purchase contracts with no offsetting sales positions. Consequently, most new generation developments in ERCOT will not be able to obtain the revenue certainty of a long-term PPA. While generation developers can procure a long-term hedge from a bank or power marketer, banks told us that since the financial crisis they have generally been unwilling to offer hedges longer than 5 years. Developers, therefore, must absorb more of the investment risk.

While all investors would prefer to have the security of a long-term buyer, the lack of long-term PPAs does not mean that generation projects cannot be built in ERCOT. It may be difficult or impossible for small, undiversified developers to debt finance their projects without a PPA, but projects can still be financed by larger, diversified IPPs that use balance sheet financing. For these reasons, investments in ERCOT will incur higher financing costs than regions with regulated retail markets, and many smaller, undiversified investors may be precluded from ERCOT's market.

2. Impact of Market and Regulatory Structure on Financeability

There is a fundamental difference in the financing and investment conditions between: (1) regulated regions, where customers bear the risks of potentially uneconomic investments; and (2) restructured markets, where suppliers take on these risks. Regulated systems use integrated planning with cost recovery to support investments; the integrated utility procuring the needed supplies through either direct investments or through long-term contracts with IPPs. Shifting risks to customers makes investing in regulated markets an attractive option for IPPs if they are able to find a long-term contracting opportunity; this also enables them to reduce their financing costs. Municipalities and cooperatives in Texas operate under this regulated model, as do state-regulated utilities throughout SPP, most of WECC, most of MISO, and the Southeastern Electric Reliability Council (SERC).

Restructured markets, on the other hand, use market-based mechanisms to attract investments, thereby placing the risk of uneconomic investment decisions on suppliers. Because long-term contracts are generally unavailable in restructured markets, suppliers bear substantially higher risks and financing costs than in regulated systems. Among restructured markets, there are a range of market designs for providing investment incentives to suppliers, from energy-only markets such as in Texas and Alberta, to forward capacity markets as in PJM and ISO-NE.[94]

Many generators in ERCOT stated in our interviews that the energy-only market is excessively volatile and uncertain, and that they would prefer that ERCOT adopt a structure with more forward price certainty, such as a forward capacity market. Interestingly, in our recent review of PJM's forward capacity market, we heard many similar concerns about capacity price volatility and uncertainty.[95] A number of PJM suppliers proposed to extend the forward period of the capacity market, or extend the capacity market into long-term products reminiscent of long-term PPAs or a regulated planning construct. After considering these market participants' arguments and analyzing the overall volatility in returns in both energy-only and capacity markets for a number of clients, our view is that energy-only markets are somewhat more volatile, uncertain, and difficult to model than capacity markets and will likely require somewhat higher projected

[94] For a comprehensive discussion of this range of market designs for resource adequacy, see Pfeifenberger, Spees, and Schumacher (2009).

[95] See Pfeifenberger, Newell, Spees, *et al.* (2011).

returns to attract investment. This view is also consistent with some of the ERCOT generators who have investment experience in both PJM and ERCOT.

Capacity markets will tend to reduce some year-to-year price volatility, largely because: (1) a portion of capital recovery is independent of power prices and weather, rather than to more-volatile realized conditions; and (2) forward capacity markets impose reliability requirements that will tend to keep the market at a more stable reserve margin and will therefore partially stabilize energy market prices. However, because both well-functioning capacity and energy-only markets tie investor returns to underlying market conditions, they place much greater risks on suppliers than do regulated market structures.

3. Estimated Cost of Capital and Range of Potential Investment Hurdle Rates

When making an investment in ERCOT, debt and equity investors will consider the minimum required return, called the cost of capital, needed to make the investment worthwhile. The project's cost of capital is the opportunity cost for a marginal investor—the rate of return that capital could be expected to earn with an alternative, equally risky investment. Because investment risks differ, investors will have a different required return for each type of project they consider. For example, an investor will have a relatively low cost of capital for a project built in ERCOT with a long-term PPA that provides a stable revenue stream. Alternatively, investors will face greater risk, and require a higher cost of capital, for a project fully exposed to market risks.

Most investments are financed by some combination of debt and equity. Generally, the riskier the investment, the less debt financing is available as a percentage of total capital. While established financial theory posits that the cost of capital of an incremental investment is independent of the capital structure of a particular asset, in practice the cost of capital is estimated by observing the proportions of debt and equity in a project or company and calculating the weighted-average cost of capital, typically on an after-tax basis.[96] Additionally, a number of smaller investors may also apply investment "hurdle rates" that exceed the cost of capital because of their greater sensitivity to volatility and more limited ability to diversify risks. While no data are publicly available on the hurdle rates for smaller, undiversified investors, they and their private equity partners have anecdotally reported higher required investment returns. Ultimately, however, such players may find it difficult to compete with larger companies who can manage and diversify risks more cost-effectively.

We estimated the after-tax weighted-average cost of capital (ATWACC) for a merchant generation project as we did in a recent study for PJM and as summarized in Table 7. We first calculated the ATWACC for a sample group of publicly-traded merchant generation companies using the capital asset pricing model and recent market data.[97] Then we added 40 basis points to the value-weighted average across companies consistent with our PJM study.[98] That adder

---

[96] See Brealey, *et al.* (2011), pp. 472-488.

[97] See Spees, *et al.* (2011), pp 35-41.

[98] For fairness opinions that ranged from 7.1% to 12.0% see Cohen and Gordon (2010a), p. 7; Mirant and RRI (2010), p. 42; Cohen and Gordon (2010b), p. 7; Mirant and RRI (2010), p. 48; Dynegy (2010), p. 48; FirstEnergy Corp. and Allegheny Energy (2010), p. 85; FirstEnergy and Allegheny (2010), p. 84; and Duke (2011), p. 102.

brought our PJM estimate closer to the middle of a range of ATWACC estimates from equity analysts rendering the reviewed fairness opinions for merger and acquisition transactions in PJM. It was also intended to reflect the business risk of a merchant investment in PJM compared to the more diversified and partially hedged portfolio of the larger companies for which we calculated an ATWACC. Consistent with financial market conditions that have lowered financing costs since we estimated the cost of capital for PJM a year ago, the result of our updated ATWACC estimate for a merchant generation project is 7.6%.

Table 7
ATWACC Estimates for a Portfolio of Merchant Generation Companies

	Cost of Equity (%) [1]	Cost of Debt (%) [2]	Debt-to Equity Ratio [3]	Tax Rate (%) [4]	ATWACC (%) [5]
Value-Weighted Portfolio Average	10.69%	7.68%	58/42	40%	7.2%
Estimated Merchant Project Parameters	**10.80%**	**7.68%**	**50/50**	**40%**	**7.6%**

Sources and Notes:
Data source is Bloomberg (2012) as of 4/2012.
Estimated merchant ATWACC represents the value-weighted portfolio average of Calpine, Mirant, RRI, and NRG. We added 40 basis points to the portfolio average, as explained above.
[1] Risk free rate for U.S. Treasuries plus a risk premium, defined as the portfolio beta multiplied by the market premium, see Brealey, *et al.* (2011), p. 193.
[2] Each company in the portfolio's bond yield, weighted by its 5-year average long-term debt
[3] Each company in the portfolio's 5-year average debt-to-equity ratio, weighted by market capitalization
[4] See KPMG (2010), p. 26.
[5] [1] x [3] + [2] x [3] x (1 – [4]), see Brealey, *et al.* (2011), p. 216.

However, some ERCOT investors we interviewed suggested that a significantly higher cost of capital may be appropriate for ERCOT due to the greater pricing uncertainty of an energy-only market. For these reasons, we also report two higher estimates. First, we considered the 9.6% ATWACC assumption that ERCOT used for long-term planning a year ago, although cost-of-capital estimates have declined since ERCOT developed its estimate.[99] To be consistent with ERCOT's assumptions and generators' filings with the PUCT, we use this as our "base case" hurdle rate for our analysis in Sections III and IV.[100] Second, through our investor interviews, we found that at least some investors claim that the investment hurdle rate for merchant generation investments in ERCOT is even higher. We therefore also provide a "high" estimate for investment hurdle rates of 11% based on these investors' statements. Table 8 summarizes these three estimates.

[99] See ERCOT (2012p).
[100] See PUCT (2012a), Item Number 87, p. 1 and ERCOT (2012p).

Table 8
Range of Investor After-Tax Hurdle Rate Estimates

Low	7.6%	2012 average for publicly held IPPs (NRG, Calpine, RRI, and Mirant)
Mid	9.6%	2011 ERCOT Long-Term Planning assumption
High	11.0%	Based on generator investor interviews

Sources and Notes:
Mirant and RRI merged in December 2010 to form GenOn. Our analysis spans the time period before and after the merger, prior to which RRI and Mirant are tracked as separate companies and after which our reported results reflect the performance of the merged company.

4. Estimated Cost of New Entry

We estimated the cost of new entry for new gas CT and CC plants based on our recent CONE study for PJM.[101] That study identified the most efficient configuration for simple-cycle and combined-cycle gas-fired plants and included a bottom-up estimate of the cost of building and operating such plants. For the purposes of the present study, we assume that the appropriate reference technologies in ERCOT are similar to those in PJM, including: (a) a 390 MW CT plant with 2 7FA.05 turbines and a 10,300 Btu/kWh heat rate; and (b) a 656 MW, 2x1 CC plant with 7FA.05 turbines, a 7,000 Btu/kWh heat rate, and duct firing.

We adjusted our overnight capital cost estimates to account for locational cost differences between Illinois and Texas using locational cost inflation factors that R.W. Beck developed in its study for the U.S. Energy and Information Administration (EIA).[102] Texas has a lower cost index than Illinois due to factors including labor productivity and rates, taxes, delivery charges, and weather-related construction interruptions. We did not adjust for differences in electrical network upgrade costs allocated to the developer. Table 9 summarizes our resulting CC and CT capital cost estimates for ERCOT.

Table 9
Gas CT and CC Capital Cost Estimates

	Total Plant Capital Cost ($M)	Net Summer ICAP (MW)	Overnight Cost ($/kW)	Fixed O&M ($/kW-y)
June 1, 2015 Online Date (2015$)				
CT Capital Cost	$260.0	390	$667.4	$15.2
CC Capital Cost	$523.6	656	$798.2	$16.0

Sources and Notes:
See Spees, *et al.* (2011) and EIA (2010).

To determine the expected annual revenues that a merchant generator in ERCOT would require on an investment, we levelized these capital and FOM costs over a 20-year economic plant life after considering the financing costs of that investment. Such levelized costs are often the basis

[101] See Spees, *et al.* (2011).
[102] See EIA (2010).

for the contract price in long-term power purchase agreements, which may be structured as annual payments that are constant over the contract duration, or as annual payments that increase over time.

We estimate the annual CONE using a "level-real" cost recovery path, which reflects levelized payments that are constant in inflation-adjusted real terms, representing an implied assumption that net market returns will increase with an estimated 2.5% annual inflation rate.[103]

Table 10 shows the resulting range of CT and CC CONE estimates in ERCOT based on the range of financing cost estimates from Section II.D.3 above. We use this range of CONE estimates to evaluate the attractiveness of incremental investments in ERCOT compared to potential project revenues in Sections III and IV.

Table 10
Range of Level-Real Cost of New Entry Estimates

		Low	Mid	High
ATWACC	*(%)*	7.6%	9.6%	11.0%
CT Cost of New Entry	*($/MW-y)*	$90,100	$105,000	$116,000
CC Cost of New Entry	*($/MW-y)*	$112,400	$131,000	$145,000

III. CURRENT MARKET OUTLOOK

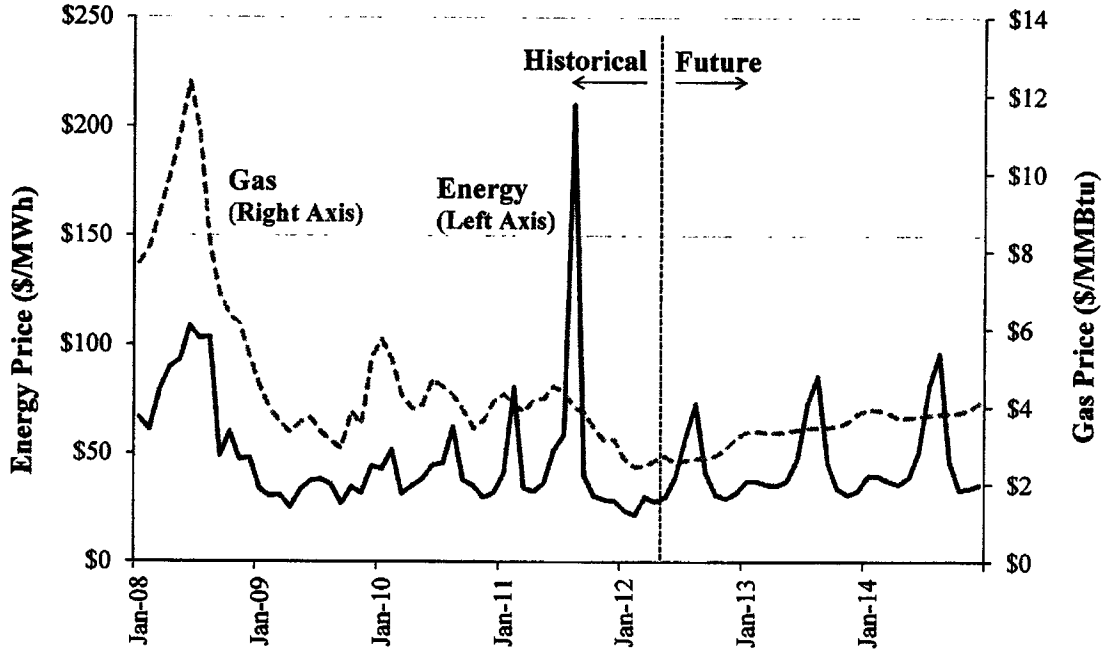
This section analyzes current market activity to assess progress in attracting new investment in ERCOT. We focus on trends in futures prices, stakeholder perceptions of the current market, and recent developments in resource additions.

A. Trends in Futures Prices

Futures prices provide valuable insights into market participants' expectations about future market conditions and about their willingness to place bets today based on those expectations. Figure 15 shows recent spot prices and futures prices for natural gas and on-peak energy. Gas prices increase only moderately from current lows to about $4/MMBtu for delivery in 2014, remaining far below prices from 2008 and earlier. Annual average on-peak energy futures also increase from $36/MWh in 2012 to about $47/MWh for delivery in 2014, with most of the increase occurring in July and August when hot weather can cause spot prices to spike. The market appears to be anticipating such spikes to increase between 2012 and 2014 as projected reserve margins tighten. However, the market does not appear to be anticipating price spikes as severe as in August 2011, when unusually extreme weather caused real-time prices to reach the $3,000/MWh cap in 19 hours and raised the monthly on-peak average price to more than $200.[104] Outside of the summer months, on-peak energy futures remain below $40/MWh.

[103] See Spees, *et al.* (2011), Section VI.A. for inflation discussion and Section VII.C. for levelization discussion.
[104] Ventyx (2012).

Figure 15
Monthly Gas and On-Peak Energy Prices



Sources and Notes:
On-peak power at ERCOT North Hub; Gas at Houston Ship Channel.
Historic gas and power prices from Ventyx (2012).
Futures prices shown as of 5/25/2012. Power futures from Platts (2012); gas futures from Ventyx (2012).

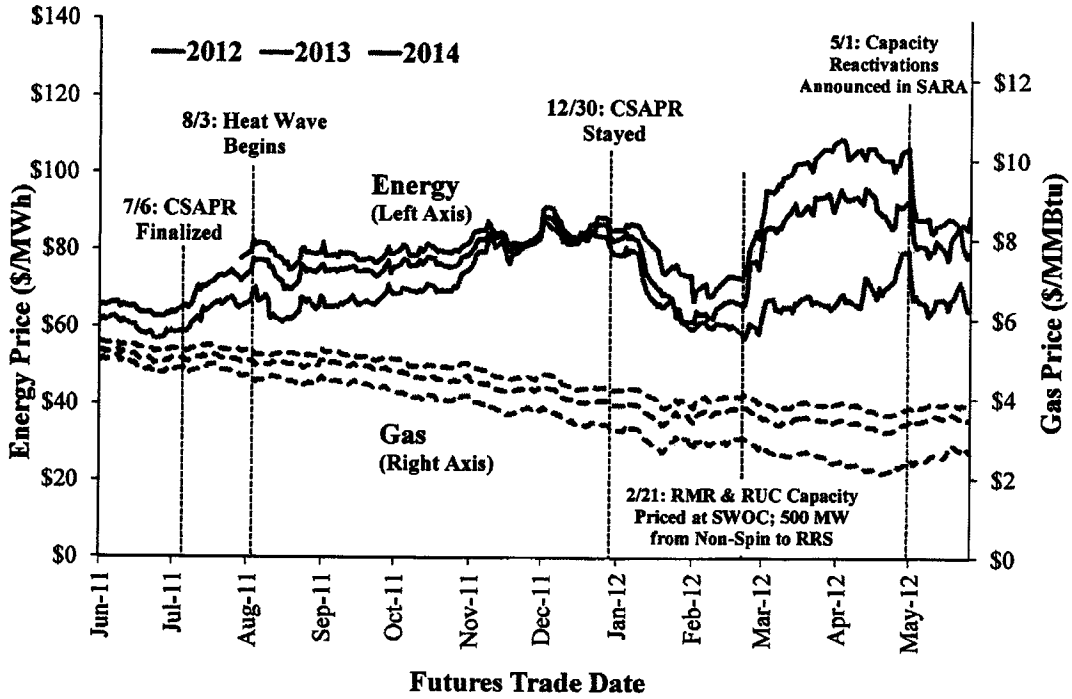
Figure 16 focuses on futures prices for just July and August delivery. By showing futures prices on the y-axis and the trade date on the x-axis, the figure illustrates how market participants' valuations of the same future product evolved over time. Prices for energy futures generally increased even as gas futures decreased. For example, power futures for delivery in July and August 2013 (in solid blue) traded at $66/MWh in June 2011, but increased to $80/MWh by May 2012, while gas futures for July and August 2013 delivery (in dashed blue) fell from $5.2/MMBtu to $3.5/MMBtu.

