of the renewable energy comes from wind generation units, as is the case in Texas.

Table 3 presents summary statistics on the annual hourly distribution of wind, solar, and combined wind and solar output for California from 2013 to 2020. Although the mean hourly output for wind and solar generation increases across the years, so does the standard deviation of hourly output. For the case of combined wind and solar generation, the standard deviation of hourly output has increased more rapidly than the hourly mean output, as evidenced by the upward trend in the coefficient of variation (CV) across the years.[18]

This increased variability in wind and solar output has characteristics that make significant investments in storage capacity necessary if the share of renewables is increased significantly beyond current levels. There can be long durations of relatively low levels of energy production from the wind and solar generation units. Table 4 presents data

Table 3
Annual Moments of Hourly Wind, Solar, and Wind and Solar Output in California

	2013	2014	2015	2016	2017	2018	2019	2020
	Hourly Wind Output (MWh)							
Mean	1033.54	1131.32	999.26	1204.73	1235.28	1597.35	1581.63	1551.73
Median	973.79	1035.19	860.06	1092.49	1074.29	1496.55	1439.55	1378.13
Standard deviation	843.79	881.27	822.59	918.41	957.56	1161.22	1148.88	1149.84
Coefficient of variation	.82	.78	.82	.76	.78	.73	.73	.74
Standard skewness	.39	.49	.53	.41	.47	.34	.42	.45
Standard kurtosis	2.03	2.29	2.18	2.05	2.08	1.92	2.07	2.1
	Hourly Solar (MWh)							
Mean	315.39	1000.38	1510.80	1910.23	2633.99	2923.06	3035.64	3214.42
Median	11.98	55.50	90.08	101.91	150.53	174.16	209.95	186.55
Standard deviation	435.64	1290.47	1906.14	2391.94	3257.65	3587.68	3761.14	3907.56
Coefficient of variation	1.38	1.29	1.26	1.25	1.24	1.23	1.24	1.22
Standard skewness	1.22	.84	.83	.73	.69	.67	.72	.66
Standard kurtosis	3.50	2.14	2.63	1.86	1.78	1.75	1.85	1.78
	Hourly Combined Wind and Solar Output (MWh)							
Mean	1348.93	2131.57	2510.06	3114.96	3869.27	4520.41	4617.28	4766.15
Median	1364.04	1971.03	2030.58	2385.57	2595.63	3255.97	3150.32	3265.43
Standard deviation	883.40	1461.08	1983.06	2426.76	3258.25	3606.08	3818.19	3894.42
Coefficient of variation	.65	.69	.79	.78	.84	.80	.83	.82
Standard skewness	.19	.45	.63	.55	.6	.55	.62	.57
Standard kurtosis	2.32	2.50	2.95	2.07	1.97	1.96	2.03	1.95

Source: California ISO OASIS website.
Note: MWh = megawatt-hours.

on the distribution of durations of wind and solar energy production below a given threshold during each year from 2013 to 2020. For a given threshold, say 1,000 MWh, the following process is applied to compute each low energy production duration. The first hour in the year that wind and solar energy production falls below 1,000 MWh starts the duration. This duration ends the first hour that wind and solar energy production is above 1,000 MWh. The second duration is defined following the same process. For example, in 2013 there were 231 durations when total wind and solar production was less than 1,000 MWh. The mean length of these durations was 13.54 hours, but there was one duration of 288 hours or 12 days. By 2020, there were roughly the same number of durations of with solar and wind energy production less than 1,000 MWh—210 durations—but the average length was 7.88 hours, and the longest was 17 hours.

Table 4
Combined Wind and Solar Output Shortfall Durations in California (Hours)

	2013	2014	2015	2016	2017	2018	2019	2020
Threshold value				1,000 MWh				
Number of durations	231	263	256	228	247	171	183	210
Mean	13.54	8.46	9.54	8.73	7.96	9.39	9.07	7.88
Standard deviation	27.43	6.08	5.70	5.79	5.49	5.65	5.33	5.31
Maximum	288	20	18	21	16	17	17	17
Threshold value				3,000 MWh				
Number of durations	53	298	356	364	388	380	396	386
Mean	160.47	21.42	15.85	14.29	12.51	10.72	10.55	10.62
Standard deviation	238.97	42.27	8.57	8.42	5.01	5.94	6.01	5.24
Maximum	1,283	684	140	141	65	44	44	21
Threshold value				5,000 MWh				
Number of durations	1	71	226	321	349	356	353	366
Mean	8,758	119.20	32.84	19.84	16.31	15.33	15.50	14.58
Standard deviation	NA	260.95	65.10	21.56	8.19	6.32	7.21	3.09
Maximum	8,758	1809	875	299	92	90	68	44
Threshold value				7,000 MWh				
Number of durations	1	1	19	131	284	318	318	349
Mean	8,758	8,759	457	61.89	23.36	19.38	19.16	16.94
Standard deviation	NA	NA	800.28	155.67	36.90	22.03	20.62	9.44
Maximum	8,758	8,759	3,177	1,363	478	226	239	116
Threshold value				10,000 MWh				
Number of durations	1	1	3	1	58	161	199	200
Mean	8,758	8,759	2918.33	8,784	146.53	48.91	37.92	37.64
Standard deviation	NA	NA	2793.92	NA	363.65	137.20	98.65	93.89
Maximum	8,758	8,759	5,583	8,784	2,145	1,173	876	849

Source: California ISO OASIS website.

In 2020 there was almost 20,000 MW of wind and solar generation capacity in California, yet for 50% of the hours of the year, 3,265.43 MWh or less energy was produced from these units. In 2020, the average length of the duration of energy production less than 5,000 MWh was 14.58 hours and the longest duration was 44 hours or slightly less than 2 days. For the 10,000 MWh threshold in 2020, the longest duration was 849 hours or more than 35 days.

Tables 5 and 6 repeat the information in tables 3 and 4 for ERCOT for 2018–20. Although ERCOT has almost 27,000 MW of wind and solar capacity in 2020, during 50% of the hours of the year less than 10,789 MWh is produced by this generation capacity. The advantage of the wind capacity

Table 5
Annual Moments of Hourly Wind and Solar Output in ERCOT (MWh)

	2018	2019	2020
Mean	8337.5	9258.6	10910.6
Median	8074.2	8996.5	10769.8
Standard deviation	4179.5	4360.7	4686.9
Coefficient of variation	.50	.47	.43
Standard skewness	.16	.16	.03
Standard kurtosis	1.99	2.00	2.02

Source: Electric Reliability Council of Texas (ERCOT).
Note: MWh = megawatt-hours.

in ERCOT is the significantly higher average capacity factors shown in figure 13 versus the average solar capacity factors shown in figure 12.

The downside of significant wind capacity in ERCOT is the substantially longer maximum durations of low output levels. For example, in 2020 the longest duration of wind and solar output less than 5,000 MWh

Table 6
Combined Wind and Solar Output Shortfall Durations in ERCOT (Hours)

	2018	2019	2020
Threshold value		5,000 MWh	
Number of durations	202	189	146
Mean	11.33	9.34	7.01
Standard deviation	13.40	10.05	8.09
Maximum	94	60	60
Threshold value		7,500 MWh	
Number of durations	222	221	242
Mean	18.21	15.62	10.30
Standard deviation	25.56	17.23	12.15
Maximum	239	133	97
Threshold value		10,000 MWh	
Number of durations	206	241	247
Mean	26.93	20.67	15.98
Standard deviation	45.26	25.16	21.30
Maximum	425	141	230
Threshold value		15,000 MWh	
Number of durations	83	143	207
Mean	99.99	53.83	32.29
Standard deviation	190.65	79.05	56.01
Maximum	1,310	428	387

Source: Electric Reliability Council of Texas (ERCOT).
Note: MWh = megawatt-hours.

is 60 hours or 2.5 days. Unlike solar energy, which relies on daily sunlight with varying levels of intensity, there are sustained periods with very low wind energy production.

