Measuring the Capacity Impacts of Demand Response

to be published in the *Electricity Journal* – pre-print version

Robert Earle
Edward P. Kahn
Edo Macan[1]

June 12, 2009

1. Introduction

Demand response is an increasing part of the energy policy agenda in the United States. The Federal Energy Regulatory Commission (FERC) has undertaken major initiatives to encourage the incorporation of demand response in the wholesale markets, the American Recovery and Reinvestment Act of 2009 (aka, "the stimulus bill") has provisions supporting demand response, and many states have instituted demand response initiatives or are contemplating whether to do so.[2]

Demand response is a reduction in demand designed to reduce peak demand or avoid system emergencies. In this regard, demand response can be a more cost effective alternative than adding peaking generation in trying to meet occasional demand spikes. The top 1 percent of hours for many electric power systems accounts for over 10 percent of the demand (measured in MW of capacity), as illustrated in Figure 1 for the system we studied.[3] In the year studied, the peak load was 44,961 MW and the top 87 hours accounted for 11 percent of the demand. In order to satisfy this demand, generation that runs infrequently must be available to meet that demand. Demand response is meant to reduce demand during those top hours, and therefore, avoid the capacity costs associated with generating units that only run a few hours out of the year. In integrated resource

---

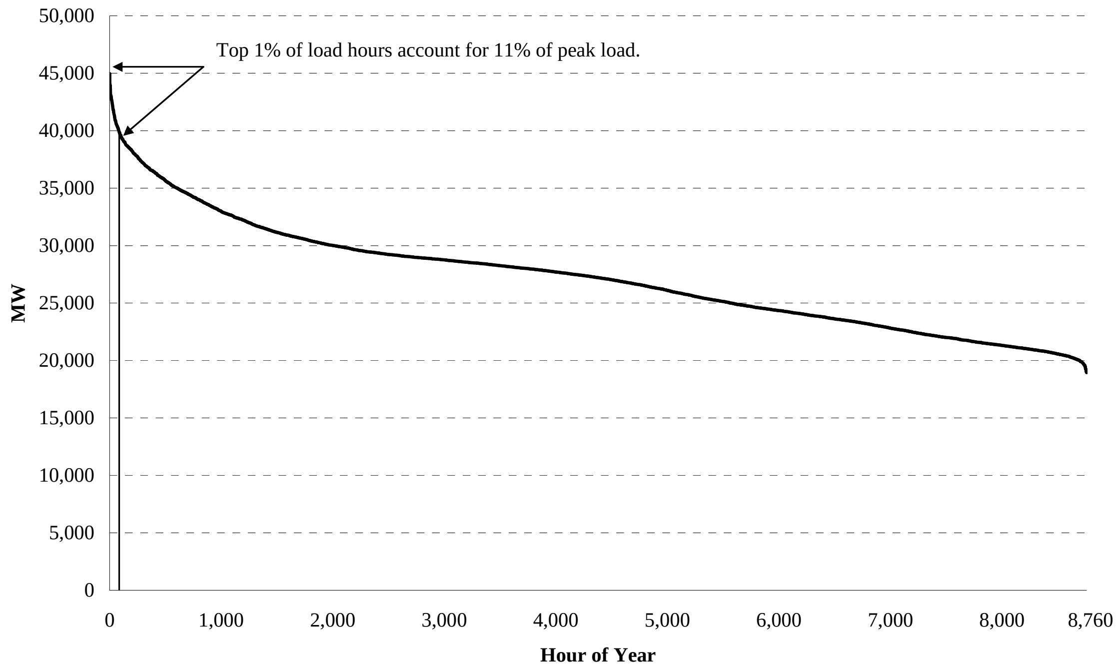
[1] The authors are employees of Analysis Group in San Francisco and would like to thank Matt Barmack, Peter Cappers, Jack Ellis, Rodney Frame, Carl Silsbee, and Susan Tierney for their helpful comments and suggestions. The views in this paper represent the views of the authors alone, not the views of Analysis Group. Questions and comments may be addressed to rearle@analysisgroup.com.

[2] In 2008, California approved default dynamic rates for electric power customers, and Colorado, Maryland, and Ohio encouraged demand response with regulatory measures or legislation. Also, the FERC issued a final rule, Order 719, in October 2008 that was intended to remove a number of barriers to demand response participation in organized markets. See FERC (2008), *Assessment of Demand Response and Advanced Metering*.

[3] The system analyzed is a combination of CAISO and SMUD using load from 2002. Complete details can be found in Kahn (2004).

planning (IRP), demand response is one method in portfolio of resources to meet peak load and reliability criteria. This paper examines the reliability contribution of certain demand response programs.

Figure 1 - Load Duration Curve for CAISO (2002)

