Market power and incentive-based capacity payment mechanisms

Shaun D. McRae* and Frank A. Wolak†

December 15, 2018

PRELIMINARY DRAFT

Abstract

Capacity markets give guaranteed payments to electricity generators for having the capacity to produce electricity, even if they do not actually do so. These markets are plagued by the weak incentives they provide for plants to be available during high-demand hours. The reliability payment mechanism in the Colombian electricity market provides market-based incentives for plants to produce during periods of system scarcity. This market 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 a 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 this capacity payment mechanism by computing a counterfactual dynamic oligopoly equilibrium for the most recent (2015-2016) El Niño event in Colombia.

*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. Email: wolak@stanford.edu.

1 Introduction

In restructured electricity markets, capacity payment mechanisms provide a fixed payment to generators for keeping their plants available, even if they do not produce any electricity. 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 both aggravated this revenue adequacy problem and increased the need to keep backup generation available. While traditional capacity markets have been very successful at providing additional revenue to generators, 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 a implicit financial penalty to generators who do not produce 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 their capacity quantity must refund the difference between the market and strike prices for their generation shortfall. Generation firms that produce more than their capacity quantity receive the difference between the market and strike prices 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 wholesale market is normally competitive, 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 the condition or not will depend on the relative magnitudes of their capacity contract quantities and their forward contract quantities. In some circumstances, we show that it is possible for the reliability option mechanism to reduce system reliability relative to a counterfactual with no capacity payment mechanism at all.

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 by far the oldest and longest-running incentive-based capacity market in the world. Recent and proposed reforms in other markets, including the Peak Energy Refund payment system, the new Pay-for-Performance rules in the New England ISO, and the capacity remuneration mechanisms in the Italian and Irish electricity markets, 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 generation capacity requirements are greatest, that generation firms have the ability and incentive to manipulate the reliability option mechanism. Such problems will potentially occur in other markets with 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 forward contract 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 faced by each firm in the market. When this lies entirely above or below the scarcity price, the scarcity condition will or will not be triggered, regardless of the generation output of the firm. However, when the inverse residual demand 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, has the choice to trigger a scarcity condition in 9.9% of the hours in our sample.

We then calculate the profitability of triggering the scarcity condition by computing the maximum net revenue for the firm each hour under both scarcity and non-scarcity conditions. These revenues are a function of the inverse residual demand, the scarcity price, the capacity contract quantity, and the net forward contract position. This calculation provides an accurate prediction of whether or not the scarcity condition is triggered. For EPM, the scarcity condition occurs in 99% of the hours in which it would maximize profits, and it does not occur in 90% of the hours for which it would not be profit-maximizing. Similar results hold for the other two large generation firms in the Colombian market.

Our analysis of plant-level bid data provides further evidence that generators are responding 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 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 bids 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. In ongoing work for this project, we will compare the reliability options to a counterfactual market structure without the mechanism, where our primary performance measure will be the probability of rationing. Given the predominance of hydro generation in the Colombian market, our analysis will account for the intertemporal constraint on firm-level generation determined by total hydro inflows (Bushnell, 2003).

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 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.

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 6 discusses the results and suggests alternative mechanisms for ensuring a reliable supply of electricity.

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.[1] The government lacked the financial capacity to invest in new thermal plants that could act as a backup for hydro gen-

[1]Fetzer et al. (2014) use satellite night lights to study the geographical variation in rationing during the 1992–93 blackouts. They show that the electricity shortages led to a short-term increase in fertility and a permanent increase in the number of children.

erators 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 41.3 terawatt-hours (TWh) in 2000 to 66.5 TWh in 2015, an average annual growth rate of 3.2 percent. Between 2000 and 2009, most of this growth in electricity demand was met by increases in hydro generation (Figure 1). However, this changed after 2010, with demand growth mostly met by increasing thermal generation. Hydro generation peaked at 48.7 TWh in 2011 and has been lower in every subsequent year. Between 2000 and 2005, hydro comprised 79 percent of total generation. This fell to 72 percent of the total between 2012 and 2015.

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 some of 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 large drop in hydroelectric generation and a large increase in thermal generation.

The market design for the industry restructuring in Colombia is different to that used in any other Latin American market (Rudnick and Montero, 2002). 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. Each generation unit may submit a single price for its output for the entire day. The quantity made available from each generation unit is allowed to vary by hour. Beginning in 2009, generators 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.[2] By contrast to the Colombian market design, the other electricity markets in the region use a cost-based dispatch, in which the "price" of each unit is set based on a regulatory cost formula.

The market also includes a system of capacity payments that are made to generators even when they are not producing electricity. The amount of the capacity payment (in $ per MW) is determined by auctions for long-term investment in new generation capacity, first

[2]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.

held in May 2008 and December 2011.[3] Both existing and new generation plants receive the payments for their assigned capacity, 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 their plants available during periods of system scarcity, in order to meet this financial obligation. In effect, the price that generators receive for providing their firm energy capacity is capped at the scarcity price, although they still receive the wholesale price for any generation in excess of their firm energy. The scarcity price is recalculated each month based on changes in the price of an international fuel oil benchmark.

For nearly all hours in the first nine years of operation of the reliability payment mechanism, the market price lay 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 start of 2016. For a six-month period, 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 different 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 (Table 1). These firms are predominantly hydroelectric, although each has a small proportion of thermal generation. Three smaller firms have 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.

3 Illustrative model

In this section we provide a simple model to illustrate the interaction of the reliability payment mechanism with the forward contract market and the potential effects on the incentives for generation firms.