The potential for multiday durations of low energy production implies the need for significant storage investments to ensure a reliable supply of energy so California and Texas can reduce significantly the amount of fossil fuel energy they consume. Although California still has the option to significantly increase its consumption of imported electricity from neighboring states during these system conditions, unless Texas interconnects with the rest of the United States this option is not available to ERCOT.

Storage generation units make money buying energy at low prices and selling it at high prices. Capacity-based long-term RA mechanisms typically suppress energy price volatility because of the mandates that retailers purchase multiples greater than 1 of their peak demand in firm capacity. Therefore, a capacity-based long-term RA mechanism provides less market revenue to the storage units necessary to manage sustained periods of low renewable energy production. Consequently, one key criterion for a long-term RA mechanism in a high renewables share market is allowing the short-term energy price volatility that will support the necessary storage investments.

V. The SFPFC Approach to Long-Term RA

As the previous sections have demonstrated, a capacity-based approach to long-term RA mechanism is poorly suited to a region with significant intermittent renewables. The primary reliability challenge is not adequate generation capacity to serve demand peaks, but adequate energy available to serve realized demand during all hours of the year. As the example of California on August 14 and 15, 2020, demonstrates, supply shortfalls do not necessarily occur during system demand peaks, but during net demand peaks.

Because of the substantial contemporaneous correlation in hourly output across locations and across renewable energy technologies, ensuring sufficient supply to meet demand throughout the year will require taking full advantage of the mix of available generation resources. Intermittent renewable resources must reinsure the energy they sell in the forward market with dispatchable generation resources and storage devices. The long-term RA mechanism must also recognize the increasing weather dependence of electricity demand with more customers heating and cooling their homes with electricity.

The SFPFC mechanism results in the realized system demand each hour of the compliance period being covered by a fixed-price forward contract. The SFPFC approach to long-term RA recognizes that a supplier with the ability to serve demand at a reasonable price may not do so if it has the ability to exercise unilateral market power in the short-term energy market. As Wolak (2000) demonstrates, an expected profit-maximizing supplier with the ability to exercise unilateral market power with a fixed-price forward contract obligation would like to minimize the cost of supplying the quantity of energy sold in forward contract. The SFPFC long-term RA mechanism takes advantage of this incentive by requiring retailers to hold hourly fixed-price forward contract obligations for energy that sum to the hourly value of system demand. The SFPFC mechanism implies that all expected profit-maximizing suppliers would like to minimize the cost of meeting their hourly fixed-price forward contract obligations, the sum of which equals the hourly system demand for all hours of the year.

To understand the logic behind the SFPFC mechanism, consider the example of a supplier that owns 150 MW of generation capacity that has sold 100 MWh in a fixed-forward contract at a price of $25/MWh for a certain hour of the day. This supplier has two options for fulfilling this forward contract: (1) produce the 100 MWh energy from its own units at their marginal cost of $20/MWh or (2) buy this energy from the short-term market at the prevailing market-clearing price. The supplier will receive $2,500 from the buyer of the contract for the 100 MWh sold, regardless of how it is supplied. This means that the supplier maximizes the profits it earns from this fixed-price forward contract sale by minimizing the cost of supplying the 100 MWh of energy.

To ensure that the least-cost "make versus buy" decision for this 100 MWh is made, the supplier should offer 100 MWh in the short-term market at its marginal cost of $20/MWh. This offer price for 100 MWh ensures that if it is cheaper to produce the energy from its generation units—the market price is at or above $20/MWh—the supplier's offer to produce the energy will be accepted in the short-term market. If it is cheaper to purchase the energy from the short-term market—the market price is below $20/MW—the supplier's offer will not be accepted and the supplier will purchase the 100 MWh from the short-term market at a price below $20/MWh.

This example demonstrates that the SFPFC approach to long-term RA makes it expected profit-maximizing for each seller to minimize the cost of supplying the quantity of energy sold in this forward contract each

hour of the delivery period. By the logic of the above example, each supplier will find it in its unilateral interest to submit an offer price into the short-term market equal to its marginal cost for its hourly SFPFC quantity of energy, to make the efficient "make versus buy" decision for fulfilling this obligation.

The incentives for supplier offer behavior in a short-term wholesale electricity market created by a fixed-price forward contract obligation are analyzed in Wolak (2000). Consider the case of a single hour in the short-term market. Let QS equal the amount of energy produced and sold in the short-term market by the supplier, PS be the short-term wholesale price, PC be the price of SFPFC energy, and QC be the quantity of SFPFCs sold by the supplier for this hour. The supplier's variable profit for the hour is:

$$\begin{aligned} \text{Profit} &= \text{PS} \times \text{QS} - C(\text{QS}) - (\text{PS} - \text{PC}) \times \text{QC} \\ &= \text{PS} \times (\text{QS} - \text{QC}) + \text{PC} \times \text{QC} - C(\text{QS}) \end{aligned} \tag{1}$$

where $C(\text{QS})$ is variable cost of producing QS. The first term in the first expression in equation (1) shows the supplier's variable profits from selling QS MWh at PS in the short-term market. The second term is the net payment to the seller of QC SFPFC contracts at price PC. The second expression in the above equation demonstrates that a supplier only has an incentive to raise the short-term price if it sells more energy in the short-term market, QS, than its fixed-price forward contract obligation, QC. This expression also demonstrates that the supplier wants the lowest possible price when it sells less energy in the short-term market than its fixed-price forward contract obligation.

Under the SFPFC mechanism, each supplier knows that the sum of the values of the hourly SFPFC obligations across all suppliers is equal the system demand. This means that each supplier of SFPFCs knows that its competitors have substantial fixed-price forward contract obligations for that hour. This implies that all suppliers know that they have limited opportunities to raise the price they receive for short-term market sales beyond their hourly SFPFC quantity.

As discussed below, a supplier's fixed-price forward quantity for an hour under the SFPFC mechanism increases with the value of hourly system demand. Therefore, the supplier that owns 150 MW of capacity in the above example has a strong incentive to submit an offer price close to its marginal cost for the capacity of its generation unit to ensure that its hourly production is higher than the realized value of its SFPFC

energy for that hour. Therefore, the SFPFC mechanism not only ensures that system demand is met every hour of the year but also provides strong incentives for this to occur at the lowest possible short-term price.

A. SFPFC Approach to RA

This long-term RA mechanism requires all electricity retailers to hold SFPFCs for energy for fractions of realized system demand at various horizons to delivery. For example, retailers in total must hold SFPFCs that cover 100% of realized system demand in the current year, 95% of realized system demand 1 year in advance of delivery, 90% 2 years in advance of delivery, 87% 3 years in advance of delivery, and 85% 4 years in advance of delivery. The fractions of system demand and number of years in advance that the SFPFCs must be purchased are parameters set by the regulator to ensure long-term RA. The SFPFCs would clear against the quantity-weighted average of the hourly locational prices at all load withdrawal locations in the short-term wholesale market.