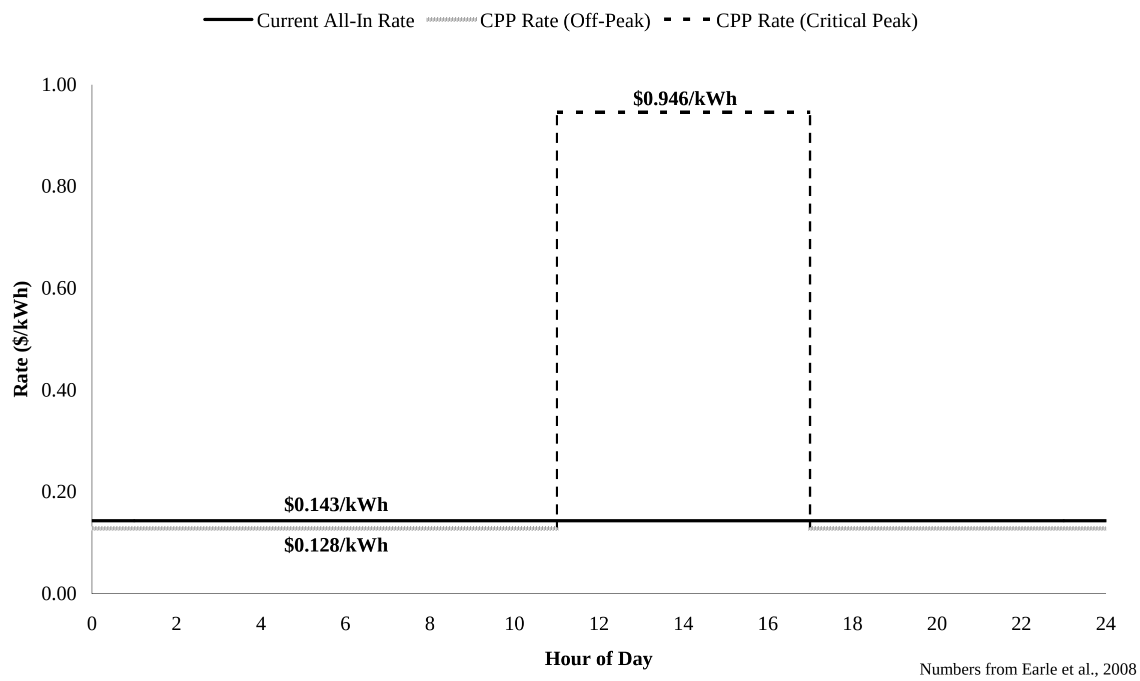


While demand response comes in many flavors, pricing programs that give consumers price signals that vary over time (to induce reduction during times of peak demand) are an increasing focus of utilities, regulators, and policy makers.[4] In the case where the time periods are not fixed ahead of time in the tariff, as with time-of-use programs, these programs are often referred to as dynamic pricing. Two typical programs under this category are critical peak pricing (CPP) and peak time rebate (PTR). Under CPP, certain hours (for example, noon to six) are designated critical peak hours. On the days that the utility designates as a critical peak day or event day, the consumer pays more than the usual tariff during critical peak hours. Which days are critical peak days are not known until the day before or the same day as the critical peak day. There is usually a limit on

[4] Different types of demand response programs include direct and indirect load control, dynamic pricing programs, etc. See Earle et al. (2008) for an overview of various types of demand response.

the number of critical peak days during the year and often they are restricted to the summer season. Figure 2 illustrates the concept.

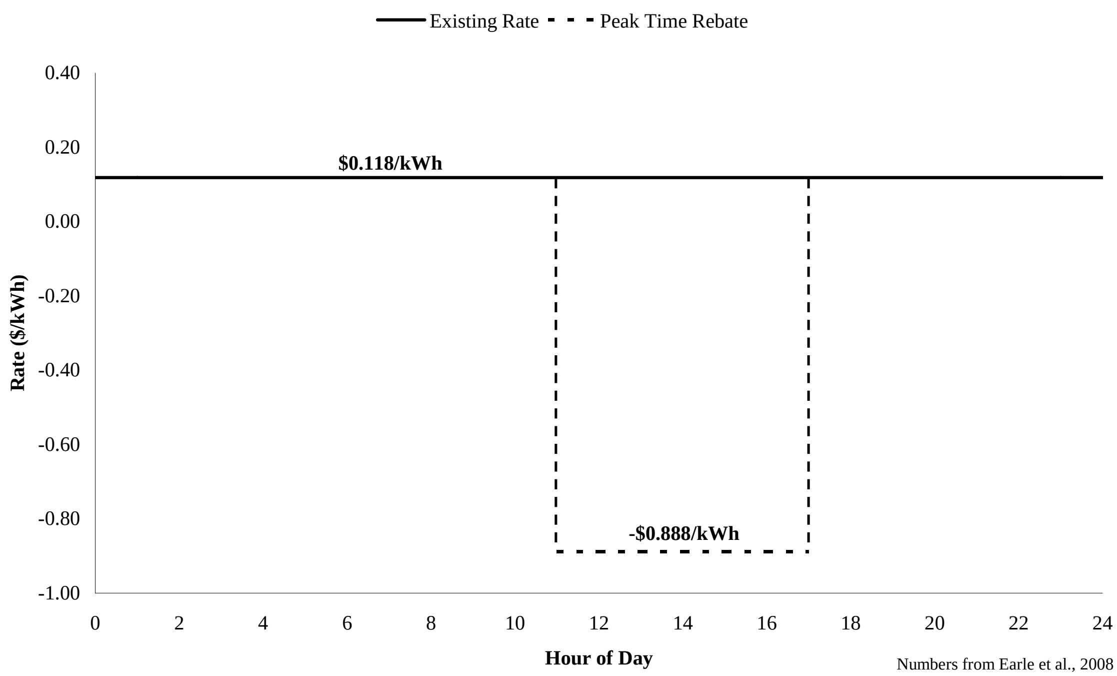
Figure 2 - Critical Peak Pricing (CPP)



PTR shares a similar design to that of CPP in that the utility selects days when the program is active during peak hours, but instead of charging the customer more for usage during peak periods, customers are given a rebate for less consumption during times selected as critical periods. Figure 3 shows how PTR essentially looks like the inverse of CPP during critical periods.[5]

[5] The roots of such programs can be traced back to long-standing programs of many utilities that offered large industrial customers a price break in all hours for signing up to a rate where their supply could be interrupted for a short duration a maximum number of times during a year. Those programs suffered from the "take-the-price-break-for-granted" problem, when the utility found that it did not need to interrupt the customers for many years at a time, and then needed to, to the dismay and grumblings of the interruptible customers who had begun to feel entitled to not being interrupted. The design of these newer approaches tends to tie the demand reduction more closely to the payment for it. Letzler (2007) gives a discussion of PTR type programs in the context of behavioral economics.

Figure 3 - Peak Time Rebate (PTR)



Under most designs of either CPP or PTR, the critical period is a fixed time of day with a maximum number of critical periods per year. For example, Pacific Gas & Electric and Southern California Edison have explored CPP/PTR designs with critical periods between noon and 6 PM that occur at most 12 times per year. That is, though there may be more than 12 times a year when stress conditions arise, 12 days is the maximum for which the customer might be exposed to "supranormal" prices. Because the value of demand response is largely due to avoided capacity costs, in order to get the most value out of CPP or PTR demand response programs, it is essential to carefully choose the periods to activate the programs in which capacity is needed the most. To examine the capacity impacts of these programs, we tested program impacts against a variety of scenarios with a reliability impacts model using California data from 2002.