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 market power of an electricity supplier by calculating its residual demand, the market demand less the quantity supplied by the firm's competitors. At any

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

price, the residual demand shows the maximum quantity the firm may supply, based on the market demand and the amount that competing firms will supply at that price.

When the firm chooses the offer curve to submit to the system operator, it is effectively choosing the point along its residual demand curve at which it would like to operate. By definition, the firm will produce the generation quantity and receive the wholesale price that are set by the price and quantity point where its offer curve crosses its residual demand curve. The firm chooses the price and quantity combination on its residual demand curve that will maximize its profits, accounting for the cost of running 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 to are the residual demand curves. However, for analytical simplicity, we assume that residual demands are linear functions.

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

$$P(Q) = 200 - 20Q \tag{1}$$

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 the top graph of Figure 3. As noted above, the firm is able to choose any price and quantity pair along this curve. We assume the generator has sufficient generation capacity to operate at any point on the curve.

Assume the marginal cost for the generator is zero. In the absence of any forward contracts, the generator will act as a monopolist off its residual demand curve and equate marginal revenue to marginal cost. In this case, $MR = 200 - 40Q$, implying the firm will maximize profits by choosing $Q = 5$ GWh. The market price corresponding to this generation quantity is \$100/MWh. This price and quantity pair is point A in Figure 3.

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 $Q_c = 4$ GW of forward contracts at a price P_c. With these forward contracts in place, the profit for the firm is now:

$$\Pi = P_c Q_c + P(Q)(Q - Q_c) - c(Q) \tag{2}$$

In this expression, $c(Q)$ is the cost of producing the generation quantity Q. P_c and Q_c are predetermined at the time the firm is choosing its generation offer. Focusing on the middle term (representing the short-run revenues or costs for the firm in the wholesale market), assuming costs are still zero, and substituting the inverse residual demand from Equation (1), gives the following expression for short-run revenues:

$$\begin{aligned}\Pi^{SR} &= (200 - 20Q)(Q - 4) \\ &= -20Q^2 + 280Q - 800 \end{aligned} \tag{3}$$

The dashed line in the bottom graph of Figure 3 shows this expression for short-run revenues as a function of generation quantity Q.

The generation firm will choose the quantity that maximizes profits: $Q = 7$ GWh. This is point D on the graph of short-run revenues (bottom of Figure 3) and the point B on the residual demand (top of Figure 3). Note that 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 point A on the residual demand curve.

Now suppose we consider the introduction of the reliability payment mechanism to this setting where generation firms have existing forward contracts. The generation firm has a firm energy contract with quantity Q_f and the firm energy payment P_f. There is an administratively-set scarcity price P_s. When the market price $P(Q)$ exceeds P_s, the generator is required to produce Q_f. If it produces less than Q_f, it will pay back the difference between $P(Q)$ and P_s for any shortfall.

For the numerical example, assume the firm energy quantity $Q_f = 1$ GW and the scarcity price is $P_s =$\$65/MWh. Note from Figure 3 that the generator controls whether or not the market is in the scarcity condition. If it restricts its generation below 5.75 GWh, the market price on its inverse residual demand will exceed \$65/MWh, triggering the scarcity condition.

Suppose the firm chooses to produce more than 5.75 GWh. The profit function will be the same as before, with the addition of the firm energy payment:

$$\Pi = P_f Q_f + P_c Q_c + P(Q)(Q - Q_c) - c(Q) \tag{4}$$

Because the short-run revenues are identical to Equation (3), point D in Figure 3 will again be locally profit-maximizing, for the case when generation exceeds 5.75 GWh.

Alternatively, suppose the firm produces less than 5.75 GWh and triggers the scarcity condition. The profit function will now be:

$$\Pi = P_f Q_f + P_c Q_c + P_s(Q - Q_c) + (P(Q) - P_s)(Q - Q_f) - c(Q) \tag{5}$$

The price that the firm will pay (or receive) in the wholesale market for generating less (or more) than the forward contract quantity Q_c is capped at the scarcity price P_s. If the firm produces more than its firm energy quantity Q_f, it receives an additional firm energy refund of the difference between $P(Q)$ and P_s, for the excess generation. Conversely, if the firm produces less than its firm energy quantity, it pays the difference between $P(Q)$ and P_s for the generation shortfall. This mechanism is designed to provide incentives for the generator to produce at least its firm energy quantity during scarcity conditions.

Substituting the above parameter assumptions into Equation (5) gives the following expression for short-run revenues in the wholesale market, for the case when the scarcity condition is triggered (that is, when generation is less than 5.75 GWh):

$$\Pi^{SR} = 65(Q-4) + (200 - 20Q - 65)(Q-1) \tag{6}$$

$$= -20Q^2 + 220Q - 405 \tag{7}$$

This expression is shown as the solid line in the bottom graph of Figure 3. It is maximized at $Q = 5.5$ GWh, or point E in the figure.

Under the reliability payment mechanism, generators with market power have the ability to choose whether or not the scarcity condition exists. For the example, this decision is a choice between point D (profit-maximizing point with no scarcity condition) and point E (profit-maximizing point with scarcity condition). Profits are higher at E (\$210) than at D (\$180). Therefore the profit-maximizing firm will choose to restrict its generation to 5.5 GWh, increasing the market price to \$90/MWh, and triggering the scarcity condition.

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.

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, it is common for Q_f to lie below Q_c.