In order to remove the effect of gas prices and isolate the effect of anticipated market tightness on power prices, Figure 17 shows implied market heat rates. Implied market heat rates are calculated as the power price divided by the gas price. The result is that implied market heat rates increased even more than energy prices.

We believe implied market heat rates increased because the resource adequacy challenge facing ERCOT became more apparent, and because the PUCT and ERCOT enacted reforms that will increase the likelihood and magnitude of scarcity pricing. Futures market heat rates were stable in June 2011; then began to increase following the finalization of CSAPR in July and the heat wave and scarcity prices in August. We heard from market participants that witnessing sustained high prices without regulatory intervention substantially increased market confidence. Market heat rates increased throughout Autumn as the PUCT considered market reforms to better reflect scarcity, fell in January following the stay of CSAPR, and then rose steeply following the PUCT's approval of several market reforms in late February. Market heat rates fell in early

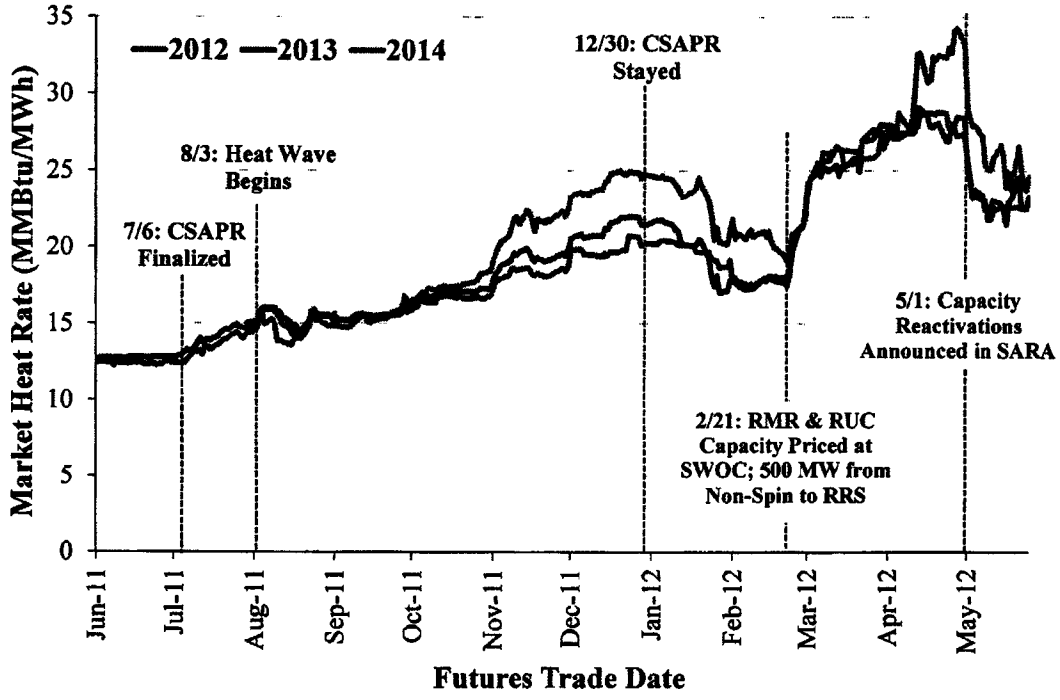
May, and some market participants attributed this to the announcement of capacity reactivations in the Seasonal Assessment of Resource Adequacy (SARA), as discussed further in Section III.D. Overall, expected heat rates for peak summer periods have nearly doubled since ERCOT's resource adequacy challenge emerged in summer 2011.

Figure 16
Futures Prices for July-August Delivery of Gas and On-Peak Energy Futures



Sources and Notes:
On-peak energy at ERCOT North Hub, from Ventyx (2012). Gas at Houston Ship Channel from Platts (2012).

Figure 17
Implied Market Heat Rates for July–August Delivery



Sources and Notes:
Market heat rate calculated as electric price divided by gas price.
On-peak power at ERCOT North Hub, from Ventyx (2012); gas at Houston Ship Channel, Platts (2012).

B. Comparison of Futures Prices to CONE

Increasing futures prices indicate an improving environment for investment but do not necessarily mean that prices will be high enough to attract new power plants. We use two approaches to evaluate whether futures prices indicate that market participants believe market prices will be high enough to support investment in new gas-fired combined-cycle (CC) power plants.[105] First, we use a simplified approach similar to that used by generation representatives in a recent PUCT filing in which they estimated a generic CC's annual energy margins based on futures prices.[106] We multiply monthly on-peak and off-peak spark spreads for a CC with a heat rate of 7•MMBtu/MWh by the number of on-peak and off-peak hours each month.[107] This approach indicates that a CC would earn $97/kW-year based on 2013 forward prices, and $95/kW-year based on 2014 prices.[108] These margins are below the levelized cost of new entry (CONE), which we estimate to be $112-145/kW-year, as calculated in Section II.D.4.

---

[105] Estimating whether futures prices support investment in a CT would require forward prices for super-peak products, for which we do not have visibility.

[106] See Group of Competitive Texas Generators (2012). Their approach assumed a constant $4/MMBtu gas price and did not consider off-peak futures, as well as other minor differences.

[107] Negative spark spreads are excluded. Based on 4,100 on-peak hours and 4,660 off-peak hours per year.

[108] Based on futures traded 5/25/2012.

Second, we estimate energy margins using a virtual dispatch against hourly prices consistent with futures. This more sophisticated approach reflects the additional market value introduced by price variations throughout the day and month that will result in greater margins for a plant that can operate during profitable hours and shut down during unprofitable hours. We also account for the negative impact of operating constraints and costs not captured in the simplified spark spread estimate described above.

To add a realistic amount of hourly price volatility around average monthly on-peak and off-peak futures, we apply heat rate shapes observed in the real-time market over 2008-2011, and in the 2011 day-ahead market (while maintaining consistency with futures prices on an on-peak and off-peak monthly-average basis). We then estimate a generic CC's energy margins using a virtual dispatch against the hourly prices, accounting for: (a) startup and ramping costs; (b) minimum up and down time constraints; (c) variable operations and maintenance costs; (d) forced and planned outages; and (e) minimum load, baseload, and maximum load with duct firing operating modes. Table 11 shows the energy margins and capacity factors estimated with each of the five modeled price shapes.

This approach indicates that a CC may be expected to earn approximately $91/kW-year based on 2013 futures, and $93/kW-year based on 2014 futures. These energy margins are slightly lower than those estimated without accounting for price volatility or operational characteristics, and are still below CONE. Overall, it appears that there is a substantial gap between market expectations about future energy prices and the prices needed to attract new combined-cycle power plants. At these levels, we would not expect market participants to invest in generic new combined-cycle capacity in ERCOT at least through 2014.

Table 11
CC Energy Margins Based on Virtual Dispatch

	2013 Futures		2014 Futures	
Virtual Dispatch Price Shape	Energy Margins *($/kW-y)*	Capacity Factor *(%)*	Energy Margins *($/kW-y)*	Capacity Factor *(%)*
2011 DA	85	67%	91	59%
2011 RT	94	53%	101	46%
2010 RT	92	56%	98	51%
2009 RT	95	59%	102	56%
2008 RT	97	61%	104	58%
Average	**93**	**59%**	**99**	**54%**

Sources and Notes:
Calculated based on futures with a trade date of 5/25/2012.
Platts (2012) and Ventyx (2012).

C. Stakeholder Perceptions of the Futures Market

In our interviews, we asked market participants to share their perceptions of the forward market in ERCOT. Potential generation investors emphasized that the two- to three-year horizon of futures prices is not long enough to support investment in an asset with a lead time of three years and an operating life of several decades. They stated that futures prices are, however, a helpful indicator showing that prices are not yet high enough to support new investment, a statement that is consistent with our own analysis. They also claimed that, while recent market reforms are a step in the right direction, future scarcity pricing conditions may be too rare and unpredictable to support investment. We note however, that these statements must be taken in context and that the true test of generators' willingness to invest will be whether they actually move ahead with projects once prices are consistent with CONE.

Some generators commented that 2013 and 2014 futures are probably underpriced relative to what actual spot prices will be. They observed that despite expectations of a lower reserve margin and a tighter market, Summer 2013 and 2014 forward heat rates are not higher than 2012 heat rates, as shown in Figure 17. A discount in futures prices could be introduced because REPs generally focus their hedging activities only on the upcoming Summer, leaving very few buyers for futures in later years. Further, the futures market may not fully reflect new rules whose implications for prices are difficult to model.

Stakeholders also stated that liquidity in futures markets is currently low, especially for super-peak products. We have heard this anecdotally but do not have data to support or refute this claim. We suspect that current regulatory uncertainty may be reducing liquidity, as both buyers and sellers are reluctant to enter transactions while the PUCT deliberates important market reforms. Some stakeholders blamed the low liquidity on high price caps, but we expect that high price caps could actually increase liquidity for peaking products as REPs become more motivated to take fully hedged positions and avoid excess exposure to extreme prices. We believe that concerns about forward markets will diminish over time as uncertainties resolve and market participants' hedging strategies adapt.

D. Observed Developments in Resource Investment

While futures prices are likely not high enough to support new greenfield generation investment, price increases expected from market reforms and declining reserve margins have attracted some low-cost, short lead-time investments such as uprates and reactivations of mothballed units. For example, generation owners have already reactivated or announced the reactivation of nearly 2,000 MW of mothballed capacity for Summer 2012.[109] Additionally, in April, Calpine announced two combined-cycle expansions totaling 520 MW.[110] Notably, these combined-cycle expansions cost less than $550/kW, giving them a substantial economic advantage over a new combined-cycle plant at approximately $800/kW.[111] The economic advantage of Calpine's two uprates is further enhanced by associated plant-wide efficiency improvements of 5% and 17% respectively that effectively lower the relative cost of the uprates even further. Based on generation owners' comments in our interviews, we expect additional low-cost uprates and

[109] See ERCOT (2012m) and (2012t).
[110] See Calpine (2012).
[111] See Calpine (2012). CC overnight cost from Section II.D.4 above.

reactivations to come online before 2014. These owners, however, may be reluctant to announce investment plans while the PUCT is actively considering whether to increase price parameters in response to an expected capacity shortfall. Beyond these reactivations and expansions, no major units are currently under construction.[112] Furthermore, several permitted projects which were expected to have already begun construction are now on hold and will not be completed until 2015 at the earliest.[113]

IV. LONG-TERM RESOURCE ADEQUACY OUTLOOK

This section describes how we estimate the "economic equilibrium" reserve margins that ERCOT's market structure is likely to achieve under current and proposed price caps. Our primary finding is that increasing the price cap to $9,000 will attract more investment, but ERCOT is still likely to fall substantially short of its current reliability target until several thousand megawatts of additional demand response are able to prevent load shedding without eliminating scarcity prices. In addition, we anticipate substantial uncertainty and year-to-year variability in the reserve margin that the current market structure will achieve.

A. METHODOLOGY

We estimate an "economic equilibrium planning reserve margin" at which generation developers would be willing to invest, using the following methodology. First, we simulate the energy margins that a new GE 7FA simple-cycle CT would earn over a range of planning reserve margins.[114] Next, we compare these energy margins to the levelized cost of new entry (CONE) that a new plant must expect to earn on average over many years to attract investment, as estimated in Section II.D.4 above. Finally, we identify the highest planning reserve margin at which energy margins exceed CONE. This represents the economic equilibrium reserve margin that the energy-only market is likely to produce on a long-term average basis.

We use outputs from ERCOT's Loss of Load Expectation (LOLE) Model to estimate the frequency of involuntary load shedding and high-priced scarcity events at each potential planning reserve margin. We estimate generators' net revenues over a range of reserve margin as a sum of three components: (1) energy margins earned during scarcity events, based on the frequency of scarcity events indicated by the LOLE model; (2) non-scarcity margins, which are a function of both reserve margins and gas prices; and (3) adders to non-scarcity margins that we expect from recently-implemented market rules, based on ERCOT's recent backcasting analyses.

---

[112] See ERCOT (2011f), p. 16, Total Future Non-Wind Resources. We exclude Sandy Creek 1, which completed construction but experienced an accident during testing in 2011.

[113] See ERCOT (2011b), p. 6. Pondera King Power Project, Las Brisas Energy Center, and Coleto Creek Unit 2 have delayed their commercial operations dates.

[114] We focus on a simple-cycle combustion turbine rather than a combined-cycle for simplicity, although we estimate that both would be economic in equilibrium. Specifically, we focus on a GE7FA-based simple-cycle combustion turbine, which has a large turbine with significant economies of scale compared to aeroderivative models, and a much lower cost per kW.

1. Generator Energy Margins During Scarcity Conditions

ERCOT's Loss of Load Expectation Model. We rely on ERCOT's LOLE model to project the frequency of load-shedding events and scarcity events that will trigger scarcity prices. This model estimates the frequency of scarcity events when extreme weather-driven load and high generation outages coincide. Probability distributions of load and generation, including the tails of those distributions, drive the simulated frequency of scarcity events. ERCOT provided simulations using several different weather-driven load profiles (each 8,760 hours), with thousands of random draws on generation outages for each load profile. A primary output is the frequency of load-shed events (when total supply is insufficient to meet demand) expected at a given planning reserve margin, which ERCOT usually uses to identify the reserve margin that achieves the reliability target. However, it is similarly useful to us for estimating the frequency of various levels of scarcity prior to load shedding. We incorporate load shedding and other metrics that we translate into scarcity pricing events (as described below) over several model runs representing a large range of planning reserve margins.

To support our study, ERCOT's planning department provided simulation results using the same modeling assumptions and input data used to develop their most recent LOLE study in 2010.[115] ERCOT staff made only one update to the LOLE model for the purposes of our study, modeling fifteen years of weather data from 1997 to 2011, rather than modeling five representative weather years selected from 1996 to 2010. Importantly, this update allows the LOLE model to account for the extreme conditions experienced in 2011. After including 2011 weather data, a 15.25% planning reserve margin is needed to achieve the "one loss of load event in ten years" (0.1 LOLE) reliability target, which is greater than ERCOT's currently approved 13.75% target. Because ERCOT is in the process of implementing several model enhancements regarding wind generation and unit availability, and because the reliability estimate is highly sensitive to assumptions regarding weather weights (as explained further below), the results we report here will not exactly match those that will be reported in ERCOT's forthcoming updated LOLE study. Our estimated reserve margin needed to achieve the 0.1 LOLE target should not be interpreted as the new target for ERCOT.[116]

The choice of weather-years and the probability weights assigned to each year strongly affect the model results. We used 15 historical weather profiles spanning 1997 to 2011 and weighted each equally with a 1-in-15 chance of occurring (sensitivity analyses using different weights are presented later). Figure 15 shows the 2012 peak load estimate consistent with each of these 15 weather profiles as well as the 15-year average and the normal weather peak load estimate.[117] The peak load estimate based on 2011 weather is substantially higher than the peak load based on other years because 2011 weather was so unusually extreme. The extreme loads lead to more frequent scarcity events that would not exist in more typical weather years.

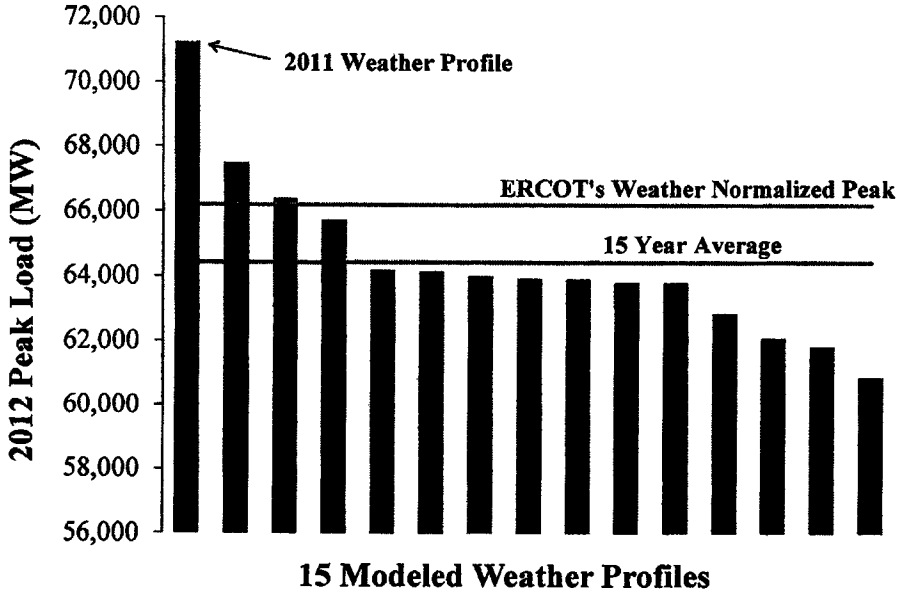
---

[115] See ERCOT (2010a).

