# Market power and incentive-based capacity payment mechanisms*

Shaun D. McRae† and Frank A. Wolak‡

July 2020

**Abstract**

Capacity markets provide guaranteed payments to electricity generation unit owners for having the ability to produce a fixed amount energy under stressed system conditions. Historically, these markets have been plagued by the weak incentives they provide for plants to be available during these system conditions. By contrast, the reliability payment mechanism in the Colombian electricity market provides market-based incentives for plants to produce during periods of system scarcity. It has served as a model for the design of capacity markets in a number of jurisdictions in North America and Europe. We demonstrate severe shortcomings of this mechanism. By adjusting their price and quantity offers, generators with the ability to exercise unilateral market power can choose whether or not a scarcity condition exists. We find that this mechanism can make it privately profitable for firms to withhold output and create a scarcity condition. We illustrate this problem using hourly data from the first ten years of operation of the reliability payment mechanism in Colombia. The mechanism not only fails to minimize the cost of meeting electricity demand but also creates perverse incentives for electricity generators that could reduce the reliability of electricity supply. We quantify the cost of the perverse incentives caused by the capacity payment mechanism by computing a counterfactual dynamic oligopoly equilibrium for the 2015-16 El Niño event in Colombia.

*We thank seminar participants at CIDE, ITAM, FGV-EPGE, PUC-Rio, University of Southern California, the IAEE International Conference, the Economics of Low Carbon Markets Workshop, POWER Conference, the International Industrial Organization Conference, the Toulouse Conference on the Economics of Energy and Climate, the From Theory to Practice conference at the Becker Friedman Institute, the Property and Environment Research Center (PERC), and the Transportation and Public Utilities Group ASSA meeting for their helpful comments and suggestions. McRae gratefully acknowledges financial support from the *Asociación Mexicana de Cultura*.

†Centro de Investigación Económica and Department of Economics, ITAM. Email: shaun.mcrae@itam.mx.

‡Program on Energy and Sustainable Development and Department of Economics, Stanford University, Stanford, CA 94305-6072. Email: wolak@zia.stanford.edu.

# 1 Introduction

In restructured electricity markets, capacity payment mechanisms pay generation plants for their ability to produce energy, even if their output is zero. The original motivation for these mechanisms was to provide an additional revenue stream for infrequently-used plants that might not cover their fixed costs from energy sales alone. The expansion of intermittent renewable generation has aggravated this revenue adequacy problem and increased the need to keep backup generation available. While traditional capacity markets have provided additional revenue to generation unit owners, their design suffers from the relatively weak incentives they provide for plants to be available during critical system conditions (Bushnell et al., 2017).

As a result, many electricity markets are changing the design of their capacity mechanisms to provide stronger incentives for generators to be available during periods of system scarcity. These market-based incentives typically take the form of reliability option contracts that provide an implicit financial penalty to generation units that do not produce sufficient energy during critical conditions. The strike price for such options is set based on the marginal cost of the highest-cost generation technology in the system. When the wholesale market price exceeds the strike price, generation firms that do not produce the "firm energy" from their generation units that was sold in the capacity market must refund the difference between the market and strike prices for their generation shortfall. Generation firms that produce more than their "firm energy" receive the market price (which exceeds the strike price) for their additional output.

In this paper, we demonstrate a potentially severe flaw in capacity mechanisms based on reliability option contracts. Generation firms with market power may have the ability to choose whether or not the scarcity condition is triggered and the option is exercised. Even if the short-term wholesale market typically yields competitive market outcomes, it is during peak demand conditions—when the generation capacity is most required—that market power problems are greatest. The incentive for generators to trigger scarcity conditions depends on the relative magnitudes of their "firm energy" quantities and their fixed-price forward contract for energy quantities. We show that the reliability option mechanism increases market prices and the average cost of thermal generation relative to a counterfactual with no capacity payment mechanism at all and the same hourly values of fixed-price forward contract obligations for energy.

We demonstrate the empirical importance of this problem using ten years of data from the reliability payment mechanism in the Colombian wholesale electricity market.

The Colombian experience is highly relevant for the design of reliability options in other markets. Having started in December 2006, it is the oldest and longest-running incentive-based capacity market in the world. Recent and proposed reforms in other markets, including the Peak Energy Rent mechanism and the new Pay-for-Performance rules in the New England ISO, the capacity payment mechanism in the Irish electricity market, and the proposed capacity payment mechanism in the Italian electricity market are based at least in part on the Colombian model (Mastropietro et al., 2018). Furthermore, the Colombian electricity market is heavily dependent on hydroelectric generation and suffers from periodic shortfalls in hydro inflows. It is during these low-water periods, when firm energy requirements are greatest, that generation firms can have the ability and incentive to withhold output to profit from the reliability option mechanism. We demonstrate that similar problems are likely to arise in other markets with a reliability option mechanism and a high share of intermittent renewable generation.

We use hourly information provided by the Colombian market operator XM for the period December 2006 to June 2016. This hourly information includes the price and quantity offers for each generation unit, the system demand, the dispatched and actual generation output of each unit, the market price, the capacity contract quantities and prices, and the fixed-price forward contract for energy positions of each firm. We supplement the hourly data with information on hydrological inflows and storage levels, as well as information on fossil fuel usage and prices.

We first show that firms have the ability to choose whether or not a scarcity condition exists (that is, whether the reliability option is exercised). We calculate the hour-by-hour inverse residual demand curve faced by each firm in the market. When this curve lies entirely above or below the scarcity price, the output of the firm does not affect whether the scarcity condition will be triggered. However, when the supplier's inverse residual demand curve crosses the scarcity price at a quantity between the firm's minimum and maximum generation output, the firm's choice of generation quantity will determine the existence of a scarcity condition. The largest generation firm in the Colombian market, EPM, can trigger a scarcity condition in 16% of the hours in our sample.

We then calculate the profitability of triggering the scarcity condition by calculating the optimal generation quantities for each hour of the day given the hourly residual demand curves faced by the firm. Because the net firm capacity position is determined for the whole day—based on the firm capacity and the total generation for the day—it is necessary to solve this optimization problem simultaneously for all 24 hours of the day.

The opportunity cost of water is the principal determinant of production decisions for hydroelectric generators. For each firm, we recover the monthly opportunity cost that minimizes the deviation between the actual and optimal generation quantities.

We find that the best response output levels by three largest suppliers in Colombia provides an accurate prediction of whether or not the scarcity condition is triggered for each firm. For EPM, the scarcity condition occurs in 90% of the hours in which EPM's best-reply output level triggers the scarcity condition, and it does not occur in 98% of the hours for which EPM's best-reply output level does not trigger the scarcity conditions. Similar results–in terms of the percent of correct predictions of scarcity and no scarcity conditions by best-reply output levels–holds for the other two large generation firms in the Colombian market.[1]

Our analysis of plant-level bid data provides further evidence that generators respond to the incentives created by the reliability option mechanism. During the hours when it is profit-maximizing to avoid triggering the scarcity condition, the distribution of accepted bid prices shows a high degree of bunching immediately below the scarcity price. Conversely, in the hours when it would be optimal to trigger the scarcity condition, the distribution of accepted bid prices lies above the scarcity price.

The observed pattern of bidding behavior had real-world effects on the reliability of the Colombian wholesale market. By keeping the bid prices of the hydro units low to avoid triggering the scarcity condition, more expensive thermal units were underutilized even when drought conditions were imminent. As discussed in McRae and Wolak (2016), the reduced level of storage in hydro reservoirs almost led to electricity rationing in early 2016.

We compare the reliability option mechanism to a counterfactual market structure without the mechanism for the May 1, 2015 to April 30, 2016 period. We use a dynamic oligopoly model of the three largest hydroelectric generation unit owners in Colombia, accounting for the intertemporal constraint on firm-level generation determined by total hydro inflows, as in Bushnell (2003). Our counterfactual market structure that eliminates the reliability option, yet retains each supplier's fixed-price forward contract obligations for energy for actual water conditions, yields significantly lower average wholesale prices, higher hydro storage levels, and lower use of thermal generation relative to the existing market structure with the reliability option mechanism.

Our paper contributes to the existing theoretical literature on strategic behavior under

1. It is important to emphasize that these results hold despite the fact that suppliers do not know the values of the hourly residual demands they face at the time they submit bids into the short-term market.

capacity payment mechanisms. Fabra (2018) develops a simple analytical framework that incorporates both generation investment and short-run pricing decisions. She studies the cases of reliability options and their potential to mitigate market power, but acknowledges the crucial role for regulators in setting the scarcity price. Her framework does not incorporate the interaction between forward contracts and reliability options. Léautier (2016) also develops an analytical model to compare reliability options with physical capacity certificates and develops conditions under which these are equivalent. Most closely related to our paper, Teirilä and Ritz (2019) construct a computational model of the Irish electricity market to study the potential exercise of market power under a system of reliability options. They model the capacity market, generator entry and exit, and the short-run wholesale market. Although the capacity market leads to new generation entry, the exercise of market power by the large incumbent generator in Ireland could increase electricity procurement costs by 40 to 100 percent relative to a competitive counterfactual.

In contrast to these existing papers, the focus of our analysis is on the interaction of reliability options and fixed-price forward contracts for energy in the short-run wholesale market. We study this issue in a dynamic oligopoly setting where strategic hydroelectric generators exercise market power by shifting their allocation of water. We abstract from the issues of the competitiveness of the capacity market auction and the level of investment in new generation.

In addition to our theoretical contribution, this paper is one of the first empirical studies of how strategic behavior in a wholesale electricity market is affected by a capacity payment mechanism. This is an urgent issue to study as more and more countries adopt an electricity market design that includes some form of capacity payment. An increasing share of final expenditure on electricity is channeled to generators through these mechanisms instead of through direct purchases of electricity. As we show in our analysis, the reliability option design is not a panacea for market power issues and may even increase the cost for consumers of market power.

Our results are directly relevant for the many wholesale electricity markets that are planning or implementing an incentive-based capacity remuneration mechanism. Reliability options may have unexpected consequences when generators with the ability to exercise unilateral market power can endogenously choose whether or not the option is exercised. The potential reduction in system reliability is particularly troublesome given that electricity consumers are paying the generation firms for these options.

We suggest an alternative approach to long-term resource adequacy that provides strong incentives for the least-cost supply of the energy necessary to serve demand in the future. This mechanism involves a standardized market for fixed-price forward contracts. Retailers must hold regulator-determined fractions of their realized energy demand in these fixed-price forward contracts at various horizons to delivery. For example, 100% of their annual demand one year in advance of delivery, 97% two years in advance, and 95% three years in advance. These forward purchases provide suppliers with the revenue streams necessary far enough in advance of delivery to build the generation capacity necessary to serve demand in the future. These standardized forward contracts also ensure that electricity retailers and large consumers are protected from short-term price volatility.

The remainder of the paper is organized as follows. Section 2 provides details on the structure of the Colombian electricity market. Section 3 uses a simple theoretical model to show the potential distortions that arise from the reliability payment mechanism. Section 4 presents a series of descriptive and analytical results for the performance of the reliability payment mechanism. Section 7 discusses the results and outlines our alternative energy contracting-based approach for ensuring long-term resource adequacy.

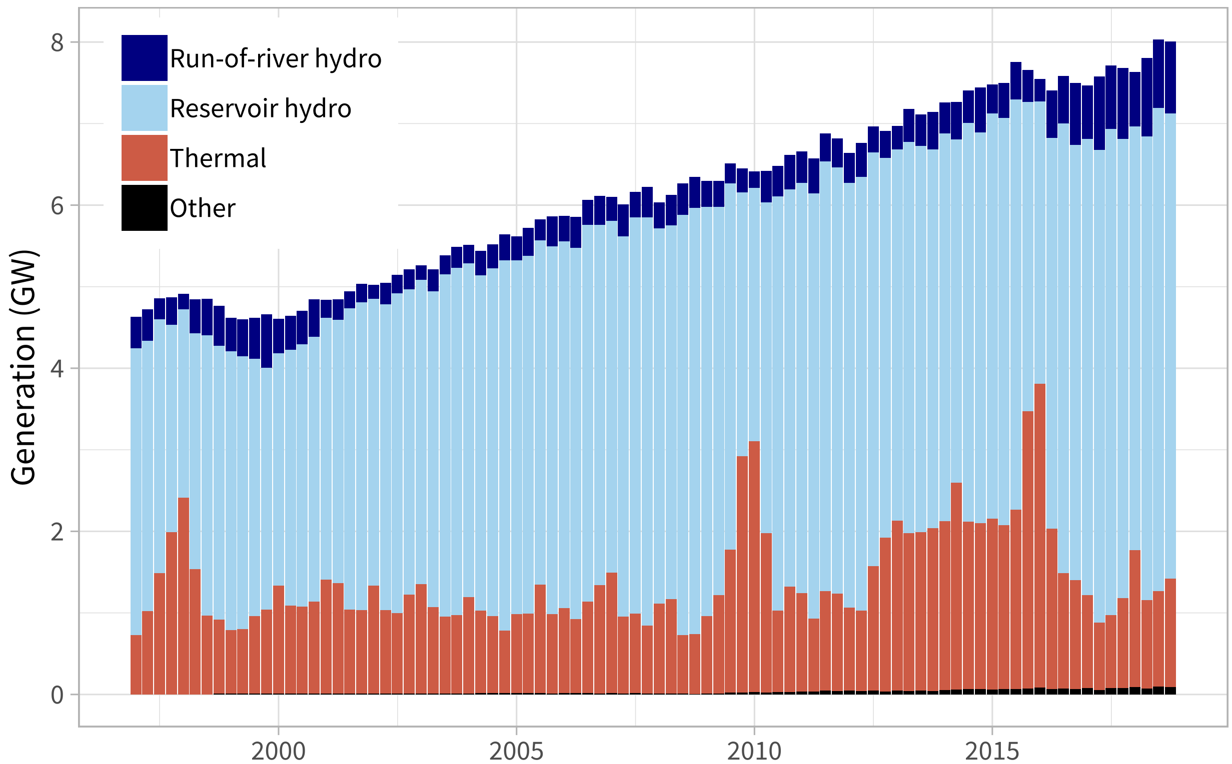
## 2 Institutional setting

Restructuring of the electricity industry in Colombia began in 1994. This process was motivated by a period of electricity rationing between March 1992 and March 1993, the result of an El Niño event that reduced inflows into hydro reservoirs.[2] The government lacked the financial capacity to invest in new thermal plants that could act as a backup for hydro generators in dry years (Dyner et al., 2006). After the reforms, there was considerable private investment in thermal capacity during the late 1990s.

Electricity generation in Colombia remains predominantly hydroelectric. Total generation increased from 4.7 gigawatts (GW) in 2000 to 7.9 GW in 2015, an average annual growth rate of 2.9 percent. Between 2000 and 2009, most of this growth in electricity demand was met by increases in hydro generation (Figure 1). Thermal generation played a larger role between 2012 and 2016, including the 2015–16 period that we study in this paper. More recently, hydro generation has regained its share of total generation. Despite the fall to 72 percent between 2012 and 2015, the share of hydro in 2018 was 82 percent of total generation.

2. Fetzer et al. (2018) use satellite night lights to study the geographical variation in rationing during the 1992–93 blackouts. They show that electricity shortages led to a short-term increase in fertility and a permanent increase in the number of children.

**Figure 1:** Quarterly electricity generation in GW, by type of generator, 1997–2018




Notes: Calculation based on plant-level hourly generation data from XM Compañía de Expertos en Mercados (2019b).

The most striking pattern of the composition of electricity generation in Colombia is the periodic reduction in hydroelectric energy associated with the climatic phenomenon known as *El Niño*. This event is characterized by an increase in water temperatures in the central Pacific Ocean. One effect of this for Colombia is a reduction in rainfall (and hence inflows into hydro reservoirs) in the major hydro-producing regions of the country. This reduction in inflows associated with El Niño occurred in 2009–10 and again in 2015–16. As seen on Figure 1, these periods were associated with a substantial drop in hydroelectric generation and a corresponding increase in thermal generation.

The market design in Colombia is different from the cost-based short-term markets used throughout Latin America.[3] It is based around a central pool in which prices are determined by daily price and quantity bids that generators submit to the system operator.

3. Galetovic et al. (2015) describe this mechanism for the case of Chile. Cost-based designs exists in Argentina, Brazil, Peru and a number of Central American countries (see Rudnick and Montero (2002)).