4 Empirical analysis

In this section we analyze the bidding behavior of the generators in the Colombian wholesale electricity market to demonstrate the real-world relevance of the stylized 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 residual demand of a generator—that is, the market demand less the supply of the competing generators—describes the possible combinations of market price and generation quantity that the firm can choose. Generators are able to pick any price and quantity combination along their residual demand curve, up to their generation capacity limit, by submitting an offer curve that crosses the residual demand at the desired point.

There are three possible configurations for the residual demand (Figure 4). The first case is when the firm's inverse residual demand 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 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 lies above the scarcity price 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 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.

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 6:00PM on July 25, 2015, EPM did not have the ability to induce the scarcity condition, for any choice of quantity. At 6:00PM on November 25, 2015, the scarcity condition would occur for any choice of generation by EPM. Finally, at 6:00PM 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 (Figure 5). For most of the six month period 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 9.9 percent of hours (top block of Table 2). In 4.8 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 substantial ability to determine the scarcity outcome, though in a smaller share of hours than EPM. Emgesa could induce the scarcity condition in 8.0 percent of hours and Isagen could do the same in 3.8 percent of hours. The smaller generation firms in the Colombian market had very limited ability to influence the scarcity outcome.

4.2 Generation firms respond to incentive to induce the scarcity condition

Although the largest three generation firms frequently have the ability to create a scarcity condition, they will not always have the incentive to do so. Equation (4) gives the short-run profits for the firm in the absence of the scarcity condition. Equation (5) gives the short-run profits once the scarcity condition has been triggered. 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 hours in which they had the ability to create a scarcity condition. For these hours, we calculate the highest possible profit that they could achieve each hour under the scarcity condition. We used a grid search along the residual demand curve between the minimum generation quantity and the critical quantity between the scarcity and non-scarcity regions (Figure 6). At each possible quantity, we calculated the short-run profits from Equation (5), assuming a zero cost for generation. We also ignored revenue from forward contract sales and firm energy payments, because these were the same in either the scarcity or non-scarcity conditions. For the residual demand curve shown in the figure, the highest hourly profit that could be achieved in the scarcity region was 637 million pesos, at a generation quantity of 2900 MW.

We then carried out the same procedure to calculate the highest possible profit that could be achieved each hour without the scarcity condition. For that case, we used a grid search along the residual demand curve between the critical quantity and the maximum generation quantity, calculating the short-run profits at each possible quantity from Equation (4). For the example hour in the figure, the highest possible profits in the non-scarcity region were 553 million pesos, for a generation quantity of 3500 MW.

Profit-maximizing firms are assumed to make a choice between the scarcity and non-scarcity regions, picking the alternative that gives the highest profit. In Figure 6, the hourly profit for EPM in the scarcity region (637 million pesos) exceeded the hourly profit in the non-scarcity region (553 million pesos), making it optimal for EPM to create a scarcity condition that hour. Indeed, the scarcity condition did occur in this hour, suggesting the EPM had responded as expected to the incentives provided by the scarcity payment mechanism.

We repeat this calculation for each firm and each hour in which they have a choice between the scarcity and non-scarcity condition. Most of the time, profits would be higher if the scarcity condition were avoided. For EPM, in only 411 out of the 8332 hours in which it had a choice (4.9 percent), profits would be higher if the scarcity condition occurred (second block of Table 2). In 99 percent of these hours, the scarcity condition did occur. This result confirms that EPM almost always created a scarcity condition when it had the ability and incentive to do so.

For the other 7921 hours (95.1 percent) in which EPM had a choice, profits would be higher if the scarcity condition were avoided. In 90 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.

The results are similar for the other two large generation firms, Emgesa and Isagen. There were 103 hours in which Emgesa had the ability and incentive to create a scarcity condition, and in 94 percent of these hours the scarcity condition occurred. For Isagen, there were 174 hours when it had the ability and incentive to create a scarcity condition, and this occurred in 97 percent of these hours. Conversely, there were 6124 hours in which Emgesa had the ability but not the incentive to create a scarcity condition, and the scarcity condition was avoided in 95 percent of these. For Isagen, the scarcity condition did not occur in 87 percent of the hours in which it had the ability but not the incentive to induce scarcity.

Overall, these results suggest that the major 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 these cases the firms bid in a manner that creates scarcity.

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. To be able to compare the bids across different months of the sample with different scarcity prices, we scale all of the price bids by the scarcity price. That is, 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 7921 hours in which EPM had the incentive to avoid creating a scarcity condition, there is a high degree of bunching of bids just below the scarcity price (Figure 7). 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 411 hours in which EPM had the incentive to create a scarcity condition, nearly all of its bids were above the scarcity price (Figure 8).

5 Counterfactual Market Outcomes

This section presents the results of a counterfactual equilibrium analysis of the annual hydroelectric cycle in Colombia that encompassed the most recent El Nio event in the absence of the capacity payment mechanism. This model assumes dynamic quantity-setting behavior by the three largest hydroelectric generation capacity-owning firms in Colombia competing against a competitive fringe similar to the model of the Western United States market in Bushnell (2003). The model will first be calibrated to the existing market outcomes with the capacity payment mechanism in place. We will then run a two sets of counterfactuals. The first will use the distribution of historical hydroelectric inflows to compute a distribution of thermal dispatch costs and the probability of an energy supply shortfall with the capacity payment mechanism in place. This will require solving the model for each realization of the time path of hydroelectric inflows. For these same hydroelectric inflows realization, We will than compute the equilibrium without the capacity payment mechanism in place. This will allow us to compare the expected thermal cost difference and the difference in supply shortfall probabilities between the market with and without the capacity payment mechanism in place.