[116] Furthermore, comparisons of our results to prior LOLE studies also must consider the following factors that have changed: the introduction of extreme 2011 weather; our recognition of 1,700 MW load over-forecast error during super-peak periods which reduces the incidence of load shedding and scarcity events; and ERCOT's new (higher) weather normalization methodology which increases weather-normalized load and therefore increases the amount of capacity implied by any given planning reserve margin.

[117] All loads modeled represent a 2012-sized economy in ERCOT. This one-year analysis is useful for our long-term equilibrium model because, although load would be higher in future years, we expect the relationship between reserve margin and scarcity events to remain approximately constant.

Figure 18
2012 Peak Load Estimate Under 15 Weather Profiles



Sources and Notes:
Data provided by ERCOT (2012a).
Consistent with load estimates under weather uncertainty provided in the *ERCOT Planning 2012 Long-Term Demand and Energy Forecast*, see ERCOT (2011h), p. 12.
ERCOT's current weather normalization methodology places its weather-normalized peak for 2012 substantially above the mean and median peak loads corresponding to the last 15 years' weather profiles imposed on a 2012 economy.

We translated the LOLE model outputs into scarcity prices in two steps. First, we recorded the frequency of hours in which the residual capacity falls below several levels of residual supply at each planning reserve margin analyzed. Second, we translated residual supply metrics into scarcity prices that are consistent with ERCOT's operations and its current and proposed scarcity pricing rules.

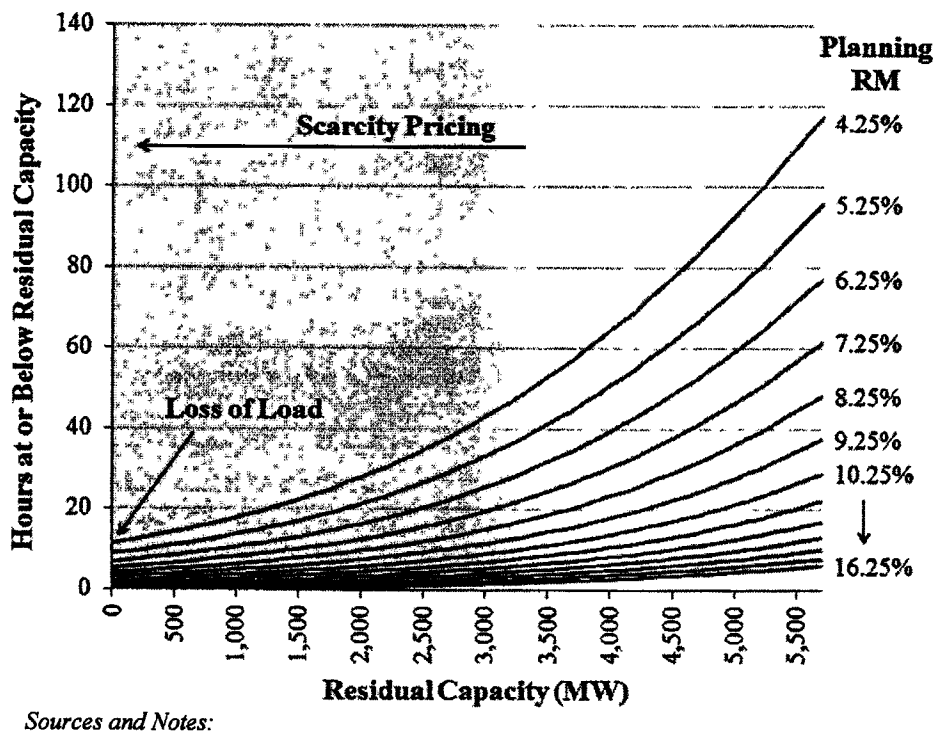
Frequency of low residual capacity. In each load hour and generation outage draw, the LOLE model calculates the residual capacity. For our purposes, we treat "residual capacity" as all available generation and controllable load response resources that ERCOT has not deployed, but could deploy to avoid shedding load. We record whenever residual capacity falls below certain low levels, at every 500 MW increment below 5,700 MW.[118] In interpreting the results, however, we adjust for demand elasticity during scarcity that is likely to occur in reality but is not accounted for by the LOLE model. Large industrial customers exposed to real-time prices are likely to reduce their consumption when prices spike, customers with four coincident peak (4CP) rate structures are likely to reduce consumption when they anticipate a 4CP interval, and the public may respond to emergency conservation appeals. Therefore, under scarcity

[118] Increments smaller than 500 MW are interpolated.

conditions, we adjusted the load downward, and therefore adjusted residual capacity upward. Our adjustment was 1,700 MW based on the observed error of ERCOT's load forecasting model at the 2011 peak load, when prices reached the $3,000/MWh cap, loads were likely anticipating a 4CP interval, and conservation appeals were in effect. This adjustment not only makes sense but also improved the calibration of our model, with scarcity event frequencies under 2011 weather and a 14% planning reserve margin close to those observed in 2011 under similar conditions.[119] However, the effects of this assumption on our analysis are tested in the sensitivity analyses presented in Section IV.

Figure 19 shows the projected number of hours with low residual capacity at varying planning reserve margins, averaged across the 15 modeled weather profiles. For example, at a 4.25% planning reserve margin, the LOLE model projects 120 hours with residual capacity less than 5,700 MW and 15 hours with zero residual capacity (*i.e.*, load shedding).[120] As expected, the frequency of load shedding and severe shortage events increases with lower planning reserve margins.

Figure 19
ERCOT's Loss of Load Expectation Model Projections

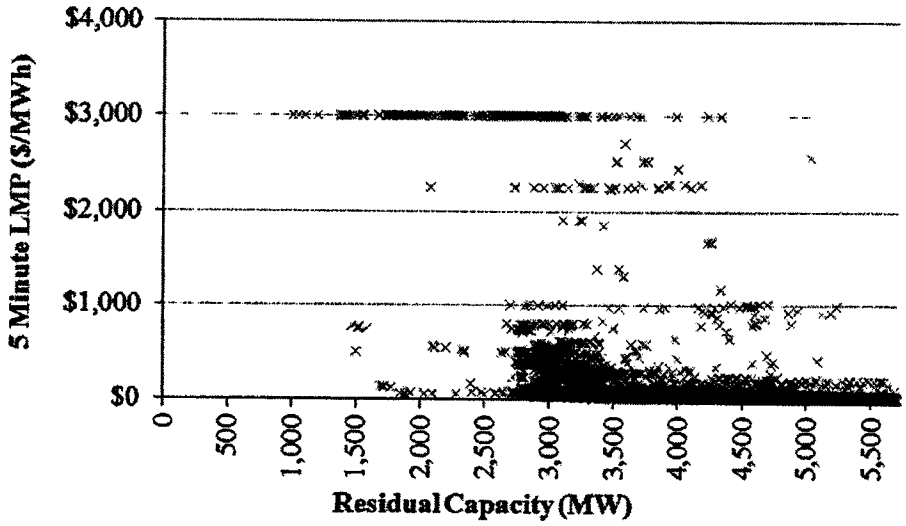


Sources and Notes:
Data provided by ERCOT (2012a), adjusted to account for demand elasticity during scarcity.

[119] In fact, a comparison against actual 2011 data showed that this adjustment is likely an under-correction.
[120] The frequency of these scarcity events varies among the 15 weather profiles; the plot shows the average frequency among these profiles.

Translation of Low Residual Capacity into Scarcity Prices. To translate low residual capacity levels into estimates of scarcity pricing, we first examine the relationship between residual capacity and scarcity pricing observed in 2011. In 2011, scarcity pricing rarely occurred when residual capacity was high, and scarcity prices were triggered with increasing frequency as residual capacity declined, as shown in Figure 20.

Figure 20
2011 Pricing and Residual Capacity Patterns



Sources and Notes:

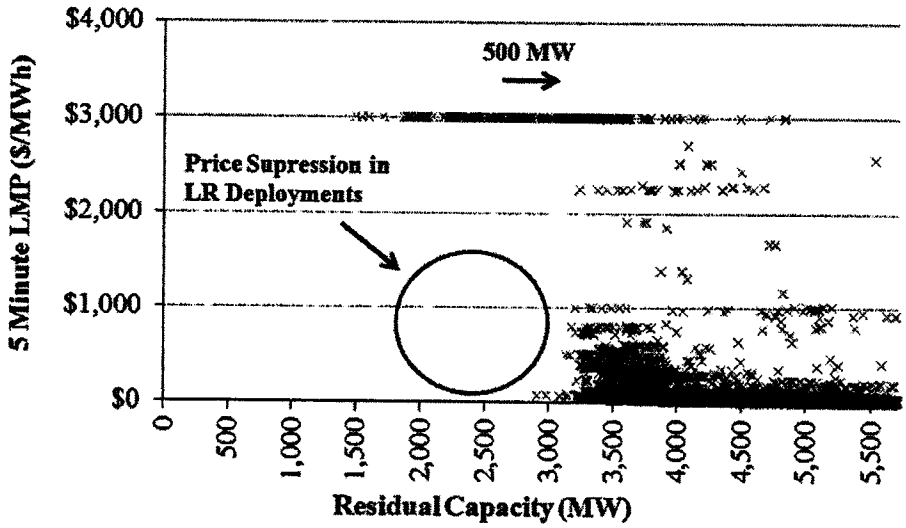
Data provided by ERCOT (2012a).

Residual capacity is calculated as Physical Responsive Capability + undeployed EILS, which is consistent with total residual capacity when all units are committed and transmission constraints are relaxed. This calculation was calibrated to check for consistency between historical data and comparable LOLE model results, and was validated by ERCOT staff to ensure consistent accounting of various resources.

We then adjusted the scarcity pricing patterns observed in 2011 to account for two reforms which will impact scarcity pricing in the future. The first reform is a 500 MW increase in responsive reserve capacity.[121] This will have the effect of triggering scarcity pricing earlier, when residual capacity is 500 MW higher, since the increase in responsive reserves will reduce the capacity available at prices below the system-wide offer cap and will essentially "widen the warning track" of scarcity pricing before load is shed. This effect is shown as a 500 MW shift to the right in Figure 21. The second reform is the proposed elimination of price suppression during load resource deployments.[122] There were several intervals observed in 2011 with very low residual capacity but with prices far below scarcity levels due to price suppression from the deployment of load resources. For example, many such intervals are apparent in the bottom-left corner of Figure 20, where prices remained below $1,000/MWh at a residual capacity of less than 2,500 MW. The effect of eliminating price suppression in these intervals is shown in the empty circle in Figure 21.

[121] See ERCOT (2012f), NPRR434, approved 02/21/2012.
[122] See ERCOT (2012f), NPRR444, pending with urgent status.

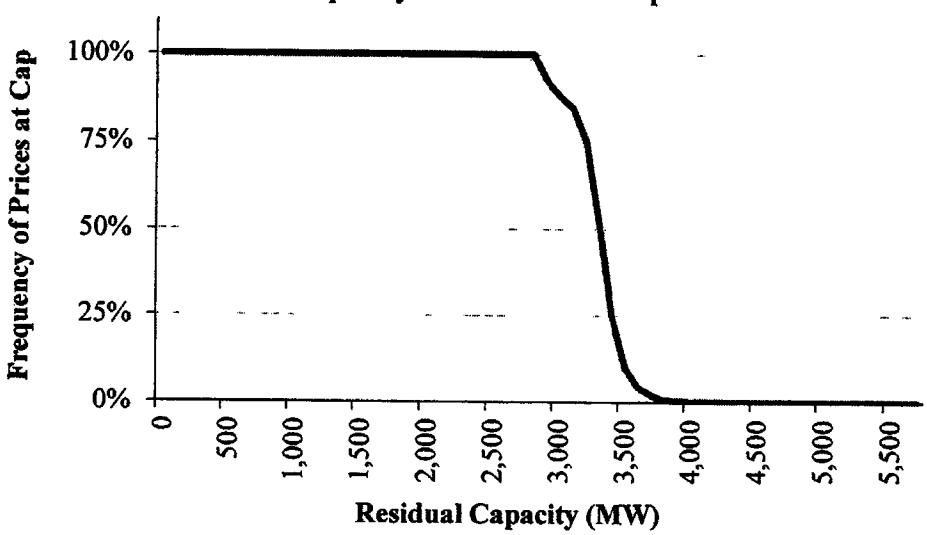
Figure 21
Impact of Reforms on Scarcity Pricing



Sources and Notes:
Data provided by ERCOT (2012a). *Brattle* Analysis.

After accounting for these reforms, we determined the frequency of prices at the system-wide offer cap for each residual capacity level, as shown in Figure 22. We then applied these frequencies to the outputs of the LOLE model to estimate the expected frequency of scarcity pricing. For example, with 3,500 MW of residual capacity, 8% of intervals have prices at the cap. Therefore, for each hour that the LOLE model projects a residual capacity of 3,500 MW, we apply an 8% chance that the price will be at the cap. (We used the same methodology to estimate the frequency of prices at various levels below the cap, down to $500/MWh).

Figure 22
Frequency of Prices at the Cap



Sources and Notes:
Data provided by ERCOT (2012a). *Brattle* Analysis.

Next, we account for the fact that generators would not be able to profit from all scarcity pricing intervals due to outages and operating constraints which can prevent them from capturing transient price spikes. We discount projected scarcity margins based on the generation patterns of actual combustion turbines in 2011 to calculate scarcity margins.

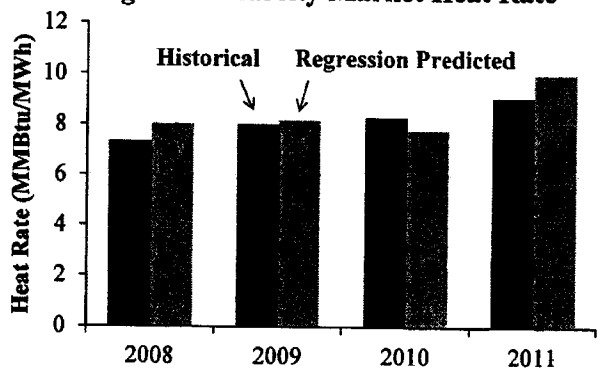
Finally, based on simulated residual capacity levels, our translation of residual capacity into prices, and our estimate of the fraction of such prices that a combustion turbine could capture, we estimate the energy margins a combustion turbine would earn. We do so at each planning reserve margin and each weather profile modeled in the LOLE model. We also calculate the *average* energy margins that a combustion turbine would earn over each of the 15 weather years.

2. Non-Scarcity Margins

The second component of our estimated energy margins is the margin a combustion turbine would earn during non-scarcity periods, when energy prices are less than $500.[123] We estimated non-scarcity margins using a regression analysis of historical data and applying the results to our forward-looking analysis at each potential reserve margin.

Our regression analysis relates hourly market heat rates to hourly reserve margins based on actual market data from 2008 through 2011. Hourly market heat rates are given by the ERCOT North real-time price divided by the daily natural gas price at the Houston Ship Channel. Hourly reserve margins are defined as the total installed capacity (each year) less the capacity unavailable due to outages each hour, divided by the actual hourly load net of wind generation.[124] We estimated this relationship using a segmented regression by partitioning the historical data based on the hourly reserve margin, and then estimating coefficients separately within each segment. This approach is a standard method,[125] and it allowed us to identify a steeper relationship between the market heat rate and reserve margins as reserve margins tighten. Figure 23 shows the result of the regression, comparing the average non-scarcity market heat rate predicted by the regression against actual historical averages.

Figure 23
Average Non-Scarcity Market Heat Rate



[123] We define non-scarcity margins as margins earned from prices less than $500. Margins earned at prices above this are defined as scarcity margins. For example, in an hour with a price of $3,000, we attribute $2,500 to scarcity margins, and $500 to non-scarcity margins. This avoids potential double-counting.

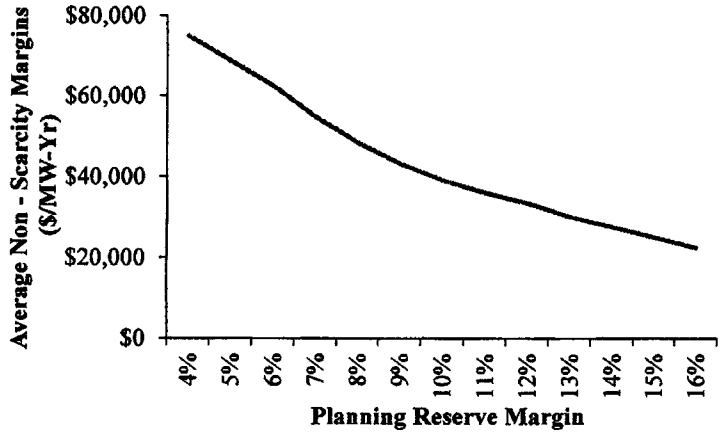
[124] ERCOT provided the hourly data on unit outages. Load data sourced from Ventyx (2012).

[125] See, for example, Boogert and Dupont (2008).

We then applied the results of the regression analysis to estimate energy margins going forward. We estimated market heat rates for each planning reserve margin and weather profile by applying the regression coefficients to hourly reserve margins (*i.e.*, total capacity less historical monthly average outage profiles, divided by forecast load net of wind generation). Based on these projected hourly market heat rates, we calculated energy margins based on a virtual dispatch against the hourly market heat rate, and a gas price of $3.5/MMBtu, which is the average forward price for 2013 delivery of natural gas at the Houston Ship Channel.[126] We accounted for start-up and variable operating and maintenance (VOM) costs.