SFPFCs are shaped to the hourly system demand within the delivery period of the contract. Figure 14 contains a sample pattern of system demand for a four-hour delivery horizon. The total demand for the four hours is 1,000 MWh, and the four hourly demands are 100, 200, 400, and 300 MWh. Therefore, Firm 1 that sells 300 MWh of SFPFC energy has the hourly system demand-shaped forward contract obligations of 30 MWh in hour 1, 60 MWh in hour 2, 120 MWh in hour 3, and 90 MWh in hour 4. The hourly forward contract obligations for Firm 2 that sold 200 MWh SFPFC energy and Firm 3 that sold 500 MWh of SFPFC energy are also shown in figure 15. These SFPFC obligations are also allocated across the four hours according to the same four hourly shares of total system demand shown in figure 14. This ensures that the sum of the hourly values of the forward contract obligations for the three suppliers is equal to the hourly value of system demand. Taking the example of hour 3, Firm 1's obligation is 120 MWh, Firm 2's is 80 MWh, and Firm 3's is 200 MWh. These three values sum to 400 MWh, which is equal to the value of system demand in hour 3 shown in figure 14.

These SFPFCs are allocated to retailers based on their share of system demand during the month. Suppose that the four retailers in figure 16 consume 1/10, 2/10, 3/10, and 4/10, respectively, of the total energy consumed during the compliance month for SFPFCs. This means that Retailer 1 is allocated 100 MWh of the 1,000 MWh SFPFC obligations

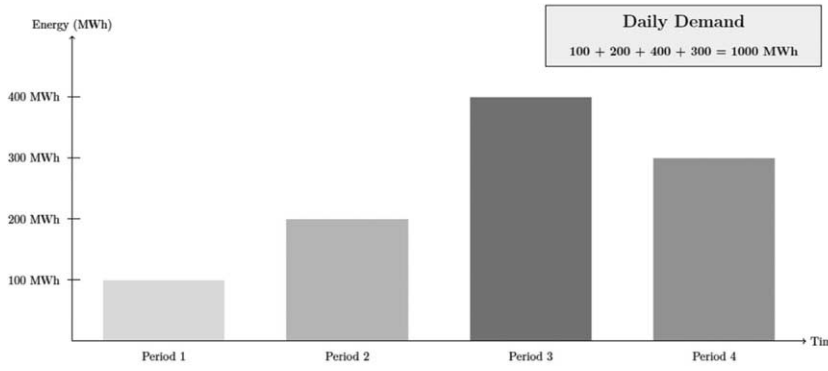

Fig. 14. Hourly system demands. Color version available as an online enhancement.

for the four hours, Retailer 2 is allocated 200 MWh, Retailer 3 is allocated 300 MWh, and Retailer 4 is allocated 400 MWh. The obligations of each retailer are then allocated to the individual hours using the same hourly system demand shares used to allocate the SFPFC energy sales of suppliers

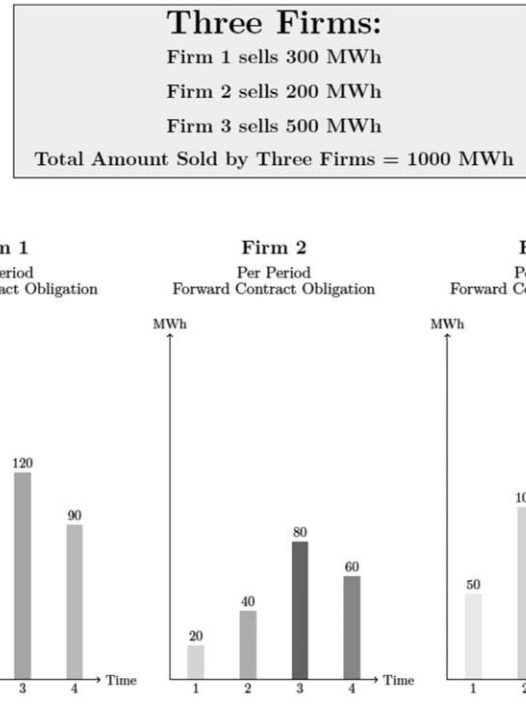

Fig. 15. Hourly forward contract quantities for three suppliers. Color version available as an online enhancement.

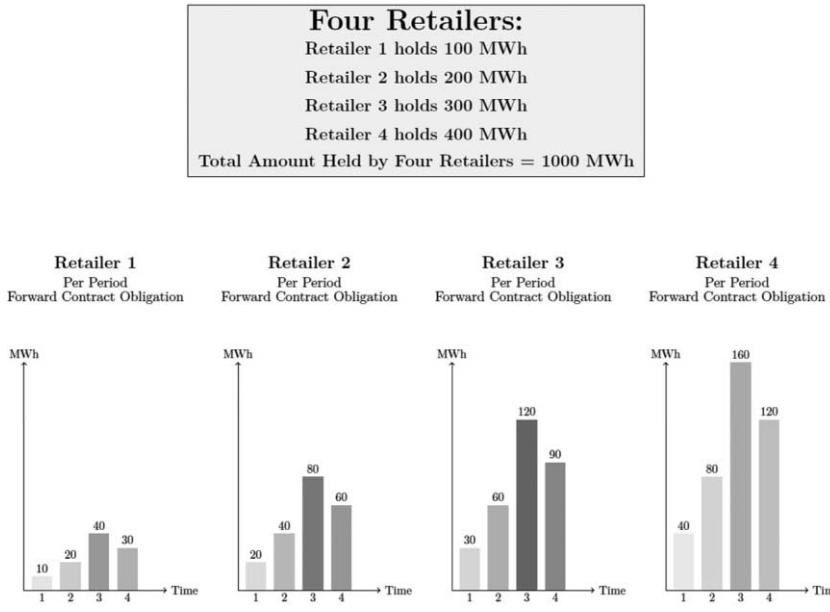

Fig. 16. Hourly forward contract quantities for four retailers. Color version available as an online enhancement.

to the four hours. This allocation process implies Retailer 1 holds 10 MWh in hour 1, 20 MWh in hour 2, 40 MWh in hour 3, and 30 MWh in hour 4. Repeating this same allocation process for the other three retailers yields the remaining three hourly allocations shown in figure 16. Similar to the case of the suppliers, the sum of allocations across the four retailers for each hour equals the total hourly system demand. For period 3, Retailer 1's holding is 40 MWh, Retailer 2's is 80 MWh, Retailer 3's is 120 MWh, and Retailer 4's is 160 MWh. The sum of these four magnitudes is equal to 400 MWh, which is the system demand in hour 3.

B. Mechanics of Standardized Forward Contract Procurement Process

The SFPFCs would be purchased through auctions several years in advance of delivery so new entrants may compete to supply this energy. Because the aggregate hourly values of these SFPFC obligations are allocated to retailers based on their actual share of system demand during the month, this mechanism can easily accommodate retail competition. If one retailer loses load and another gains it during the month, the share of the aggregate hourly value of SFPFCs allocated to the first retailer falls and the share allocated to the second retailer rises.

The wholesale market operator would run the auctions with oversight by the regulator. One advantage of the design of the SFPFC products is that a simple auction mechanism can be used to purchase each annual product. A multiround auction could be run where suppliers submit the total amount of annual SFPFC energy they would like to sell for a given delivery period at the price for the current round. Each round of the auction the price would decrease until the amount suppliers are willing to sell at that price is less than or equal to the aggregate amount of SFPFC energy demanded.

The wholesale market operator would also run a clearinghouse to manage the counterparty risk associated with these contracts. All US wholesale market operators currently do this for all participants in their energy and ancillary services markets. In several US markets, the market operator also provides counterparty risk management services for long-term financial transmission rights, which is not significantly different from performing this function for SFPFCs. Both buyers and sellers would be required to post collateral with the wholesale market operator to ensure that each market participant finds it unilaterally profit-maximizing to meet its financial commitments for the SFPFC energy that it has purchased or sold.