Each generation unit may in supplier's portfolio must submit a single bid price for the unit output for the entire day. The quantity made available from each generation unit is allowed to vary by hourly throughout the day. After 2009, generators also were allowed to submit the startup costs associated with each unit, and the dispatch algorithm used by the system operator ensured that plants were only turned on if they would recover these costs.[4]

The market also includes a system of firm energy payments that are made to generators even when they are not producing electricity. The amount of the paid for firm energy (in $ per MW) is determined by auctions for long-term investment in new generation capacity, first held in May 2008 and December 2011.[5] Both existing and new generation plants receive the payments for their firm energy, known as the firm energy obligation. During periods when the wholesale price exceeds a regulated "scarcity price", the generators who received these payments are required to pay the difference between the wholesale price and the scarcity price, multiplied by their firm energy quantity. This creates a financial incentive for the generators to make at least their firm energy available during periods of system scarcity, in order to meet this financial obligation. The scarcity price is recalculated each month based on changes in the price of an international fuel oil benchmark.

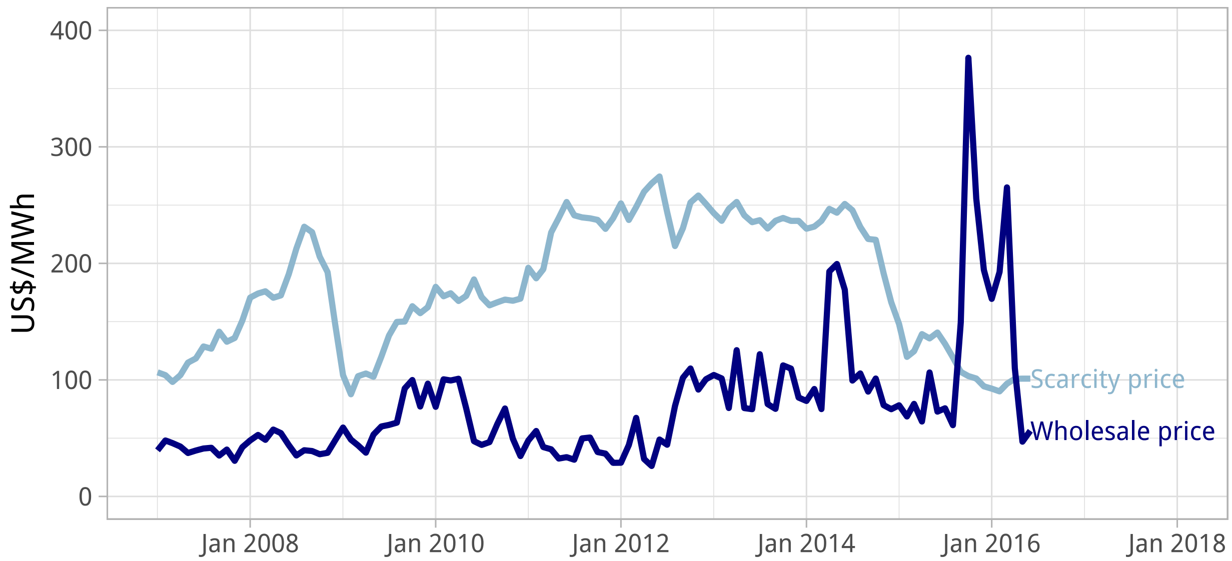
For nearly all hours during the first nine years of operation of the reliability payment mechanism, the market price was below the scarcity price, meaning that the scarcity condition was not triggered (Figure 2). This changed during the El Niño event at the end of 2015 and the start of 2016. For six months, the market price exceeded the scarcity price. Generation firms that did not produce their firm energy quantity were required to refund the shortfall, multiplied by the difference between the market and scarcity prices.

The three largest firms in the Colombian market are Empresas Públicas de Medellín (EPM), Emgesa, and Isagen, with a combined generation capacity of 60 percent of the total. These firms are predominantly hydroelectric, although each has a small proportion of thermal generation. Three smaller firms also own significant amounts of hydroelectric generation capacity: Celsia, AES Chivor, and Urrá. Ownership of thermal generation capacity is less concentrated, and there are several small firms that own or operate a single thermal plant.

4. Riascos et al. (2016) study the effect of including startup costs in the generation bids in the Colombian market. They find that it led to a reduction in production costs but this was not passed through to lower wholesale prices. Reguant (2014) studies the use of these complex bids in the Spanish wholesale electricity market.

5. Harbord and Pagnozzi (2012) review the design, outcome and performance of these auctions.

**Figure 2:** Wholesale market prices and scarcity prices, 2007–2016




Notes: The figure shows the monthly mean wholesale market price and the monthly scarcity price, for each month from 2007 to June 2016. For those hours in which the market price exceeds the scarcity price, generation firms are required to produce at least their firm energy quantity. This condition occurred in almost every hour between October 2015 and March 2016.

# 3 Illustrative model

In this section, we provide a simple model to illustrate the interaction of the reliability payment mechanism with the fixed-price forward contract for energy market that illustrates the incentives the reliability mechanism creates for suppliers to withhold output to cause scarcity conditions.

Market power is the ability to profitably raise and maintain prices at higher than competitive levels—what they would be if every firm submitted its marginal cost curve as its offer curve. We measure the ability to exercise market power by an electricity supplier by calculating its residual demand curve—the market demand less the quantity supplied by the firm's competitors at each possible market price. McRae and Wolak (2014) demonstrate that the four large generation unit owners in New Zealand submit offer prices closer to their marginal cost of producing energy when they face flatter residual demand curves.

It is important to emphasize that a generation unit owner does not know the precise form of the residual demand curve it will face when it submits its bids into the short-term market because all firms submit their bids at the same time. In addition, the realized demand is unknown at the time supplier submit their offers. The market operator then aggregates these bids and crosses them with the realized demand to compute the market-clearing price paid to all generation units each hour of the day. Firms in Colombia are able to observe the bids submitted by their competitors and water levels of hydroelectric resources with a two-week delay, which helps them to predict the residual demand curve they are likely to face.

When the firm chooses the offer curve to submit to the market operator, it is effectively choosing the point along its realized residual demand curve that it will operate. By definition, the firm will produce the generation quantity and receive the wholesale price that are set by the price and quantity pair where its offer curve crosses its realized residual demand curve. As discussed in Wolak (2000) and Wolak (2003) the firm chooses the bid price and quantity increment combinations that make up its aggregate willingness-to-supply curve given the distribution of possible residual demand curves that it faces to maximize its expected profits given the variable cost of operating its generation units.

In most electricity markets, including the Colombian wholesale market, the offer curves submitted by generators are step functions. Each step is a price and quantity pair representing the additional generation quantity that the firm is willing to supply at that price. Because the offer curves are step functions, so too are the residual demand curves. However, for analytical simplicity, we assume that residual demands are linear functions for our illustrative model. In our empirical analysis, we utilize the step-function residual demand curves.

Suppose a generator faces a downward-sloping inverse residual demand curve:

$$P(Q) = 400 - 100Q \tag{1}$$

The variable $Q$ in this expression is the generation quantity of the firm and $P(Q)$ is the corresponding market price. This inverse residual demand curve is shown in each graph of Figure 3. For this analysis, we assume that the firm can observe its residual demand curve and choose any price and quantity pair along it, and that it has sufficient capacity to operate at any point on the curve. In our empirical analysis we use the firm's actual residual demand curve and only allow it to produce energy up to the amount of capacity it owns.

In the absence of any forward contracts, the generator will act as a monopolist off its residual demand curve and produce where marginal revenue is equal to marginal cost. Assume the marginal cost for the generator is zero. In this case, $MR = 400 -$

$200Q = 0$, implying the firm will maximize profits by choosing $Q = 2$. The market price corresponding to this generation quantity is \$200.

Fixed-price forward contracts reduce the incentive for electricity generators to restrict their output and increase the market price. Suppose the generator in the example has signed $Q_c = 3$ of forward contracts at a price $P_c$. With these forward contracts in place, the profit for the firm is now:

$$\Pi = \underbrace{P_c Q_c}_{1} + \underbrace{P(Q)Q}_{2} - \underbrace{P(Q)Q_c}_{3} - \underbrace{c(Q)}_{4} \tag{2}$$

There are four components in the profit equation:

1. the forward contract revenue $P_c Q_c$, which is predetermined when the firm chooses its generation offer;

2. the revenue from selling the quantity $Q$ at the wholesale market price $P(Q)$;

3. the cost of fulfilling the supplier's forward contract obligation, $Q_c$, at the short-term market price; and

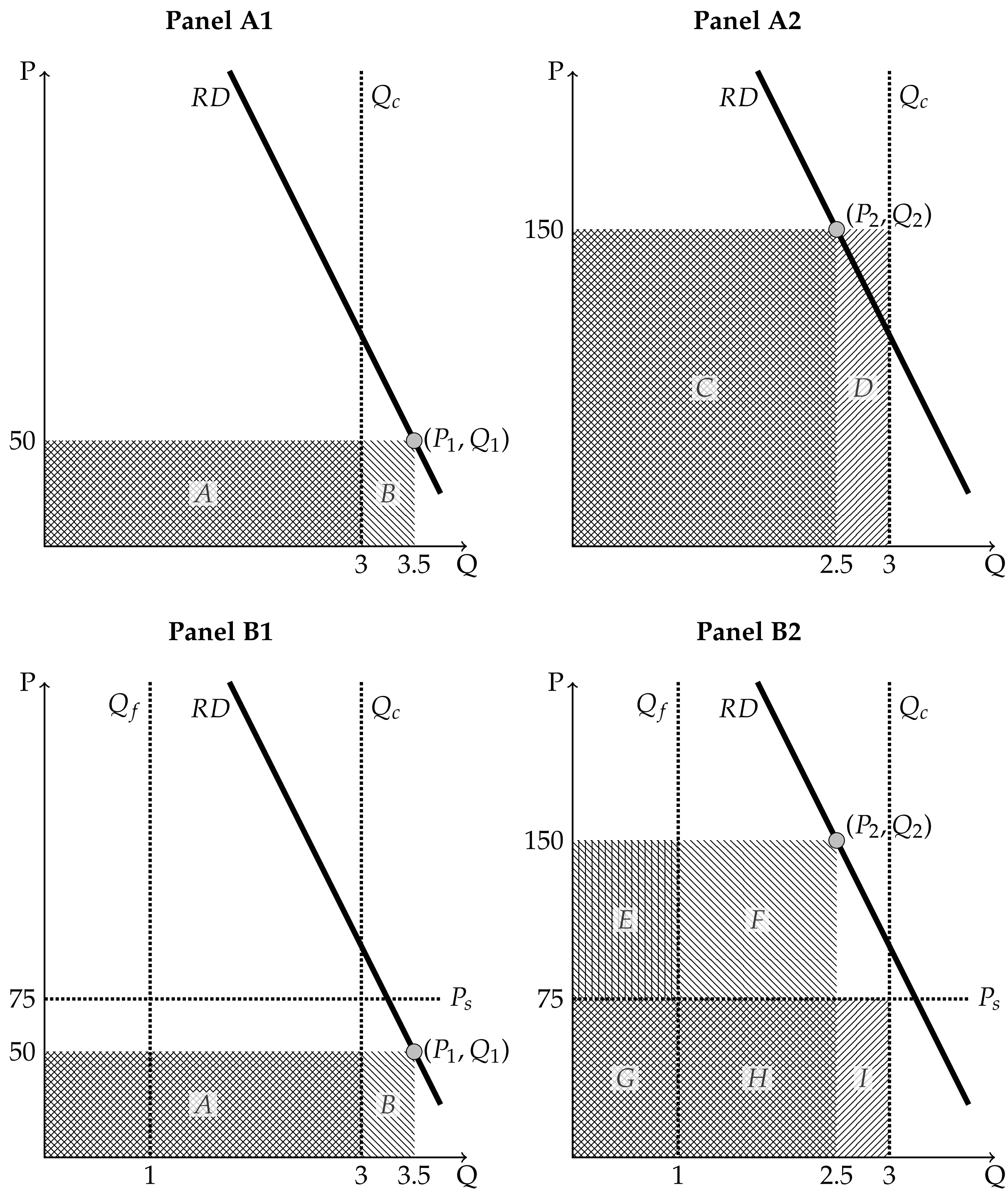
4. the generation cost $c(Q)$.

The forward contract obligation creates an incentive for the firm to increase its production above 2, because of the additional term $-P'(Q)Q_c$ in the marginal revenue. This incentive is shown graphically in Panels A1 and A2 of Figure 3. Given its residual demand, the firm still has the ability to increase the market price to \$150 by restricting its generation to 2.5, as shown in Panel A2. Its wholesale market revenue is the area $C$. However, its forward contract obligation is the area $C + D$, creating a net revenue loss of the area $D$.

Instead, the firm can maximize its profits by increasing its generation to 3.5 (Panel A1). The wholesale price falls to \$50 and reduces the wholesale market revenue to the area $A + B$. The forward contract obligation is the area $A$, leaving the firm with a positive net revenue of the area $B$, plus its forward contract revenue. With the forward contracts in place, the firm has less **incentive** to withhold generation to push up the wholesale market price, even though it still has the **ability** to produce at any point along its residual demand curve.

Now suppose we introduce the reliability payment mechanism to this setting where generation firms have existing forward contracts (Panels B1 and B2 of Figure 3). The generation firm has a firm energy contract with a quantity $Q_f = 1$ and the firm energy

**Figure 3:** Effect of the reliability payment mechanism on incentives for generation firms to withhold, in the presence of forward contracts




Notes: Panel A shows the calculation of net revenue with a forward contract quantity of 3. Net revenue will be maximized by producing 3.5 (A1). Panel B shows the calculation with the addition of a firm energy quantity of 1 and a scarcity price of 75. Net revenue will be maximized by producing 2.5 (B1).

payment $P_f$. There is an administratively-set scarcity price $P_s = \$75$. The profit for the firm is given by Equation (3).

$$\Pi = \underbrace{P_c Q_c + P_f Q_f}_{1} + \underbrace{P(Q)Q}_{2} - \underbrace{\min(P(Q), P_s)Q_c - \max(P(Q) - P_s, 0)Q_f}_{3} - \underbrace{c(Q)}_{4} \tag{3}$$

There are still four components in the profit equation:

1. the forward contract revenue $P_c Q_c$ and the firm energy revenue $P_f Q_f$, both of which are predetermined when the firm chooses its generation offer;
2. the revenue from selling the quantity $Q$ at the wholesale market price $P(Q)$;
3. the cost of fulfilling the supplier's forward contract obligation, $Q_c$, and the cost of fulfilling the firm energy obligation, $Q_f$; and
4. the generation cost $c(Q)$.

Suppose the market price $P(Q)$ is below the scarcity price $P_s$. In that case, Equation (3) simplifies to Equation (2), with the addition of the firm energy payment $P_f Q_f$ (Panel B1 of Figure 3). At a price of \$50, the wholesale market revenue and forward contract obligation are identical to Panel A1. The firm will have a net revenue of the area $B$, plus its forward contract and firm energy revenue.

The reliability payment mechanism changes the incentive for the firm to restrict its generation and increase the market price (Panel B2). As before, by producing a quantity of 2.5, the firm has the ability to increase the market price to \$150. The wholesale market revenue is the area $E + F + G + H$, identical to the area $C$ in Panel A2.

The reliability payment mechanism caps the price for purchases in the wholesale market at the scarcity price $P_s$. Electricity retailers and large consumers pay a charge of $P_f$ each period, and in return they are guaranteed to pay no more than $P_s$ for their purchases from the short-term wholesale market. This cap limits the cost of meeting the forward contract obligation to the area $G + H + I$.

The mechanism also creates an obligation for generators to pay the difference between the market price and the scarcity price for their firm energy quantity, when this difference is positive. Generators receive a payment of $P_f$ each period in exchange for accepting this obligation. This obligation creates an incentive for the firm to produce at least their firm energy quantity during periods with high prices, in the same way that the firm had an incentive to produce at least its forward contract quantity in Panel A. The cost of the firm energy obligation for the firm in Panel B2 is the area $E$. The overall net revenue is calculated as the area $(E + F + G + H) - (G + H + I) - E = F - I$.

Under the reliability payment mechanism, generators with the ability to exercise unilateral market power can often choose whether or not the scarcity condition exists by their output choice. In Panel B1, in which the firm avoids the scarcity condition, the profit-maximizing quantity is 3.5, with a net revenue of 25 (the area $B$). In Panel B2, the firm triggers the scarcity condition, with a profit-maximizing quantity is 2.5 and net revenue of 75 (the area $F - I$). Choosing between these two alternatives, it will be optimal in this example for the firm to restrict its output to 2.5.

This is a striking result. The reliability payment mechanism provides an additional revenue stream $P_f Q_f$ to the generator, funded by a charge on electricity consumers. The implicit promise of the mechanism is that it will provide incentives for the generator to make its capacity available during periods when it is most required. Instead, for this example, the generator has an incentive to withhold generation relative to what it would have produced in the absence of the reliability payment mechanism but with its fixed-price forward contract for energy obligations in place.