6 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 occurence 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.[4]

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

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

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.

References

Bushnell, James, "A mixed complementarity model of hydrothermal electricity competition in the western United States," *Operations research*, 2003, *51* (1), 80–93.

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

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

Fetzer, Thiemo, Oliver Pardo, Amar Shanghavi et al., "More than an Urban Legend: The long-term socio-economic Effects of Unplanned Fertility Shocks," 2014.

Harbord, David and Marco Pagnozzi, "Second Review of Firm Energy Auctions in Colombia," 2012.

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

McRae, Shaun and Frank Wolak, "Diagnosing the Causes of the Recent El Nino Event and Recommendations for Reform," 2016.

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

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

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

Figure 1: Quarterly actual electricity generation in TWh, by type of generator

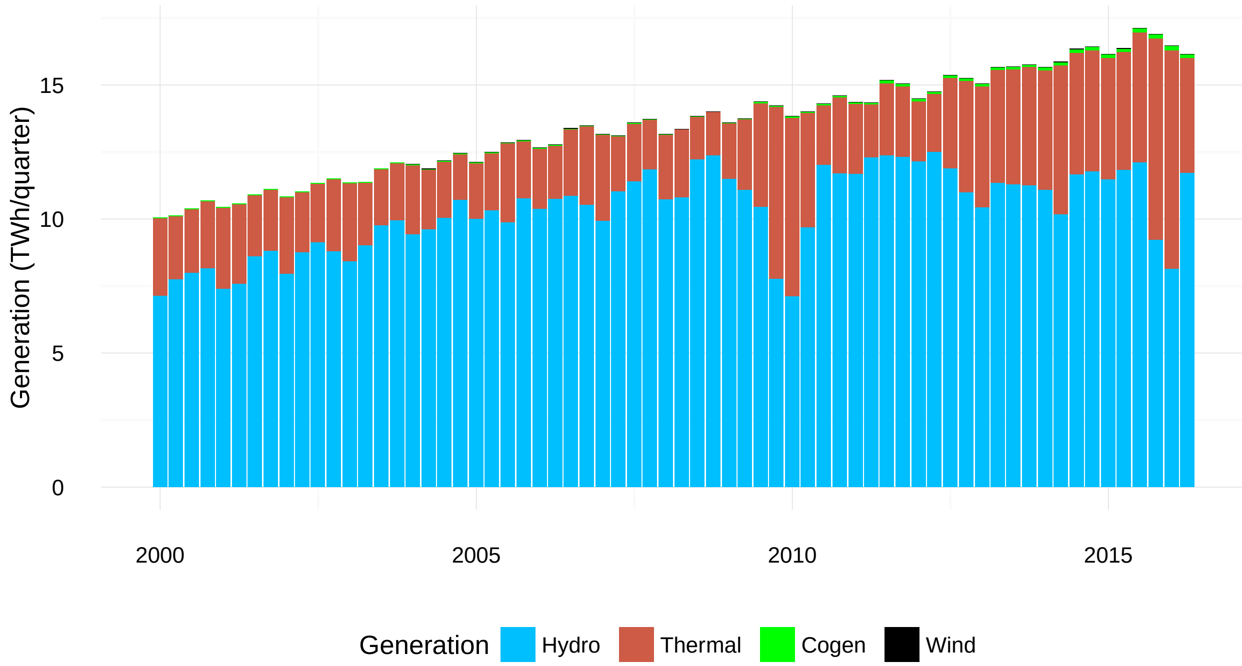
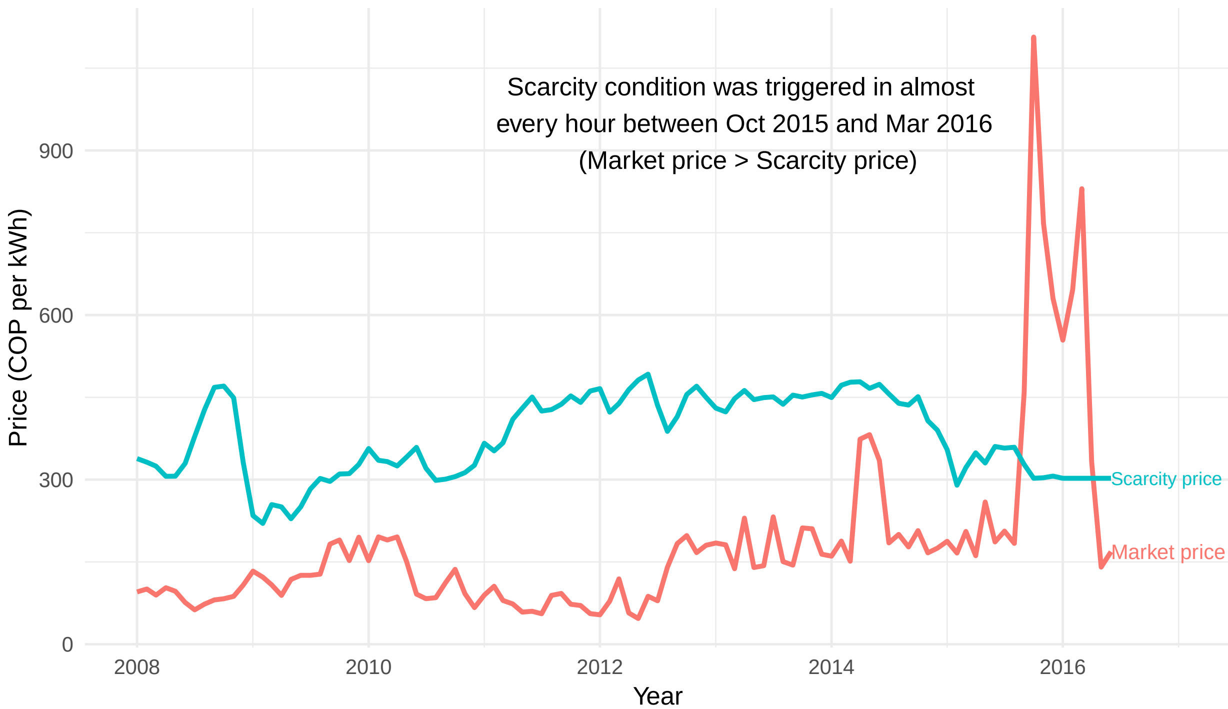