SFPFCs auctions would be run on an annual basis for deliveries starting 2, 3, and 4 years in the future. In steady state, auctions for incremental amounts of each annual contract would also be needed so that the aggregate share of demand covered by each annual SFPFC could increase over time. The eventual 100% coverage of demand occurs through a final true-up auction that takes place after the realized values for hourly demand for the delivery period are known.

C. True-Up Auctions and Settlement of SFPFCs

The vast majority of SFPFC contracts will be purchased in advance of delivery. However, because the mechanism requires that the total quantity of SFPFC energy sold during the compliance period equal the realized demand during that same period, after each compliance period there needs to be true-up auctions to buy back unused SFPFC energy or purchase additional SFPFC energy.

These true-up auctions serve the same role as the real-time market in a two-settlement short-term energy market, with two key differences. First, the aggregate demand for additional SFPFC energy is known when suppliers submit their offers into incremental true-up auction or the total

amount of SFPFC energy to be purchased by suppliers is known when they submit their bids to the true-up auction. Second, suppliers to the true-up auction know the quantity-weighted average hourly short-term market price during compliance period that all SFPFC energy sold or purchased in the true-up auction will clear against when they submit their bids or offers to these auctions. This implies that a supplier would be unlikely to be willing to sell SFPFC energy at a price less than or equal to this quantity-weighted average short-term price or buy SFPFC energy at a price above this quantity-weighted average short-term price.

It is also important to emphasize that the true-up auctions are very unlikely to trade significant quantities of energy given the relatively small rate of growth of energy demand in California. Table 7, taken from the 2017 and 2019 versions of the California ISO's *Annual Report on Market Issues and Performance*, shows the average load (total annual energy demand divided by the number of hours in the year) and annual peak load in the California ISO control area from 2013 to 2019.

The typical rate of growth of the annual demand for energy is substantially less volatile than the rate of growth in annual peak demand. Moreover, total annual energy demand growth is negative for 2018 and 2019. It is also likely to be negative for 2020 because of COVID-19. The volatility of annual peak demand emphasizes the importance of allocating the SFPFC energy using the actual hourly pattern of demand throughout the quarter rather than a forecast of these magnitudes. This precommitment ensures that during all hours of the year the total shortfall of suppliers producing less than their total SFPFC commitments is equal to the total surplus of suppliers producing more than their total SFPFC commitments

Table 7
Annual System Load in California ISO Control Area 2013–2019

Year	Annual Total Energy (GWh)	Average Load (MW)	% Change	Annual Peak Load (MW)	% Change
2013	231,800	26,461	–1.0	45,097	–3.7
2014	231,610	26,440	–.1	45,090	0
2015	231,495	26,426	0	46,519	3.2
2016	228,794	26,047	–1.4	46,232	–.6
2017	227,749	26,002	0	50,116	8.4
2018	220,458	25,169	–3.2	46,427	–7.4
2019	214,955	24,541	–2.5	44,301	–4.6

Note: ISO = Independent System Operators, GWh = gigawatt-hours, MW = megawatts.

for that hour, which means that final consumers have no net exposure to short-term prices.

The most straightforward approach to running the quarterly SFPFC auctions would be to run them as 12 independent auctions, one for each future quarter at least 3 years in the future. However, to facilitate a 3-year future revenue stream that could finance investment in new generation capacity, the 12 quarterly auctions could be run simultaneously so that a potential new entrant could sell prespecified quantities of SFPFC energy in all 12 auctions or nothing at all. For example, the new entrant could submit offers to sell the same amount of energy in all auctions.

The appendix contains several examples using the four-period model to illustrate how the true-up auctions would work. These examples demonstrate that the SFPFC obligation of a supplier provides a strong financial incentive to offer in at least as much energy at its marginal cost as it expects will be its final SFPFC allocation for that hour of the compliance period. If the realized value of the total system demand for the compliance period is higher than expected and the supplier sells SFPFC energy in the true-up auction, its final SFPFC allocation for each hour of compliance period will be higher than its initial SFPFC allocation.

Failure to account for the possibility of selling energy in the true-up auction can result in the supplier purchasing energy from the short-term market at a price that is substantially higher than the marginal cost of the generation capacity that the supplier does not offer into the short-term market. In this sense, the SFPFC obligation provides a supplier with a must-offer obligation for at least its final allocation of the SFPFC energy after the true-up auction for that hour of the compliance period, because the SFPFC mechanism requires the supplier to purchase any shortfall in output from its generation resources relative to this hourly SFPFC allocation at the hourly short-term price.

D. Incentives for Behavior by Intermittent Renewable and Controllable Resources

Because all suppliers know that all energy consumed every hour of the year is covered by a SFPFC in the current year and into the future, there is a strong incentive for suppliers to find the least-cost mix of intermittent and controllable resources to serve these hourly demands. To the extent that there is concern that the generation resources available or likely to be available in the future to meet demand are insufficient,

features of the existing capacity-based RA mechanism can be retained until system operators have sufficient confidence in this mechanism leading to a reliable supply of energy. The firm capacity values from the existing capacity-based long-term RA approach can be used to limit the amount of SFPFC energy a supplier can sell.

The firm capacity value multiplied by number of hours in the year would be the maximum amount of SFPFC energy that the unit owner could sell in any given year. Therefore, a controllable thermal generation unit owner could sell significantly more SFPFC energy than it expects to produce annually, and an intermittent renewable resource owner could sell significantly less SFPFC energy than it expects to produce annually. This upper bound on the amount of SFPFC energy any generation unit could sell enforces cross-hedging between controllable in-state generation units and intermittent renewable resources. This mechanism uses the firm capacity construct to limit forward market sales of energy by individual resource owners to ensure that it is physically feasible to serve demand during all hours of the year.

Cross-hedging between a controllable resource and an intermittent resource implies that in most years, the controllable resource owner would be producing energy in a small number of hours of the year but earning the difference between the price at which it sold the energy in the SFPFC auction and the hourly short-term market price times the hourly value of its SFPFC energy obligation for all the hours that it does not produce energy. Intermittent renewables owners would typically produce more than their SFPFC obligation in energy and sell any energy produced beyond this quantity at the short-term price. In years with low renewable output near their SFPRC obligations, controllable resource owners would produce close to the hourly value of their SFPFC energy obligation, thus making average short-term prices significantly higher. However, aggregate retail demand would be shielded from these high short-term prices because of their SFPFC holdings.

E. Assessment of the SFPFC Approach to Long-Term RA

This mechanism has several advantages relative to a capacity-based approach. There is no regulator-mandated aggregate capacity requirement. Generation unit owners are allowed to decide both the total MW and the mix of technologies to meet their SFPFC energy obligations. There is also no prohibition on generation unit owners or retailers engaging in other hedging arrangements outside of this mechanism. Specifically, a retailer

could enter into a bilateral contract for energy with a generation unit owner or other retailer to manage the short-term price and quantity risk associated with the difference between their actual hourly load shape and the hourly values of their retail load obligation.

This mechanism provides a nudge to market participants to develop a liquid market for these bilateral contract arrangements at horizons to delivery similar to the SFPFC products. Instead of starting from the baseline of no fixed-price forward contract coverage of system demand by retailers, this mechanism starts with 100% coverage of system demand, which retailers can unwind at their own risk.

This baseline level of SFPFC coverage of final demand is a more prudent approach to long-term RA in a region such as California where the vast majority of customers purchase their electricity according to a fixed retail price or price schedule that does not vary with real-time system conditions. A baseline 100% SFPFC coverage of final demand provides the retailer with wholesale price certainty for virtually all its wholesale energy purchases (except for the small true-up uncertainty described above), which significantly limits the financial risk retailers face from selling retail electricity at a fixed price and purchasing this energy from a short-term wholesale market with increasingly volatile wholesale prices.