The reliability payment mechanism may provide an incentive to withhold generation whenever the firm energy quantity $Q_f$ is less than the forward contract quantity $Q_c$. This is because the $Q_c$ becomes irrelevant for production decisions when the scarcity condition is triggered. As a result, forward contracts lose their moderating role on the incentives for firms to exercise market power. The following sections will show that this is not just a theoretical possibility. For some generation firms in the Colombian market, particularly those that own a substantial amount hydroelectric generation capacity, it is common for $Q_f$ to lie below $Q_c$.

The example presented in Figure 3 uses a simplification of the firm energy calculation in order to highlight the strategic incentives. In practice, the firm energy quantity is only defined on a daily level. The net position for the firm is calculated as the difference between its total generation for the day and the firm energy quantity. Firms with excess generation (as in Panel B2) will have their hourly firm energy obligations (area $E$ in Panel B2) determined based on an allocation of their firm energy across hours, proportional to their generation. Firms with a generation shortfall will have their daily firm energy obligation determined based on their share of the total shortfall for all generators. Appendix A provides a detailed example of the firm energy calculation.

# 4 Empirical analysis

In this section, we analyze the bidding and operating behavior of the generators in the Colombian wholesale electricity market to demonstrate the real-world relevance of the predictions from model in Section 3.

The data for our analysis was provided by the Colombian market operator XM. We use hourly information on the operation of the market for the period January 2008 to June 2016. This hourly information includes the price and quantity offers for each generation unit, the system demand, the dispatched and actual generation output of each unit, and the market price. We supplement the hourly data with information on hydrological inflows and storage levels, as well as information on fossil fuel usage and prices.

## 4.1 Large generators have the ability to create scarcity condition

In some hours, the largest generators in the Colombian electricity market have the ability to unilaterally determine whether or not a scarcity condition exists. The realized residual demand of a generator—that is, the realized market demand less the aggregate willingness-to-supply curve of competing generators—describes the possible combinations of market price and generation quantity pairs that the firm can choose. Under the assumption that the generation unit owner observes the residual demand curve it will face, it can pick any price and quantity combination along this residual demand curve, up to its generation capacity limit, by submitting an offer curve that crosses the residual demand curve at the desired point.[6]

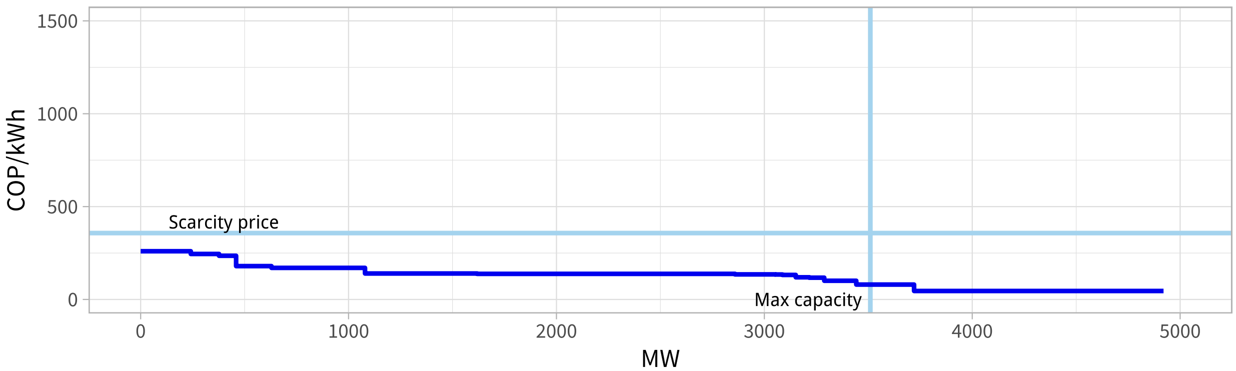
There are three possible configurations for the residual demand curve as shown in Figure 4. The first case is when the firm's inverse residual demand curve lies below the scarcity price for all feasible generation quantities. The maximum generation quantity is determined by the nameplate capacity of the generation units. The minimum generation quantity may be greater than zero for hydroelectric generators if there are environmental regulations on downstream water flows. With the inverse residual demand curve lying below the scarcity price, for any choice of generation quantity, there will not be a scarcity condition.

The second case is when the inverse residual demand curve lies above the scarcity price

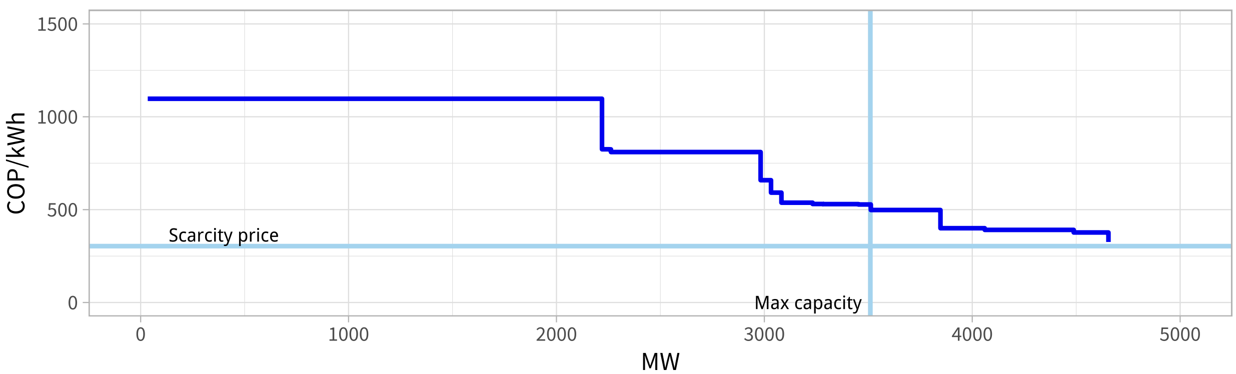
6. As noted earlier, because the bids of other suppliers and the realized value of system demand is unknown at the time the unit owner submits it bid curve, it is unlikely that the supplier's bid curve will intersect its realized residual demand curve at the *ex post* profit-maximizing price and quantity pair.

**Figure 4:** In certain system conditions, generation firms have the ability to choose whether or not the scarcity condition occurs

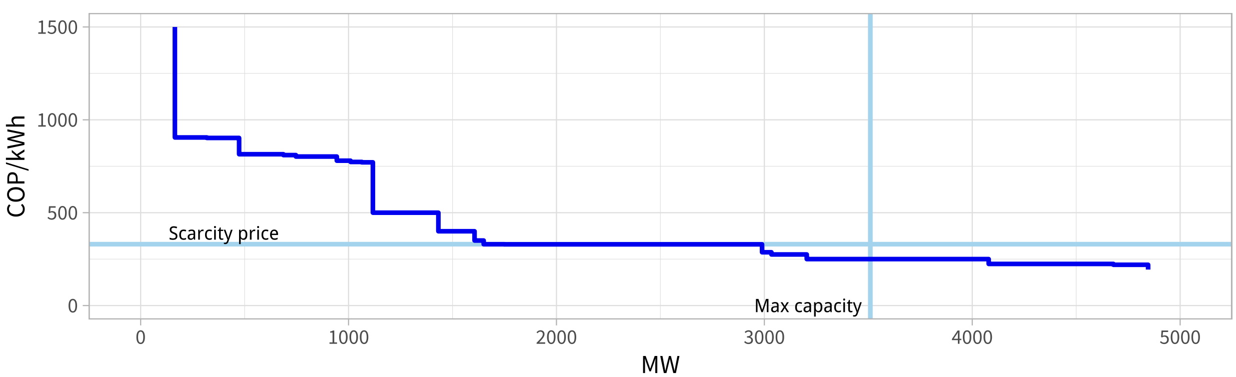
**Case 1:** Generation firm does not have ability to induce scarcity condition
*Residual demand for EPM on 25 July 2015, at 6:00 PM.*




**Case 2:** Scarcity condition will occur regardless of the generation quantity
*Residual demand for EPM on 25 November 2015, at 6:00 PM.*




**Case 3:** Generation firm can choose whether or not the scarcity condition occurs
*Residual demand for EPM on 25 May 2015, at 6:00 PM.*



over the entire range of feasible generation quantities. In that case, the scarcity condition will occur regardless of the generation quantity of the firm.

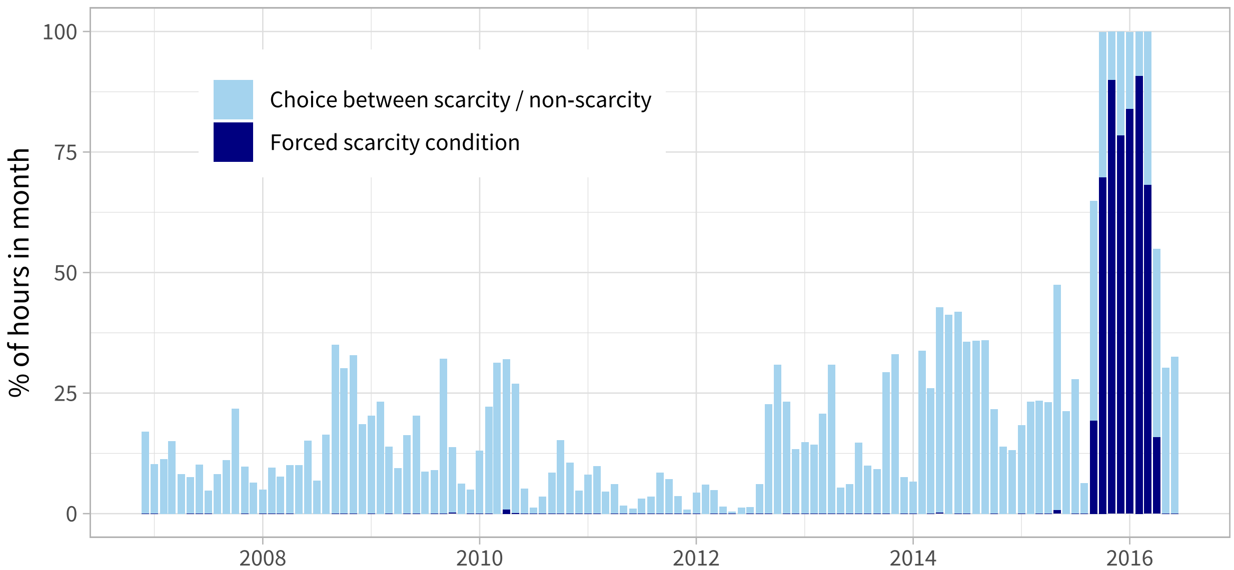
The final case is the one in which the inverse residual demand curve crosses the scarcity price at a quantity that lies within the range of feasible generation quantities. In that case, if the generator chooses a quantity that is less than the crossing point, then the scarcity condition will occur. If the generator chooses a quantity that is more than the crossing point, then there will not be a scarcity condition. For this case, because it is feasible to generate quantities that are either greater than or less than the crossing point, then the generator has the ability to choose whether or not the scarcity condition occurs. Because the residual demand curve a supplier faces is unknown at the time it submits its offer curve, there is no guarantee that its desire to create or avoid a scarcity condition will be successful.

Changes over time in the residual demand and scarcity price mean that the ability of a generator to determine the scarcity condition will vary across days and hours (Figure 4). At 6:00 PM on July 25, 2015, EPM did not have the ability to induce the scarcity condition, for any choice of quantity. At 6:00 PM on November 25, 2015, the scarcity condition would occur for any choice of generation by EPM. Finally, at 6:00 PM on May 25, 2015, EPM could have induced a scarcity condition by producing less than 1600 MW or could have avoided a scarcity condition by producing more than that quantity.

Throughout most of the sample period, EPM had the ability to induce the scarcity condition during at least a few hours of each month as shown in Figure 5. For most of the six months at the end of 2015 and beginning of 2016, the scarcity condition would have occurred regardless of the price and quantity bids by EPM. However, even in this extreme period, EPM had the ability to determine the scarcity outcome in a small proportion of the hours each month.

Over the entire sample period, EPM had the ability to choose between scarcity and non-scarcity conditions in 16 percent of hours (top block of Table 1). In 4.5 percent of hours, all during 2015 and 2016, the scarcity condition was forced to occur for any choice of bids for EPM. The other two large generation firms also had a substantial ability to cause scarcity conditions, though in a smaller share of hours than EPM. Emgesa could induce the scarcity condition in 11 percent of hours and Isagen could do the same in 6 percent of hours. The smaller generation firms in the Colombian market had limited ability to cause scarcity conditions.

**Figure 5:** Proportion of hours each month in which EPM could choose to induce scarcity condition




Notes: The graph classifies the residual demand of EPM for each hour of the sample period. Light bars show the hours where the residual demand crossed the scarcity price within the range of feasible generation quantities for EPM. Dark bars show the hours where the residual demand lay above the scarcity price. For the excluded hours, the residual demand lay below the scarcity price.

## 4.2 Generation firms respond to incentive to induce the scarcity condition

Although the largest three generation firms frequently had the ability to create a scarcity condition, they do not always have the incentive to do so. Equation (3) shows how the short-run profits for the firm depend on whether or not the wholesale price exceeds the scarcity price. When a profit-maximizing firm has the ability to create the scarcity condition, it will only do so if the profits under the scarcity condition are greater than profits without the scarcity condition.

We empirically analyze the choices made by the largest generation firms during the days and hours in which they had the ability to create a scarcity condition. For this analysis, we calculate the best response of the firm to the generation bids of the other firms on that day. This best response will depend on the firm's hourly forward contract position, its daily firm energy allocation, and its generation costs. We ignore revenue from forward contract and firm energy sales, because the prices and quantities of these contracts are fixed and do not depend on the short-term market decisions.

The calculation of the best response is complicated by the non-separability of the firm energy payments across hours of the day. When the scarcity condition is triggered, the firm energy refund or payments depend on the generation in every hour of the day, not just the scarcity hours. The refund or payment also depends on the net firm energy position of the other firms in the market. Figure 7 illustrates the profit-maximization problem for one day for Emgesa. There were three hours in which it would have been optimal for Emgesa to withhold generation and induce the scarcity condition. In these hours, the optimal generation was less than the forward contract quantity—however, this did not matter because the forward contract obligation was capped at the scarcity price. Emgesa could have further increased its profits by increasing its generation in the non-scarcity hours and reducing its firm energy obligation in the scarcity hours. The actual market outcomes on this day showed that scarcity condition was triggered in the three hours for which it would have been optimal for Emgesa.

The profit-maximizing choice of quantities depends on the marginal generation costs. Most of the generation capacity of the three largest firms in the Colombian market is hydroelectric. Although there is no direct monetary cost of using water, the firms' production decisions are determined by the opportunity cost of water usage. An extra megawatt-hour of water released from a reservoir will not be available to produce electricity at a later date, potentially when the price is higher. The challenge for our analysis of profit-maximizing behavior is that the opportunity cost of water used by the firms is unobserved.

For an assumed value of the opportunity cost $c_w$, we calculated the marginal cost curve based on the plant-level capacity data, the heat rates of the thermal plants, and confidential data on the thermal fuel costs. Given this marginal cost curve, we solved a non-linear optimization problem to find the profit-maximizing combination of hourly quantities, as shown in Figure 7. This optimization used the Subplex optimization algorithm, a variant of the derivative-free Nelder-Mead algorithm (Ypma, 2014). We repeated this procedure using a grid of opportunity cost values. Each opportunity cost gave a vector of the optimal hourly thermal and hydro generation quantities for that day.

Using the optimal generation costs, we recovered an estimate of the monthly opportunity cost of water implied by the observed thermal and hydroelectric generation during the month. For each firm, month, and opportunity cost $c_w$, we calculated the sum of squared deviations between the hourly optimal ($opt$) and actual ($act$) generation quantities, both hydro ($H$) and thermal ($F$):

$$SSD(c_w) = \sum_{k=H,F} \sum_{t=0}^{T} (q_{kt}^{opt}(c_w) - q_{kt}^{act})^2 \quad (4)$$

The opportunity cost for each firm and month is the $c_w$ that minimizes Equation (4).