Though the year 2002 was chosen as a test year primarily because of the availability of data, and while our exact results are dependent on the particular system and program designs studied, the general nature of the results is likely to hold for systems that are relatively more stressed than the one we have examined. As a result, we are able to draw some general conclusions. First, at relatively low levels of demand response, CPP and PTR programs can provide close to 100% of the capacity value relative to the amount of demand reduction. In other words, at low levels of demand response, these programs provide almost a MW of capacity value for a MW of demand response. Second, as demand response levels increase, the capacity value of demand response can decrease significantly. Thus, the capacity value of the studied demand response programs is subject to decreasing returns. Third, as the amount of actual response increases (rather than the enrolled MW or number of participants), flexibility in program design becomes increasingly important. However, there are important tradeoffs between having the simplicity of rate design that results in transparent and understandable rates, and dynamic prices that meets its full potential. Fourth, because there is not yet full implementation of these programs, little is currently known about the variability of the level of response. Demand response that provides 1000 MW of reduction with certainty during a critical period will likely have a greater impact than demand response that has an average response of 1000 MW but may be more or less than that average.

The rest of this paper is organized as follows. Section 2 discusses the model used to study the capacity impacts of mass market dynamic pricing, as well as presents basic results demonstrating the decreasing returns to scale for CPP and PTR programs. Section 3 discusses sensitivity runs of the model that explore the importance of program design choices and flexibility in the deployment of demand response. This is followed by Section 4 which examines the impacts of the variability and uncertainty of demand response on its capacity value. Finally, Section 5 concludes the paper with some policy implications that arise from the work presented in this paper.

2. Modeling Demand Response Capacity Impacts and Decreasing Returns to Scale

Just as with thermal units, the reliability impacts of demand response programs depend on a variety of factors. Thermal units with higher forced outage rates naturally contribute less to reliability than do units with smaller forced outage rates. Moreover, large thermal units contribute less to reliability on a MW per MW basis than smaller units with the same forced outage rate. Power system engineers have developed reliability indices and applied them to making incremental assessments of new capacity additions. The literature on reliability measurements of power systems goes back over 50 years with probability measures introduced in the late 1940s.[6] One of these indices, the loss of load expectation (LOLE) index, measures the expected number of hours with loss of load within a year. The often used "one day in ten years" criterion commonly cited as a planning objective for LOLE means that LOLE should be about 2.4 hours per year.[7]

The effective load carrying capacity (ELCC) is defined by as the amount of new load that can be added to a system at the initial LOLE after a new unit is added.[8] ELCC can be expressed in terms of the percentage of the rated capacity of a unit and has been used to measure the capacity contribution of wind resources.[9] The box insert, **Reliability Indices**, gives a more detailed description of ELCC.

We can measure the ELCC of a demand response program in a similar manner, and have adapted the approach in Kahn (2004) for this purpose. Load data from 2002 for CAISO is used along with publicly available data on thermal unit capacities and forced outage rates. As new thermal units were added throughout the year, they were added to the model. Publicly available hydro generation data was available from the year 2000 and this generation was applied in a deterministic fashion to the 2002 load. The hydro generation was applied to the load by matching the highest hydro generation hour against the highest load hour and so on through to the lowest hydro generation hour against the lowest load hour. Kahn (2004) uses this method and also tests another method of

[6] Calabrese (1947).
[7] There is some confusion about precise interpretation of the tradition "one day in ten years" criterion. It was originally developed when computational resources were relatively expensive and LOLP calculations represented load as one peak demand value per day. Using an hourly load model results in something less than 2.4 hours per year when the same system is evaluated with an hourly load model as opposed to assuming that load in every hour of the day is the peak load. One test of these relationships found an LOLE expressed in hours/year of 0.25 corresponding to 0.1 days/year using the daily peak model (see Jamali (1979)). For ELCC calculations in large systems, the precise interpretation may not matter much since the LOLE versus peak load profile has an almost linear shape on log-linear scales for LOLE values below 0.3 days/year (see Billinton and Allan, 1984, Figure 2.10).
[8] Garver (1966).
[9] Kahn (2004).

matching hydro generation against load and finds that the results do not differ significantly. Imports are also treated deterministically in our calculations.

The reliability measure LOLE is sometimes criticized because it only measures the expected amount of time that outages will occur as opposed to the expected amount of unserved energy (EUE) or load not met. For a given system, an hour with 100 MW of unmet load is logically worse than an hour with only 5 MW of unmet load. In general, this is a valid concern. However, for the system we have examined, the LOLE and EUE correspond closely and using one measure or the other to define ELCC makes no material difference to our results.

For the initial results, the simulation assumed that the critical peak period was between noon and 6 PM with 12 critical peak days to be called during the summer months (June through September). An assumption was made that the program would be deployed to maximum effect for reliability purposes. In other words, the operator of the program would choose with perfect foresight the days that would be best to declare critical peak days, which in turn increases the measured ELCC of the program.[10] This assumption is relaxed in the next section to show the importance of choosing the right days to declare to be critical peak days. Another key assumption is that the demand response amount is perfectly knowable and predictable. In reality, this is certainly not the case, and is examined in Section 4. The effect of this assumption is to raise the measured ELCC, thus giving more capacity credit to the demand response program than it is likely to achieve. Finally, many anticipate that mass market dynamic pricing programs will achieve little net energy savings.[11] In other words, while consumers may reduce their demand during the critical peak hours, they increase their consumption during the non-critical peak hours, so, on an energy basis, consumption is not reduced. Consumers seem to substitute consumption in one time period for another.

---

[10] Other than these assumptions, the simulation is agnostic with respect to other aspects of program design such as when the price signal is issued, what type of price signal, or how the response is achieved (through a customer behavior or through automation).