Figure 24 shows the results of this projection at varying reserve margins, averaged across the 15 weather profiles. To account for the degree of uncertainty inherent in this type of projection, as well as uncertainty regarding future gas prices, Section IV.B.4 will present sensitivity analyses examining the effect of different non-scarcity margins on reliability outcomes.

Figure 24
Estimated Non-Scarcity Margins a CT Would Earn



3. Adders to Non-Scarcity Margins from New Rules

The third component of energy margins we estimated is "adders to non-scarcity margins from recently-implemented market rules." The impact of the recently approved ONRUC & non-spin price floors is not captured in historical non-scarcity margins, nor can it be estimated with the LOLE model because deployments of these resources are triggered by ramping and unit commitment issues, not true capacity shortages.

Non-Spin Price Floors — An ERCOT backcast analysis of 2011 estimated the impact of the non-spin price floors at less than $2,000 MW-year.[127] We assume a GE7FA-based combustion turbine would capture this amount every year, in addition to the scarcity and non-scarcity margins calculated above.

ONRUC Price Floor — We project margins for a GE7FA-based simple-cycle combustion turbine, which has a large turbine and is not as flexible as aeroderivative turbines. We

[126] Ventyx (2012). Futures traded 5/25/2012.
[127] ERCOT (2012j).

therefore assume they do not capture the RUC floor impact due to its brevity and unexpectedness. It is quite possible that this assumption is incorrect, and that a GE7FA combustion turbine would in fact capture some or all of the price floor impact. We find that assuming it would capture the price floor impact estimated by ERCOT's 2011 backcast analysis does not significantly affect the results of our simulations.

Based on the Independent Market Monitor's net revenue estimates, CC plants earned $20,000 – $30,000/MW-year more than combustion turbines from 2009 through 2011.[128] The difference in margins is similar to the difference in levelized cost of new entry between a CC and a CT, about $26,000/MW-year.[129]. A substantial rise in gas prices above 2009 – 2011 levels could improve the relative economics of a CC, as illustrated by CCs' revenue advantage of more than $50,000 MW-year during the high gas prices of 2008. However, both types of plants are likely to be economic in a long-term equilibrium.

[128] Potomac Economics (2012), and Potomac Economics (2011c).

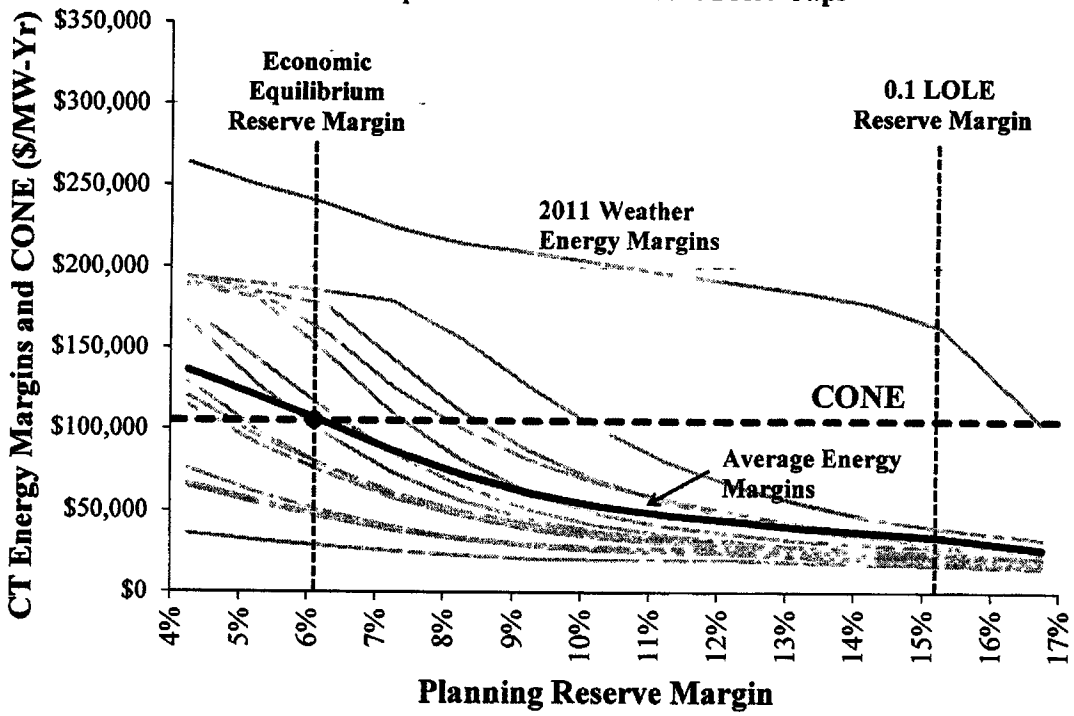
[129] Estimated CC CONE is $131,000 MW-Yr, estimated CT CONE is $105,000 MW-Yr. See calculations in Section II.D.4.

B. Estimated Reserve Margins this Market will Support

1. Economic Equilibrium Reserve Margins at Current Price Caps

To determine the economic equilibrium planning reserve margin, we combine scarcity, non-scarcity, and additional margins to estimate total energy margins, and then compare these to CONE. Figure 25 plots this comparison, with energy margins estimated under current price caps consisting of a $3,000/MWh High Cap (HCAP), a $500/MWh Low Cap (LCAP), and a $175,000/MW-year Peaker Net Margin (PNM) threshold which triggers the LCAP. At each planning reserve margin, the figure shows total energy margins for each of the 15 weather profiles (in light blue), average energy margins across the 15 profiles (in dark blue), and CONE (in red). Energy margins are greater at lower planning reserve margins, and are consistently greater than average with the extreme weather of 2011.

Figure 25
Economic Equilibrium with Current Price Caps



The economic equilibrium is the highest planning reserve margin at which total energy margins exceed CONE. With current price caps, the projected economic equilibrium planning reserve margin is 6.1%, which is more than 9 percentage points short of the 15.25% planning reserve margin needed to achieve the 0.1 LOLE target, assuming each of the past 15 years' weather patterns are equally likely. At this equilibrium planning reserve margin of 6.1%, ERCOT is projected to experience 2.2 loss-of-load events, 7.0 loss-of-load hours, and 35 hours with prices at the cap, on an annual average basis. In a worst-case year with the extreme weather of 2011, 27 loss-of-load events, 92 loss-of-load hours, and 248 hours with prices at the cap are projected. This extreme possibility strongly affects the average even though it is only assigned a 1-in-15 probability. In fact, across the fourteen weather profiles excluding 2011, only 0.5 loss-of-load events and 0.9 loss-of-load hours are projected on average.

2. Economic Equilibrium Reserve Margins at Higher Price Caps

Higher price caps increase generators' margins during scarcity intervals and therefore reduce the number of intervals needed to provide sufficient energy margins to support investment. We examine the impact of three sets of increased price caps on the economic equilibrium and long-run resource adequacy.[130] These price caps represent the range of proposals recently put forth by the PUCT and are summarized in Table 12.

Table 12
Range of Price Caps and PNM Thresholds

Scenario	HCAP	LCAP	PNM Threshold
	($/MWh)	*($/MWh)*	*($/MW-y)*
Current	$3,000	$500	$175,000
Mid	$4,500	$2,000	$262,500
High	$6,000	$3,000	$300,000
Highest	$9,000	$4,500	$300,000

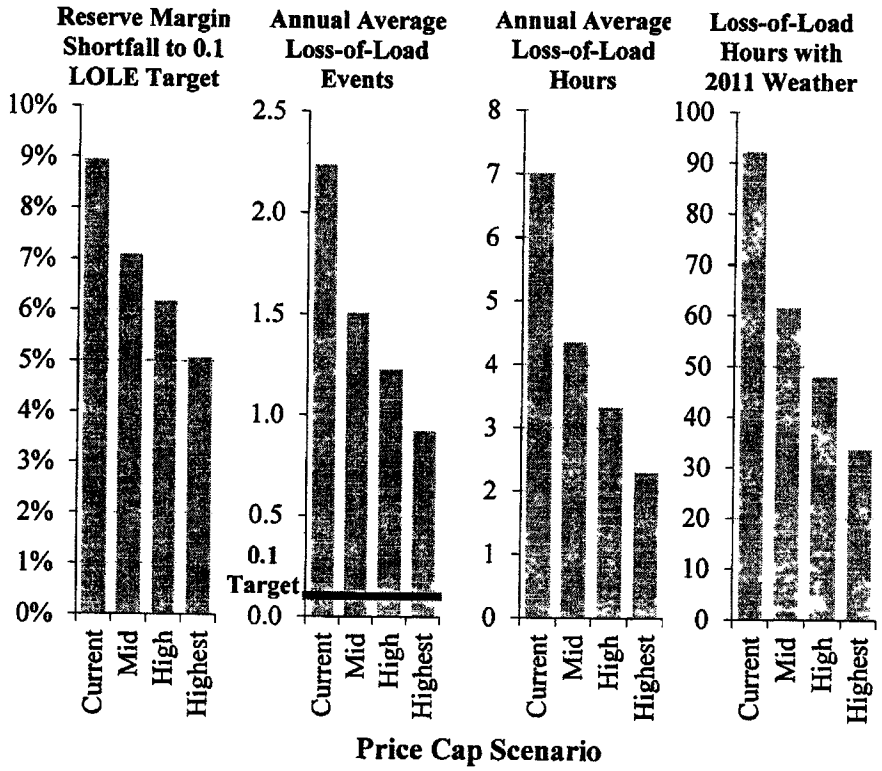
Figure 26 presents our estimates of long-run reliability achieved at the economic equilibrium with these price cap scenarios. Our key finding is that all scenarios fall short of the 0.1 LOLE target reserve margin. Even with the highest price caps, ERCOT is projected to experience an annual average of 0.9 loss-of-load events, and is exposed to the risk of experiencing more than 30 loss-of-load hours under extreme 2011 weather conditions.[131] We leave it to the regulators and stakeholders to determine whether such reliability outcomes would be adequate. Section VI.A of this report provides a framework for thinking about reliability objectives.

We also examine the impact of the PNM threshold and LCAP by simulating a variation of the highest price cap scenario without these caps. At equilibrium, this scenario is projected to achieve an annual average of 0.6 LOLE rather than the 0.9 LOLE achieved under the standard highest price cap scenario. We caution, however, that investors may be reluctant to trust that future regulators will not interfere with the prolonged high spot pricing patterns that would have to be allowed to prevail in order to support this improved reliability outcome, as discussed in Section VI.

[130] Our analysis of higher price caps does not account for potential changes in market behavior, exercise of market power, or additional demand response during scarcity.

[131] With the highest price caps, an average of 0.04 loss-of-load events and 0.06 loss-of-load hours are projected on average across the fourteen non-2011 weather profiles.

Figure 26
Reliability at the Economic Equilibrium with Higher Price Caps



3. Reliability and the Frequency of Scarcity Pricing Conditions

Generation in ERCOT currently faces low non-scarcity margins due to a flat supply curve and low gas prices, as discussed in Section I. Without capacity payments, potential investors therefore require a moderate number of scarcity pricing hours on average to recover the costs of a new plant. For example, if a combustion turbine earns non-scarcity margins of $30,000/MW-year and has a levelized CONE of $105,000/MW-year, eight hours of scarcity pricing are needed to support investment even with a price cap of $9,000/MWh.

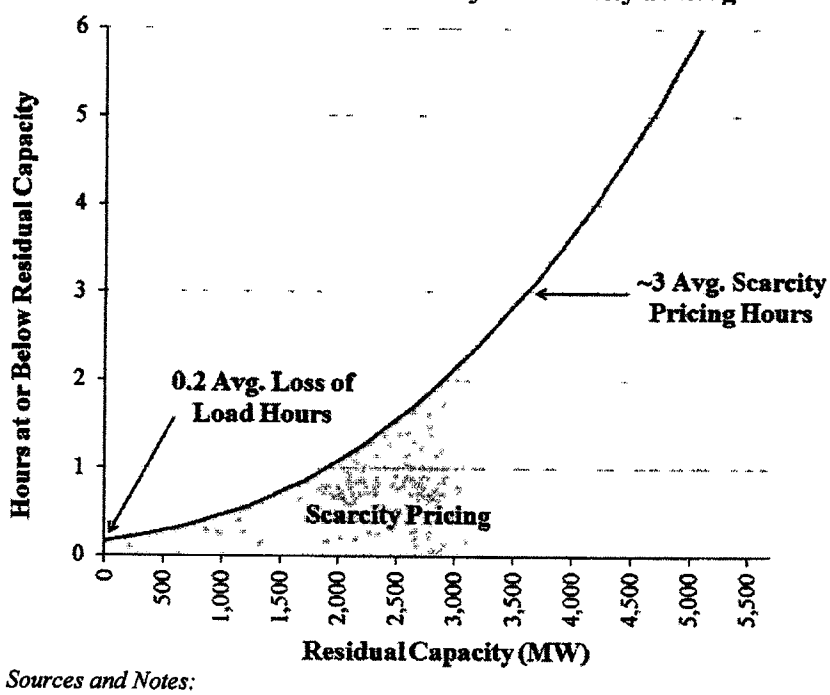
Such reliance on scarcity pricing to ensure long-term reliability presents a challenge simply because higher planning reserve margins increase reliability but also decrease the frequency of scarcity pricing. The shortfall between the economic equilibruim and the 1-in-10 LOLE target planning reserve margin we projected largely reflects the fact that very few scarcity pricing hours are likely to occur if the planning reserve margin is high enough to achieve the 1-in-10 LOLE target. In other words, the high reserve margin needed for reliability eliminates the very scarcity that is required for recovering investment costs.

Figure 27 illustrates this fundamental challenge. It shows the expected annual frequency (in red) of low residual capacity levels at the 15.25% planning reserve margin which achieves the 1-in-10 LOLE target. At this reserve margin, there is an average of 0.2 loss-of-load hours, as shown by

the expected frequency of 0 MW residual capacity. For simplicity, we show expected scarcity pricing intervals (in green) occurring at residual capacity levels lower than 3,500 MW.[132] Based on this, only about 3 hours of scarcity pricing are expected on average, substantially less than the number of scarcity hours per year that would support generation investment.

Thus, it is very difficult to achieve a 1-in-10 reliability target through scarcity pricing unless large amounts of demand response are able to avoid load shedding without eliminating scarcity prices. The expected reserve margin shortfall would decrease if price caps were increased even higher than $9,000, but the magnitude of the shortfall would be uncertain on average and year-to-year, as we show in the next section. In addition, we heard from many stakeholders that much higher caps might create prohibitively higher credit requirements.

Figure 27
Illustration of Reliability vs. Scarcity Pricing



Sources and Notes:
Data provided by ERCOT (2012a), adjusted to account for demand elasticity during scarcity. Average of 1997 – 2011 weather at a 15.25% reserve margin.

4. Sensitivity to Weather, Regulatory, Cost, and Modeling Uncertainties

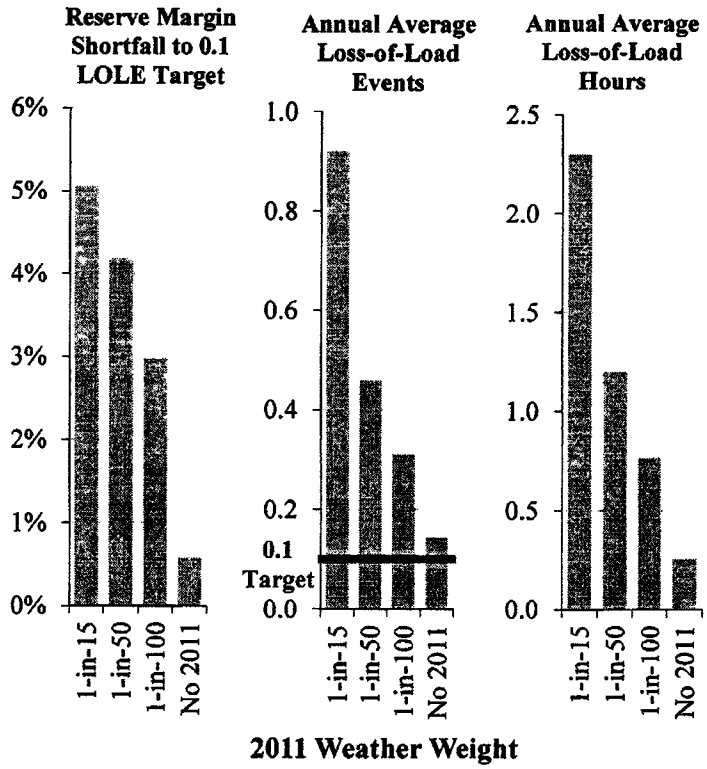
The analysis presented above involves numerous assumptions and modeling techniques. Reasonable analysts could differ in their assumptions and approach and arrive at different results. More importantly, *investors'* beliefs and willingness to place bets could differ from our analysis, and their decisions ultimately determine the level of investment. To test the sensitivity of our estimated "equilibrium reserve margin" and reliability outcomes to reasonable variation in assumptions, we analyzed the following five different cases:

[132] Section IV.A.1 describes the relationship between residual capacity and scarcity pricing in more detail.

Variation in Weather Distribution. While extreme weather is a major driver of scarcity margins and loss of load, its frequency is highly uncertain. To highlight the impact of extreme weather, we projected the economic equilibrium with various likelihoods of 2011 weather, ranging from 1-in-15 to zero.[133] Note that the base case described above is based on a 1-in-15 chance of 2011 weather, using a simple average of the 1997 – 2011 profiles.