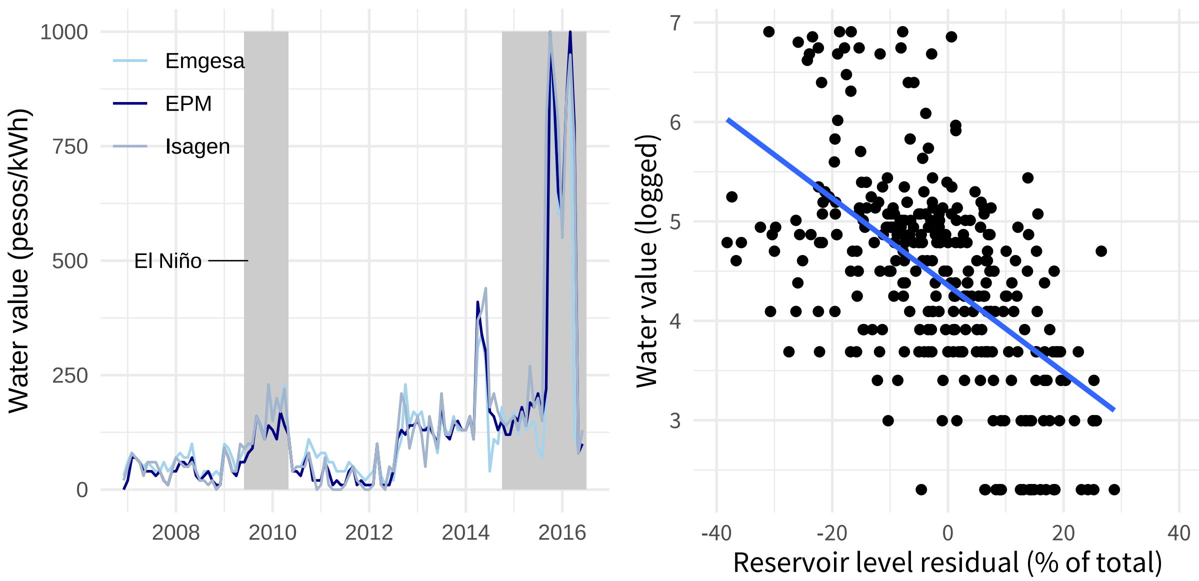
We solved the optimization problem to find the monthly opportunity cost of water and hourly profit-maximizing generation quantity, for each day from December 1, 2006 to June 30, 2016, for the three firms EPM, Emgesa, and Isagen. Figure 6 shows the implied monthly water values. These are correlated across the three firms, reflecting the commonalities in their hydrological and market conditions. The opportunity cost is highest during the two El Niño periods when water was scarcest (left panel). There is a negative correlation between the opportunity cost for a firm and its seasonally-adjusted reservoir levels (right panel).

The optimization procedure gives the best-response prices and quantities for each hour for the three firms, accounting for the short-run incentives of the reliability payment mechanism. We can compare these profit-maximizing outcomes to the observed prices and quantities. In particular, we focus on the triggering of the scarcity condition by withholding sufficient generation so that the market-clearing price exceeds the scarcity price. Whether this is optimal will depend on the firm energy and forward contract positions and the shape of the residual demand in each hour.

During the 115 months in the data, there were 13,575 hours when EPM could choose between the scarcity and non-scarcity condition. Most of the time, profits would be higher if the scarcity condition were avoided. For EPM, in 1,274 of the hours in which it had a choice, profits would be higher if the scarcity condition occurred (second block of Table 1). In 90 percent of these hours, the scarcity condition did occur. This result confirms that EPM usually created a scarcity condition when it had the ability and incentive to do so.

For the other 12,301 hours (90.6 percent) in which EPM had a choice, profits would be higher if the scarcity condition were avoided. In 98 percent of these hours, the scarcity condition did **not** occur. That is, in most of the hours in which EPM had the ability but not the incentive to create a scarcity condition, EPM ensured that the scarcity condition did not occur. These two sets of results are remarkable because EPM did not know the exact residual demand curve it would face during each of these hours. Nevertheless, it was able to make the *ex post* profit-maximizing choice between inducing scarcity conditions or avoiding scarcity conditions with at least 90 percent accuracy.

**Figure 6:** Monthly opportunity cost of water for the three largest firms, 2006–2016

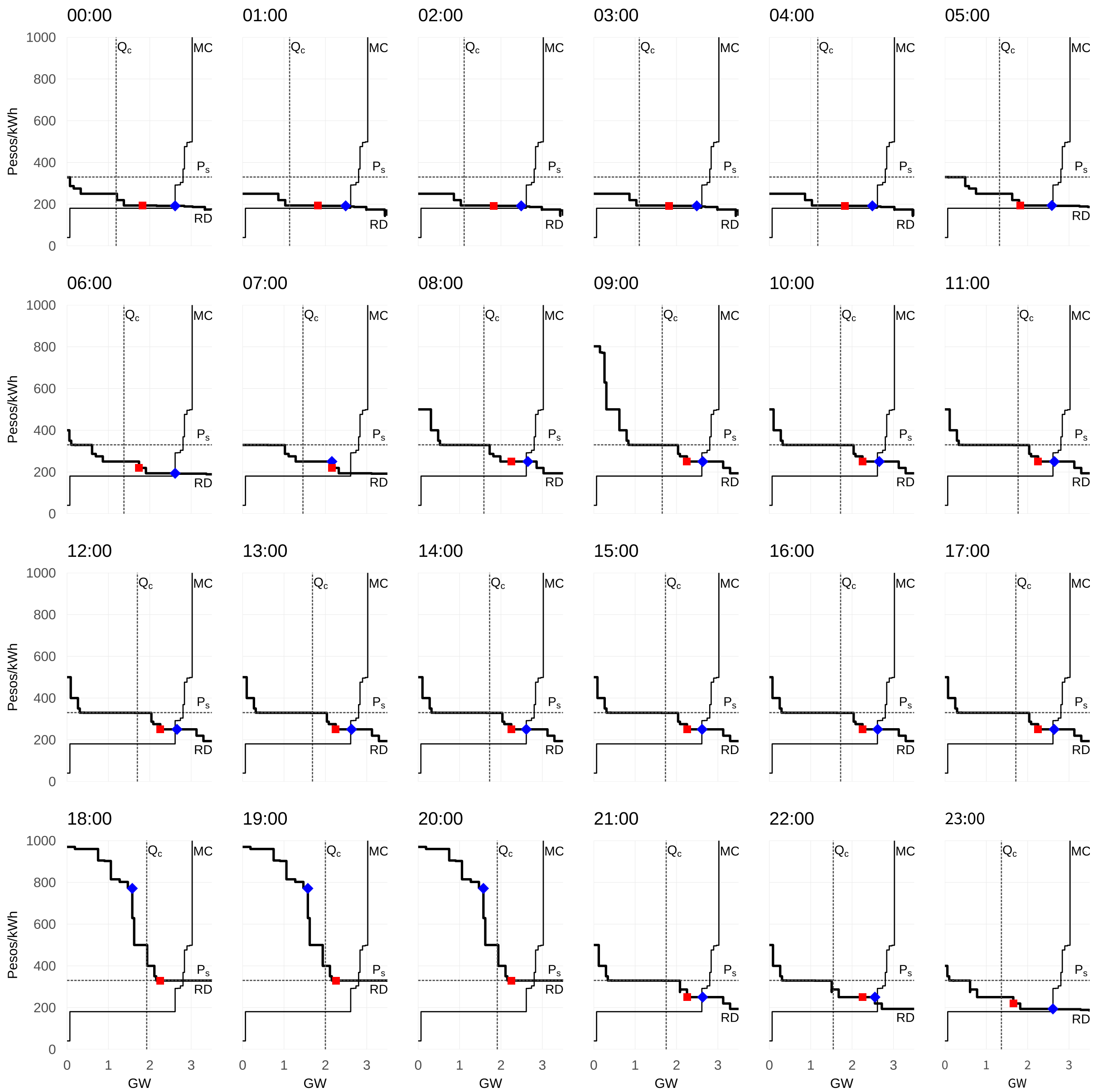



Notes: The left figure shows the monthly opportunity cost of water for each firm, calculated as the $c_w$ that minimizes Equation (4). The shaded dates are the El Niño periods when hydro inflows are reduced. The right figure shows the correlation between the logged opportunity cost and the reservoir levels for each firm, as a percentage of total capacity. The reservoir levels are seasonally-adjusted using a regression of the reservoir levels on firm-by-month dummies for the period 2000 to 2018.

The results are similar for the other two large generation firms, Emgesa and Isagen. There were 447 hours in which Emgesa had the ability and incentive to create a scarcity condition, and in 89 percent of these hours the scarcity condition occurred. For Isagen, there were 871 hours when it had the ability and incentive to create a scarcity condition, and this occurred in 75 percent of these hours. Conversely, there were 8,971 hours in which Emgesa had the ability but not the incentive to create a scarcity condition, and the scarcity condition was avoided in 97 percent of these. For Isagen, the scarcity condition did not occur in 96 percent of the hours in which it had the ability but not the incentive to induce scarcity.

Overall, these results provide strong evidence that the largest generation firms recognize the incentives created by the scarcity mechanism and respond to these incentives in their bidding behavior. Most of the time, profits would be lower under the scarcity condition, and in these hours the firms bid in a manner to avoid crossing the scarcity threshold. In a small number of hours, profits would be higher under the scarcity condition, and in these cases the firms bid in a manner that attempted to create scarcity.

**Figure 7:** Profit incentive for choosing between scarcity and non-scarcity condition




Notes: The figure shows the residual demand *RD* faced by Emgesa for each hour on May 25, 2015, plus the hourly contract position $Q_c$, the scarcity price $P_s$, and the marginal cost curve *MC*. Noise has been added to the marginal cost curve to mask confidential cost information. The diamond points show the best-response quantities and prices that would maximize profits for the day. The square points show the actual quantities and prices in each hour. There are three hours for which it would be profit-maximizing for Emgesa to withhold generation and create a scarcity condition: 18:00, 19:00, and 20:00. The realized prices in these hours were above the scarcity price.

**Table 1:** Ability and incentive to choose between scarcity and non-scarcity conditions, for the three strategic firms

| | Emgesa | EPM | Isagen |
|---|---|---|---|
| Non-scarcity hours | 70018 | 66639 | 74548 |
| Forced scarcity hours | 4564 | 3786 | 4386 |
| Scarcity/non-scarcity choice hours | 9418 | 13575 | 5066 |
| Total hours | 84000 | 84000 | 84000 |
| Hours when scarcity condition was optimal | 447 | 1274 | 871 |
| % which were scarcity | 88.8 | 90.1 | 75.4 |
| % which were non-scarcity | 11.2 | 9.9 | 24.6 |
| Hours when scarcity condition was not optimal | 8971 | 12301 | 4195 |
| % which were scarcity | 3.0 | 2.4 | 4.4 |
| % which were non-scarcity | 97.0 | 97.6 | 95.6 |

Notes: The top section of the table shows the classification of hourly residual demand into the three categories shown in Figure 4, for the three strategic firms. The bottom section of the table focuses on the hours in which the firm had the ability to choose between scarcity and non-scarcity. These hours are classified based on the profit-maximizing choice for the firm between the scarcity and non-scarcity condition. For each choice, the percentage of hours in the two categories are shown.

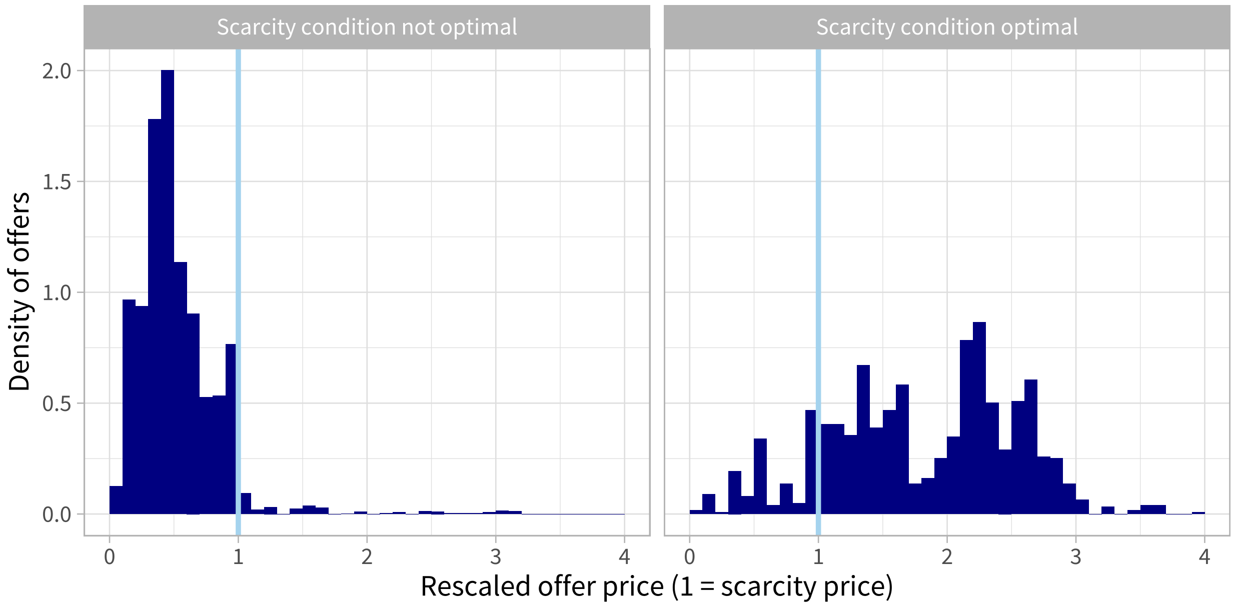
## 4.3 Bidding behavior reflects the incentives of the scarcity mechanism

In the previous subsection, we showed that the market outcomes—whether or not the scarcity condition occurred—were strongly associated with the profit-maximizing incentives for the generation firms. In this section, we show direct evidence of the firms' responses to these incentives in their bidding behavior.

For each firm, we focus again on the hours in which it had the ability to choose whether or not the scarcity condition occurred. We then compare the distributions of generation bid prices for the hours when the firm did and did not have the incentive to induce scarcity, as defined above. In each hour, we calculate the highest accepted bid (that is, the highest bid with positive dispatched generation). To compare the bids across different months with different scarcity prices, we scale all of the price bids by the scarcity price. A price of 1 would be a price bid that exactly equals the scarcity price in effect at the time of the bid. A scaled price greater than 1 would be a bid above the scarcity price, potentially inducing the scarcity condition. A scaled price less than 1 corresponds to a bid below the scarcity price.

For the 12,301 hours in which EPM had the incentive to avoid creating a scarcity condition, there is a high degree of bunching of the accepted bids just below the scarcity

**Figure 8:** Generation price offers for EPM respond to incentives to induce or avoid scarcity condition




price (Figure 8). This distribution of bids is consistent with EPM recognizing its incentive to avoid scarcity and submitting generation bids that would do so. Conversely, for the 1,274 hours in which EPM had the incentive to create a scarcity condition, nearly all of its bids were above the scarcity price.

## 5 Counterfactual market outcomes

The empirical results in the previous section demonstrate that the firm energy mechanism in Colombia creates incentives for generation unit owners with the ability to exercise unilateral market power to withhold output to create scarcity conditions. In this section, we assess the cost to Colombia electricity consumers of these incentives for inefficient behavior by suppliers by computing a number of counterfactual market outcomes.

We formulate a dynamic Cournot oligopoly model of quantity-setting behavior by strategic electricity generators facing a competitive fringe based on Colombian market. The model is similar to the model of the Western United States electricity market in Bushnell (2003). We calibrate this model to the annual May to April of the following year hydroelectric cycle in Colombia and present results for the most recent El Niño event in 2015 and 2016. We assume there are three strategic firms: EPM, Emgesa, and Isagen. The remaining generators behave competitively. The dynamic aspect of the model arises from the constraint on total water availability over the annual cycle. This introduces a shadow price for water each hour of the year into the model, with firms choosing their optimal generation from each generation unit they own each hour of the year.

We present model results for two main cases. The first approximates the reality of Colombian electricity market with the interaction between forward contract and firm energy incentives discussed in Section 4. The second case presents a counterfactual without the incentives created by the firm energy mechanism, holding fixed the existing forward contract obligations. By comparing the results from these two cases, we can describe the potential distortions that arise from the firm energy mechanism. We show how the magnitude of these distortions vary based on water availability. In particular, it is during dry years with scarce water that the firm energy incentives cause the greatest shift away from prudent reservoir management.

## 5.1 Model structure

We develop a stylized model of the Colombian wholesale electricity market. There are $T$ periods and $S$ strategic firms. The strategic firms have both hydro and thermal generation assets. In addition, there is an aggregated price-taking fringe firm that also has both hydro and thermal generation. All firms face a constraint on their aggregate hydro production across the $T$ periods and each firm $i$ chooses its optimal hydro allocation in period $t$, $q^h_{it}$.

The market price, $p_t$, depends on total generation in period $t$. The inverse demand is assumed to be linear and is given by Equation (5). Here $Q_t$ is the aggregate generation in period $t$, summing across the strategic and fringe firms. The demand parameters $a_t$ and $b_t$ may vary each period.

$$p(Q_t) = (a_t/b_t) - Q_t/b_t \tag{5}$$

Profits for the strategic firms depend on their hourly forward contract position, $q^c_{it}$, and their firm energy allocation, $q^f_{it}$. Both forward contracts and firm energy are assumed to be predetermined outside the model. The scarcity price, $p^s_t$, is also determined exogenously outside the model.