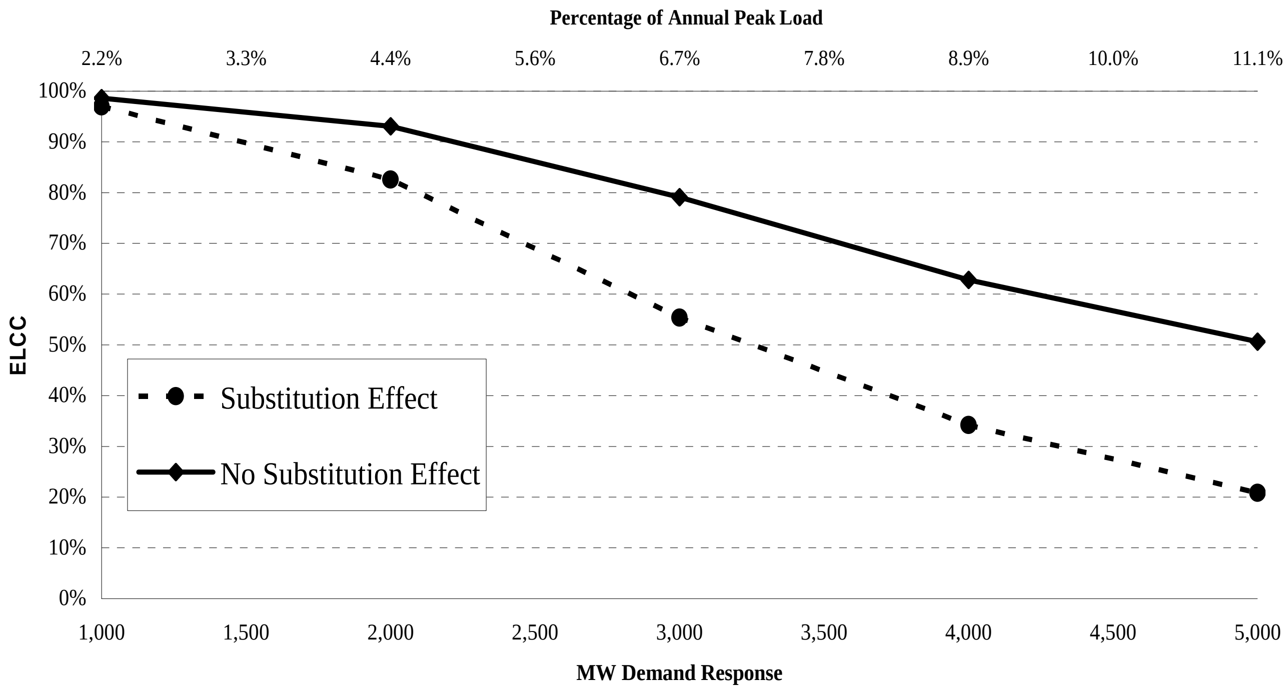
[11] See, for instance, Faruqui and Wood (2008) for a discussion of the general level of various types of impacts from dynamic pricing. York and Kushler (2005) discusses the nexus between energy efficiency and demand response and how one measure may reinforce effects in the other. IBM (2007) and Summit Blue (2006) find some conservation effects.

The initial results present two scenarios of the ELCC measurement with respect to this substitution. In the first, there is no substitution effect, so that reductions during critical peak result in net energy savings of the same amount. The second scenario produces a very mild substitution effect in which the reduction during critical peak hours results in an increase in consumption over all the other 18 hours of the critical peak day. So that, if there is 300 MW of reduction during the critical peak hours, noon to 6 PM, all the other hours in the critical peak day have an increase in consumption of 100 MW. This very mild substitution effect is used in order to be conservative in our estimates of ELCC. It would stand to reason that the decreased consumption between noon and 6 PM would tend to increase consumption on hours closer to the critical period such as 6 PM to 10 PM rather than 2 AM as in our assumption. Nevertheless, it appears there is very little or no empirical evidence for how the substitution effect plays itself out over the hours of a day. More study of the timing of potential substitution effects will be important as systems increase their reliance on demand response and as new end uses like PHEV (plug-in hybrid electric vehicles) enter the system.[12]

The initial results under the assumptions outlined above are shown in Figure 4. The figure shows the MW of demand response (actual amount of reduction as opposed to the enrolled MW) versus the ELCC as a percentage of the MW of demand response. The top axis gives the demand response as a percentage of the annual system peak.

[12] In this regard, the work of NERC to establish standards for gathering data on demand response may be critical. See NERC, *Data Collection for Demand-Side Management.* Centolella and Ott (2009) suggest some methods to better understand the intra-day effects of demand response.

Figure 4 - ELCC as Percentage of Demand Response MW



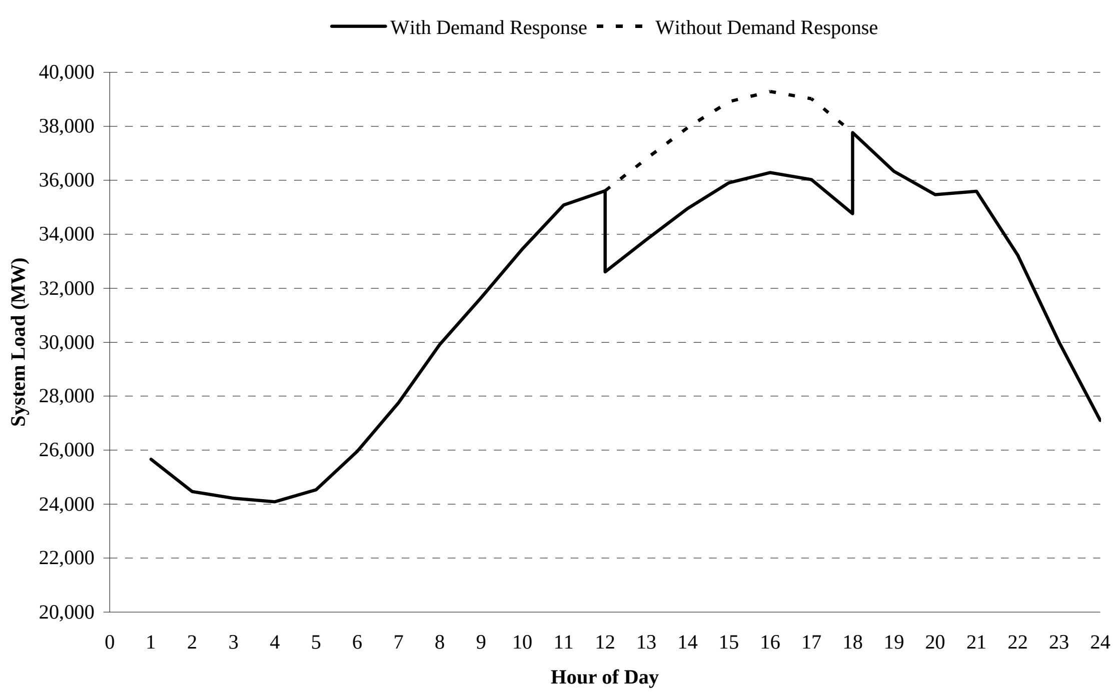
At relatively low levels of response, the ELCC of demand response is nearly 100%. The selected hours are able to effectively pick off the most critical hours and increase reliability nearly to the same amount of the response. As the MW response increases, however, a decline is seen in the value of response. At a demand response level of 2,248 MW, which is roughly equal to 5% of peak demand, demand response has an ELCC of about 90 percent with no substitution effect and less than 80% with the mild substitution effect. 5000 MW of demand response shows significantly degraded ELCCs of about 50% and 20% for the no substitution scenarios and the substitution scenarios respectively.[13]