Figure 28 presents estimates of reliability at the economic equilibrium with these weather distributions under the $9,000 offer cap scenario. As already presented above, the equilibrium reserve margin achieved assuming the base case weather distribution would result in 0.9 loss-of-load events per year on average. However, under an alternative assumption that 2011 weather will never occur again, the equilibrium reserve margin would very nearly achieve the 0.1 LOLE target. This highlights that the extreme events at the tails of the weather distribution are a major reason why reliance on scarcity prices in unlikely to achieve ERCOT's current reliability objectives. If it were not for these extremes, an energy-only market with high price caps would have a much better chance of achieving the current objectives.

Figure 28
Reliability at the Economic Equilibrium with a Range of Weather Distribution
Under the Highest Price Cap Scenario



[133] Our projections under varying weather distributions account for the impact of weather on both reliability and generators' scarcity margins. For example, a projection based on a zero chance of 2011 weather would have lower scarcity margins than the base case, as well as a lower loss of load frequency.

Investors' Weather Expectations Differ from Actual. Estimating the likelihood of extreme weather is highly speculative, and any disconnect between investor expectations and actual weather conditions will substantially affect the economic equilibrium. The base case assumes investors' weather expectations are perfectly accurate—meaning that they expect a 1-in-15 chance of 2011 weather and this is exactly what materializes on a long-run average basis. We examine a sensitivity where investors are overly optimistic and expect a 1-in-5 chance of 2011 weather, but it materializes at a 1-in-15 rate. This case would increase reliability at the economic equilibrium because generators would invest expecting that they would frequently earn the substantial scarcity margins caused by extreme weather, but ERCOT would not actually suffer frequent extreme weather and its associated reliability problems. We also examine a sensitivity where investors are pessimistic and expect a 1-in-100 chance of 2011 weather, but it materializes at a 1-in-15 rate. This case would decrease reliability at the economic equilibrium. Our goal in presenting these sensitivities is not to suggest that investors' expectations of extreme weather will be persistently optimistic or pessimistic, but rather is to illustrate how sensitive ERCOT's long-term reliability is to these highly speculative judgments.

Regulatory Uncertainty. Generators may discount any expectations of high scarcity margins earned under high price caps if they fear that the future regulators will reinstate lower caps. To illustrate the impact of such uncertainty, we conducted a sensitivity analysis in which investors expect future regulators will reduce price caps to $1,000 in 5 years. This future regulatory uncertainty may partially undermine the current Commission's ability to attract generation and improve reliability by increasing price caps. Even if the current Commission signals a firm commitment to letting the market work, it is difficult for them to guarantee that future commissions and legislatures would not reduce price caps in response to extreme outcomes.

Variation in Cost of New Entry (CONE). Variation in investors' cost of capital and the construction cost of generation affects the levelized cost of new entry, which will in turn affect the economic equilibrium. We determine the economic equilibrium assuming a High CT CONE of $116,000/MW-year (based on an 11% cost of capital), and a Low CT CONE of $90,100/MW-year (based on a 7.6% cost of capital). The Base CT CONE is $105,000/MW-year, with a 9.6% cost of capital to be consistent with ERCOT assumptions.[134]

Variation in Non-Scarcity Margins. Variation in non-scarcity margins could be caused by a change in gas prices, fleet composition, and also uncertainty surrounding our projections based on recent historical data. We therefore included one sensitivity analysis with 40% higher non-scarcity margins, and another with 40% lower scarcity margins than the base case.

Other Modeling Uncertainties. Both scarcity pricing and load shedding are highly uncertain, not only because they are stochastic, which we accounted for in our modeling, but also because of several factors which are difficult to model accurately or at all, including: the possibility of common mode failures; the correlation between outages, load and wind generation; the variation in operational practices; the imperfect calibration of our forward-looking model based on a limited set of past conditions; the adjustments for demand elasticity not modeled explicitly; and the uncertainty about how actual investors would model this problem. We test two cases designed to reflect a broad range of modeling uncertainties.

---

[134] See Section II.D for more detail on CONE and cost of capital calculations.

High Equilibrium Modeling Parameters: (1) Scarcity pricing is triggered at residual capacity levels 500 MW higher than estimated based on 2011 patterns and new rules; (2) An additional 1000MW of demand elasticity prior to load shed reduces loss-of-load frequency but does not depress prices.

Low Equilibrium Modeling Parameters: (1) Scarcity pricing is triggered at residual capacity levels 500 MW lower than estimated based on 2011 patterns and new rules; (2) Load is shed when 500 MW of residual capacity remain available.

The results of these five sensitivities are shown in Table 13 and Table 14.

Table 13
Economic Equilibrium Sensitivity to Uncertainties
Under The Highest Price Cap Scenario

Sensitivity	Annual Average Loss-of-Load Events	Annual Average Loss-of-Load Hours	Loss-of-Load Hours with 2011 Weather	Reserve Margin Shortfall to 0.1 LOLE Target
Base Case	0.9	2.3	34	(5.1%)
Variation in weather distribution	0.1 - 0.9	0.3 - 2.3	NA - 56	(0.6%) - (5.1%)
Investors overly optimistic or pessimistic about weather	0.2 - 1.4	0.3 - 3.9	4 - 56	(0.8%) - (6.7%)
Price caps reduced to $1,000 after 5 years (Regulatory Uncertainty)	1.4	4.1	58	(6.9%)
Low CONE - High CONE	0.7 - 1.0	1.7 - 2.7	25 - 39	(4.3%) - (5.5%)
Variation in non-scarcity margins	0.7 - 1.1	1.7 - 2.9	25 - 43	(4.3%) - (5.8%)
Range of other modeling uncertainties	0.4 - 1.3	0.9 - 3.5	13 - 50	(2.8%) - (6.3%)

Table 14
Economic Equilibrium Sensitivity to Uncertainties
Differences Shown Relative to the Base Case

Sensitivity	Annual Average Loss-of-Load Events	Annual Average Loss-of-Load Hours	Loss-of-Load Hours with 2011 Weather	Reserve Margin Shortfall to 0.1 LOLE Target
Base Case	0.9	2.3	34	(5.1%)
Variation in weather distribution	(0.8) - 0	(2.0) - 0	NA - 22	(4 5%) - 0%
Investors overly optimistic or pessimistic about weather	(0.7) - 0.5	(2.0) - 1.6	(30) - 22	(4.3%) - 1.6%
Price caps reduced to $1,000 after 5 years (Regulatory Uncertainty)	0.5	1.8	24	1.8%
Low CONE - High CONE	(0.2) - 0.1	(0 6) - 0.4	(9) - 5	(0.8%) - 0.4%
Variation in non-scarcity margins	(0.2) - 0.2	(0.6) - 0.6	(9) - 9	(0.8%) - 0.7%
Range of other modeling uncertainties	(0.5) - 0.4	(1.4) - 1.2	(21) - 16	(2.3%) - 1.2%

5. Impact of Additional Demand Response Penetration

'Price-setting' demand response which can set market prices or support price formation at high levels corresponding to its willingness-to-pay can play a very beneficial role in the energy market. It can also support resource adequacy by reducing the loss-of-load frequency while preserving the high energy prices which support generation investment. Demand response that curtails at lower prices, however, partially erodes the high prices needed to support generation investment, and therefore greater quantities are needed to achieve the same beneficial impact on resource adequacy.

We assessed the impact of additional demand response (ERS and LR are already modeled in our base case projections) that could set market prices by applying the curtailment price of the demand response to the residual capacity levels where the demand response would set prices. As listed in Table 15, 3,600 – 5,600 MW of price-setting demand response could achieve ERCOT's 1-in-10 LOLE reliability target at the economic equilibrium under the highest price cap scenario, depending on the curtailment price of the demand response.

Table 15
Amount of Price-Setting DR Needed
to Achieve 0.1 LOLE Target at Equilibrium
Under The Highest Price Cap Scenario

Curtailment Price of Price-Setting DR *($/MWh)*	Price-Setting DR Quantity *(MW)*
$4,500	3,600
$3,000	4,300
$1,500	5,600

Achieving so much DR that can set or support prices at these levels will require a major increase in DR development as well as wholesale market design changes, as discussed in Section V.B.

C. Implications for Resource Adequacy

Administrative scarcity pricing is unlikely to achieve ERCOT's 1-in-10 reliability target, even with aggressive increases in scarcity pricing parameters. Further, the level of long-run reliability achieved by adjusting scarcity pricing parameters is highly sensitive to uncertainties regarding weather, plant costs, fuel prices, modeling assumptions, and investors' expectations of weather and regulatory stability. Moreover, the reserve margin in any particular year could diverge substantially from the long-run equilibrium, for example following the retirement of a large unit.

Many lower cost resources, such as reactivations and uprates, will enter before new greenfield generation, as we have seen recently. There is a limited supply, however, of these low cost resources, and durable, long-run resource adequacy will only be achieved under a market construct which supports new entry.

Our analysis shows that major increases in the penetration of demand response resources that can set prices could support ERCOT's 1-in-10 LOLE target when combined with increased price caps. However, DR penetration on the scale needed is likely multiple years away, as explained in Section V.B.

V. REVIEW OF ENERGY MARKET DESIGN ELEMENTS

This section focuses on the economic efficiency of ERCOT's wholesale energy prices. Ideally, during most hours, energy prices should be at or near the locational marginal system production cost, including variable operating costs, start-up and shut-down costs, opportunity costs, and performance risks. When there is insufficient supply to meet load and maintain the full amount of operating reserves, prices should still reflect marginal system costs, but the marginal cost should also account for the possibility of shedding load. When supplies become insufficient to serve all load, prices should be set at customers' willingness to pay as they economically ration scarce supplies. We first examine the efficiency of a number of wholesale pricing mechanisms, and then address how to facilitate greater *demand response* participation in efficient price formation.

The market design elements we examine in this section will affect not only prices and system operations, but also investment signals. However, achieving fully efficient wholesale price signals will not necessarily support any particular reserve margin target. In Section VI below, we describe high-level policy options for supporting resource adequacy targets, including some options that deviate from the design ideals we discuss in this section.

A. Wholesale Pricing Mechanisms

In this section we examine the efficiency of ERCOT's *wholesale pricing* rules and proposed reforms regarding: (1) the accuracy of energy and A/S prices during normal conditions; (2) the accuracy of prices during scarcity conditions; (3) the impact of the Peaker Net Margin threshold and the low system offer cap; (4) the locational scarcity pricing mechanisms; and (5) the market power mitigation rules.

1. Accuracy of Prices during Normal Conditions

Most of the time, the ERCOT market will be in normal operating conditions that are not affected by capacity shortages. Even during normal system conditions, there may be short-term supply scarcity related to ramping constraints, unit commitment, or under-forecasted load. For our purposes, however, we do not characterize these transient scarcity events as representing scarcity conditions, because they are just as likely to occur whether or not the market has excess capacity resources. For our purposes, we define "scarcity conditions" as those hours when administrative interventions are required in response to capacity shortages, and where a contributing cause of the capacity shortage is a low planning or realized reserve margin.

Although prices during normal conditions are not directly related to capacity shortages, they do influence long-term resource adequacy. Energy and A/S margins that generators earn during non-scarcity conditions will contribute to their return on capital and will therefore affect suppliers' willingness to invest. For this reason, any systematic overpricing or underpricing during normal conditions could adversely affect investment signals and resource adequacy. We examine three aspects of price-setting during normal conditions that may be improved: (a) pricing inefficiencies introduced through unit commitment processes, which may tend to suppress prices overall; (b) pricing inefficiencies related to imperfect coordination between energy and A/S prices, which may suppress or inflate prices; and (c) new pricing mechanisms that could inefficiently increase prices to "scarcity" levels during non-scarcity conditions.

a. Price Inefficiencies Related to Unit Commitment, including RUCs

Prices during normal market conditions are conceptually easy to set if unit commitment decisions can be ignored, because one can economically dispatch the lowest-cost resources first until supply is sufficient to meet demand. In this simplified case, the efficient price is just the marginal production cost from the highest-cost resources dispatched to satisfy load, transmission, or operational constraints.

However, supply discontinuities related to unit commitment complicate price formation. Unit commitment determines which resources to turn on over a multi-period timeframe to minimize system costs, including startup costs, operating costs, and minimum load costs. Economic dispatch determines which units operate at a given instant considering only those units that are already committed, and the energy price generally reflects the incremental offer of the last unit dispatched without regard to startup or minimum load costs. This leads to an inconsistency in

which costs are considered between unit commitment processes used to determine start-up and shut-down schedules, and economic dispatch processes used to determine short-term output and market prices. These inconsistencies can create small, systematic underpricing effects and can therefore necessitate uplift payments for units that are committed but that do not earn enough energy margins to cover their startup or minimum load costs. In 2011, Day-Ahead make whole uplift payments were less than $0.01 per MWh of load, and RUC uplift payments were $0.05 per MWh of load.[135]

Although these underpricing effects can be caused by "in-market" resources that are committed in the day-ahead market, market participants have been particularly concerned about price-suppression impacts on the real-time market from units that ERCOT commits out-of-market, through its Reliability Unit Commitment (RUC) process.[136] The RUC process is intended to commit incremental generation on a day-ahead or hour-ahead basis to meet projected load and operating reserve requirements and support local reliability.[137] We understand that nearly all RUCs are committed through the hour-ahead, not day-ahead process; RUC units may be committed to replace units that trip offline, support local reliability, support upcoming ramp requirements, or meet other system operating constraints.

Whenever RUC units require uplift payments, this indicates that market prices were artificially low over that dispatch period. However, ERCOT recently implemented the first part of a solution to prevent RUC units from suppressing prices by releasing RUC-dispatched generation (above minimum load) as available to SCED only at the offer cap. An urgent-priority NPRR will also address the "0-LSL" issue, which refers to price suppression from RUC and some other types of units operating at minimum generation.[138] The NPRR proposes to conduct one SCED run that accounts for the minimum generation constraint on these RUCs and other units for dispatch purposes, and a separate SCED run that relaxes this constraint for market pricing purposes.[139] However, we believe these solutions are inefficient, especially in combination.

RUC units that do not receive uplift payments are not depressing prices below a competitive level because these units would have opted to run on an in-market basis had they had perfect foresight of market prices. For RUC units that would require uplift payments, a more appropriate approach to preventing price suppression would be to inflate their incremental cost offer by an amount that more closely reflects their commitment costs or estimated uplift payments. Raising a RUC unit's offer to the system-wide offer cap is more than necessary. Some market participants claim that any generation by these units below the cap depresses prices, since the market did not elicit their output. We disagree because if there were an hour-ahead market, the unit would presumably have entered in-market.

---

135 See ERCOT (2012a). RUC uplift payments are net of RUC clawback charges.

136 The day-ahead market includes a unit commitment process followed by a day-ahead dispatch and pricing process. This day-ahead unit commitment process is separate from any subsequent day-ahead RUC that ERCOT may require. However, units committed in the day-ahead market that fail to earn sufficient revenues based on day-ahead prices and dispatch will also receive an uplift payments, see ERCOT (2012s).

137 See ERCOT (2010j).

138 These types of administrative commitments include RUC units, ERCOT commitment of RMR units, and Offline Non-Spin, as well as quickstart generation dispatch by SCED. See ERCOT (2012f), NPRR444.

139 The NPRR also includes a provision to add load curtailed from load resources and ERS in the pricing run. See ERCOT (2012f), NPRR444.

The one exception is RUCs committed for local reliability reasons. If they are not dispatched above LSL, they need uplift payments. The price depressing effect of their minimum load output should be mitigated if it is very significant. However, we are cautious of solutions involving separate SCED runs for dispatch and price-setting because this approach will result in some plants being dispatched in a way that deviates from what pricing signals would support.

Starting after implementation in June 2012, Phase 1 of ERCOT's "Look-Ahead SCED" initiative could potentially reduce the number of RUC units required, by providing a 40- to 50-minute forward indicative price. Forward indicative prices will be informative to suppliers that are self-committing generation and to demand-side participants interested in efficiently reducing their load.[140] However, these self-commitment actions would not be coordinated by ERCOT. In future phases of Look-Ahead SCED, ERCOT may better coordinate these actions by explicitly integrating unit commitment considerations into the process, although the details of these options have not yet been developed.

On a longer-term basis, we recommend addressing similar commitment-related price suppression impacts whenever ERCOT or stakeholders identify a particular issue as introducing substantial uplift payments. For example, it may be desirable to create a mechanism for enabling block-loaded resources to set day-ahead and real-time energy prices.[141] There are also more ambitious options for incorporating commitment costs and other discontinuities into dispatch and pricing software, such as moving toward convex hull pricing.[142] However, such options could be very expensive to implement and should only be pursued if simpler fixes are insufficient and the benefits can be shown to exceed the associated software upgrade costs. The ideal result of any future pricing enhancements would be that suppliers self-dispatching against these prices with perfect foresight would exactly match the least-cost system result. Like other markets, ERCOT can make steps toward this ideal, but likely will not fully be able to achieve it.

b. Co-Optimizing Energy and Ancillary Services

In its day-ahead market, ERCOT already fully co-optimizes market clearing and price-setting between its energy and A/S markets. As a longer-term component of its broad Look-Ahead SCED initiative, ERCOT and stakeholders have also considered the option of co-optimizing energy and A/S in real-time markets as is done in NYISO.[143] This revised design would readjust which suppliers are providing energy and A/S services at any given time in response to real-time system conditions and changing economics. This proposal is in its early stages and no specific design construct has been developed, but it would be an improvement that could increase price

---

140 See ERCOT (2012f), NPRR444.

141 Block loaded resources have minimum generation equal to maximum generation and so have no variable output range over which to calculate the marginal cost of incrementally more power. For any interval when a block-loaded resource is the highest-cost unit dispatched, the market will end up being underpriced without such a change and the block-loaded resource will require uplift payments.