Following Equation (3), in the periods when $p_t$ is less than $p^s_t$, the profit for strategic firm $i$ is given by:

$$\Pi(q_{it}) = P_{it}^f q_{it}^f + P_{it}^c q_{it}^c + p(Q_t)q_{it} - p(Q_t)q_{it}^c - c(q_{it}) \tag{6}$$

The first-order condition for profit-maximization during non-scarcity periods is given by Equation (7):

$$\Pi'(q_{it}) = p'(Q_t)(q_{it} - q_{it}^c) + p(Q_t) - c'(q_{it}) \tag{7}$$

The alternative is that the price $p_t$ is greater than $p_t^s$. In these scarcity periods, the profit for the strategic firm $i$ is:

$$\Pi(q_{it}) = P_t^f q_{it}^f + P_{it}^c q_{it}^c + p(Q_t)q_{it} - p_t^s q_{it}^c - (p(Q_t) - p_t^s)q_{it}^f - c(q_{it}) \tag{8}$$

Based on this equation, the first-order condition for profit maximization during scarcity periods is given by Equation (9). This expression depends only on the firm energy quantity $q_{it}^f$ and not on the forward contract quantity $q_{it}^c$.

$$\Pi'(q_{it}) = p'(Q_t)(q_{it} - q_{it}^f) + p(Q_t) - c'(q_{it}) \tag{9}$$

For computational purposes we smooth the transition between the first-order conditions in Equations 7 and 9 at the threshold $p_t^s$. Equation (10) defines the marginal revenue of firm $i$ in period $t$.

$$MR(q_{it}) = p'(Q_t)\{q_{it} - q_{it}^c + 0.5(1 + \tanh(p(Q_t) - p_t^s))(q_{it}^c - q_{it}^f)\} + p(Q_t) \tag{10}$$

The hyperbolic tangent function is 1 for large positive arguments and -1 for large negative arguments. The formulation in Equation (10) switches the contract quantity in the first-order condition from the forward contract quantity to the firm energy quantity when the price $p(Q_t)$ exceeds the scarcity price $p_t^s$.

For strategic hydro generator $i$, the hydro generation in period $t$, $q_{it}^h$, satisfies Equation (11), where the marginal revenue is given by Equation (10).

$$MR(q_{it}) - \lambda_i - \overline{\phi}_{it} + \underline{\phi}_{it} = 0 \tag{11}$$

The total generation by firm $i$ in period $t$, $q_{it}$, is the sum of hydro generation $q_{it}^h$ and the generation from each of its thermal generation plants, $q^o_{ikt}$. The Lagrange multiplier on the water constraint faced by firm $i$ is $\lambda_i$. The total hydro generation over the $T$ period must be no greater than the amount of water available for the firm, $W_i$. As shown by the complementarity condition in Equation (12), $\lambda_i$ is 0 if the water constraint does not bind, and strictly positive otherwise.

$$\sum_t q^h_{it} \leq W_i \perp \lambda \geq 0 \tag{12}$$

The other Lagrange multipliers in Equation (11) are from the constraints on the minimum and maximum hydro output each period for generator $i$. If hydro output is exactly equal to the minimum generation then $\underline{\phi}_{it}$ will be strictly positive, otherwise it will be zero. If hydro output is exactly equal to the maximum generation then $\overline{\phi}_{it}$ will be strictly positive, otherwise it will be zero.

For the strategic thermal generator $i$, thermal generation from plant $k$ in period $t$, $q^o_{ikt}$, will satisfy Equation (13).

$$MR(q_{it}) - MC(q^o_{ikt}) - \overline{\gamma}_{ikt} + \underline{\gamma}_{ikt} = 0 \tag{13}$$

Marginal cost of plant $k$ is assumed to be a non-decreasing linear function of output $q^o_{ikt}$. As in the case of hydro generation, $\overline{\gamma}_{ikt}$ and $\underline{\gamma}_{ikt}$ are multipliers on the maximum and minimum output constraints for the plant.

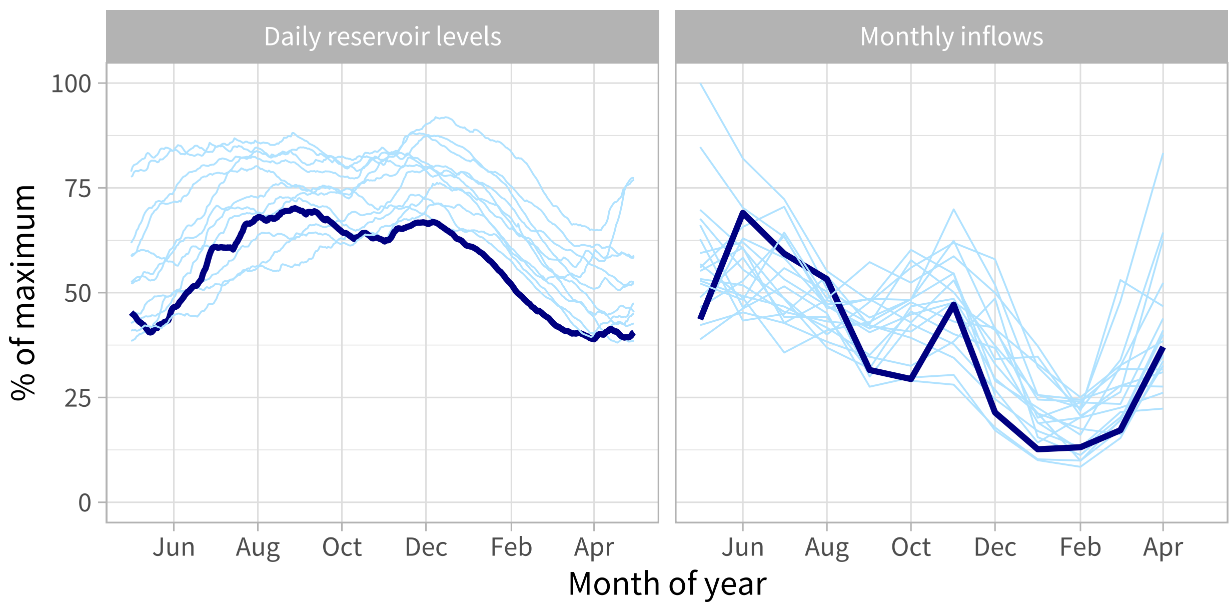
Hydro and thermal generation by the competitive fringe has exactly the same structure as Equations (11) and (13), except that the marginal revenue term $MR(q_{it})$ is replaced by the market price $p(q_{it})$. In particular, the price-taking fringe hydro generator still faces a constraint on total water availability and allocates the scarce water across the $T$ periods. However, instead of equalizing marginal revenue across periods, the competitive firm allocates water to equalize price across periods, subject to constraints on its maximum and minimum generation each period. Similarly, the competitive thermal producer will choose its level of production to equate price with marginal cost, again subject to constraints on its hourly minimum and maximum generation.

The above conditions define a nonlinear complementarity problem, with the nonlinearity introduced by the transition between $q^c_{it}$ and $q^f_{it}$ in Equation (10). The solution to the problem will be an equilibrium of the dynamic oligopoly problem of strategic producers facing a competitive fringe, subject to firm-level constraints on hydro availability. We formulate this problem in the AMPL modelling language (Kernighan et al., 1993) and solve it using the PATH solver (Dirkse and Ferris, 1995).

## 5.2 Calibration

We calibrate the model to a hydrological year from May 1, 2015 to April 30, 2016. The choice of period was based on the observed pattern of hydrological conditions in Colombia (Figure 9). In a typical year, the lowest levels in the hydro reservoirs occur in April. Inflows are greatest between April and August. Reservoirs are usually at their highest levels in September or December. Finally, inflows are lowest between January and March, leading to a rapid run-down in storage levels during that period. For hydro generators, the choice to use water or to store it between May and August, when rainfall is plentiful, will affect the quantity that they can generate between December and March.

**Figure 9:** Historical reservoir levels and hydro inflows, 2006–2017




Notes: Each line represents one year. The bold lines highlight data for 2015–16. Inflows are defined as the change in reservoir level, plus generation and spill over the period. Monthly inflows are shown as the percentage of the maximum observed monthly inflow in the data. Reservoir levels are shown as the percentage of the reservoir capacity. Data on reservoir levels and spill is from XM Compañía de Expertos en Mercados (2019a).

We use data on the hourly generation from each plant in the Colombian market, including small facilities that do not participate in the market dispatch. For hydro generation, we separate reservoir and run-of-river plants. In Colombia, there are 23 hydro reservoirs that supply water to 21 plants. We aggregate the run-of-river generation to the firm level and treat the hourly generation as fixed at its observed value. We also include the output from wind and cogeneration facilities in this non-dispatched aggregate.

The choice to model the hydro reservoirs as a firm-level aggregate is appropriate in Colombia because there are few hydro chains that operate along a single river. For markets such as Brazil that have rivers with multiple hydro facilities, the output of one plant will be the input for another plant. This creates strategic interdependencies in the production decisions that might complicate aggregation (Moita, 2008; Moita and Monte, 2020).

We have the nameplate capacity for each of the reservoir hydro plants which we aggregate to give the maximum hourly output by firm. For minimum output, there are a small number of plants that have permission from the regulator to submit minimum output levels as part of their price and quantity bids.[7] In practice, the observed minimum hydro output in the data exceeds the minimum output in the bids. For example, this may be because of ancillary services provided by the large hydro generation plants. For each month, we use the minimum aggregated output in that month as our estimate of the minimum hourly output by firm.

For each of the hydro reservoirs, we have daily data on the storage levels, maximum usable storage, inflows, and spilled water (XM Compañía de Expertos en Mercados, 2019a). These variables are reported in both cubic meters and in kilowatt-hour equivalents. Because of the complexity involved in the conversion of water volumes to energy equivalents, there are often substantial discrepancies in the hydro balance identities. We avoid this problem by imputing the daily reservoir inflows from the change in storage levels, plus the daily generation and the reported spill. For each firm, we sum the reservoir inflows over the year for each of the plants, to obtain a single number for the hydro energy available for the year. In the model, the firm reallocates this energy across the hours in the year, subject to the constraints on the minimum and maximum hourly generation.

For thermal generation, we have daily data on the fuel consumption of each plant and unit in MMBTU (XM Compañía de Expertos en Mercados, 2019b). We combine the fuel data with unit-level generation to calculate an implied average heat rate for each unit and year. For costs, we obtained confidential daily plant-level data on the energy and transportation components of the fuel cost, in pesos per MMBTU. Combining the fuel prices with the heat rates provides the marginal fuel cost in pesos per kWh.

We obtained the other costs for thermal generators from the daily system information

7. This is required because, unlike most wholesale electricity markets, firms in Colombia submit single price and quantity bids in the day-ahead market. Quantities may vary each hour but the bid price is the same for all hours of the day.

reports (XM Compañía de Expertos en Mercados, 2016). We use the regulated values from these reports for thermal operating and maintenance costs, separately for plants using coal, natural gas, and other fuels. In addition, we include the taxes and other charges paid by generators. These include charges for ancillary services, a tax to support grid expansion to unconnected regions, a tax for funding environmental regulators, and a charge to fund the capacity payments. The capacity charge is about US$16/MWh. We add the fuel cost, operating and maintenance costs, and the taxes and charges, to calculate the daily marginal cost for each thermal generating unit. For the counterfactual scenarios without the capacity payment mechanism, we exclude the capacity payment charge from the thermal marginal cost.

We use the nameplate capacity as the maximum potential generation from each unit and assume that the minimum generation from each unit is zero. In our model, we do not consider possible restrictions on fuel availability. We also do not model the ramping up and down of thermal generators would introduce dynamic constraints on the feasible program of hourly generation (Wolak, 2007; Cullen, 2011).

The three strategic firms own between one and three thermal plants. We model generation output from each of these plants individually. For the plants with multiple units, we rank the units in order of increasing marginal costs. The minimum marginal cost is the marginal cost of the lowest-cost unit. The slope of the marginal cost curve for each plant is taken from a straight line between the minimum and maximum marginal cost units, for generation quantities between zero and the maximum generation capacity of the plant.

We aggregate the thermal generation plants owned by the competitive fringe producers. For each month, we rank the plants and units in order of increasing marginal cost, allowing for changes in capacity and marginal cost from month to month. We then approximate the thermal marginal cost curve using a piecewise-linear function with three segments and two breakpoints (Muggeo et al., 2008). Maximum generation output is set to the total nameplate capacity of the plants.

The quantity of electricity demanded is set to the hourly aggregate generation. For each month, we rank the hours by the level of demand, and split the data into twelve evenly-sized bins based on percentiles of hourly demand. In the model, each bin represents between 56 and 62 hours of real-world generation. For each hour, we calibrate a linear inverse demand function. Following Bushnell (2003), we keep the slope of the inverse demand equal across all hours, with the slope calibrated to give a price elasticity of demand of -0.05 for the highest-demand hour. Given this assumption on the slope, the intercept of the inverse demand is calibrated in each binned hour to match the mean generation and mean price for that hour.

For the strategic producers, we use hourly data on the net forward contract position and their firm energy obligation. Unlike in most other wholesale electricity markets, the quantity of forward contract obligations in publicly available, by firm and hour (XM Compañía de Expertos en Mercados, 2019c).[8] The annual firm energy obligation for each plant is also publicly available. We use confidential data from the market operator with the calculation of the hourly firm energy obligation by firm.[9] We calculate the mean contract position and firm energy obligation for each of the binned hours in our sample. Finally, we use data from the market operator with the monthly scarcity price.

Table 2 provides a summary of actual generation data for 2015–16, aggregated to the 144 representative periods used in the model calibration. Most of the generation output of the three strategic firms is from reservoir hydro, supporting the modelling decision to focus on the water allocation problem. The maximum observed hydro generation from each of the three firms is about 2 GW, compared to the maximum system generation of 9.6 GW. For thermal generation, the three major hydro producers have limited thermal capacity, relative to their hydro capacity and to the total thermal capacity in the system. Ownership of the remaining thermal generation is divided among multiple small firms, several of whom own just one generation plant.

## 5.3 Results

We show results of two scenarios for the 2015–16 hydrological year (May 2015 to April 2016). First, we use the existing firm energy obligations and forward contract obligations for each of the strategic generators. We show the results for the equilibrium of the dynamic game, given these obligations and the exogenous path of the scarcity price during 2015–16. Second, we solve a counterfactual scenario with no firm energy and with only the forward contract obligations.

The crucial determinant of the results is the quantity of hydro inflows during the year. These determine how much water the strategic firms have available to allocate across the

8. The three large generation firms own electricity retailing businesses that sell electricity to commercial users, typically at fixed prices. EPM also owns several distribution networks and electricity retailers serving regulated (residential) users. We calculate overall net forward contract positions of the vertically-integrated groups, including the fixed-price retail load obligations.

9. There is a small secondary market that allows generators to trade their firm energy obligations. This means that the hourly shares of firm energy obligations may differ slightly from the annual share.

**Table 2:** Summary of generation data for 2015–16 hydrological year

| | Generation (GW) | | | Storage (TWh) | |
|---|---|---|---|---|---|
| Firm | Min | Mean | Max | Min | Max |
| **Reservoir hydro** | | | | | |
| Emgesa | 0.75 | 1.33 | 2.01 | 2.87 | 5.71 |
| EPM | 0.12 | 1.11 | 1.89 | 2.66 | 3.84 |
| Isagen | 0.31 | 0.98 | 1.84 | 0.30 | 1.20 |
| Others (fringe) | 0.13 | 0.80 | 1.47 | 0.02 | 1.08 |
| **Run-of-river hydro** | | | | | |
| Emgesa | 0.00 | 0.05 | 0.19 | | |
| EPM | 0.08 | 0.13 | 0.23 | | |
| Isagen | 0.02 | 0.05 | 0.08 | | |
| Others (fringe) | 0.30 | 0.43 | 0.59 | | |
| **Thermal** | | | | | |
| Emgesa | 0.07 | 0.21 | 0.33 | | |
| EPM | 0.07 | 0.26 | 0.41 | | |
| Isagen | 0.16 | 0.23 | 0.27 | | |
| Others (fringe) | 1.08 | 2.02 | 2.71 | | |
| Total generation | 5.53 | 7.61 | 9.61 | | |

Notes: Summary statistics are based on the aggregation of generation data to the 144 representative hours used in the model (twelve hours for each of twelve months). Run-of-river hydro includes a small amount of non-dispatchable wind and cogeneration.

months of the year. The water availability determines whether the scarcity price will be binding. To illustrate the effects of the hydro inflows, we consider counterfactual scenarios with higher inflows (an increase of 10 percent or 20 percent compared to 2015–16 levels) and with lower inflows (a decrease of 10 percent or 20 percent). For these inflow scenarios, we hold everything else in the model constant: firm energy and contract obligations, thermal costs, electricity demand, and the scarcity price.