For economists, as dismal scientists, these decreasing returns to scale are not surprising as a matter of principle. But why do they occur in this case? One answer can be found

[13] By comparison, one of the few public references on the ELCC of demand response can be found in the Arizona Public Service Company's *Resource Plan Report* (p. 97). For a variety of commercial and industrial load management programs they report an ELCC between 70 and 80 percent. They do not, however, seem to report the percentage of peak demand that the load management programs comprise.

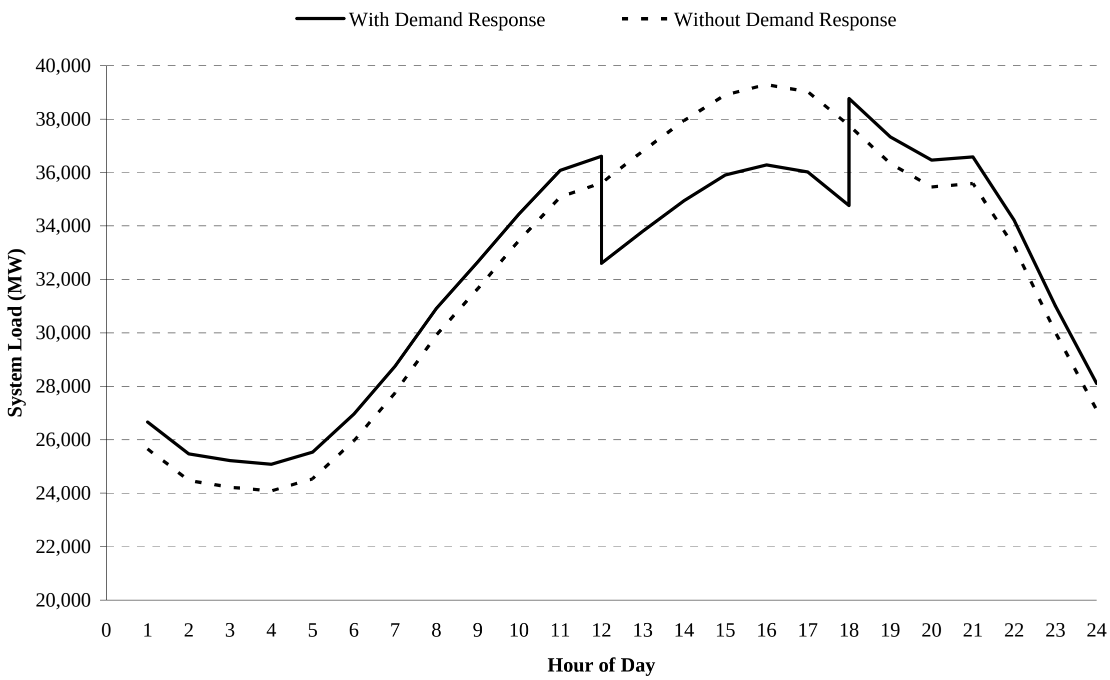
by going back to Figure 1 of the system load duration curve. The demand response program modeled has 72 hours of impact that occur over noon to 6 PM during twelve days. Through examination of the data on when peak demand occurs, we see that the top 72 hours of the year accounting for 10% of system peak are actually spread over 17 days rather than 12. This means that the program as designed cannot capture all of the top 72 hours. Moreover, even if the program could be spread over 17 days rather than 12, 5 of the top 72 hours occur outside the noon to 6 PM window. More importantly, however, is that at a certain point, reductions in demand during the critical peak window do little to increase reliability. Figure 5 shows a demand response of 3000 MW on 08/01/02 with no substitution effect.

Figure 5 - Example of a Demand Response Program with No Substitution Effect



While the pre-existing peak is eliminated, the shoulder hours now constitute a new peak. With the substitution effect, demand response "fangs" are created with the increase of load in the non-critical peak hours. Figure 6 illustrates this effect.