142 The convex hull pricing approach is mathematically complex and requires substantial software upgrades, but represents a best practice approach to minimizing uplift requirements and imperfect price incentives. Convex hull pricing replaces a blocky and discontinuous supply curve with a smoother incremental cost curve that can be used to set more efficient and stable prices. We do not attempt to fully explain this approach here, which is conceptually and mathematically complex with many variations. Instead we refer readers to Gribik (2009).

143 See ERCOT (2012q), p. 6.

and operational efficiency. This design improvement would result in better-behaved, smoother pricing, least-cost dispatch and higher operating reliability. However, it would also involve a major and potentially costly software change and so should only be pursued after careful consideration of the costs and benefits.

c. Avoiding Mechanisms for "Scarcity" Prices during Non-Scarcity Conditions

In addition to pricing revisions that will increase the efficiency of wholesale prices, some mechanisms may inefficiently increase prices to "scarcity" levels even when resources are plentiful. We recommend avoiding such changes because they would increase prices in a way that is unrelated to an underlying need for new investments.

The recent 500 MW increase in the RRS requirement is an example of a change that could inefficiently introduce scarcity prices during non-scarcity conditions.[144] This increase in operating reserves does not necessarily reflect an operational system need, and will therefore unnecessarily increase system operating costs all of the time whether there is a scarcity event or not (*i.e.*, with more capacity spinning than operationally needed for 8,760 hours per year). This will increase prices and returns to suppliers as intended, but will unfortunately also inefficiently increase customer costs. We recommend that operating reserves requirements instead be determined based on analysis of contingency risks, ramping needs, wind balancing requirements, load balancing requirements, or other operating considerations.

The new RUC mechanism described above also is likely to introduce scarcity prices during non-scarcity events. This mechanism will add RUC units to the SCED pricing run at the offer cap of $3,000/MWh. The likely result is that prices may be driven to very high levels during high-ramp or under-forecast conditions. These ramping and forecasting considerations represent real system operating needs, but are not related to resource adequacy or the realized reserve margin. High-price events caused by this RUC mechanism will be just as likely to occur with a 30% reserve margin as with a 10% reserve margin. We do recognize, however, that the old RUC mechanism inefficiently suppressed prices by failing to incorporate commitment costs into pricing. Balancing these concerns, we recommend a different approach to preventing price suppression from RUC units using approaches similar to those discussed in Section V.A.1.a above.

2. Accuracy of Prices during Scarcity Conditions

During scarcity conditions, as during normal conditions, the efficient market price will reflect the marginal system cost of power. Ideally, it would be best to rely on high-priced DR curtailment bids (or offers, if DR is included on the supply side) to set prices in scarcity conditions, as discussed in Section V.B. However, without substantial DR resources, it is difficult to determine the marginal system cost during scarcity events because typical dispatch and price-setting mechanisms are not sufficient to bring supply and demand into balance. Instead, out-of-market actions must be implemented, including deploying operating reserves, relaxing transmission constraints, deploying backstop resources, or shedding firm load. These out-of-market actions

[144] However, the "Energy-only with adders" policy option described in Section VI.B.2 could include increased operating reserve requirements and other elements to support resource adequacy through elevated prices.

can sometimes inappropriately suppress market prices when high prices are most needed, although ERCOT and the PUCT have already changed market rules to prevent price suppression in many cases.

If market-based supply offers and demand bids cannot be used to determine the marginal cost of power during scarcity events, then the price must be administratively determined. Marginal system cost is difficult to calculate during scarcity, but would include, among other things, the system risks introduced by operating with low reserves and customer costs incurred during load-shed events. We examine here various components of ERCOT's scarcity pricing mechanisms, including: (a) the "small fish swim free" rule's relationship to scarcity pricing; (b) the price cap and high system offer cap; and (c) administrative scarcity pricing mechanisms.

a. Small Fish Swim Free Mitigation Rule

In its original scarcity pricing framework, the PUCT developed a construct that relied on the "small fish swim free" mitigation rule to produce high prices during shortage conditions. Under this mechanism, small suppliers with less than 5% market share are always allowed to offer into the wholesale market at high prices up to the offer cap.[145] The offer cap would usually be the High System Offer Cap (HCAP), which is currently $3,000/MWh.[146]

This framework was intended to strike a balance between allowing wholesale prices to reach high levels during scarcity conditions and limiting the potential for exercise of market power during non-scarcity conditions. During normal conditions when the efficient market-clearing price is low, these small suppliers would rationally offer close to their marginal cost to ensure that they would clear. However, during scarcity conditions small suppliers would become pivotal and could still clear even if they offered at high prices. They could create high scarcity prices only during true shortage conditions.

Unfortunately, the small fish swim free approach has not proven to be a reliable scarcity pricing mechanism. The most important problem is that small suppliers may incorrectly predict scarcity conditions, thereby inefficiently pricing themselves out of the market during non-scarcity conditions or under-bidding during shortages. The risk of inadvertently pricing out of the market is a substantial burden to place on small suppliers. The market monitor has examined historical prices during shortage events and noted that "relying exclusively upon the submission of high-priced offers by market participants was generally not a reliable means of producing efficient scarcity prices."[147] Another unfortunate result of this mechanism is that small suppliers who offer their generation at very high prices face the risk of public reprisal, as occurred in 2008 when a small supplier's above-cost offers were featured in a *Wall Street Journal* article that drew comparisons to the California Electricity Crisis.[148] Finally, the mechanism is not functionally tied to any measure of the severity of a scarcity event, and so is likely to produce similarly high prices during moderate and severe events.

---

[145] See PUCT (1999).

[146] The offer cap would drop to the Low System Offer Cap if the PNM threshold were exceeded as discussed further in Section V.A.2.b below. To date, the PNM threshold has never been reached, see ERCOT (2012j).

[147] See Potomac Economics (2011b), p. 54.

[148] See Smith (2008).

Overall, it appears that this mechanism has failed to introduce sufficiently high prices reflective of scarcity conditions to meet long-term resource adequacy needs. While it does not appear that this mechanism must be revised, it does appear that ERCOT will require supplemental mechanisms to produce needed scarcity premiums. We believe that this observation is consistent with the observations and recent market design activities of the PUCT, ERCOT, and stakeholders.

b. Price Cap and High System Offer Cap

ERCOT's high system offer cap (HCAP) is set to $3,000/MWh; while ERCOT does not have any enforced price cap, it would be unusual for prices to rise above the offer cap.[149] Commissioners of the PUCT have stated plans to further increase the offer cap to possibly $4,500 to $9,000/MWh, motivated by concerns that the current cap is too low to attract a desired level of investment.[150] Neither the current offer cap nor the proposed offer cap increases are based on an analysis of customers' VOLL or an analysis of the price cap needed to sustain investments.

We recommend creating a locational marginal price (LMP) cap set at the average customer VOLL, which would also impose a maximum limit on other parameters such as the offer caps and the Power Balance Penalty Curve (PBPC) shadow price. This is the efficient price level during severe scarcity conditions when ERCOT must enact involuntary load shedding, because this is the price that the average customer would have been willing to pay to avoid curtailment. A VOLL-based price cap approximates what the demand curve would have been had customers been actively bidding to avoid curtailment. Setting the price cap at VOLL is supported by a rich theoretical literature demonstrating the economic efficiency of this approach.[151]

Determining an accurate estimate of VOLL is difficult, however, and could range from a few thousand to tens of thousands of dollars depending on customer class. For example, in its 2006 review of VOLL studies, MISO found that VOLL ranged from $1,500-$3,000/MWh for residential, $10,000-$50,000/MWh for commercial, and $10,000-$80,000/MWh for industrial customers.[152] Ultimately, MISO decided to set its price cap at the low end of $3,000/MWh, consistent with residential VOLL estimates.[153] As another example, Australia's National Energy Market (NEM) price cap is at a VOLL of $12,500/MWh AUD ($12,200 USD), with the parameter subject to periodic study and updating.[154] The VOLL estimate appropriate in ERCOT is likely in the same range as VOLL estimates elsewhere, but a study would need to be conducted to estimate the number accurately. In particular, the study would have to consider: (1) the VOLL of different classes of customers; (2) the likely ratio of load shed events that would be imposed on each customer class, including considering that utility protocols may result in more load shedding for residential rather than large C&I customers; and (3) that certain very high-VOLL customers should be excluded from the analysis because they will already have

---

149 Some nodal prices may rise above the offer cap if, for example, the penalty factor on a certain system constraint had a very high shadow price.

150 See, for example, PUCT (2012a), Item Number 106.

151 See Hogan (2005), pp. 9-11; and Joskow and Tirole (2004) p. 14.

152 See MISO (2006).

153 See MISO (2012a), Section 5.

154 See AEMO (2012). Exchange rate assumed is USD/AUD = $1.02 from Bloomberg (2012).

invested in backup generation or dual distribution feeds and will therefore not experience a full outage even during a load shed event.

Another way to set the price cap would be to derive it, along with other administrative scarcity pricing parameters, based on an estimate of the price levels needed to attract a desired level of investment. We more fully examine this option under Section VI.B.2 below, although we do not recommend this as a dependable way to achieve a particular reserve margin.

Finally, we recommend creating a functional distinction among: (1) ERCOT's price cap, which is currently undefined, meaning that prices may exceed the offer cap depending on transmission constraints; (2) the high, low, and other offer caps created for market mitigation purposes and implementing the small fish rule; and (3) administrative scarcity pricing thresholds used to set prices during scarcity events. Each of these mechanisms has a different purpose, and so they should not be forced to have identical values in all cases. The purpose of imposing a price cap at VOLL is to prevent LMPs from exceeding customers' willingness to pay to avoid outages during load-shed events.[155] The high and low offer caps used under the small fish swim free rule might be set to a separate, lower level based on PUCT and market monitor analyses of market power mitigation concerns. Administrative scarcity pricing thresholds might be set to different levels as discussed in the next Section.

Increasing the offer and price caps would introduce some risks associated with potential defaults. We have not analyzed all of the credit requirements, qualifications, and other provisions that might be required to ensure that market participants are able to cover their day-ahead and forward bilateral positions without defaulting. However, we are concerned that as reserve margins tighten and offer caps increase, an unscrupulous REP with little to lose might find ways to exploit asymmetric risk exposures, if any exist. Such a REP could under-hedge in order to make money in the likely event that realized spot prices are lower than forward prices, while ignoring the risk that spot prices could spike to levels they cannot pay in the unlikely event of 2011-like weather. Instead of paying the cost of such an extreme event, they could simply default and exit the retail electric business, and ERCOT's other customers would have to pay. Given risks such as these, we recommend that the PUCT revisit its credit and qualification provisions for REPs, as we understand ERCOT is already doing for settlements under their purview.

c. *Administrative Price-Setting during Scarcity Events*

There are three key objectives when developing price-setting mechanisms during scarcity events: (1) ensuring that administrative reliability interventions do not artificially suppress prices during scarcity events; (2) incorporating DR into price-setting as much as possible as discussed in Section V.B.4 below; and (3) developing administrative price-setting mechanisms that will accurately reflect marginal system costs.

Price suppression during administrative reliability interventions is a risk in any market because these interventions make incremental supplies available for dispatch. If those actions add supply at a low offer price (or reduce demand), then the typical result will be to reduce prices just when

[155] Note that in the absence of a price cap, increasing the offer cap to $9,000/MWh means that actual realized prices could exceed $9,000/MWh and the VOLL at specific nodes, depending on system constraints.

very high prices are most needed. Over recent months, ERCOT has implemented, or is developing, corrective measures to prevent this outcome from reliability interventions including RMR units' dispatch, deploying responsive reserves for energy, and Online and Offline Non-Spin deployments.[156] However, there are a few types of reliability interventions that could suppress market prices but have not yet been addressed. We recommend that ERCOT develop protections to prevent price suppression from these actions, including during: (1) Emergency Response Service (ERS) and Load Resource (LR) deployments, which a current NPRR is intended to address; (2) calling on emergency imports; (3) relaxing internal transmission constraints; and (4) any other type of reliability intervention that stakeholders or ERCOT may identify in the future.[157] Additionally, we recommend that ERCOT periodically examine price outcomes during all scarcity-related reliability events to confirm that no unexpected low prices occurred during those events.

Setting prices at an *efficient* level during these scarcity events is just as important as ensuring that prices are not inadvertently suppressed. To date the price corrections that ERCOT and PUCT have pursued have been tied to the objectives of preventing price suppression or attracting new investments. For this reason, some interventions including depleting regulation reserves at an initial price of $200/MWh may be underpriced relative to their system cost impact, while other interventions such as deploying responsive reserves only at $3,000/MWh even for the first MW deployed may be overpriced.

We recommend developing price correction mechanisms that tie all administrative pricing mechanisms to the marginal system costs of these interventions. Table 16 summarizes the principles that could be used to set efficient prices during each type of scarcity event and compares these pricing mechanisms with those that are currently in place. For example, if ERCOT can avoid shedding load by making an administrative off-system power purchase at $600, then we would recommend setting the price to $600 during that intervention. As another example, it may be possible to estimate the marginal system costs of operating with reduced levels of reserves by accounting for the increased system contingency risks and loss of load probability (LOLP) introduced by operating with lower reserves. Depleting operating reserves will increase the likelihood of load shedding from contingencies and so introduces a greater risk to customers as the scarcity event becomes more severe. For that reason, the efficient price during these events will also increase with the severity of the event and ultimately reach VOLL when load must be shed. Note that setting prices to VOLL when there are still enough operating reserves to operate reliably could result in customers' unnecessarily reducing high-value uses of power.[158]

A gradual approach to administrative scarcity pricing will result in a continuum of high-price outcomes related to the severity of each scarcity event. Under current rules with most scarcity interventions priced at the cap of $3,000, prices can jump back and forth between $200 and $3,000. A more continuous scarcity pricing approach will better-enable price-responsive

---

[156] See ERCOT (2012f), NPRR442, NPRR427, and NPRR428.

[157] ERS and LR resources were formerly referred to as "EILS" and "LaaR," respectively. For the current NPRR 444 regarding re-pricing for these resource deployments, see ERCOT (2012f), NPRR444.

[158] If the PUCT opted to move ahead with an approach to adjust price parameters to achieve a particular reserve margin, it may introduce this type of inefficient load shedding in order to achieve that desired reserve margin objective. We discuss this issue further in Section VI.B.2 below.

demand to contribute to price formation even if it is not incorporated into SCED. For example, a customer valuing energy at $1,500 can reduce consumption if the price rises above $1,500 without causing prices to plummet back to non-scarcity levels. Graduated scarcity pricing will also serve as a guide to ensure that the lowest-cost reliability interventions are implemented first before costlier measures.

The concept of introducing scarcity prices gradually is already implied by the PBPC, which starts at $200/MWh and increases to the offer cap over 50 MW. The 50 MW range of the PBPC is based on the quantity of regulating reserves that can be deployed for energy before substantial reliability concerns arise. A new PUCT proposal to implement a more gradual PBPC over 200 MW is a move in the right direction and would require RRS deployments to make up the required energy.[159] We recommend something simple and gradual, such as stretching the entire scarcity pricing curve from $500 when first depleting operating reserves, then increasing to $9,000 or some similar VOLL-based level when close to shedding load. The shape of the increasing curve could be a simple linear function or a more complex function approximating the shape of system cost increases as operating reserves are deployed. As an alternative, if not all reliability interventions can be incorporated into one scarcity pricing function, these interventions could be treated as re-priced units in SCED similar to the current treatment of non-spin, RMR, and other types of reliability interventions. However, these re-priced units would have increasing marginal cost curves that approximate the smoothed scarcity pricing function.

[159] See PUCT (2012a), Item Number 125.

Table 16 summarizes how the marginal system costs of any one reliability intervention might be calculated to inform the shape of the scarcity pricing function. We recommend that all types of reliability interventions be incorporated into this scarcity pricing curve, which will extend the graduated scarcity pricing effects over a wider range of MW from low-cost interventions to high-cost interventions.