Table 3 shows the base case results for the 2015 inflows, comparing the scenario with both firm energy and forward contracts to a counterfactual scenario with only forward contracts. Subsequent sections of the table show this comparison for the cases with higher and lower hydro inflows. For each case, we show the maximum price, mean price, and the share of hours for which the price exceeds the scarcity price. To illustrate the changes

**Table 3:** Summary results from dynamic oligopoly model to show effect of firm energy incentives on wholesale market indicators

| | Price (US$/MWh) | | | Storage | Hydro | Thermal |
|---|---|---|---|---|---|---|
| | Max | Mean | % > $P_s$ | Max % | Wet TWh | Max GW |
| **Base case: 2015 water** | | | | | | |
| Firm energy + contracts | 110.75 | 104.73 | 66.39 | 80.30 | 11.84 | 1.54 |
| Forward contracts only | 92.82 | 87.30 | 0.00 | 86.06 | 10.92 | 1.49 |
| **High water (+20%)** | | | | | | |
| Firm energy + contracts | 85.76 | 76.40 | 0.00 | 95.20 | 13.04 | 1.11 |
| Forward contracts only | 75.25 | 64.39 | 0.00 | 96.37 | 12.90 | 1.14 |
| **High water (+10%)** | | | | | | |
| Firm energy + contracts | 95.83 | 89.55 | 0.00 | 87.13 | 12.53 | 1.19 |
| Forward contracts only | 84.17 | 76.34 | 0.00 | 87.59 | 12.45 | 1.19 |
| **Low water (-10%)** | | | | | | |
| Firm energy + contracts | 125.13 | 117.64 | 74.86 | 79.40 | 10.21 | 1.96 |
| Forward contracts only | 104.20 | 100.03 | 58.20 | 82.10 | 9.78 | 1.86 |
| **Low water (-20%)** | | | | | | |
| Firm energy + contracts | 135.17 | 127.60 | 83.24 | 74.69 | 9.18 | 2.55 |
| Forward contracts only | 115.91 | 110.35 | 74.86 | 75.90 | 8.99 | 2.43 |

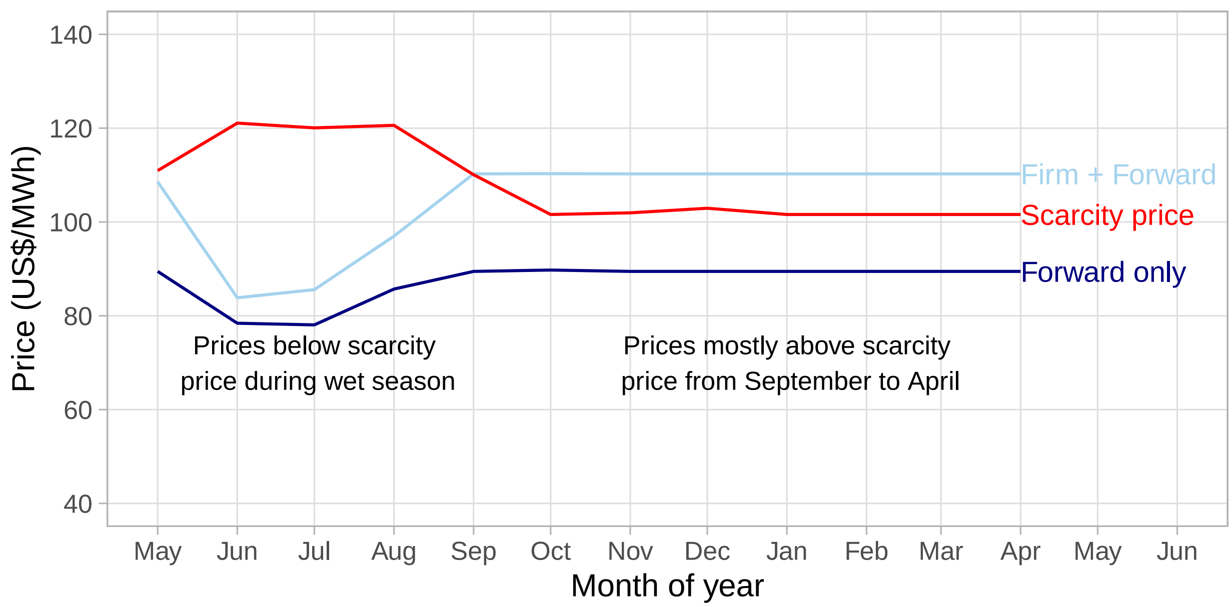
Notes: Each line represents a summary of the results from a separate case. For the "firm energy and contracts" cases, firm energy obligations and net forward contracts positions for the three strategic firms are set to their observed levels for 2015–16. For the "forward contracts only" cases, firm energy is ignored and the forward contract position is set to its 2015–16 levels. The bottom four groups show to changes to the 2015–16 inflows, either higher or lower, keeping everything else the same.

in hydrological risk, we report the maximum level of the reservoirs of the strategic firms during the year. We also report the hydro generation of the strategic firms during the months with the highest rainfall, between May and August. These two measures capture the extent to which strategic firms are conserving water during the wet season. Finally, we report the average cost of highest hourly thermal generation observed during the year.

In all water scenarios, both the maximum and mean market prices are considerably lower without the firm energy mechanism. Most of the difference in price is caused by the exclusion of the firm energy charge from the marginal costs of thermal generators. Figure 10 illustrates the change in prices for the base case scenario. With only forward contracts, the price would be consistently lower throughout the year, and there would be no periods in which it exceeds the scarcity price.

For hydro storage, the maximum reservoir level over the hydrological cycle is lower with firm energy. For the base case, the maximum storage level would be 80 percent with firm energy and 86 percent without firm energy. The reason for the lower storage levels is apparent from the hydro generation totals for the period from May through August. With firm energy, hydro generation is higher and total thermal generation is lower. In the base case scenario, strategic hydro generation is 11.84 TWh between May and August with firm energy and 10.92 TWh with only forward contracts.

**Figure 10:** Comparison of monthly mean prices between counterfactual scenarios

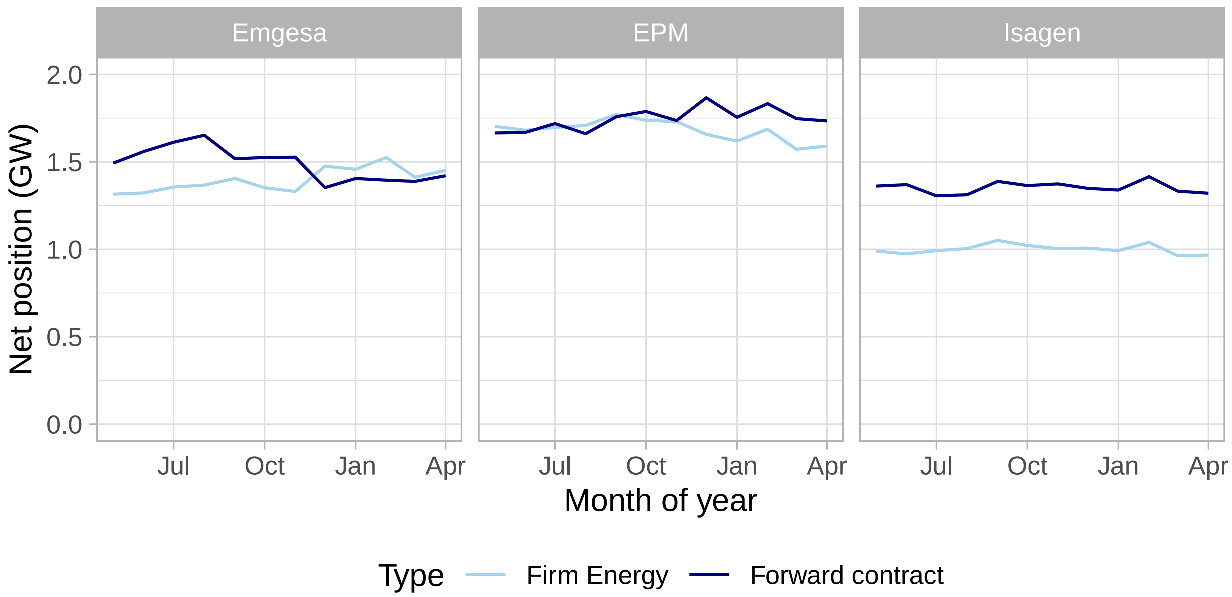



Notes: The two price lines show the mean monthly prices output from our model for the base case, with and without the firm energy incentives. The third line on the graph shows data for the monthly scarcity price. Scarcity periods occur when the market price exceeds the scarcity price.

In the firm energy scenario, higher hydro generation implies lower thermal generation during the wet season. Because the total amount of water available is assumed to be fixed across the year, thermal generation will therefore be higher during the dry season. The highest hourly thermal generation during the year is 1.54 GW with firm energy and 1.49 GW with only forward contracts (final column of Table 3). Holding fixed total demand, the total annual cost of thermal generation will be higher with firm energy, given the convex marginal cost curve for thermal plants. In our model, the effect of reallocating thermal generation between periods is partially offset by the lower price and higher quantity demanded, given our assumption that demand is not perfectly inelastic.

The incentives for each of the strategic firms in the two cases differ based on the relative magnitude of their firm energy and forward contract obligations. Figure 11 shows data for the monthly mean firm energy and forward contract positions for the strategic firms for 2015–16. We treat these quantities as fixed parameters in the modelling analysis. Figure 12 shows the monthly reservoir hydro generation for the three firms in the two model scenarios.

**Figure 11:** Comparison of monthly mean firm energy and forward contract obligations

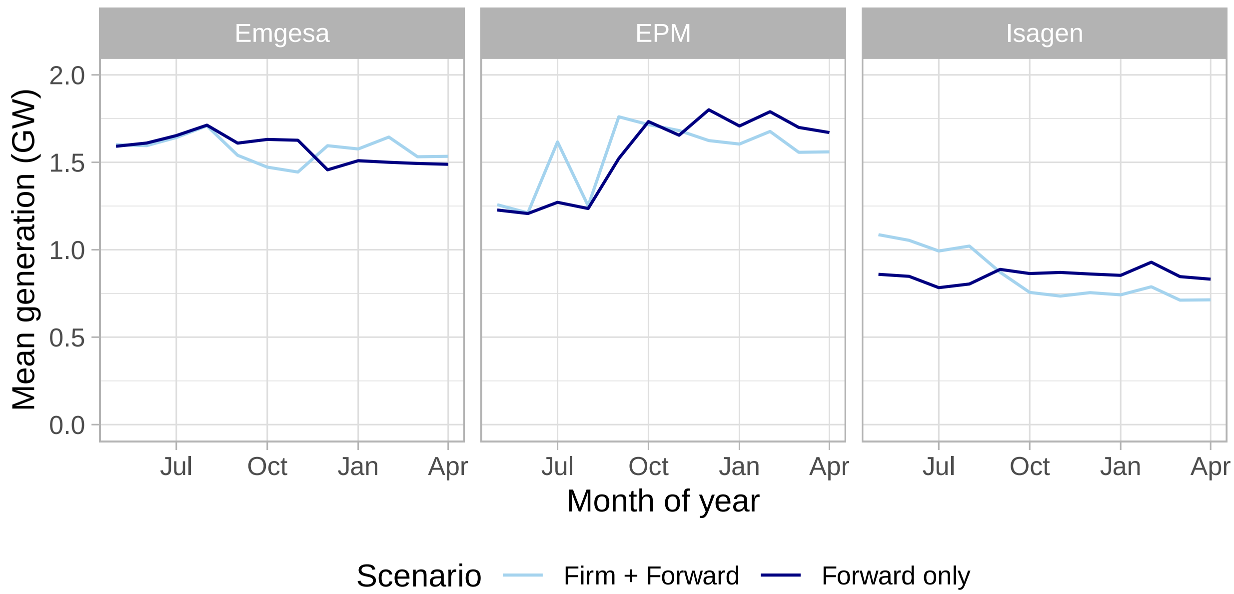



Notes: Each graph shows data for the monthly mean net forward contract position and the monthly mean firm energy obligation, for the three strategic firms. Hourly contract data is from XM Compañía de Expertos en Mercados (2019c). XM also provided data on firm energy obligations.

For Isagen, its forward contract obligations are greater than its firm energy quantities for all months of the year. In the scenario with firm energy and forward contracts, it is the firm energy obligation that binds during scarcity periods. This obligation is lower than the forward contract quantities that binds for the scenario without firm energy. This implies that in the firm energy scenario, Isagen has a greater incentive to exercise market power during the scarcity months than in the counterfactual world without firm energy. The optimal generation output for Isagen reflects its response to these incentives (Figure 12). With firm energy, Isagen exercises market power by reallocating its water out of the scarcity period, lowering generation and increasing prices.

For EPM, its firm energy and forward contract obligations were similar in the first half of the year. After November, it had a forward contract obligation that exceeded its firm

**Figure 12:** Comparison of hydro generation output for strategic firms between counterfactual scenarios for the base water case




Notes: Each graph shows the mean monthly hydro generation output from our model for the two cases, with and without the firm energy incentives.

energy quantities. In the firm energy scenario, EPM had more incentive to exercise market power during the scarcity periods after November. Just as for Isagen, the optimal response would be to allocate less water for generation during the scarcity periods, compared to the counterfactual without firm energy. The modeled hydro generation for EPM matches this prediction. With firm energy, EPM reallocated water from the scarcity to the non-scarcity periods.

The net position of the third strategic firm, Emgesa, is more balanced than for Isagen. The forward contract obligation is greater for the first part of the year and then slightly lower than the firm energy obligation for the second part of the year. The greater similarity between the firm energy and forward contract obligations suggests that there will be less difference in the optimal generation path between the firm energy and forward contract scenarios. This prediction matches the model results, which show similar time paths of generation output in the two scenarios.

The reallocation of water created by the firm energy mechanism, illustrated in Figure 12, is greatest during low-water years such as 2015–16. If hydro inflows are sufficiently low, then during some hours of the year the market price will exceed the scarcity price. This leads to the distortion in production incentives described above. However, if hydro inflows are high, then the value of water and the market price will be low. Few hours will be scarcity hours. In that case, there is no change in strategic incentives. This can be observed in Table 3, where there is less difference in hydrological outcomes between the two cases in the high-water years.

We emphasize that the distortion in hydro generation during low-water years is not a quirk of our model or calibration exercise. It is created by a central feature of the firm energy mechanism. By design, the firm energy allocations for large hydro generators are conservative. Market regulators calculate the allocations from worst-case scenarios based on historical inflow data. Risk-neutral generators might reasonably take on forward contract obligations exceeding their firm energy allocation. In aggregate, this is what we see in the data: the three largest generation firms have more forward contract than firm energy obligations.

The observed reallocation of water across the year follows from the relative magnitude of the firm energy and forward contract quantities. When inflows are low—and when the scarcity price is sufficiently low compared to the marginal costs of thermal generation—then there will be scarcity hours during the year. During these hours, the constraint on the exercise of market power provided by the forward contract obligation is replaced by the laxer constraint from the firm energy obligation. In aggregate, the strategic producers will have a greater incentive to exercise market power during scarcity hours than they would have had during the same hours in the absence of the firm energy mechanism. This incentive leads to the reallocation of water away from the scarcity hours.

A claimed feature of the firm energy mechanism is that it protects electricity consumers against the exercise of market power by large generation firms (Cramton and Stoft, 2008). The scarcity price caps the electricity price paid by retailers, passed on to electricity consumers through their regulated tariffs. The high market prices during scarcity prices only change the transfers between generation firms based on whether they have a short or long firm energy position.

However, for hydro-dominated systems, it is a dangerous misconception that the only costs of market power are higher prices that create transfers from consumers to producers. Hydro generators exercise market power by reallocating their scarce water. The firm energy mechanism provides an incentive to shift water in the opposite direction to what we might expect under prudent reservoir management: using more water during the wet season and storing less water for the dry season. As shown in Table 3, this leads to productive inefficiency from an increase in generation by the most expensive and most polluting thermal generation during the dry season.[10] For consumers, this reallocation also reduces reservoir levels at the start of the dry season and increases the risk of catastrophic system blackouts if there are lower-than-expected inflows or unanticipated generation unit failures.

As in any modeling exercise, our dynamic model of strategic behavior in the Colombian electricity market contains several simplifications of reality. The most important assumption is that of perfect information. We assume reservoir owners know the total amount of water they have available and the residual demand they will face. This implies that at the start of the year they can choose their optimal allocation of water in each hour of the year. In reality, there is uncertainty about future inflows, demand, and supply from other firms. In our model context, firms might face uncertainty when they are making their generation decisions in May about whether a scarcity period will occur in October.

Nonetheless, the low hydro inflows that occurred between December 2015 and March 2016 could have been anticipated at the start of our model in May 2015. The consensus forecast in May 2015 attached a higher than 80 percent probability to an El Niño event occurring in December 2015 (International Research Institute for Climate and Society, 2015). Hydro inflows after that date, between June and August 2015, were among the highest of the previous decade (Figure 9). Given the information on the high probability of low inflows eight months later, prudent reservoir managers had an opportunity to store more water earlier in the year to reduce the risk of a future shortfall.