Figure 6 - Example of a Demand Response Program with Substitution Effect

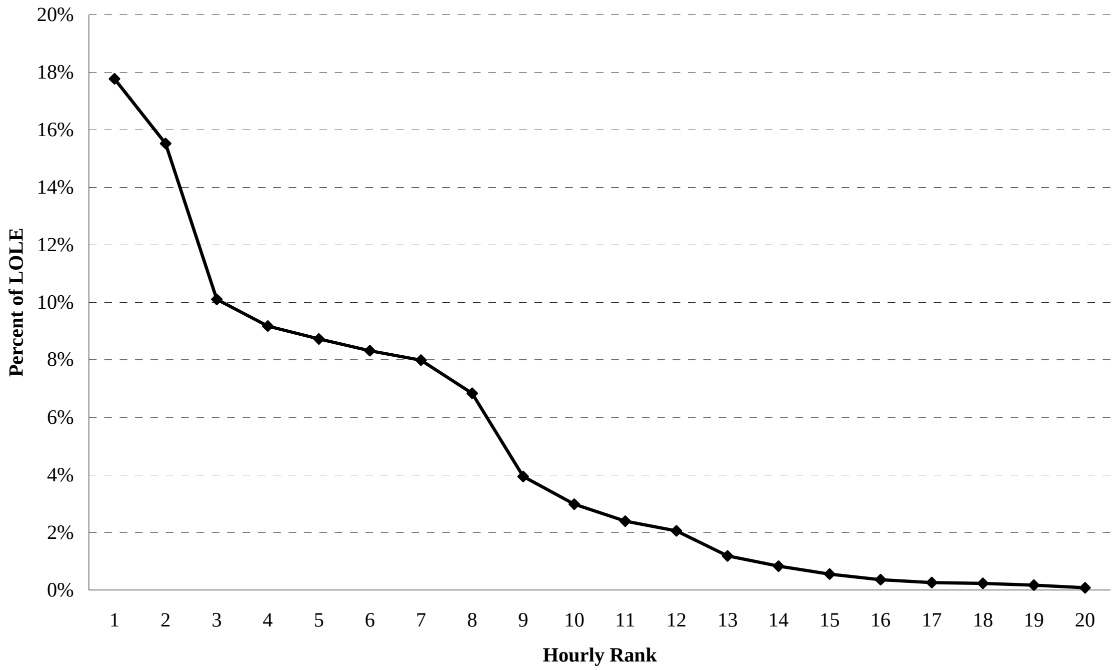


As a result, to the degree that there is substitution, that is, increased consumption in non-critical peak hours to match decreased consumption in critical peak hours, the capacity value of CPP and PTR type demand response programs will be much less than with no substitution. Substitution effects that are more likely to be realistic (such as substitution only to the hours close to the critical peak period) would show degradation of the capacity value of the demand response program that is even more severe.

What drives these results? Most of the LOLE (or EUE) occurs within just a few hours out of the year. The top 20 hours in the year account for over 99% of the LOLE with these hours all occurring in just four summer days as shown in Figure 7.[14]

[14] The numbers are from our calculation of LOLP for each day in the simulation.

Figure 7 - Top 20 Hours LOLP as Percentage of LOLE



As a result, targeting those hours is key. However, once those hours are targeted – and they seem to be targeted fairly well with the particular year examined and the program design that was simulated – reducing the LOLP or EUE for a particular hour by reducing demand quickly exhausts itself. Other hours now start to dominate the LOLE calculations and programs that can target those hours help retain the effectiveness of demand response programs as measured by ELCC. Hence, flexibility of program design and hours targeted can be important in achieving effectiveness. These results depend, of course, on the system, but given the "peaky" nature of many systems in the U.S. it seems quite possible that these results are not peculiar to the particular system and year we have examined. The next section addresses these issues of flexibility and targeting of the right hours.

To the degree that other systems do not have most of their LOLE concentrated in just a few hours, as in Figure 7, then the results of our simulation would differ. However, such systems would have a much less "peaky" load shape than the one we have examined and thus would not need demand response as much as the system we have examined.

3. How much does flexibility in demand response program design and deployment affect its capacity value?

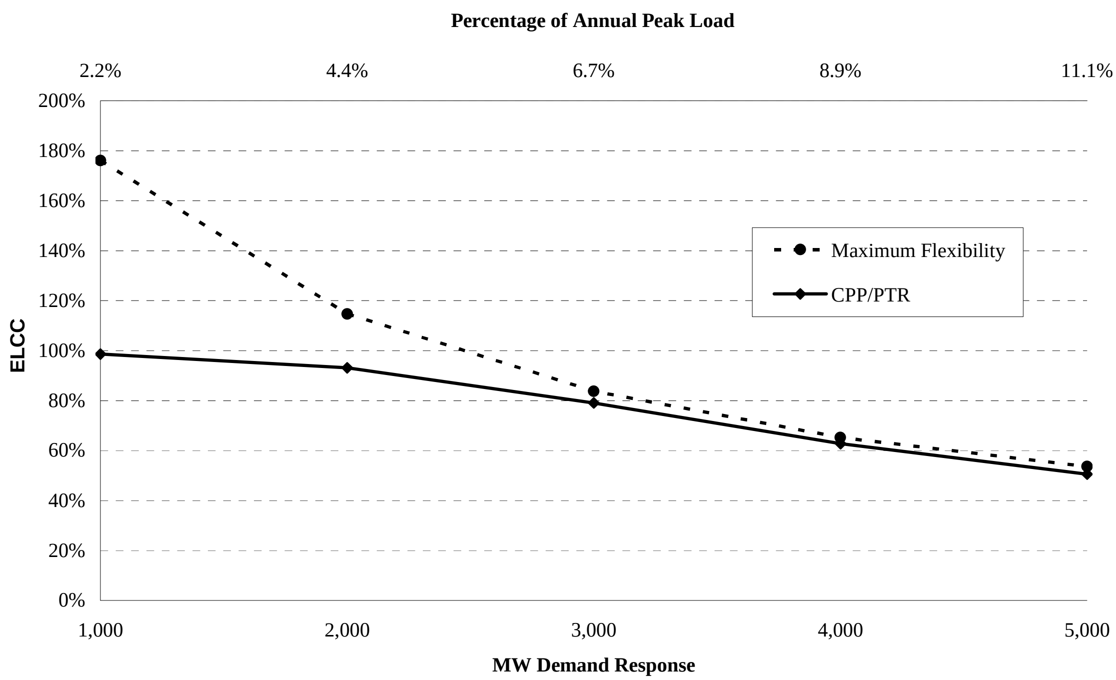
The previous section showed that demand response from CPP and PTR type programs degrades in its reliability or capacity value as the amount of demand response provided increases. These results suggest that increased flexibility in program design as well as programs complementary to CPP/PTR could greatly increase their value. Discussions of CPP/PTR programs tend to have fixed hours for critical peak periods and limited number of days in which they occur. This is because rate designers and regulators are concerned whether consumers are able to understand and react to dynamic prices.[15] In the simulation from the previous section, for instance, critical events were limited to 12 days during the summer and only between noon and 6 PM. The ability to target other days and other hours would enhance the effectiveness of the demand response program, but at the cost of making the rules of the program more complicated for consumers. Prices that have overly complicated rules make it difficult for consumers to react to them so that both inefficiency of response as well as unhappiness with the tariff regime arise. There are many alternative rate designs for dynamic pricing and some of these take into account the need for flexibility. CPP-V (CPP Variable), for example, is a variant on CPP that allows for a flexible critical peak period.