Figure 2: Wholesale market prices and scarcity prices, 2007–2016



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.

Figure 3: Illustrative example of the effect for generator incentives of forward contracts and the reliability payment mechanism

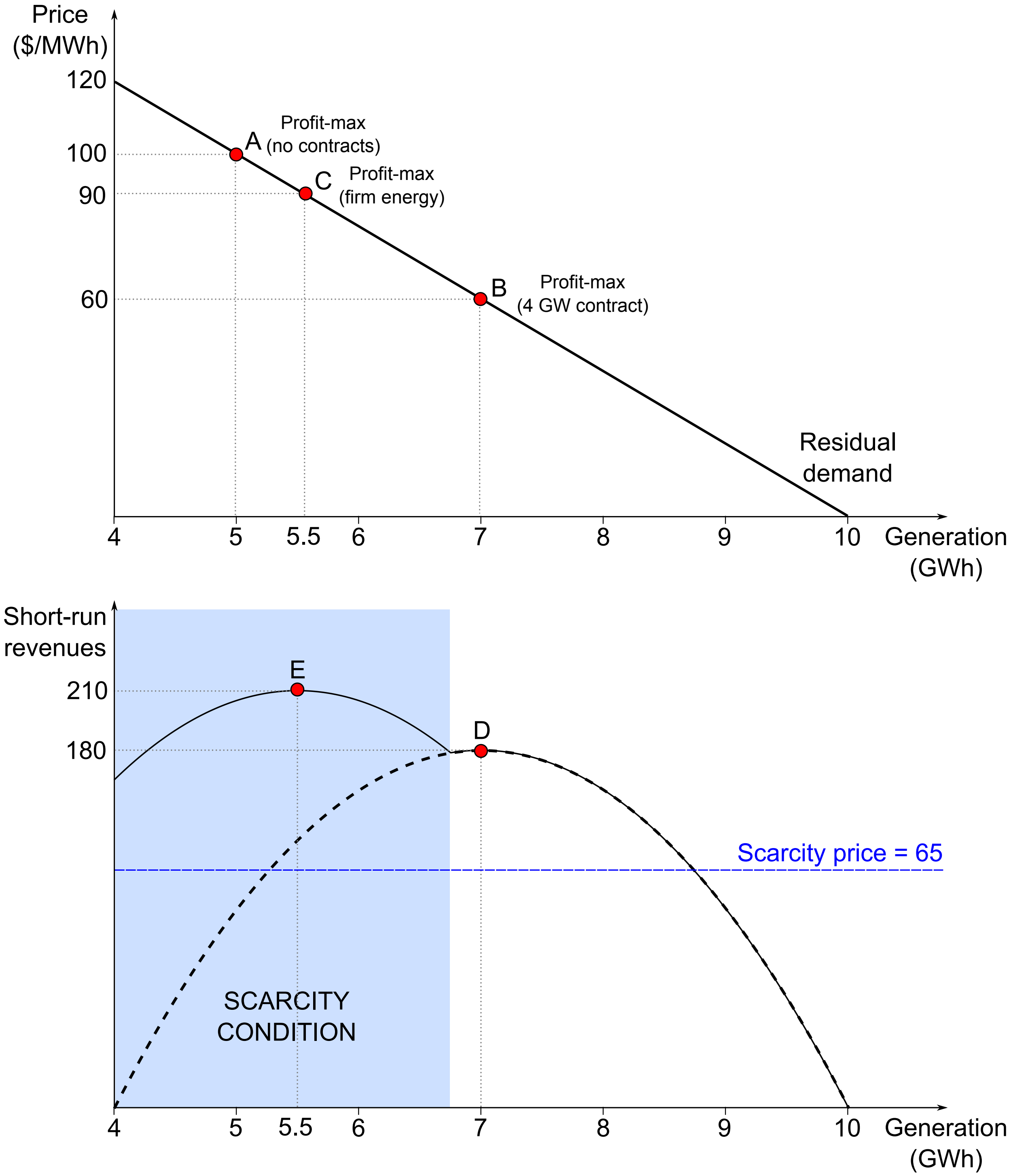
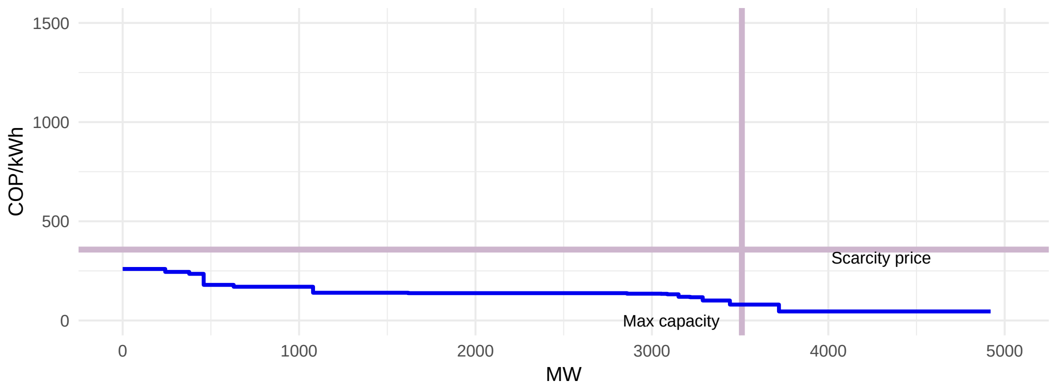
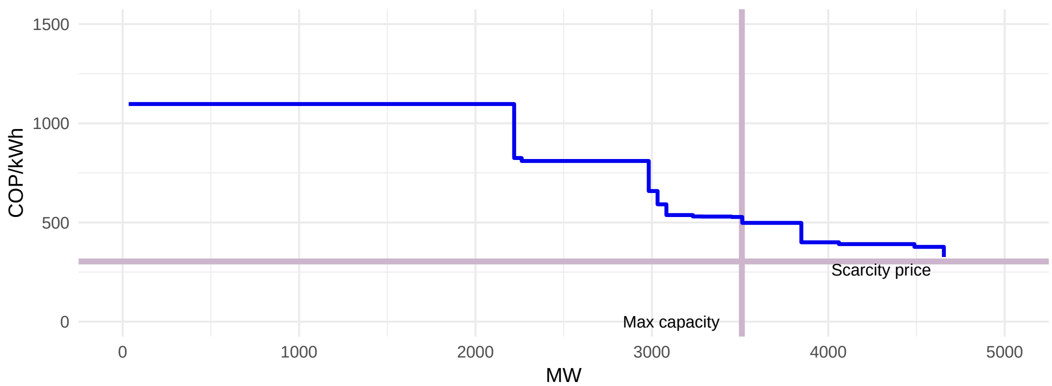