Table 16
Administrative Scarcity Pricing Mechanisms
(As Currently Implemented and Under Potential Marginal System Cost Mechanisms)

Scarcity Intervention	Current Pricing Mechanism	Potential Marginal Cost-Based Pricing Mechanism
Emergency Imports	- Reduces demand, suppresses prices	- Purchase price of emergency power
Call RRS	- Gen: Add to SCED at $3,000/MWh - Load: Reduces demand, which can reverse prices. Reform under consideration to add LR back to the demand for settlement.	- Estimate marginal system cost of operating at low reserves, including LOLP * VOLL, or some proxy thereof that varies with megawatts
Relax Transmission Constraints	- Shadow price caps for relaxing constraints - Some constraints not passed to SCED	- Estimate marginal LOLP impact, plus - O&M cost impact on transmission elements
Call Reg-Up for Energy (Not for Balancing)	- 0-50 MW used for PBPC at $200-$3,000 - Rest added to SCED at $3,000/MWh	- Estimate marginal LOLP impact, plus - Cost of reduced load balancing efficiency
Call RMR	- Add to SCED at $3,000/MWh	- Ideally, market would exclude RMRs (never entirely possible due to transmission security concerns) - If called, marginal cost is total availability payments divided by expected number of call-hours
Call ERS	- Reduces demand and can suppress prices. Reform under consideration	- Ideally, a pure energy-only market would exclude ERS - If kept, marginal cost could be estimated based on total payments divided by expected number of call-hours
Load Shed	- SWCAP of $3,000 ($500 if PNM exceeded) - Increased high & low caps under consideration	- Price at VOLL - Study end users to estimate average VOLL in Texas (possibly ~$10,000/MWh)

3. Level of Peaker Net Margin Threshold and Low System Offer Cap

The Peaker Net Margin (PNM) threshold and low system offer cap (LCAP) are a combined mechanism intended to prevent extreme, excessively high prices in any one year. ERCOT calculates the accumulated PNM over each calendar year as the operating margins of a gas CT with a heat rate of 10 MMBtu/MWh.[160] ERCOT's estimate of such a unit's operating margins excludes variable operating and maintenance (VOM) costs, start-up and shut-down costs,

[160] See ERCOT (2012k), Section 4.4.11.
See also, PUCT (1999).

emissions costs, and imperfect dispatch. If the PNM over the year increases above the threshold of $175/kW-year, then the offer cap is reduced from the $3,000 HCAP to the LCAP.[161] This lower LCAP is the greater of: (a) $500/MWh; or (b) the theoretical operating cost of a gas-fired unit with a heat rate of 50 MMBtu/MWh. In light of recent resource adequacy concerns about the ability of the market to attract sufficient investments, commissioners and stakeholders have proposed increasing the PNM threshold and LCAP as a means of achieving a more favorable overall investment environment.

The PNM threshold is ultimately a regulated safety valve to prevent extreme one-year results, meaning that the parameter has no one "correct" level that can be derived from economic principles. However there are a number of considerations to incorporate into the determination of PNM threshold:

- First, it makes sense to set the PNM threshold as some multiple of the cost of building a new peaking plant, in order to relate it to the overall investment returns suppliers may expect to earn. If CONE is the average net revenue needed over many years to attract investments, then the PNM threshold can be considered an approximate soft cap on returns that will be allowed in any given year. The PNM is an imperfect measure because actual peaking generators' are able to earn only 60-85% of PNM due to imperfect dispatch and various operating costs.[162] Further, it is a "soft" cap because generators will continue to earn some incremental returns at a reduced rate even after the PNM threshold is exceeded.
- Second, it is important to consider the frequency and magnitude of price spikes.[163] For example, if substantial scarcity is expected only once every five years and energy margins in non-scarcity years are only half of CONE, then the returns and PNM in the scarcity year would need to be approximately three times CONE to attract investment. There is only limited value to such a simplified formula, however, if prices become extremely high extremely infrequently. For example, if prices were very low in most years but high enough for a generator to earn 10 times CONE once every ten years, we would expect very few investments. The difficulty of modeling severe outlier outcomes and the risk of potential political interventions would likely cause most suppliers to discount the potential for such high prices. If suppliers do not expect such extreme pricing events to be allowed to persist, discounting these events through an administrative rule such as PNM (rather than ad hoc interventions) makes prices more predictable.
- Finally, regulators should consider the public's tolerance for withstanding years with extremely high prices, with the average wholesale price being a relevant metric for

[161] Id.

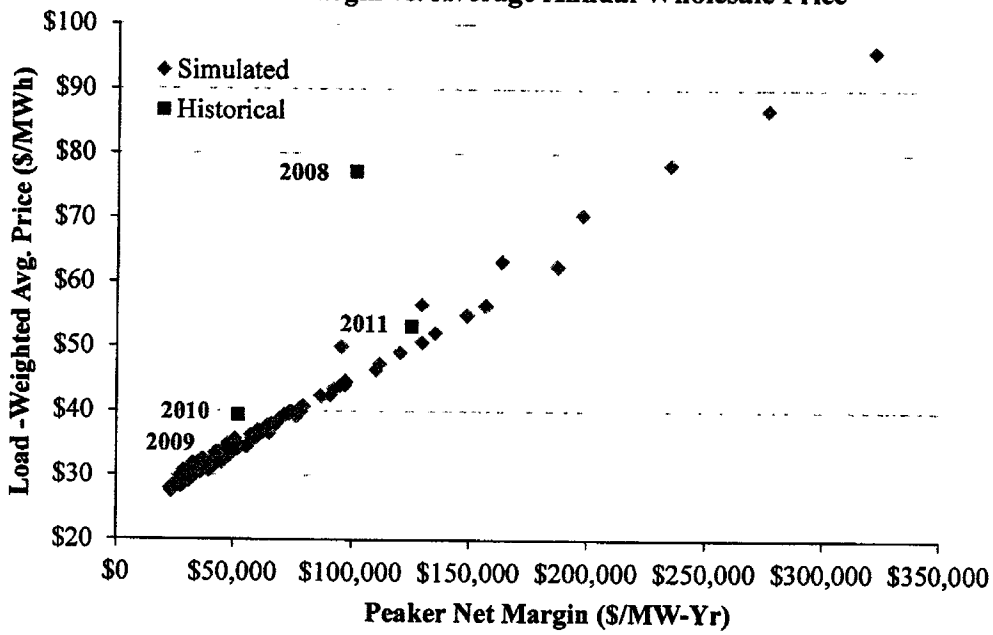
[162] Calculated based on 2008-2011 net revenue estimates from Potomac Economics (2011c) and (2012). For PNM data, see ERCOT (2012j).

[163] Although suppliers generally sell most of their power on a forward basis, these forwards will reflect the market's expectations of spot market price spikes adjusted for risk. Therefore, spot prices need to be allowed to rise to levels sufficient to support investment. Forwards will also incorporate the expected impact of PNM in preventing certain very high-price outcomes in the real-time market. For these reasons and for simplicity, we discuss the impacts of PNM only on real-time prices and presume that the effects will translate to generator returns regardless of their hedging strategies.

determining whether customers will demand political intervention.[164] Figure 29 below shows the relationship between PNM and the average wholesale energy price in recent years, based on the simulation methodology used in Section IV above.[165] For example, if the public is intolerant of prices that rise to twice the average level, this would indicate an appropriate PNM threshold of $300/kW-year or approximately three times CONE.

Overall, we stress that there is no "correct" level for the PNM threshold. In fact, the stability and predictability of the parameter over a number of years may be more important than the exact level. After considering all of these factors, we would recommend a PNM threshold in the range of $250-350/kW-year that increases in some predictable way over time, commensurate with the increasing costs of construction. For example, the PNM threshold may be set at a specific multiple of CONE and inflated annually according to a standard index such as Handy-Whitman.[166]

Figure 29
Peaker Net Margin vs. Average Annual Wholesale Price



Sources and Notes:
Historical data from ERCOT (2012j) and Potomac Economics (2011c).
Simulated results based on methodology outlined in Section IV, with a high cap of $3,000/MWh and no PNM threshold. Some results with a PNM greater than $350/kW-year are not shown.

The LCAP is a related parameter because it is the offer cap imposed after the PNM has been exceeded. The purpose of the LCAP is to assist in preventing excessively high prices on a

[164] We note that this metric is one indicator of potential customer concern, but is not perfect for a number of reasons including that: (1) much of retail customers' load is hedged on a forward basis and not exposed to spot prices; (2) it may be monthly rather than annual extremes in bills that would cause most consumer or REP concerns; and (3) wholesale prices are only a portion of customer bills and so do not translate proportionally.

[165] The simulation reflected includes all 15 weather year profiles as described in Section IV.

[166] See Handy-Whitman (2011).

continuous basis over the year, so it makes intuitive sense to keep this cap at a relatively low level as long as it does not introduce excessively inefficient price distortions. For example, an LCAP of only $100/MWh would introduce excessive inefficiencies because it would preclude a large number of peaking generators from being dispatched. The current LCAP of $500 (or higher in high fuel price circumstances) may be reasonable if one considers only a generation market and ignores the potential for demand response. We would recommend increasing this LCAP to a higher level if any generation resources in the fleet have a marginal cost (including opportunity costs) above the cap. Further, as demand response grows in Texas, it will be important to raise the LCAP to a level that ensures that most load reductions would be achieved at prices below the LCAP. Determining this level could be informed by an econometric study to evaluate the level of demand reductions achieved at various price levels. We further discuss how such a study could be conducted in Section V.B.4.c below.

4. Locational Scarcity Pricing

Resource adequacy can be a regional or sub-regional concern, depending on the nature of transmission constraints. Even if the overall RTO has sufficient generation supplies, this does not necessarily mean that all locations will achieve the reliability target because the system may have: (1) load pockets within which there is insufficient local generation or import capability to meet peak demands; or (2) generation pockets with excess supply but insufficient export capability to meet peak demands in other locations.

In markets with resource adequacy requirements, locational reliability concerns are directly defined and addressed, for example, through local capacity requirements within load pockets.[167] In energy-only markets, it is more difficult to address locational resource adequacy concerns so directly, particularly in ERCOT, the first nodal energy-only market. Other energy-only markets have relied solely on system-wide prices as in Alberta or on zonal prices as in Australia.[168] In a nodal energy-only market, it will be a challenge to achieve an effective scarcity pricing mechanism that is: (1) location-specific enough that it will attract investments to where incremental generation is most needed; and (2) not so focused on a small number of nodes that a more regional or sub-regional resource adequacy need fails to be reflected in the broader price. The IMM highlighted this challenge in the 2008 State of the Market Report, which cautioned that the move to nodal pricing could focus scarcity pricing into too-small clusters of nodes.[169]

To date, ERCOT's scarcity pricing mechanisms have not been developed in a way that explicitly considers the potential for locational resource adequacy concerns as opposed to system-wide resource adequacy concerns. We recommend assessing the need to revise these mechanisms for locational relevance. While a number of approaches could be used to achieve this result, one option would be to revise administrative scarcity pricing mechanisms around new "A/S Regions" that may or may not coincide with ERCOT's current Load Zones. The mechanisms could be conceptually similar to the Reserve Zone approach implemented by MISO that expresses: (1)

[167] See Pfeifenberger, Spees, and Schumacher (2009), Section VII.B.3.
[168] See AEMO (2010), and AESO (2011).
[169] See Potomac Economics (2009), pp. 83-87.

system-wide scarcity prices when depleting system-wide reserves; and (2) scarcity prices specific to that Reserve Zone when reserves drop below that location's requirement.[170]

Implementing this type of concept in ERCOT might require the RTO to:

Define A/S Regions — We recommend that in its LOLE study or transmission planning processes, ERCOT evaluate whether there are load pockets or generation pockets relevant for resource adequacy. Load pockets would be identified as regions within which LOLE is concentrated due to import constraints; generation pockets would be defined as regions with excess supply that is generally unavailable to the rest of the system during peaking conditions. While this question has not previously been analyzed in ERCOT, it appears that the Houston Load Zone is a candidate for evaluation as a potential load pocket relevant for locational resource adequacy; however, we note that such load or generation pockets would be defined based on transmission topology and would not necessarily coincide with a current Load Zone. For the purposes of our discussion here, we presume that the boundaries of these load and generation pockets would be equally relevant for defining new boundaries in the A/S markets and so we term these locations as "A/S Regions."[171] To the extent that no such A/S Regions are needed now or are expected within the coming years, we would not recommend pursuing any of the other following mechanisms at this time. However, if locational resource adequacy concerns are identified, then we recommend refining scarcity pricing mechanisms in a way that ensures that locational scarcity will be reflected in realized prices in those defined regions.

Define A/S Penalty Curve by A/S Region — All supply resources in SCED, including the virtual resource represented by the PBPC, must be assigned to a specific node. The current PBPC is defined at the reference bus, meaning that it has a distributed "location" across all load nodes. This also means that scarcity pricing outcomes related to the PBPC will be tied to system-wide but not location-specific scarcity conditions. However, locational scarcity may be better reflected if each identified A/S Region had its own A/S Penalty Curve that affected prices only at the group of nodes defined within that region. However, system-wide shortages could still be reflected in scarcity prices driven by the system-wide PBPC.

Evaluate Each Administrative Scarcity Mechanism for Locational Relevance — Several of the scarcity pricing mechanisms developed by ERCOT rely on administratively re-pricing certain types of resources and adding them into SCED, including RMR, RRS, Non-Spin, and RUC resources. Because each of these resources represents a real generation unit, they are all tied to a specific node and may have the effect of increasing prices in that location but not in others, depending on transmission constraints. It is not clear whether or under what circumstances these mechanisms are likely to introduce scarcity pricing signals where they are most needed. We recommend individually evaluating each mechanism for this purpose. For example, if a load pocket exhibits incremental A/S needs or requires an RMR for capacity, then we would recommend that any scarcity

[170] See MISO (2012a), Sections 3.3, 5.1.1, and 5.2.

[171] To the extent that it is more appropriate to define A/S market boundaries separately or differently from load and generation pockets relevant for resource adequacy, the mechanisms we propose here would have to be revised at least to some extent.

pricing related to those associated resources' deployments be developed in a way that impacts all nodes in that A/S Region. Deploying these resources would only impact RTO-wide node prices in the case of an ERCOT-wide shortage.

Align Load Settlements by A/S Region — Customer prices are defined based on Load Zone prices, which could create an economic disconnect for sub-zonal load pockets. This means that potential price-responsive demand within these small regions may go undeveloped due to uneconomically low load prices there; similarly, too much price-responsive demand might be incented outside the load pocket where it is not helpful for resolving the transmission constraints. To the extent that such sub-zonal resource adequacy zones exist, we would recommend re-defining Load Zones and settlement according to the boundaries of that A/S Region. This would create the most efficient price for price-responsive loads to respond to for resource adequacy purposes.

Align Real-Time Mitigation Procedures with A/S Regions — Under certain circumstances, ERCOT's real-time mitigation procedures could prevent locational scarcity prices from materializing. For example, high offer prices in SCED from small fish, or administratively-priced RMR, RRS, Non-Spin, or RUC units could be re-priced down to marginal cost if those resources are behind a "non-competitive" constraint.[172] We suspect that in many cases these mitigation procedures would not result in underpricing relative to locational resource adequacy needs because these units may still set locational scarcity prices to the extent that they are behind "competitive" constraints. However, we do recommend that ERCOT examine the extent to which the definitions of competitive and non-competitive constraints could prevent locational scarcity prices from materializing.

More generally, as ERCOT's scarcity mechanisms are refined or revised, we recommend that they be developed in a way that explicitly considers how well they will perform to reflect both locational and system-wide resource adequacy shortages.

5. Offer Monitoring and Mitigation Rules

There are two levels of market monitoring and mitigation affecting market prices in ERCOT. The first is the well-defined mitigation construct that ERCOT enforces in its real-time market to prevent suppliers that are behind non-competitive constraints from artificially inflating prices in small, constrained locations.[173] ERCOT does not impose any type of mitigation procedures in its day-ahead market.

The second level of market monitoring and mitigation is implemented by the IMM under PUCT mandate, which is governed by three general principles set out by the PUCT: (1) that "small fish" with less than 5% market share will be allowed to offer energy at any price up to the offer cap; (2) that larger entities may not offer their power at levels that are "substantially above its marginal cost"; and (3) that generators who wish to confirm that their approach to offering into the market is not in violation can request that the PUCT, with IMM input, approve a voluntary

[172] See ERCOT (2012r), Section 6.5.7.3.
[173] See ERCOT (2012r), Section 6.5.7.

mitigation plan for compliance.[174] The extent to which market participants have completed the process of having a voluntary mitigation plan approved is not public.

In our interviews, a number of suppliers proposed relaxing or clarifying these offer mitigation rules. In terms of clarifying the rules that the IMM will enforce, market participants have expressed uncertainty about what constitutes offer prices that are "substantially above marginal cost" as indicated in the PUCT Substantive Rule. While it may not be possible to strictly define this principle in all circumstances, we do agree that it would benefit suppliers to have a better understanding of how the IMM will generally interpret and enforce it. For example, publishing a document outlining monitoring and mitigation guidelines similar to the one published by Alberta's Market Surveillance Administrator would better inform suppliers' current activities and voluntary mitigation plan proposals.[175] Some of the specific questions that such a guideline could address include: (1) whether marginal cost should be interpreted as short-run or long-run; (2) what constitutes "substantially above" marginal cost and is it situation-specific; (3) whether the IMM will monitor or mitigate based on offer prices in the day-ahead market, or whether it considers this market entirely voluntary as it is treated under ERCOT's protocols; and (4) how suppliers can reflect in their marginal cost calculations various types of opportunity costs (*e.g.*, those related to environmental constraints that limit annual run-hours, self-imposed operating constraints intended to postpone maintenance cycle limits, or fuel-related opportunity costs).

In terms of potentially relaxing the PUCT Substantive Rules regarding monitoring and mitigation, the primary rationale would be to allow market participants to increase offer prices as a means of providing needed investment signals to the market. While there are a number of options for relaxing these rules, the approach would be to enable larger market participants to exercise their market power to a greater but still limited extent. This would increase prices and market returns. We also note that this type of limited, measured approach to market mitigation has worked effectively in Alberta's energy-only market to help attract enough merchant generation to sustain resource adequacy over the past decade.[176] Similarly, more permissive mitigation approaches are applied in European energy-only markets and in the conduct and impact mitigation approaches applied by the MISO and NYISO.[177] A somewhat relaxed monitoring and mitigation construct could contribute to restoring price signals without requiring any substantial market design changes.