An additional benefit of this mechanism is that the retail market regulator, in this case the CPUC, can use the purchase prices of SFPFCs to set the wholesale price implicit in the regulated retail price over the time horizon that the forward contract clears. This would provide retailers with a strong incentive to reduce their average wholesale energy procurement costs below this price through bilateral hedging arrangements, storage investments, or demand response efforts.

There are several reasons why this mechanism should be a more cost-effective approach to long-term RA than a capacity-based mechanism in a zero marginal cost intermittent future. First, the sale of SFPFC energy starting delivery 2 or more years in the future provides a revenue stream that will significantly increase investor confidence in recovering the cost of any investment in new generation capacity.

Second, because retailers are protected from high short-term prices by total hourly SFPFC holdings equaling actual system demand, the offer cap on the short-term market can be raised to increase the incentive for all suppliers to produce as much energy as possible during stressed system conditions. Third, the possibility of higher short-term price spikes can finance investments in storage and load-shifting technologies and

encourage active participation of final demand in the wholesale market, further enhancing system reliability in a market with significant intermittent renewable resources.

If SFPFC energy is sold for delivery in 4 years based on a proposed generation unit, the regulator should require construction of the new unit to begin within a prespecified number of months after the signing date of the contract or require posting of a substantially larger amount of collateral in the clearinghouse with the market operator. Otherwise, the amount of SFPFC energy that this proposed unit sold would be automatically liquidated in a subsequent SFPFC auction and a financial penalty would be imposed on the developer. Other completion milestones would have to be met at future dates to ensure the unit is able to provide the amount of firm energy that it committed to provide in the SFPFC contract sold. If any of these milestones were not met, the contract would be liquidated.

F. Empirical Evidence on the Performance of the SFPFC Mechanism

Although the SFPFC mechanism in the form described above does not exist in any currently operating electricity supply industry, the long-term RA mechanisms in Chile and Peru create the same set of incentives for supplier behavior as the SFPFC mechanism by assigning system-wide short-term price and quantity risk during all hours of the year to suppliers. Both Chile and Peru operate a supplier-only, cost-based short-term wholesale electricity market. The system operator employs regulated variable cost estimates for each generation unit and an opportunity cost of water for hydroelectric generation units to dispatch generation units to meet demands throughout each country. All consumers or their retailers are required to purchase full-requirements contracts from suppliers to meet their retail load obligations. Suppliers financially settle imbalances between the amount of energy they produce and the amount of energy their customers consume under these full-requirements contracts. Suppliers that produce more energy than their customers consume receive payments from the suppliers that produce less energy than their customers consume.[19]

To see the equivalence of the incentives created for supplier behavior under the market designs in Chile and Peru and the SFPFC mechanism, let QR_i equal the consumption of customers served by the supplier i and PR_i the quantity-weighted average price paid for full-requirements contracts by customers served by supplier i. Let system demand equal QD,

which is also equal to $\Sigma_{i=1}^{N} QR_i$, the sum of the consumption of all customers served by the N suppliers. The variable profit of supplier i is equal to

$$\begin{aligned} \text{Profit}_i &= PS \times QS - C(QS) - (PS - PR_i) \times QR_i \\ &= PS \times (QS - QR_i) + PR_i \times QR_i - C(QS), \end{aligned} \tag{2}$$

which is identical to equation (1) presented earlier by setting QR_i equal to QC and PR_i equal to PC. Moreover, because $QD = \Sigma_{i=1}^{N} QR_i$; all short-term price and quantity risk is borne jointly by the N suppliers that have sold full-requirements contracts.

The long-term RA mechanisms in Chile and Peru have delivered a reliable supply of electricity for at least the past 15 years in each country in the face of significant hydroelectric energy supply uncertainty and an increasing share of the energy consumed coming from intermittent wind and solar generation units. This outcome has been achieved in two countries with average annual load growth rates that are three to four times that in regions in the United States with formal wholesale electricity markets. Consequently, the experience of Chile and Peru provides a strong argument in favor of the SFPFC mechanism for regions of the United States with significant intermittent renewable energy goals.

VI. Final Comments

Wholesale market design is a process of continuous learning, adaptation, and, hopefully, improvement. As the analyses of Sections II and III have shown, a capacity-based long-term RA mechanism designed for an industry based on dispatchable thermal generation units is poorly suited to an industry with a significant share of energy coming from intermittent renewable generation capacity. These analyses demonstrate that future supply shortfalls similar to those that occurred in California during August 2020 and in Texas during February 2021 are likely in regions with significant intermittent renewable generation capacity without a change in the paradigm for ensuring long-term RA.

These analyses demonstrate that the major system reliability challenge with a significant amount of renewable generation capacity changes, from having sufficient generation capacity to meet annual system demand peaks to the ability to meet the hourly net demands (system demand less intermittent renewable output) for energy throughout the year. Particularly in an electricity supply industry with a summer annual peak demand and significant installed solar generation capacity, meeting daily system demand peaks is relatively straightforward because demand peaks when

there is significant solar energy production. The new focus on meeting net demand peaks implies a system-wide focus on energy adequacy where intermittent renewable resources have a financial incentive to hedge their short-term and production quantity risk with dispatchable generation resources to cover these net demand peaks.

The standardized energy contracting approach to long-term RA described in this paper delivers this outcome by allowing dispatchable resources to sell significantly more energy in these standardized forward contracts than they expect to produce to provide the revenue necessary to keep sufficient amounts of this generation capacity available to meet these hourly net demands throughout the year, even though these thermal units operate at smaller annual average capacity factors. Intermittent renewable resources are allowed to sell significantly less energy in these standardized forward contracts than they expect to produce annually to ensure that sufficient dispatchable generation capacity will be available to meet the intermittent net demand peaks throughout the year. The experience of Chile and Peru over the past 15 years, each of which has a market design that creates the same set of incentives for supplier behavior as the SFPFC mechanism, provides encouraging empirical evidence in favor of its adoption in regions with significant intermittent renewable energy goals.

Appendix

I. Examples of Positive and Negative True-Up Auction Outcomes

A compliance auction would be run far in advance of the compliance period to purchase 1,000 MWh of energy for the four time periods shown in figure 14. Suppose this auction cleared at a price of $60/MWh. Figure 15 shows the quantities sold in the auction for the three suppliers and their hourly SFPFC obligations, assuming the pattern of aggregate demand in figure 15 is realized for the four time periods. Figure 16 shows the hourly SFPFC holdings of the four retailers for the four time periods. The total demand across the four periods for each retailer is shown at the top of figure 16.

Now suppose that the realized demand for the compliance period turns out to be 10% higher in each of the four periods. The new demands for the four periods are shown in figure A1. This implies the need for an ex post true-up auction for 100 MWh. Because demand is 10% higher in each of the four periods, the shares that allocate this additional

100 MWh across four time periods to the four retailers are the same as those used to allocate the original 1,000 MWh across the four time periods. The incremental allocations to each of the four retailers are shown in figure A2 and the total realized demands for the four periods for each retailer are shown at the top of the graph. The period-level obligations for the incremental SFPFC energy purchased in the true-up auctions depend on which suppliers sell this energy. If each firm sells 10% more SFPFC energy in the true-up auction and system demand increases by 10% in each of the four periods, the period-level allocations of the additional SFPFC energy for each supplier are shown in figure A3. In this example, we assume that the true-up auction cleared at \$70/MWh and the demand-weighted average short-term price for the four periods is \$55/MWh.