Figure 4: In certain system conditions, generation firms have the ability to choose whether or not the scarcity condition occurs

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.

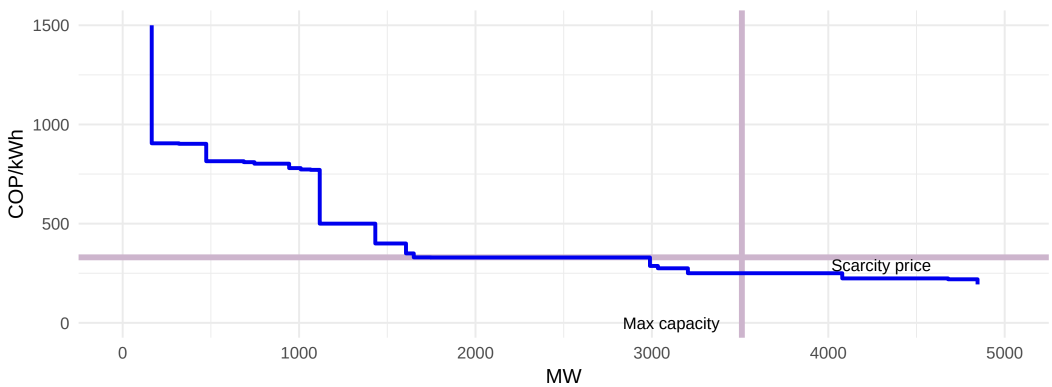


Figure 5: Proportion of hours each month in which EPM could choose to induce scarcity condition