Automation on a large scale through efforts such as the smart grid may well be the answer that allows for more flexible and therefore more valuable demand response. The so-called “Ron Popeil” effect is where customers can set their preferences ahead of time and not have to pay direct attention in real time to fluctuating prices. That is, as Popeil is famous for saying about many of his kitchen appliances, “Set it, and forget it.” Whether smart grid, home automation, and the like can make demand response much more flexible while keeping rate structures acceptable to consumers and whether it is cost-effective to do so are questions beyond the scope of this study. However, in order to

[15] See Bonbright et al. (1988) for the classic list of ratemaking criteria which include simplicity and understandability as important attributes.

examine the potential capacity value of such efforts, we calculated what the maximum capacity impact of a completely flexible demand response portfolio would be. For a given CPP/PTR program size we assumed that the equivalent number of MWh of demand reduction would be available whenever needed across all hours. So, for example, for a CPP/PTR program with 1000 MW of impact for 6 critical peak hours on 12 days, we calculated the impact of 72,000 MWh[16] of demand reduction. To simulate what the maximum effectiveness of completely flexible demand response might be, the demand reduction was allocated to the hours by shaving peak so that the hours with the highest demand are reduced first.[17] The results measuring the capacity value of demand response flexibility are shown in Figure 8.

Figure 8 - The Value of Flexibility as Percentage of Demand Response



The CPP/PTR line (the lower of the two) shows our basic results for the CPP/PTR program discussed above (12 critical days, noon to 6 PM, no substitution effect). The maximum flexibility line shows the ELCC with the same number of MWh of demand response are deployed in order to shave peak. At low levels of demand response (MW

[16] 72,000 MWh = 1000 MW x 12 days x 6 hours/day.

[17] While this approach might not result in the precise maximum ELCC for completely flexible demand response, given the close relationship for the system studied between the level of demand and LOLP, it is likely very close.

reduction), the flexible demand response provides more than 100% ELCC. This is possible, of course, because more than the program amount can be deployed in a given hour. For instance, even though the program size may be 1000 MW, the very top peak hour can be shaved by more than 1000 MW. The impact of flexibility starts to decline rapidly, however, as the program size increases. This is not surprising since as we shave the load shape more and more, the additional MWh get spread across more hours resulting in less net gain from additional MWh of demand response.

The difference between the two lines in Figure 9 shows the maximum that can be gained above the CPP/PTR baseline by making demand response flexible. For relatively small program sizes, the value of flexibility is quite high, but it quickly decreases as the program size increases. While the exact value of flexibility from a reliability/capacity point of view is system dependent, and the pace of decrease of that value as program size increases is also system dependent, it seems likely that the general picture we have found for the system studied in this paper holds more generally. As the program size increases returns to flexibility in deployment of demand response decrease because the hours that contribute the most to LOLP or EUE have already been addressed.[18]

4. Uncertainty of Demand Response

Although they accept that demand response programs can provide peak reduction, there are many who are still concerned that the level of demand response is uncertain. For example, 1000 MW of reduction might be expected, but the actual amount of reduction realized could be some number either greater or smaller than 1000 MW. For this reason they are reluctant to count demand response programs in resource adequacy. While there is little evidence about the variability of response, the California Statewide Pricing Pilot (SPP) provided some insights.[19] In that pilot, various demand response programs were tested and it was possible to calculate the variability of response. Ninety-five percent confidence intervals ranged from plus or minus 6% to 18% of the mean response. So, for

[18] It should also be noted that we have assumed perfect foresight so that the capacity value implied by each of the lines in the figure is likely overstated.
[19] See Faruqui and Woods for a summary of the SPP.

example, if the expected response level was 1000 MW, using 10% as the variability of the response, the range 900 MW to 1100 MW covered the actual response with 95% probability.[20] To test the impacts of the uncertainty of demand response, we introduce a probability distribution in our model to make demand uncertain.[21] We estimate the ELCC for various levels of demand response uncertainty as measured by the coefficient of variation (i.e. standard deviation divided by the mean). The programs tested in the California SPP, for example, had coefficients of variation of around 3 to 9%. A generating unit with a forced outage rate of 5% has a coefficient of variation of 23%.[22]

The results from the simulation of uncertain demand response are shown in Figure 9.

[20] See, Earle et al. (2008). The pilot tested a variety of CPP style programs.

[21] A discrete distribution with a triangular shape was used for ease of computation. It seems unlikely that the particular form of the distribution is of much consequence in the results. We also assume that the amount of realized demand response is independent of system load. This might not be true as there may be more air conditioning load to reduce on a day that is particularly hot meaning a positive correlation between load and demand response. On the other hand, on a particularly hot day, consumers may be more reluctant to reduce their air conditioning load resulting in a negative correlation between demand response and load.

[22] Simply comparing the coefficients of variation of a generating unit and demand response programs does not give an adequate comparison as the shapes of their distributions are very different. The coefficient of variation of a generating unit with a 5% effective forced outage rate is high because it cannot produce anything for 5% of the time.