In addition to the variable profits they would earn from selling the energy they produce from their own generation units in the short-term market, the three suppliers would receive the following difference payments to settle their SFPFC contract positions:

Firm 1 = (\$60 – \$55)300 + (\$70 – \$55)30
Firm 2 = (\$60 – \$55)200 + (\$70 – \$55)20
Firm 3 = (\$60 – \$55)500 + (\$70 – \$55)50.

Besides the variable profits they would earn from purchasing energy from the short-term market and selling to their retail customers at the retail price, the four retailers would pay the following difference payments:

Retailer 1 = (\$60 – \$55)1,000(110/1,100) + (\$70 – \$55)(110/1,100)100
Retailer 2 = (\$60 – \$55)1,000(220/1,100) + (\$70 – \$55)(220/1,100)100
Retailer 3 = (\$60 – \$55)1,000(330/1,100) + (\$70 – \$55)(330/1,100)100
Retailer 4 = (\$60 – \$55)1,000(440/1,100) + (\$70 – \$55)(440/1,100)100.

Both the original and true-up aggregate SFPFC purchases are allocated to individual retailers based on their actual share of total demand served during the four demand periods.

If this 100 MWh total demand increase is instead shared equally between periods 1 and 2, period 1 demand would now be 150 MWh and the period 2 demand would now be 250 MWh. Demand in periods 3 and 4 are unchanged from those in figure 14. In the final settlement, 150 MWh of the SFPFCs would be allocated to retailers in period 1, 250 MWh in period 2, 400 MWh in period 3, and 300 MWh in period 4. Suppose that Retailer 1 consumed the entire additional 100 MWh of energy during the compliance period. Retailer 1 would now be assigned 2/11 = (200/1,100)

of the above period-level values of SFPFCs as opposed to the values shown in figure 16. Retailer 2, 3, and 4 would be also be assigned 2/11, 3/11, and 4/11, respectively, because their demand totals for the four periods did not change.

Suppose that the entire 100 MWh true-up auction quantity was all sold by Firm 1 at a price of $65/MWh and as a result of a different pattern of demands throughout the four periods, the demand-weighted average short-term price is $50/MWh. Now, in addition to the variable profits they would earn from selling energy in the short-term market produced by their generation units, the three suppliers would receive the following difference payments to settle their SFPFC contract positions:

Firm 1 = ($60 – $50)300 + ($65 – $50)100
Firm 2 = ($60 – $50)200
Firm 3 = ($60 – $50)500.

Besides the variable profits they would earn from purchasing energy from the short-term market to sell to their customers at the retail price, the four retailers would pay for the following difference payments:

Retailer 1 = ($60 – $50)(1,000)(2/11) + ($65 – $50)100(2/11)
Retailer 2 = ($60 – $50)(1,000)(2/11) + ($65 – $50)100(2/11)
Retailer 3 = ($60 – $50)(1,000)(3/11) + ($65 – $50)100(3/11)
Retailer 4 = ($60 – $50)(1,000)(4/11) + ($65 – $50)100(4/11).

Again, both the original and true-up aggregate SFPFC purchases are allocated to individual retailers based on their actual share of total demand served during the four demand periods.

What price clears the true-up auction depends on the extent of competition among suppliers to provide this additional energy. Clearly, suppliers are extremely unlikely to offer to supply this energy below the demand-weighted average short-term price over the compliance period because their overall profits would decline. However, if there are a substantial number of suppliers willing to sell this additional SFPFC energy, the price is unlikely to be significantly above the demand-weighted average short-term price.

It is important to note that the lower the demand-weighted average short-term price, the larger are the difference payments that suppliers receive. This is another way of demonstrating that all suppliers have an incentive to minimize the cost of meeting their SFPFC obligations by offering to supply this energy at their marginal cost of production in the short-term market.

The true-up auction for excess SFPFC energy operates in an analogous manner. Suppose that demand is 10% lower in every period as shown in figure A4. Suppose each firm buys back 10% of its SFPFC quantity in the true-up auction. This yields the period-level SFPFC quantities for each supplier in figure A5. If all retailers reduce their consumption in each of the four periods by 10%, their hourly SFPFC allocations and their total demands for the four periods are those shown in figure A6. Suppose that the demand-weighted average short-term price is $45/MWh and true-up auction clears at $40/MWh.

In addition to the variable profits they would earn from selling energy produced by their generation units in the short-term market, the three suppliers would now receive the following difference payments to settle their SFPFC contract positions:

Firm 1 = ($60 – $45)300 – ($40 – $45)30
Firm 2 = ($60 – $45)200 – ($40 – $45)20
Firm 3 = ($60 – $45)500 – ($40 – $45)50.

Besides the variable profits they would earn from purchasing energy from the short-term market to sell to at the retail price to their customers, the four retailers would pay the following difference payments:

Retailer 1 = ($60 – $45)(90/900)1,000 – ($40 – $45)(90/900)100
Retailer 2 = ($60 – $45)(180/900)1,000 – ($40 – $45)(180/900)100
Retailer 3 = ($60 – $45)(270/900)1,000 – ($40 – $45)(270/900)100
Retailer 4 = ($60 – $45)(360/900)1,000 – ($40 – $45)(360/900)100.

Once again, the price that clears the true-up auction depends on the extent of competition among suppliers to purchase the excess energy. Clearly, suppliers are extremely unlikely to bid a price for this energy above the demand-weighted average short-term price over the compliance period. However, if there are a substantial number of suppliers willing to buy this excess SFPFC energy, the auction price is unlikely to be significantly below the demand-weighted average short-term price.

Now suppose that the entire 100 MWh true-up auction quantity was purchased by Firm 1 at a price of $35/MWh and this 100 MWh reduction in demand across the four periods came entirely from period 3 and only from Retailer 3. Suppose that as a result of a different pattern of demand throughout the day, the realized demand-weighted average short-term price is $40/MWh. This implies the following realized system load shares for the four periods: 1/9, 2/9, 3/9, and 3/9. The total realized demands for each retailer are now 100, 200, 200, and 400, so

portions of both aggregate SFPFC purchases are allocated to retailers using the following shares: 1/9, 2/9, 2/9, and 4/9.

Now, in addition to the variable profits they would earn from selling the energy produced by their generation units in the short-term market, the three suppliers would receive the following difference payments to settle their SFPFC contract positions:

Firm 1 = ($60 – $40)300 – ($35 – $40)100
Firm 2 = ($60 – $40)200
Firm 3 = ($60 – $40)500.

Besides the variable profits they would earn from purchasing energy from the short-term market to sell to their retail customers, the four retailers would pay for the following difference payments:

Retailer 1 = ($60 – $40)(1,000)(100/900) – ($35 – $40)100(100/900)
Retailer 2 = ($60 – $40)(1,000)(200/900) – ($35 – $40)100(200/900)
Retailer 3 = ($60 – $40)(1,000)(200/900) – ($35 – $40)100(200/900)
Retailer 4 = ($60 – $40)(1,000)(400/900) – ($35 – $40)100(400/900).

The original and true-up aggregate SFPFC purchases are allocated to individual retailers based on their actual share of total demand served during the four demand periods. More details on the SFPFC mechanism and examples of true-up auctions are given in California Public Utilities Commission (2021).