The second simplification in our model is the assumption of a perfectly competitive thermal fringe with a generation capacity set to nameplate capacity and constant marginal cost. In reality, Colombia has a small and illiquid market for generation fuel, especially natural gas. This creates an upward-sloping and potentially vertical short-run supply curve for fuel inputs. Regulators are concerned about shortfalls in thermal fuel availability during periods with low hydro generation.

# 6 Alternative Approach to Long-Resource Adequacy

Virtually all jurisdictions with formal wholesale electricity markets have regulatory mandates aimed at maintaining an adequate long-term supply of energy at a reasonable price.

10. Holding demand fixed, total thermal generation costs would be lower for the year from running more low-cost generation throughout the year, saving more water and avoiding the need to run the most expensive plants later. Asker et al. (2019) provide a related example of inefficiency costs from production misallocation in a dynamic oligopoly context, for the case of the world oil market.

In this United States, these are called long-term resource adequacy mechanisms. Mandating that retailers hold standardized forward contracts for energy that clear against the hourly short-term price at long enough horizons to delivery to allow new entrants to compete to supply these contracts can be used as an alternative mechanism for ensuring long-term resource adequacy. Wolak (2020) describes the reliability externality rationale for a long-term resource adequacy mechanism and why a standardized forward contract approach is likely to be least cost approach to addressing it in regions with significant intermittent resources.

There is increasing dissatisfaction in the United States with the capacity-based long-term resource adequacy processes. This is particularly the case for regions with significant intermittent renewable energy goals. The firm capacity of a generation unit is typically defined as the amount of energy that the generation unit can produce under extreme system conditions, which makes defining the firm capacity of intermittent renewable generation units difficult, if not impossible. In addition, capacity-based resource adequacy processes procure firm capacity up to a pre-specified multiple of the peak demand, typically around 1.15, which limits wholesale price volatility and the incentive for investments in storage and active participation of final consumers in the wholesale market.

An energy-based long-term resource adequacy process has the potential to reduce the total amount of generation capacity required to serve the annual demand for energy, which can allow consumers to pay lower average wholesale prices, despite an increase in wholesale price volatility. Consumers can be protected from a significant fraction of this wholesale price volatility by holding standardized fixed-price long-term contracts for energy. A liquid market for standardized forward contracts provides a mechanism for providing the necessary hedges against wholesale price volatility, as well as a mechanism for ensuring long-term resource adequacy if the contract purchases are made far enough in advance of delivery to allow new entry to occur.

The wholesale market regulator would mandate that all load-serving entities in the region hold until delivery fixed-price forward contracts purchased from this standardized market equal to a pre-specified fraction of their final demand. For example, the mandate could be that all retailers purchase 97 percent of their realized final demand 1 year in advance, 95 percent 2 years in advance, and 92 percent 3 years in advance. As discussed in Wolak (2020), these purchases would be made through periodic auctions and allocated to electricity retailers according to their shares of system load.

These mandates should be a sufficiently large fraction of system demand and continue far enough into the future to give the regulator sufficient confidence that energy adequacy will ultimately be achieved in the delivery year. To the extent the regulator is concerned that adequate generation capacity and other resources will be available to meet demand in the future, the regulator can increase the number of years in the future that the mandate exists from 3 years to, say, 5 years, and increase the percentage of demand that must be purchased in standardized fixed-price forward contracts in each year in the future, for example 97.5 per cent 1 year in advance, 95 percent 2 years in advance, 92.5 percent 3 years in advance, and 90 percent 4 years in advance.

As discussed in Wolak (2020), the hourly quantities for these forward contracts are tailored to the realized system load shape. This implies that retailers in the aggregate bear little short-term wholesale price risk. Assume that $w_h$ is the share total energy delivered to the system during of hour $h$ of the delivery period of the contract. If the total amount of energy purchased under the contract is $QF$, then the delivery quantity during hour $h$ of the delivery period is $w_h \times QF$.

The requirement that all retailers hold these standardized forward contracts is straightforward to implement. These contracts would be purchased through periodic auctions run by the market operator and oversee by the regulator and allocated to retailers based on their actual load shares. These contracts provide consumers with wholesale price certainty for virtually all of their final demand far in advance of delivery to obtain a competitive price and provides a revenue stream to generation unit owners far enough in advance of delivery to allow them to bring on line sufficient resources to meet demand. Finally, the regulatory mandate that all retailers hold these contracts, ensures liquidity in the futures market at the mandated horizons to delivery.

This mandate to hold these contracts to "delivery" does not rule out market participants entering into other bilateral hedging arrangements. For example, a renewable resource owner might enter into a cap contract with a thermal resource owner where the thermal resource owner provides price spike insurance for a fixed quantity of energy each hour in exchange for an up-front payment. For example, a 50 MW solar resource might purchase insurance against prices above \$100/MWh during the night-time hours of the day for the capacity of this resource, to hedge the risk of a price spike when its unit is unable to operate. In this case, the solar resource would pay an up-front fee to the seller of the contract in exchange for the payment stream $\max(0, P(spot) - \$100/\text{MWh}) * 50\text{MW}$ from the seller of the contract during each night-time hour during the contract period.

Using a standardized futures markets for energy as the basis for a long-term resource adequacy process has the following advantages. First, it is technology and capacity neutral. There is no need for the regulator to determine the firm capacity of a generation unit or set an overall capacity requirement. It leaves decisions about what is the least-cost mix of generation capacity, demand response, and storage needed to meet the demand for energy in the future to market participants, which are likely to be the entities best able to make these decisions. Second, it allows wholesale prices to reflect scarcity conditions that can make storage investments and active demand-side participation economic. This will increase the capacity factor of existing generation units, which allows the same annual demand to be met with less generation capacity, thereby reducing annual average wholesale prices.

The prices of these futures contracts can also be used to set the wholesale price component of the regulated retail price. For example, if the regulator would like to set the wholesale component of the regulated retail price for the coming year, it can use the weighted average price of these contracts delivering in the following four quarters. Because the retailer has been allocated this mix of contracts to meet its regulatory mandate, the regulator knows that the retailer can at least supply energy at a retail price that includes this wholesale price. In this way, the regulator is able to set the regulated retail price for a vertically integrated electricity retailer. It simply uses the average futures prices for the relevant delivery horizon as the wholesale energy price component of the retail price.

A final very favorable property of this mechanism is that it is ideally suited to an electricity supply industry with a significant share of intermittent renewable generation capacity where the firm capacity construct makes very little sense. The firm capacity of a generation unit is the amount of energy that can be produced from a generation unit under extreme system conditions. For a thermal resource, this is a relatively well-defined concept. It is typically equal to the capacity of the unit times its availability factor. However, the amount of energy a wind unit can produce on an extremely hot high-demand day with no wind is clearly zero, and the amount of energy a solar unit can produce at dusk on an extremely hot high-demand day is close to zero. Consequently, determining the firm energy of these resources is more of a political decision than a technical engineering decision. Consequently, paying for firm capacity from these units when they are very likely not to be available during stressed system conditions is costly for consumers because they are paying for something they are not getting (firm capacity from the intermittent units) and paying for more firm capacity from dispatchable units to replace the firm capacity they are not getting from the intermittent units.

The regulator-mandated standardized market for long-term contracts approach to long-term resource adequacy avoids this issue by focusing on ensuring there is sufficient energy to meet demand in the future. As discussed in McRae and Wolak (2016) and Wolak (2020), this mechanism creates incentives for intermittent renewable resources to re-insure their forward energy sales with dispatchable thermal resources so that their forward commitment for energy in the future will be met.

## 7 Discussion

One rationale for the new capacity payment mechanism set up in 2006 was to provide financial support for new and existing thermal generators, in order to keep them available as backup for El Niño years. However, as illustrated by the market outcomes during the 2015–16 El Niño event, the mechanism has not been completely successful at achieving this goal (McRae and Wolak, 2016). Several new thermal generation plants that were assigned firm energy in the auctions were never built or were completed far behind schedule. Some existing thermal plants failed to procure sufficient fuel in order to operate at capacity during the scarcity period. In one case, a thermal plant walked away from its firm energy obligations and refused to produce electricity, in spite of having received the firm energy payment during the previous nine years. For hydroelectric generations, the mechanism placed regulatory restrictions on the management of reservoirs, which limited the ability of these firms to optimally manage their water resources.

A second rationale for the capacity payment mechanism was to limit the incentive of generation firms to exercise market power during scarcity periods. The firm energy obligation had a similar effect to a forward contract: during scarcity conditions, generation firms receive the fixed scarcity price for output up to their firm energy obligation. Output in excess of the firm energy obligation received the wholesale market price. However, unlike an ordinary forward contract, generation firms have control over the occurrence of a scarcity condition, because their market power gives them the ability to set the wholesale price (recall that scarcity conditions are defined as the wholesale price exceeding the regulated scarcity price). Furthermore, during scarcity conditions, the settlement price for existing forward contracts held by generation firms is capped at the scarcity price. This means that for wholesale market prices above the scarcity price, the forward contract quantity no longer reduces the incentive of firms to increase the market price. As a result, the capacity mechanism creates a complex set of incentives for firms to exercise market power by either increasing or reducing the market price, depending on whether the firms are short or long relative to both their firm energy obligation and their forward contract position.

The capacity mechanism did limit the extent to which final end users were affected by the exercise of market power during the 2015–16 El Niño event. The maximum price that unregulated customers had to pay for electricity was capped at the scarcity price. However, the high wholesale market prices still had important financial implications for generation firms. The generators with a long position relative to their firm energy obligations earned high profits during this period, at the expense of those generators with a short position relative to their firm energy obligations. In addition, the lack of price signals to electricity users created additional inefficiencies in the market. Consumers had no reason to adjust their consumption in response to the scarcity conditions.[11]

Because it "solves" the incentive problem, market designers regard the reliability payment mechanism as a best-practice model for capacity markets. However, our analysis demonstrates that the mechanism not only fails to minimize the cost of meeting electricity demand but also creates perverse incentives for electricity generators that reduce the reliability of electricity supply. This result is of broad interest, especially because several wholesale electricity markets are considering the adoption of the Colombian capacity market model.

11. In early 2016, the government introduced an ad hoc rebate system to provide an incentive for regulated users to reduce their electricity consumption.

# References


Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2019. "(Mis)Allocation, Market Power, and Global Oil Extraction." *American Economic Review* 109, no. 4 (April): 1568–1615. doi:10.1257/aer.20171438.

Bushnell, James. 2003. "A mixed complementarity model of hydrothermal electricity competition in the western United States." *Operations Research* 51 (1): 80–93. doi:10.1287/opre.51.1.80.12800.

Bushnell, James, Michaela Flagg, and Erin Mansur. 2017. "Capacity markets at a crossroads." *Report to the Department of Energy, Office of Energy Policy and Systems Analysis, Washington, DC.*

Cramton, Peter, and Steven Stoft. 2008. "Forward reliability markets: Less risk, less market power, more efficiency." *Utilities Policy* 16 (3): 194–201. doi:10.1016/j.jup.2008.01.007.

Cullen, Joseph A. 2011. "Dynamic Response to Environmental Regulation in the Electricity Industry."

Dirkse, Steven P, and Michael C Ferris. 1995. "The path solver: a nommonotone stabilization scheme for mixed complementarity problems." *Optimization Methods and Software* 5 (2): 123–156. doi:10.1080/10556789508805606.

Dyner, I, S Arango, and ER Larsen. 2006. "Understanding the Argentinean and Colombian electricity markets." Chap. 17 in *Electricity Market Reform: An International Perspective,* edited by Fereidoon P. Sioshansi and Wolfgang Pfaffenberger, 595–616. Oxford: Elsevier Ltd.

Fabra, Natalia. 2018. "A primer on capacity mechanisms." *Energy Economics* 75:323–335. ISSN: 0140-9883. doi:10.1016/j.eneco.2018.08.003.

Fetzer, Thiemo, Oliver Pardo, and Amar Shanghavi. 2018. "More than an urban legend: the short- and long-run effects of unplanned fertility shocks." *Journal of Population Economics* 31:1125–1176. doi:10.1007/s00148-017-0685-5.

Galetovic, Alexander, Cristián M Muñoz, and Frank A Wolak. 2015. "Capacity payments in a cost-based wholesale electricity market: the case of Chile." *The Electricity Journal* 28 (10): 80–96. doi:10.1016/j.tej.2015.10.011.

Harbord, David, and Marco Pagnozzi. 2012. "Second Review of Firm Energy Auctions in Colombia." http://www.csef.it/IMG/pdf/firmenergauctionsfinal18december2012.pdf.

International Research Institute for Climate and Society. 2015. *IRI ENSO Forecast: 2015 May Quick Look.* https://iri.columbia.edu/our-expertise/climate/forecasts/enso/2015-May-quick-look. [Online; last accessed February 28, 2019].

Kernighan, Brian W, Robert Fourer, and David M Gay. 1993. *AMPL: a modeling language for mathematical programming.* Scientific Press, San Francisco.

Léautier, Thomas-Olivier. 2016. "The Visible Hand: Ensuring Optimal Investment in Electric Power Generation." *Energy Journal* 37 (2). doi:10.5547/01956574.37.2.tlea.

Mastropietro, Paolo, Fulvio Fontini, Pablo Rodilla, and Carlos Batlle. 2018. "The Italian capacity remuneration mechanism: Critical review and open questions." *Energy Policy* 123:659–669. doi:10.1016/j.enpol.2018.09.020.

McRae, Shaun, and Frank Wolak. 2014. "How Do Firms Exercise Unilateral Market Power? Evidence from a Bid-Based Wholesale Electricity Market." Chap. 18 in *Manufacturing Markets: Legal, Political and Economic Dynamics,* edited by Eric Brousseau and Jean-Michel Glachant, 390–420. Cambridge: Cambridge University Press.

———. 2016. *Diagnosing the Causes of the Recent El Nino Event and Recommendations for Reform.*

Moita, Rodrigo MS. 2008. "Entry and externality: Hydroelectric generators in Brazil." *International Journal of Industrial Organization* 26 (6): 1437–1447. doi:10.1016/j.ijindorg.2008.02.006.

Moita, Rodrigo, and Daniel Monte. 2020. "Hydroelectric Generators Competing in Cascades." *Revista Brasileira de Economia* 74 (1): 49–63. doi:10.5935/0034-7140.20200003.

Muggeo, Vito MR, et al. 2008. "Segmented: an R package to fit regression models with broken-line relationships." *R News* 8 (1): 20–25.

Reguant, Mar. 2014. "Complementary Bidding Mechanisms and Startup Costs in Electricity Markets." *The Review of Economic Studies* 81 (4): 1708–1742. doi:10.1093/restud/rdu022.

Riascos, Alvaro, Miguel Bernal, Luciano de Castro, and Shmuel Oren. 2016. "Transition to Centralized Unit Commitment: An Econometric Analysis of Colombia's Experience." *Energy Journal* 37 (3). doi:10.5547/01956574.37.3.aria.

Rudnick, Hugh, and Juan-Pablo Montero. 2002. "Second generation electricity reforms in Latin America and the California Paradigm." *Journal of Industry, Competition and Trade* 2 (1): 159–172. doi:10.1023/A:1020887106859.

Teirilä, Juha, and Robert A Ritz. 2019. "Strategic behaviour in a capacity market? The new Irish electricity market design." *The Energy Journal* 40 (The New Era of Energy Transition). doi:10.5547/01956574.40.SI1.jtei.

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. doi:10.1080/10168730000000017.

———. 2003. "Identification and Estimation of Cost Functions Using Observed Bid Data: An Application to Electricity Markets." In *Advances in Econometrics: Theory and Applications, Eighth World Congress, Volume II, Electricity Market Reform: An International Perspective,* edited by Lars Peter Hansen Mathias Detwatripont and Stephen J. Turnovsky, 133–169. New York: Cambridge University Press.

———. 2007. "Quantifying the supply-side benefits from forward contracting in wholesale electricity markets." *Journal of Applied Econometrics* 22 (7): 1179–1209. doi:10.1002/jae.989.

———. 2020. *Market Design in a Zero Marginal Cost Intermittent Renewable Future.* http://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/wolak_ieee_v4.pdf. [Online; last accessed July 13, 2020].

XM Compañía de Expertos en Mercados. 2016. *TDIASYS reports: Información total diaria del sistema.* http://www.xm.com.co/Paginas/Home.aspx. [Online; last accessed December 16, 2016].

———. 2019a. *Información Inteligente > Hidrología.* http://informacioninteligente10.xm.com.co/hidrologia/paginas/default.aspx. [Online; last accessed February 28, 2019].

———. 2019b. *Información Inteligente > Oferta.* http://informacioninteligente10.xm.com.co/oferta/Paginas/default.aspx. [Online; last accessed February 28, 2019].