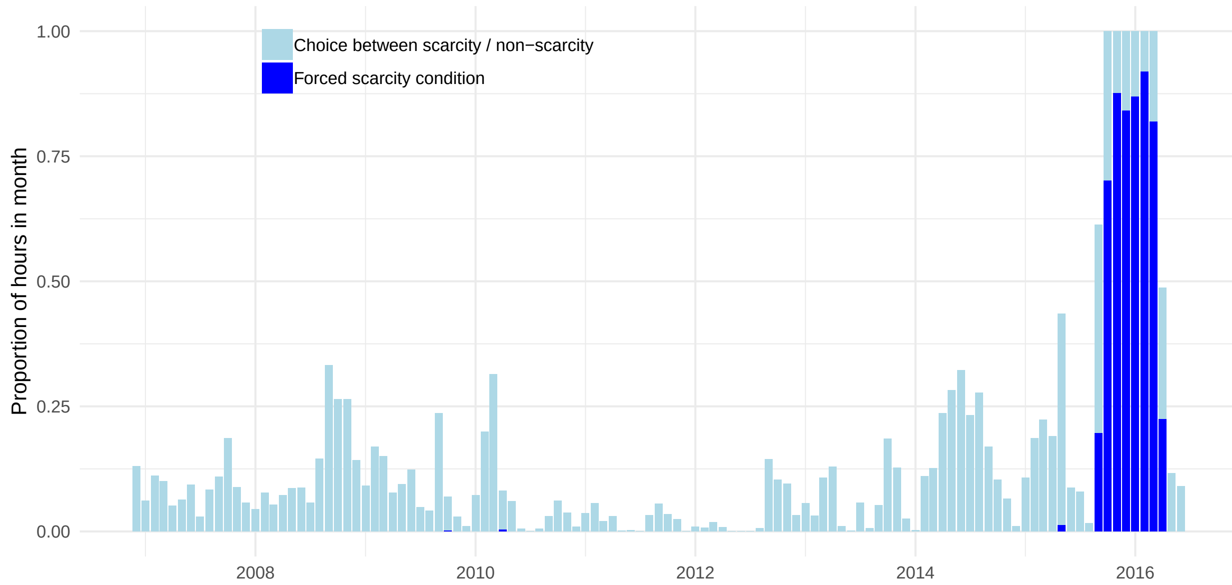
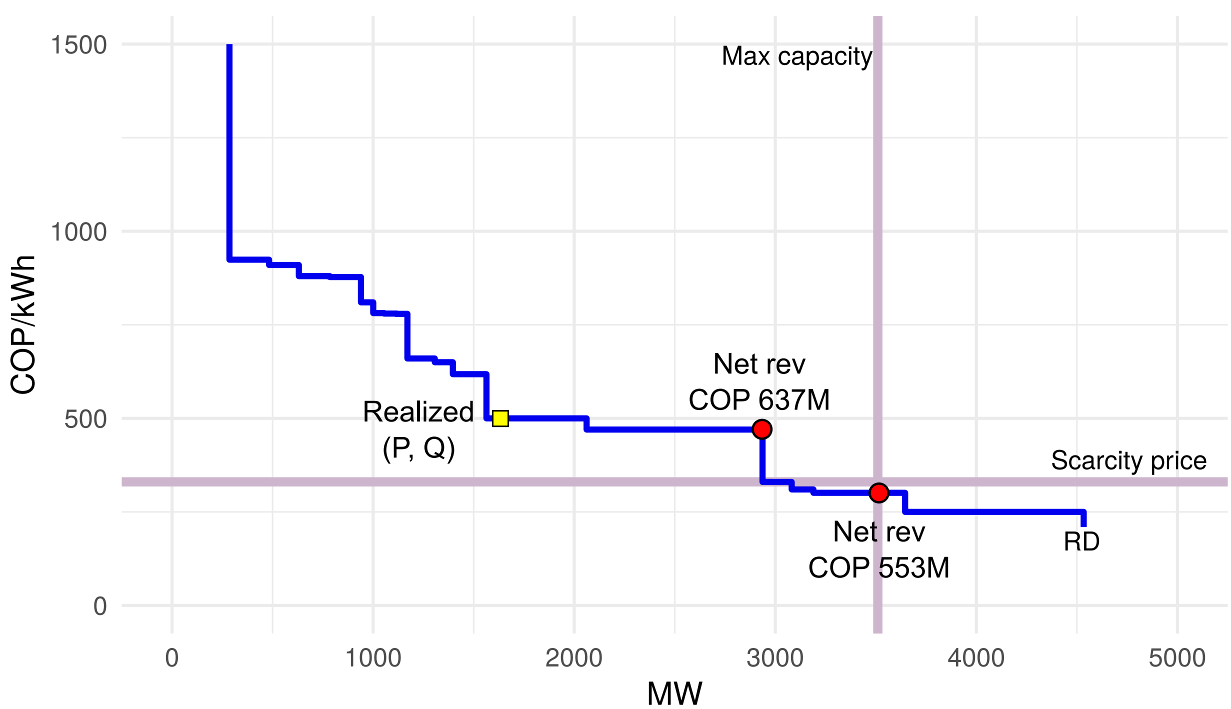


Figure 6: Profit incentive for choosing between scarcity and non-scarcity condition



Residual demand for EPM on 15 May 2015, at 8:00 AM

Figure 7: For hours when non-scarcity is optimal, generation price offers for EPM exhibit bunching below the scarcity hours