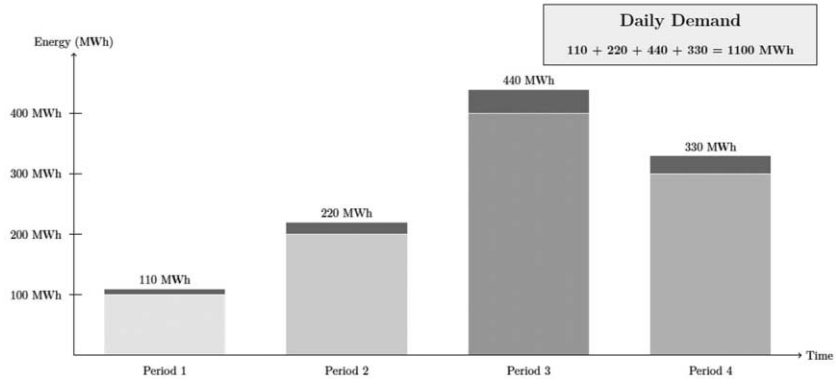

Fig. A1. Hourly system demands (10% higher). Color version available as an online enhancement.

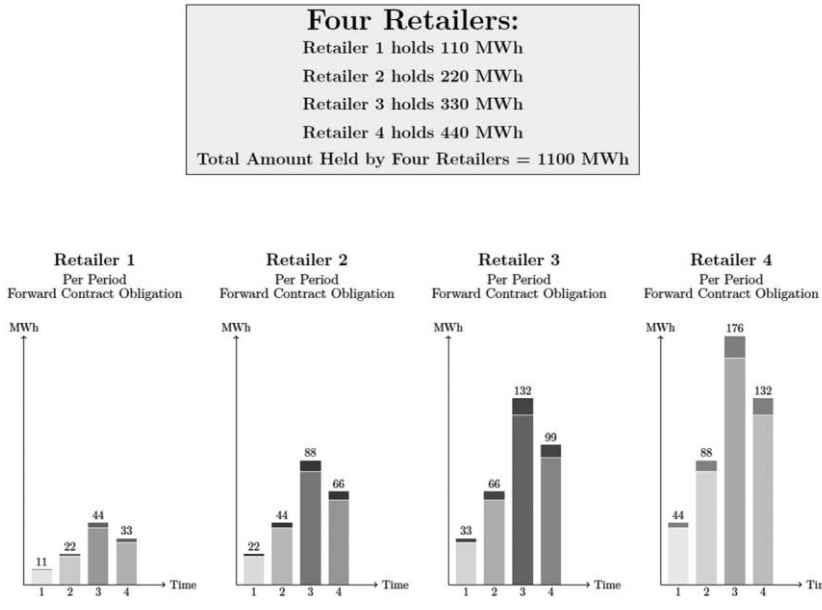

Fig. A2. Hourly forward contract quantities for four retailers (10% higher). Color version available as an online enhancement.

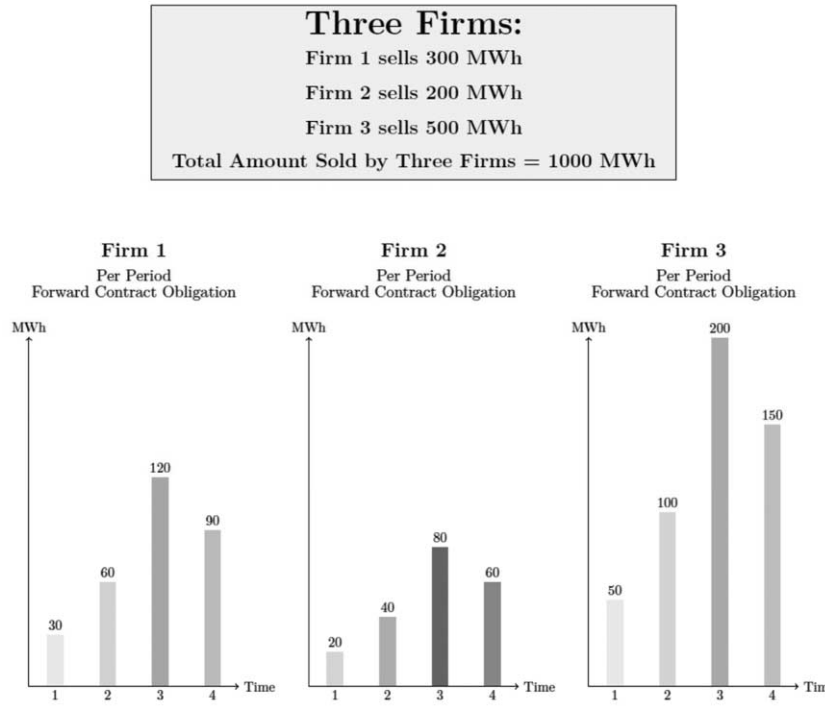

Fig. A3. Hourly forward contract quantities for three suppliers (10% higher). Color version available as an online enhancement.

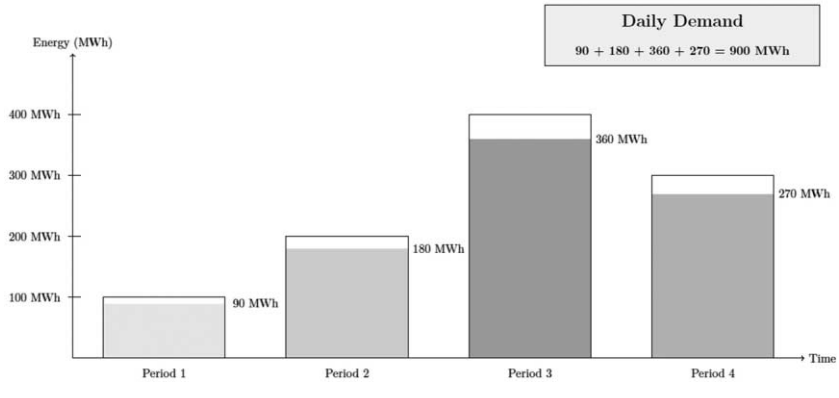

Fig. A4. Hourly system demands (10% lower). Color version available as an online enhancement.

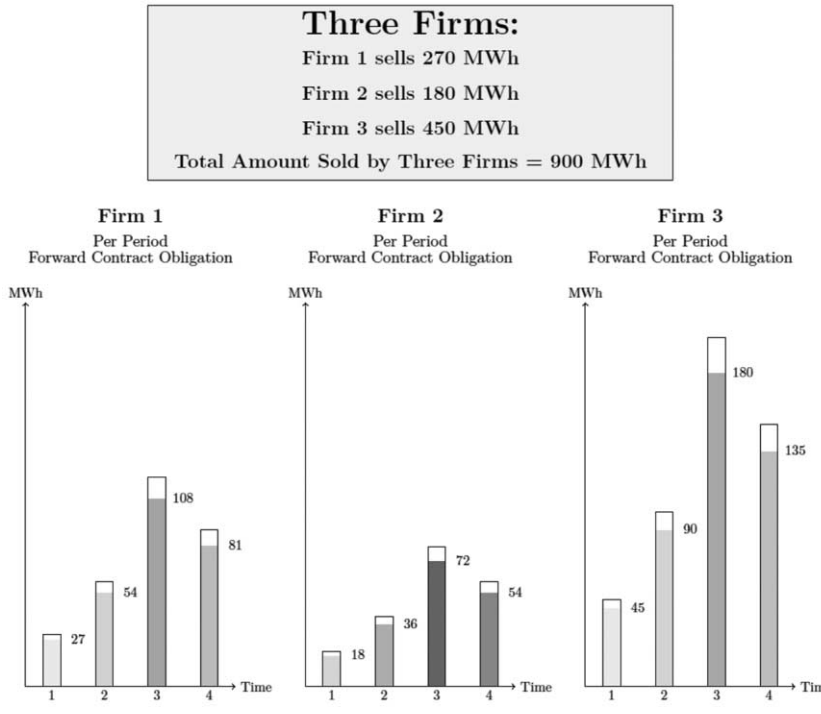

Fig. A5. Hourly forward contract quantities for three suppliers (10% lower). Color version available as an online enhancement.