XM Compañía de Expertos en Mercados. 2019c. *Información Inteligente > Transacciones.* `http://informacioninteligente10.xm.com.co/transacciones/Paginas/default.aspx`. [Online; last accessed February 28, 2019].

Ypma, Jelmer. 2014. *Introduction to nloptr: an R interface to NLopt.* Technical report.

# A Additional details of firm energy calculation

## A.1 Daily firm energy calculation

Let $Q_{jd}(deviation)$ be the daily firm energy deviation (*Desviación Diaria de la Obligación de Energía Firme* or *DDOEF* in Spanish) for generator $j$ on day $d$. This is the difference between the daily ideal generation for generator $j$ and the daily firm energy for generator $j$.[12] Both quantities are summed across the plants $i$ belonging to generator $j$ ($I_j$).

$$Q_{jd}(deviation) = \sum_{i \in I_j} \sum_{h=1}^{24} Q_{ijhd}(ideal) - \sum_{i \in I_j} Q_{ijd}(firm) \tag{14}$$

If $Q_{jd}(deviation)$ is positive, meaning that generator $j$ has an ideal generation exceeding its firm energy obligation, then during scarcity hours the generator will be paid for the excess generation. The hourly firm energy for this calculation, $Q_{jhd}(firm)$, is determined by a pro rata assignment of the daily firm energy using the share of ideal generation in each hour.

$$Q_{jhd}(firm) = Q_{jd}(firm) \frac{Q_{jhd}(ideal)}{\sum_{h=1}^{24} Q_{jhd}(ideal)} \tag{15}$$

In this expression, $Q_{jd}(firm)$ is the daily firm energy for generator $j$ and $Q_{jhd}(ideal)$ is the hourly ideal generation for generator $j$, in both cases summing across all of the plants $i \in I_j$. The hourly firm energy is only calculated for the generators with a positive firm energy deviation.

The generators with positive $Q_{jd}(deviation)$ receive an hourly firm energy refund (*Desviación Horaria de la Obligación de Energía Firme* or *DHOEF* in Spanish) during scarcity hours.[13] This refund is the difference between the wholesale price and the scarcity price, multiplied by the difference between the ideal generation and the hourly firm energy.

$$R_{jhd}(firm) = \max(0, P_{hd} - P_d(scarcity))(Q_{jhd}(ideal) - Q_{jhd}(firm)) \tag{16}$$

12. The system operator in Colombia solves two generation dispatch problems, with and without accounting for transmission constraints. The ideal generation is the generation calculated under an assumption of no transmission constraints.

13. Note that the terminology in the regulation is inconsistent. The daily firm energy deviations (*DDOEF*) are measured in kWh. The hourly firm energy refunds are measured in pesos but are labelled as "hourly deviations" (*DHOEF*).

The total daily firm energy refunds are the sum of the hourly firm energy refunds for the generators with positive firm energy deviations ($J_+$).

$$R_d(firm) = \sum_{j \in J_+} \sum_{h=1}^{24} R_{jhd}(firm) \tag{17}$$

The firm energy refunds $R_d(firm)$ are assigned to the generators with negative firm energy deviations ($J_-$). They need to make firm energy payments ($S_{jd}(firm)$). The assignment is based on the share of the daily firm energy deviation out of the total of the negative firm energy deviations.

$$S_{jd}(firm) = R_d(firm) \frac{Q_{jd}(deviation)}{\sum_{j \in J_-} Q_{jd}(deviation)} \tag{18}$$

By construction the total payments by the generators with negative firm energy deviations will equal the total refunds to the generators with positive firm energy deviations.

We note that the obligations for the generators with ideal generation below their firm energy obligation depend only on their total ideal generation for the day. The intraday pattern of generation is irrelevant. In particular, for days in which the wholesale price exceeds the scarcity price in only some hours, it does not matter whether the generator produced more or less of its output during the scarcity hours.

## A.2 Example of firm energy calculation

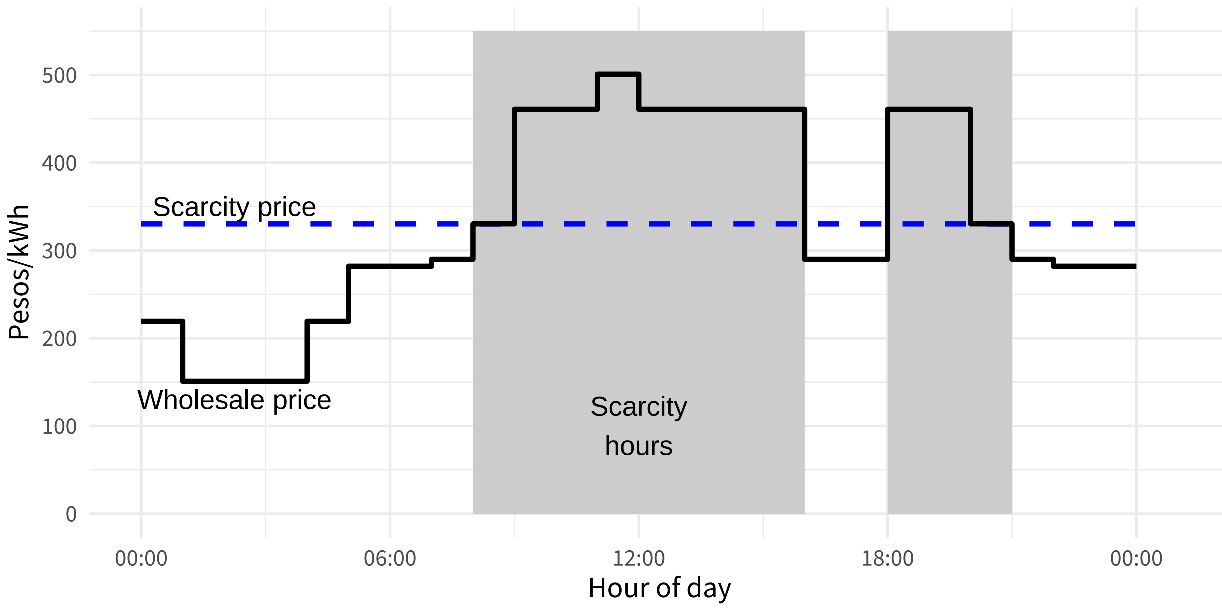
We illustrate the calculation in Section A using data for one day: May 28, 2015.

The wholesale price exceeded the scarcity price on 11 hours of the example day (Figure A1). The scarcity price of 330.27 pesos/kWh was constant for all hours in May 2015. The wholesale price reached a daily maximum of 500.94 pesos/kWh at 11:00AM on May 28. There were two hours (8:00AM and 8:00PM) with a wholesale price of 330.34 pesos/kWh, a fraction of a peso above the scarcity price.

The daily firm energy obligation of each generator was scaled so that the aggregate firm energy was exactly equal to the aggregate ideal generation.[14] The unadjusted firm energy on May 28 was 197.8 GWh, divided between 187.4 GWh of dispatched generation and 10.4

14. More specifically, the aggregate firm energy is scaled to equal the total domestic demand, including both self-consumption by generators and allocated transmission losses. Electricity exports to Venezuela and Ecuador are excluded from domestic demand. These exports comprised 0.4 percent of total electricity demand on May 28, 2015. We ignore this additional adjustment for the purpose of this example.

**Figure A1:** Hourly wholesale prices and scarcity price on May 28, 2015

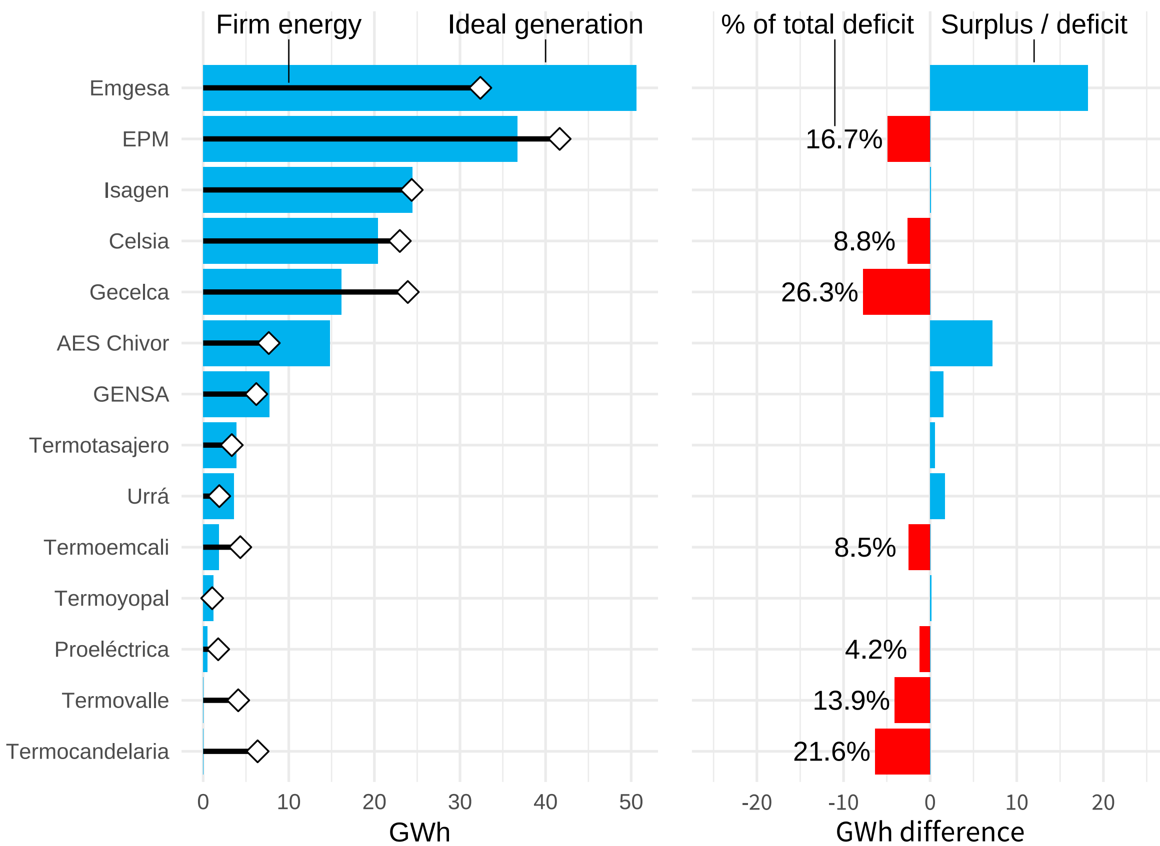



GWh of non-dispatched generation. The ideal generation on May 28 was 175.9 GWh for the dispatched plants and 10.4 GWh for the non-dispatched plants. The adjustment factor to scale the firm energy of the dispatched plants was $175.9/187.4 = 0.939$.

Because of the scarcity condition for at least one hour on May 28, the scaled firm energy of each generator was compared with its ideal generation (left panel of Figure A2). Emgesa had a positive daily firm energy deviation: its ideal generation of 50.6 GWh exceeded its firm energy obligation of 32.4 GWh. In contrast, EPM had a negative daily firm energy deviation, with its ideal generation of 36.7 GWh below its firm energy obligation of 41.6 GWh. There were seven generators with positive deviations and seven with negative deviations (right panel of Figure A2). By construction, the sum of the positive and negative deviations is exactly equal to zero.

The hourly firm energy obligation is calculated only for the generators with positive deviations. The top panel of Figure A3 shows the calculation of the hourly positive deviations for Emgesa on May 28. The top line shows the hourly ideal generation for Emgesa. The bottom line shows the proportional allocation of the daily firm energy, based on the share of ideal generation each hour out of the total ideal generation. Emgesa received a firm energy refund for the 11 scarcity hours. This refund was calculated as the difference between its ideal generation and the allocated firm energy, multiplied by the difference between the scarcity price and the wholesale price. The bottom panel of Figure

**Figure A2:** Daily generation and firm energy by firm for May 28, 2015




A3 shows the hourly refund. The hourly refund was almost zero at 8:00AM and 8:00PM, because the difference between the wholesale price and the scarcity price in those hours at only 0.08 pesos/kWh.

We repeated this calculation for the six other generators with positive firm energy deviations (left panel of Figure A4). Emgesa had the largest daily refund of 1009 million pesos. Isagen had a very small refund (6 million pesos) because its daily ideal generation was very similar to its daily firm energy obligation. The sum of the positive firm energy refunds was 1726 million pesos.

The positive refunds were allocated to the seven generators with negative firm energy deviations (right panel of Figure A4). This allocation was based on the share of each generator's negative firm energy deviation of the total negative firm energy deviations. This share is shown on the right panel of Figure A2. For example, EPM had a generation shortfall of 4.9 GWh, which was 16.7 percent of the total negative firm energy deviations of 29.4 GWh. As a result, EPM had to make a firm energy payment of 289 million pesos, equal to a 16.7 percent share of 1726 million pesos. By construction, the total payments for negative firm energy deviations were equal to the total refunds for positive firm energy deviations.

**Figure A3:** Hourly generation, firm energy allocation, and firm energy refund for Emgesa on May 28, 2015

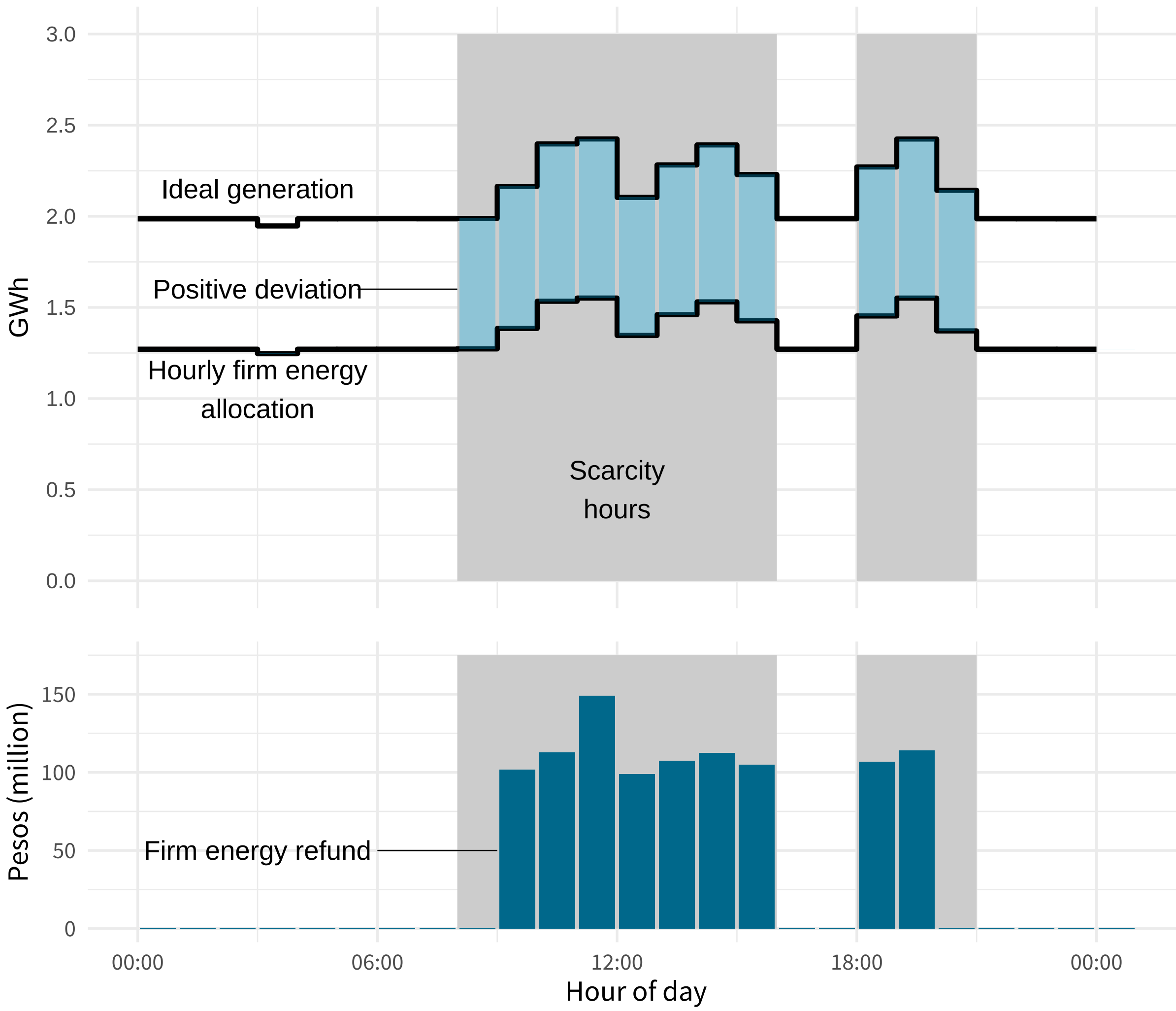


**Figure A4:** Firm energy obligations and payments on May 28, 2015