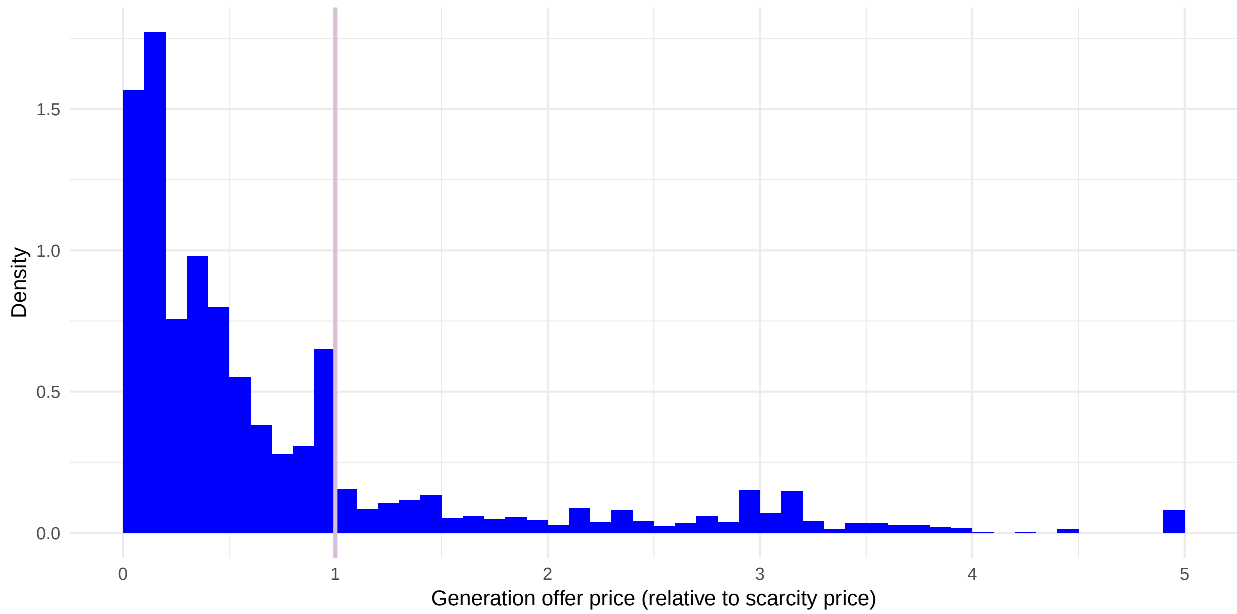


Figure 8: For hours when scarcity condition would be optimal, most generation price offers for EPM lie above the scarcity price

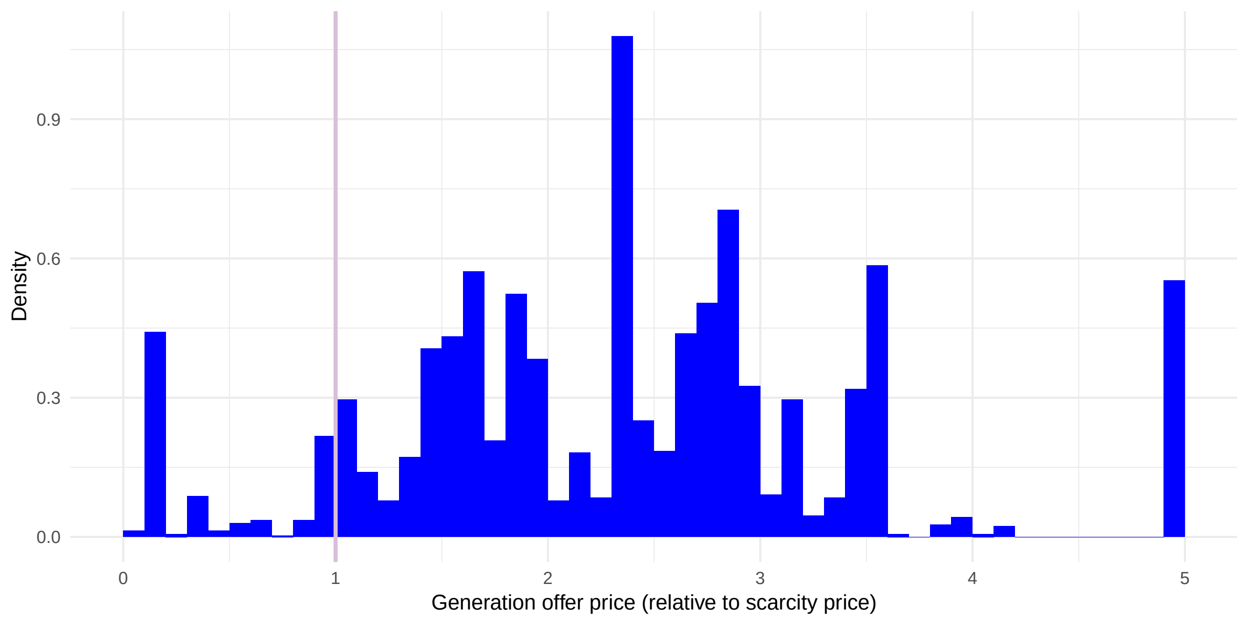


Table 1: Ownership or control of generation capacity in GW, as of 30 June 2016

Firm	Hydro	Thermal	Cogen	Wind	Total	Percent
EPM	2.99	0.54	0.01	0.02	3.56	21.48
Emgesa	3.02	0.41	0.01	0.00	3.44	20.73
Isagen	2.73	0.26	0.00	0.00	2.99	18.02
Celsia	1.08	0.78	0.05	0.00	1.90	11.47
SCLEA consortia	0.00	1.22	0.00	0.00	1.22	7.34
AES Chivor	1.00	0.00	0.00	0.00	1.00	6.03
Gecelca	0.00	0.46	0.00	0.00	0.46	2.80
Urra	0.34	0.00	0.00	0.00	0.34	2.04
Colgener	0.00	0.33	0.00	0.00	0.33	1.98
Gensa	0.00	0.32	0.00	0.00	0.32	1.93
Termocandelaria	0.00	0.32	0.00	0.00	0.32	1.90
ContourGlobal	0.00	0.21	0.00	0.00	0.21	1.28
Other	0.33	0.11	0.06	0.00	0.50	3.01
Total	**11.48**	**4.97**	**0.13**	**0.02**	**16.59**	**100.00**

Table 2: Generation firm results for optimal choice between scarcity and non-scarcity conditions

	EPM	Emgesa	Isagen	Celsia
Non-scarcity hours	71673	66425	76223	78997
Forced scarcity hours	3995	4772	4577	4826
Scarcity choice hours	8332	6227	3200	177
Total hours	84000	77424	84000	84000
Hours with scarcity condition profitable	411	103	174	30
Scarcity proportion of hours	0.99	0.94	0.97	0.97
Non-scarcity proportion of hours	0.01	0.06	0.03	0.03
Hours with non-scarcity condition profitable	7921	6124	3026	147
Scarcity proportion of hours	0.10	0.05	0.13	0.80
Non-scarcity proportion of hours	0.90	0.95	0.87	0.20