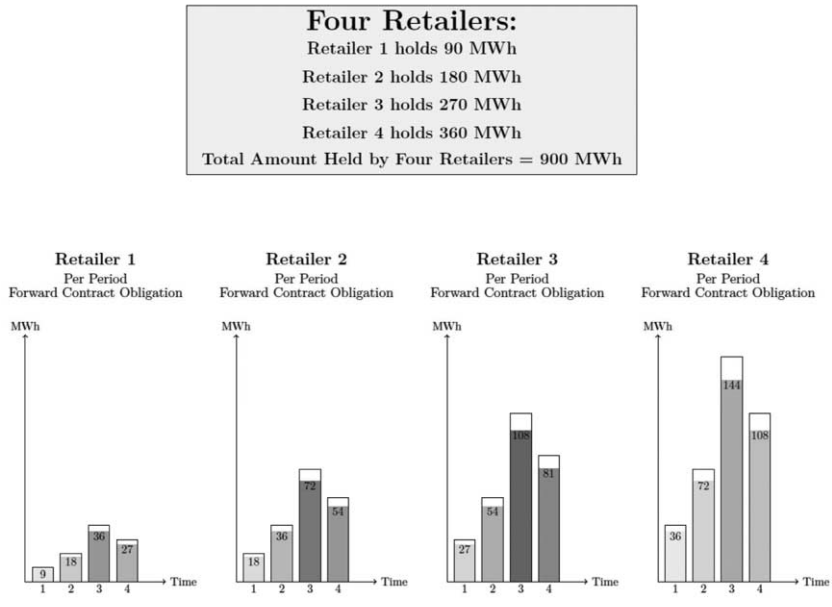

Fig. A6. Hourly forward contract quantities for four retailers (10% lower). Color version available as an online enhancement.

Endnotes

Author email address: Wolak (wolak@zia.stanford.edu). For acknowledgments, sources of research support, and disclosure of the author's material financial relationships, if any, please see https://www.nber.org/books-and-chapters/environmental-and-energy-policy-and-economy-volume-3/long-term-resource-adequacy-wholesale-electricity-markets-significant-intermittent-renewables.

1. McRae and Wolak (2019) demonstrate the difficulty of ensuring that supply equals demand in the hydroelectric-dominated electricity supply industry in Colombia because of El Niño weather events. Wolak (2003) argues that a key causal factor in the California electricity crisis of 2000–01 was the low levels of hydroelectric energy available in the Pacific Northwest, which typically supplies a substantial amount of electricity to California each year. Wolak (2009) demonstrates that the two supply shortfall periods in 2001 and 2003 in the New Zealand wholesale electricity market were also due in large part to low water availability.

2. See http://www.ercot.com/content/wcm/key_documents_lists/225373/Urgent_Board_of_Directors_Meeting_2-24-2021.pdf.

3. The number of generation units available is a binomial random variable with probability $p = .9$ and with number of trials N = the number of generation units. The probability of meeting the demand peak is the probability the available capacity is greater than or equal to the peak demand.

4. The annual average hourly capacity factor for a generation technology first computes the total production by that technology during each hour of the year divided the total of installed capacity of that technology during that hour. It then computes the annual average of these hourly values over all hours of the year.

5. For a map of all North American Interconnections, see https://www.eia.gov/electricity/data/eia411/images/nerc_old.jpg.

6. Wood, Wollenberg, and Sheblé (2013) provide an accessible introduction to Kirchoff's laws and transmission and distribution grid operation.

7. The monthly firm capacity figures for 2020 are listed in the file NetQualifyingCapacityList-2020.xls, available at http://www.caiso.com/planning/Pages/ReliabilityRequirements/Default.aspx.

8. There are limited direct current interconnections with neighboring states that sell limited amounts of electricity to ERCOT or export energy from ERCOT.

9. See https://data.census.gov/cedsci/table?q=ACSDT1Y2019.B25040.

10. See http://www.ercot.com/content/wcm/key_documents_lists/225373/Urgent_Board_of_Directors_Meeting_2-24-2021.pdf.

11. See https://www.bloomberg.com/news/articles/2021-02-15/texas-power-retailers-in-face-of-freeze-please-leave-us.

12. Ibid.

13. See https://www.houstonchronicle.com/politics/texas/article/Simple-paperwork-blunder-Texans-cold-winter-storm-16032163.php.

14. See https://www.spglobal.com/platts/en/market-insights/blogs/electric-power/041521-texas-electricity-market-february-freeze-power-outages.

15. See https://www.eia.gov/todayinenergy/detail.php?id=46896.

16. See https://www.houstonchronicle.com/politics/texas/article/Simple-paperwork-blunder-Texans-cold-winter-storm-16032163.php.

17. See https://www.texasmonthly.com/news-politics/texas-blackouts-natural-gas.

18. CV = (Standard Deviation)/(Mean).

19. See Section 3.2 of Wolak (2021) for more details on this settlement mechanism.

References

California Public Utilities Commission. 2021. "Addendum to Staff Draft Straw Proposal for Consideration in Track 3B.2 of Proceeding R.19-11-009." February 26. https://docs.cpuc.ca.gov/PublishedDocs/Efile/G000/M372/K082/372082582.PDF.

Carden, Kevin, Alex Krasny Dombrowsky, and Chase Winkler. 2020. "2020 Joint IOU ELCC Study, Report 1." https://www.astrape.com/2020-joint-ca-iou-elcc-study-report-1.

Doss-Gollin, James, David J. Farnham, Upmanu Lall, and Vijay Modi. 2021. "How Unprecedented Was the February 2021 Texas Cold Snap?" *Environmental Research Letters* 16 (6): 064056.

Garver, Leonard L. 1966. "Effective Load Carrying Capability of Generating Units." *IEEE Transactions on Power Apparatus and Systems* 8:910–19.

Kahn, Edward P. 2004. "Effective Load Carrying Capability of Wind Generation: Initial Results with Public Data." *Electricity Journal* 17 (10): 85–95.

McRae, Shaun D., and Frank A. Wolak. 2016. "Diagnosing the Causes of the Recent El Niño Event and Recommendations for Reform." http://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/diagnosing-el-nino_mcrae_wolak.pdf.

———. 2019. "Market Power and Incentive-Based Capacity Payment Mechanisms." http://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/2019-03-mcrae-wolak-capacity.pdf.

Wolak, Frank A. 2000. "An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market." *International Economic Journal* 14 (2): 1–39.

———. 2003. "Diagnosing the California Electricity Crisis." *Electricity Journal* 16 (7): 11–37.
———. 2009. "An Assessment of the Performance of the New Zealand Wholesale Electricity Market." Report for the New Zealand Commerce Commission. http://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/new_zealand_report_redacted.pdf.
———. 2013. "Economic and Political Constraints on the Demand-Side of Electricity Industry Restructuring Processes." *Review of Economics and Institutions* 4 (1): 42.
———. 2016. "Level versus Variability Trade-offs in Wind and Solar Generation Investments: The Case of California." *Energy Journal* 37 (Bollino-Madlener Special Issue).
———. 2021. "Final Report on Thematic Line 2: Transformation of the Peruvian Wholesale Electricity Market." Report for the Ministry of Mines and Energy of Peru. http://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/report_wolak_June_2021_draft_0.pdf.
Wood, Allen J., Bruce F. Wollenberg, and Gerald B. Sheblé. 2013. *Power Generation, Operation, and Control*. Hoboken, NJ: Wiley.