There are two important drawbacks to relaxing market monitoring and mitigation rules. First, it invites offers that may deviate substantially from short-run marginal cost, which could introduce pricing and dispatch inefficiencies under some circumstances—although dispatch inefficiencies will also result if opportunity costs cannot be reflected in suppliers' bids. Second, and possibly more importantly, there is no clear way to determine how much the mitigation rules would need to be relaxed to achieve any particular desired level of investment. Given these drawbacks, we recommend relaxing mitigation rules but recognize that doing so may not be the most effective or direct way to restore investment signals, as we discuss further in Section VI below.

---

[174] See PUCT (1999), 25.505c-e.
[175] See Market Surveillance Administrator (2011).
[176] See Pfeifenberger and Spees (2011).
[177] See Reitzes, *et al.* (2007)

B. Facilitating Demand Response

In this section, we address how to facilitate greater demand response participation in efficient price formation, including: (1) a discussion of the importance of demand response in electricity markets; (2) the current level of demand response penetration in ERCOT compared to the estimated potential; (3) the wholesale factors affecting demand response development; and (4) how to efficiently accommodate demand response in wholesale markets so that it can support both reliability and efficient price formation.

1. The Importance of Demand Response in Electricity Markets

Electricity wholesale markets are more volatile than most commodity markets because they have neither storage capability nor sufficient short-term demand elasticity to moderate consumption during shortages. Most customers are on fixed rates and have no incentive to use less when spot prices are high, even on the hottest day of the year. When demand exceeds available supply, real-time prices rise to the offer cap, and system operators initiate involuntary load shedding. Therefore, load's traditional inelasticity creates reliability challenges without a large generation reserve margin. It is also economically inefficient, with some high-value end-uses being curtailed involuntarily while other relatively low-value uses continue.

In contrast, if a portion of demand is exposed and responsive to real-time prices, involuntary load shedding may be avoided. Demand response can thus reduce the amount of generation reserves needed to maintain a given level of reliability. Moreover, demand response can enable energy-only markets to support sufficient generation investment to reliably serve the residual load. Because demand response tends to occur at strike prices exceeding the offers of generation, its participation in the market can yield relatively high clearing prices, but only if it is able to set the price at its strike price. Realizing a continuum of high prices related to demand reductions at varying levels of scarcity would create a more robust and predictable distribution of peaking prices. This translates into a more stable revenue stream for generators than reliance on rare excursions to a high price cap. Achieving more DR participation would also displace some generation investments, but would achieve the same level of reliability. Our "High DR" simulations presented in Section IV demonstrate this effect.

Achieving this ideal requires widespread demand response and market structures that enable loads to contribute to efficient price formation. However, the ERCOT market is only part way toward that ideal. The ERCOT market has some demand response, although substantially less than the likely economic potential, and it is largely unable to contribute to efficient price formation. Nor is the reliability value of all types of DR fully recognized in ERCOT's loss-of-load studies used to set the target reserve margin.

2. Demand Response Penetration in ERCOT versus Estimated Potential

There are many forms of demand response in ERCOT: some is administered by ERCOT, some by the transmission and distribution service providers (TDSPs), and some by the load serving entities (LSEs); some is triggered by under-frequency and emergency conditions (which generally coincide with very high prices); and some is triggered by high prices directly. In this section, we quantify the current penetration of demand response in ERCOT in terms of peak load reductions, and we compare it to the achievable potential.

a. Current Demand Response Penetration

ERCOT currently administers two DR programs: (1) the Load Resources program (previously known as "LaaR"), where large customers with loads that are controllable via telemetry and under-frequency relays can provide up to 1,400 MW of responsive reserves on a day-ahead basis and can be deployed for energy in system emergencies;[178] and (2) the Emergency Response Service (ERS) program, where approximately 365 MW of medium-large commercial and industrial (C&I) customers are paid a capacity payment to be callable as a last resort in system emergencies.[179] ERCOT also allows price-sensitive demand bids in its day-ahead market. In fact large quantities of day-ahead demand bid as price-responsive, but it appears that most of these bids reflect load-side hedging rather than price-responsive load, *i.e.*, LSEs will defer some purchases to real-time if day-ahead prices become too high. ERCOT does not admit load reductions to offer and be paid as energy supply in either the day-ahead or real-time markets as in the FERC-regulated ISOs.[180]

The TDSPs pay for approximately 310 MW of load to be curtailable in emergencies as part of their "energy efficiency" programs.[181] The TDSPs also stimulate some peak load reduction through their four coincident peak (4CP) rate structures, which charge municipally-owned utilities, electric cooperatives, and large customers for transmission based on their metered load during the highest system load intervals in each of the four summer months (June through September). ERCOT does not directly observe the quantity of peak load reductions associated with 4CP rates, but a 2007 survey of LSEs indicates that it results in about 223 MW of load reductions.[182] Other analyses have found higher levels of response (600 MW or more) although an empirical analysis has not been conducted.

Price-based demand response is currently provided only by LSEs, but not through ERCOT. REPs and public power entities can create incentives for price-based DR by providing lower rates to customers who use less or curtail when spot prices are highest. We understand from our interviews with REPs that many large industrial customers are on "block-and-index" pricing, where all consumption above a certain amount is exposed to real-time prices. We also understand that few smaller customers are exposed to prices or engaged in any type of demand response.

Unfortunately, the extent of price-response programs is difficult to quantify exactly because pricing arrangements are a private contractual matter between REPs and their customers. Price-based load reductions were likely a major contributor to the 1,700 MW ERCOT load forecasting error in 2011 when prices reached $3,000/MWh. (The error may also be attributable in part to

---

[178] Although 2,500 MW are registered to participate as Load Resources, much less has been available at any given time. In the summer, as little as 850-900 MW participates during peak conditions, when registered Load Resources seek to preserve their ability to respond to high prices, to reduce their metered load during 4CP intervals, and to reduce their likelihood of being curtailed. ERCOT has not yet hit the 1,400 MW LR procurement ceiling since the ceiling was raised in April along with the PUCT's expansion of the RRS requirement, see ERCOT (2012a).

[179] See ERCOT (2012a).

[180] See Newell and Hajos (2010), p. 9.

[181] See ERCOT (2012a).

[182] See ERCOT (2007), pp. 18-19. The survey of LSEs did not account for 4CP load reductions that were triggered in response to signals from third parties or undertaken unilaterally by customers.

4CP response, voluntary public response to conservation appeals, and load forecast model error.) Other estimates indicate less price responsive demand. ERCOT's 2007 survey identified only 431 MW of curtailable load on real-time pricing.[183] Another study of industrial customers' 2002–2005 consumption found very little responsiveness to wholesale price signals.[184] We believe that quantifying price-responsive demand in ERCOT is an important area for further study.

Based on the above, we estimate the total quantity of peak load reductions from demand response in ERCOT to be approximately 1,600 MW in various registered ERCOT and TDSP programs. We also attribute approximately 1,000 MW of ERCOT's 2011 load forecasting error to customers' responses to high prices and 4CP rate structures. Including both registered and price-based load reductions, we estimate a total demand response penetration of approximately 4% of peak load.[185]

b. Comparison to Demand Response Potential Estimates

We estimate a total achievable potential of 8-15% of peak load reductions, which implies that DR penetration could grow by another of 4-11% of peak load. Our sense is that the potential is much closer to the high end of this range, given how few small customers are yet engaged and how much more valuable demand response will become as generation reserve margins become tighter and offer caps rise.

This 8-15% total potential estimate is based on a review of recent studies estimating the possible peak demand reduction achievable in Texas through expanded DR activities. The first study, FERC's 2009 *National Assessment of Demand Response Potential*, was led by *Brattle* staff and is the only study to develop a bottom-up DR potential estimate for each state.[186] It estimates DR potential by 2019 under three scenarios representing different program adoption rates. The second study is ACEEE's 2007 *Potential for Energy Efficiency, Demand Response and Onsite Renewable Energy to Meet Texas's Growing Electricity Needs.*[187] It estimates the potential impact of all cost-effective DR by 2023.

Basic assumptions behind the scenarios in these studies include:

ACEEE "Economic Potential" — Residential and commercial participation in direct load control are 43% and 36%, respectively; with expanded industrial participation in the Load Resources program.

FERC "Expanded Business as Usual" — Existing programs are expanded to national "best practices" levels of enrollment, with roughly 25% of residential and industrial customers and 5% of commercial customers enrolled in DR programs.

FERC "Achievable Participation" — Dynamic pricing and load control technologies are deployed on an opt-out basis, with roughly 75% of customers participating.

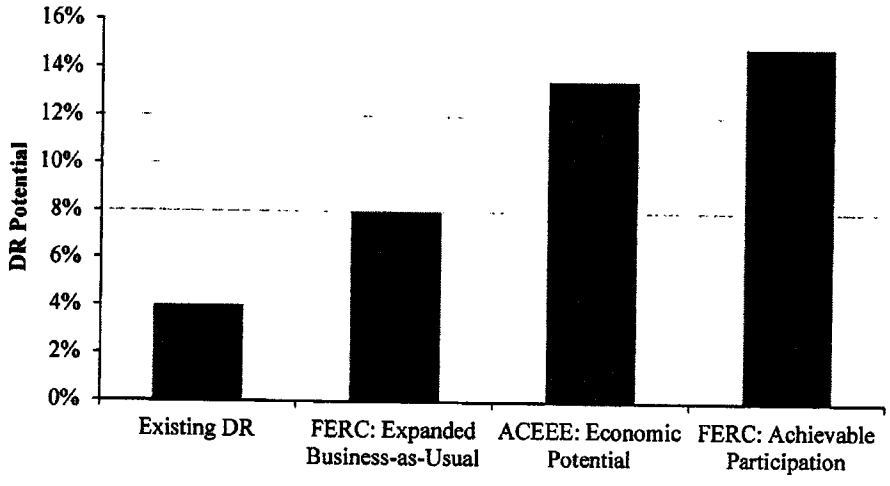
---

183 See ERCOT (2007), p. 12.
184 See Zarnikau and Hallet (2008).
185 The 1,600 MW of registered DR includes 900 MW LR, 365 MW ERS, and 310 MW in TDSP programs. Total penetration percentage calculated from 68,379 MW 2011 peak load; see ERCOT (2012h), p. 3.
186 FERC (2009).
187 Elliot, *et al.* (2007).

FERC "Full Participation" — Dynamic pricing and load control are deployed on a mandatory basis; we did not include this scenario in our analysis because we do not consider it plausible over a ten-year horizon, and because a mandatory deployment under a default tariff would not be possible under current regulations.

Figure 30 shows the range of DR potential estimates of 8%-15% peak load reductions from the two studies, relative to the current penetration of approximately 5%.

Figure 30
ERCOT DR Peak Demand Reduction Potential



Sources and Notes:
Existing DR estimated by *Brattle* to include large industrial price response.
ACEEE Economic Potential is adjusted from the figure in ACEEE report to include only DR potential and exclude impacts from other sources.

Some parts of the country are closer to meeting their potential, such as in PJM, where DR can reduce peak load by 10%.[188] We and others attribute that success primarily to the attractiveness of capacity payments to third-party curtailment service providers (CSPs), who are able to aggregate customers' load reduction capabilities and sell them into PJM's capacity market. In fact, over the nine delivery years since the capacity market was implemented, PJM's DR penetration has increased from 1% of peak load in 2006/07 up to 10% for 2015/16.[189]

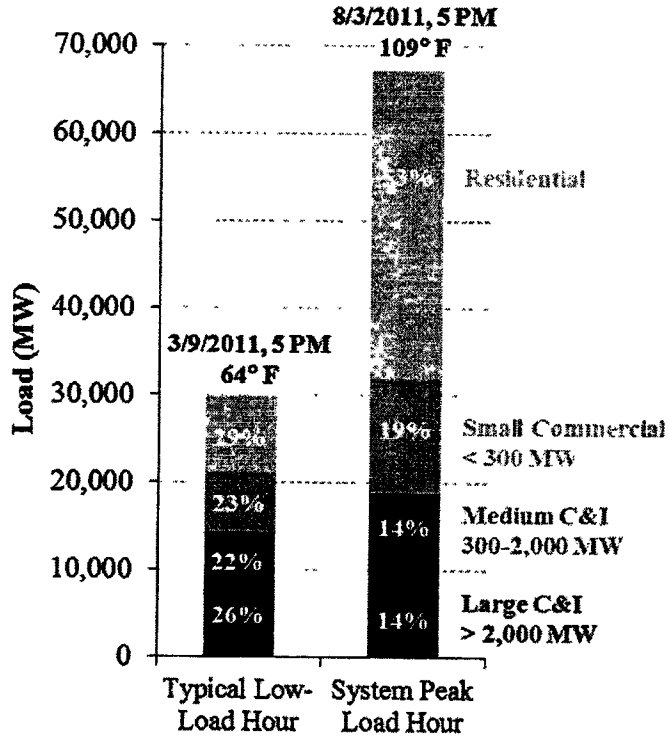
c. *Opportunity Areas for Achieving Greater Penetration*

Current penetration and achievable potential vary by customer segment, as do the types of opportunities most suitable for each. We examine large C&I customers, medium-large C&I customers, and residential and small commercial customers. The most important DR opportunities are those that are achievable during summer peak load conditions, although some less substantial cost savings would also be available during other parts of the year. Figure 31

[188] Based on 15,755 MW of cleared demand resources in the 2015/16 Base Residual Auction, see PJM (2012a), p. 11. The PJM 2015/16 peak load forecast is 163,168 MW, see PJM (2012b), p. 4.
[189] Based on 1,200 MW of Active Load Management in 2006/2007, see PJM (2012a), p. 12. The PJM 2006 peak load was 145,951 MW, see PJM (2007), p. 56.

shows that residential and small commercial customers accounted for 72% of load during the 2011 system peak, even though they made up only 52% of load during low-load conditions. Residential and small commercial customers show very high loads during peaking conditions relative to their average load. By comparison, large C&I customers show a peak load that is only marginally above off-peak loads.

Figure 31
Peak and Off-Peak Load by Customer Segment



Source:
Based on ERCOT data, ERCOT (2012a).
Reported temperatures are from Dallas.
Customer class breakdown is for competitive choice areas.

Largest Commercial and Industrial

The largest customers with peak loads greater than a few MW are already quite engaged in programs to exercise their operating flexibility to manage their substantial energy costs. We heard in our interviews that large C&I customers who have any flexibility are already responding to wholesale prices, managing 4CP load to reduce T&D charges, and providing some LR and ERS in ERCOT programs.

Given their active participation, the potential for engaging additional large C&I *customers* may be limited, but the magnitude of their load reductions may increase in response to higher prices. The most interesting incremental opportunity with large customers is to enhance their role in price formation, by entering them directly into SCED at a strike price, as we discuss further below. Because Load Resource providers are already controllable with telemetry, they are the best candidates to participate in SCED and set real-time prices.

Medium Commercial and Industrial

Most customers below a few MW of peak demand do not spend enough on energy to have developed sophisticated load management practices. They usually do not have their own energy management staff, although they may engage consultants or their REPs for this purpose. We understand from our interviews with REPs that this segment is not yet fully engaged in dynamic pricing or curtailable rates. So far, they have done little to reduce peak demand because super-peak prices have not been high enough to generate widespread interest among customers for whom energy is a small percentage of their total costs. The REPs we interviewed expressed growing interest as their exposure to higher price caps increases, but they cautioned that penetration will be slow.

Customers with peak loads exceeding 300 kW may be attractive targets for CSPs if they have the opportunity to aggregate their load reductions and sell them as wholesale capacity supply. In fact, this is the segment that has provided most of the 450 MW of ERS to date. Other ISOs with capacity payments have elicited substantially more DR and done so in a way that forces the DR to compete with generation for providing both capacity and energy. While we believe that ERCOT could similarly procure substantially more DR through its ERS program, we caution against this avenue because of its out-of-market nature, as we discuss in Section VI.

Residential and Small Commercial

As Figure 31 above shows, small customers account for more than 70% of peak load, and they currently provide little demand response, especially in the retail-choice areas of the ERCOT region. Large numbers of small customers are inherently difficult to engage for DR because the potential savings are relatively small compared to recruitment, equipment, setup, and ongoing management costs. Mass market penetration will be slow and possibly never very deep due to limited interest and lack of economies of scale. However, penetration can be much higher than zero, and the PUCT has laid the groundwork for willing customers to engage.

The TDSPs will soon complete the deployment of the Advanced Metering Infrastructure (AMI) in all investor-owned utility territories, achieving capabilities beyond all other states.[190] This AMI will enable REPs to offer time-varying and dynamic rate options to the mass market, but it is unclear how many customers will enroll in dynamic rates. There is limited empirical evidence on adoption rates, since dynamic tariffs are just being rolled out to residential customers in places like California, Maryland, and Washington, DC.[191] However, post pilot surveys find very high satisfaction levels among participants.[192] Recent pilots around the world have consistently shown that customers will reduce peak demand when faced with a higher peak price or a rebate for load curtailment, even without installing any enabling equipment such as smart thermostats. Peak load reductions generally range from 5% to 10% for time-of-use rates, and from 10% to 30% for rates with stronger peak price signals.[193]

---

190 See PUCT (2011).

191 For example, Pepco is in the process determining the details of its residential dynamic pricing rollout with Maryland stakeholders; see MD PSC (2012). Also, PG&E has begun offering dynamic pricing on a voluntary basis to all eligible residential customers; see PG&E (2012).

192 See Faruqui (2009).

193 See Faruqui and Palmer (2012), and Faruqui and Sergici (2009).