Figure 9 - ELCC as Percentage of Demand Response

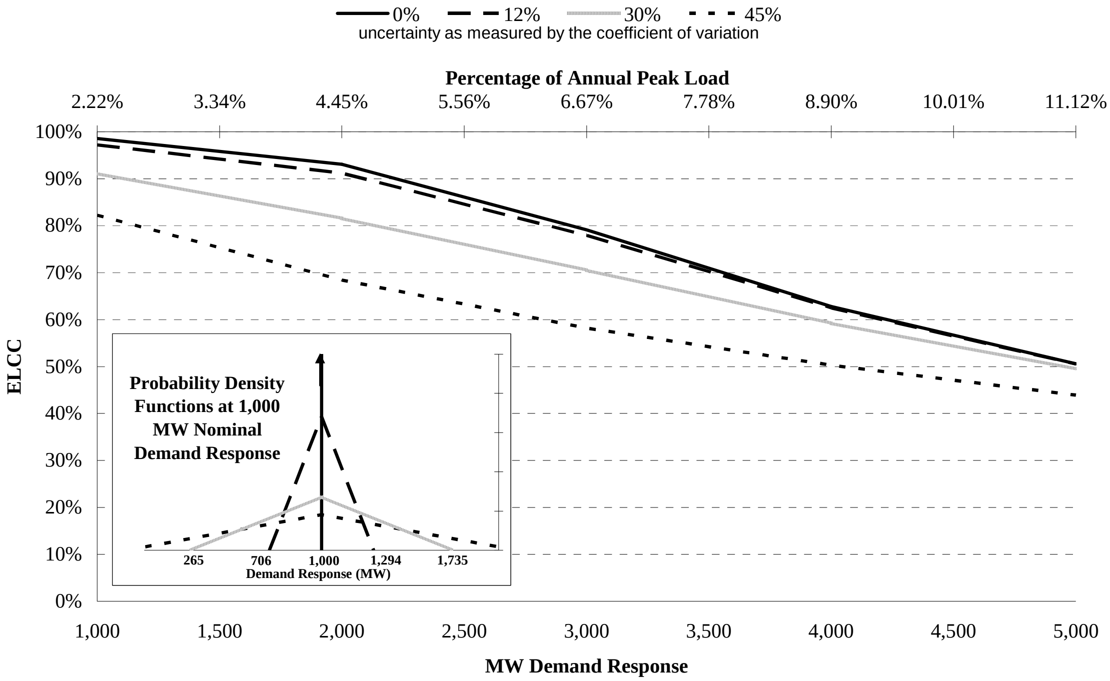


Figure 9 shows the ELCC as a percentage of expected or mean demand response at a series of expected response levels, ranging from 1000 MW to 5000 MW, at different levels of uncertainty of the response as measured by the coefficient of variation. The line with a coefficient of variation of 0% is our basic result with no substitution effect, as reported in Section II above. As shown the Figure 9, a coefficient of variation of 12% makes only a small difference in the reliability effectiveness of the demand response program. At 30% and 45% coefficient of variation, however, the demand response program suffers degradation due to its variability. One of the interesting aspects of the chart is that uncertainty effects start to disappear at higher demand response levels. For example, at 30% coefficient of variation, the demand response program suffers in effectiveness relative to the certain demand scenario (0% coefficient of variation) at relatively low levels of expected demand response (1000 to 3000 MW). At higher levels of demand response, however, the results are close to the certain case. While this sounds good for higher demand response levels, it must be remembered that at those high levels the capacity value of demand response is significantly degraded. One possible explanation is that, as the program size increases, neither the upside of having more demand response by chance helps much nor losing a little hurts much since those are not likely to have high capacity value anyway.

An important conclusion from the examination of uncertain demand response follows if one thinks that the available evidence from the California SPP is an accurate measurement of the potential uncertainty in CPP/PTR type programs. In this situation, the uncertainty of the level of response is likely to have little effect on the capacity and reliability value of these demand response programs. The California SPP evidence suggests a coefficient of variation of between 3 and 9%. At 12% coefficient of variation response shows little difference from certain response as seen in the Figure 9. As with the other results in this paper, while they would seem likely to hold for a broad category of cases, the results have only been simulated on one particular system configuration. Simulations particular to the system in question should be conducted before policy conclusions are made for that particular system.

5. Policy Implications and Conclusions

This paper has examined the capacity impacts of particular types of demand response programs using data from California during 2002. While CPP/PTR programs have constraints in how they are operated, in terms of the number of days and limitations on the hours of operation (*i.e.,* the peak period), they offer benefits by increasing system reliability, and therefore, reducing capacity needs of the electric power system. These benefits, however, decrease substantially as the size of the programs grows relative to the system size. More flexible schemes for deployment of demand response can help address the decreasing returns to scale in capacity value, but more flexible demand response has decreasing returns to scale as well. The apparent good news for demand response from this study is that the little evidence that there is on the uncertainty of the level of demand response suggests that uncertainty does not reach a level that greatly impairs its ability to contribute to reliability. All of these conclusions depend on the types of programs studied and the particular system examined. However, as discussed above, there are good reasons to think that the same conclusions would hold true for other systems as well.

Several potential policy implications emerge out of these conclusions. First, when measuring demand response potential or demand response program cost effectiveness, it is typically assumed that the reliability effectiveness or capacity value of the program does not decrease with added demand response relative to the overall size of the system. This assumption seems unlikely to be true, and the probable decreasing returns to scale may have rather large impacts on overall program potential and its cost effectiveness. As policy makers look towards massive deployment of demand response, these impacts will become increasingly important.

Second, the results on the added capacity value of flexible demand response above that of CPP/PTR type programs suggests that the capacity value in smart appliances, home automation, and the like, perhaps comes in two forms. As enablers of CPP/PTR programs, they will increase the average response rates in the programs by making it easier for consumers on dynamic rates to respond.[23] Automation may also make it easier for more flexible definition of dynamic rates by targeting those demand reduction hours that do not necessarily fall within a pre-defined critical peak window. Both of these effects of automation will tend to decrease as the amount of demand response in a system increases. Policy makers should incorporate these effects and their decreasing returns to scale in their ratemaking decisions and smart grid policies.

Third, there is little evidence available on the uncertainty of demand response from mass market programs such as CPP/PTR. As we demonstrate in our simulations, uncertainty at fairly high levels can have a dramatic impact on the capacity value. While the small bit of evidence from the California SPP suggests that the level of uncertainty is not great enough to materially decrease the capacity value of mass market demand response, more evidence is needed to better understand the uncertainty of demand response.

[23] In this regard, it is possible to imagine that automation might decrease the uncertainty of demand response by removing the behavioral component in the response other than the initial system configuration.

References

Arizona Public Service Company, *Resource Plan Report*, January 29, 2009.

R. Billinton. and R. Allan. *Reliability Evaluation of Power Systems*. New York: Plenum Press, 1984.

J. Bonbright, A. Danielsen, D. Kamerschen, *Principles of Public Utility Rates*, Public Utility Reports, Inc., Arlington, Virginia, 1988.

G. Calabresse, Generating Reserve Capacity Determined by the Probability Method, AIEE Trans., 1947, 66 at1439–1450.

P. Centolella and A. Ott, "The Integration of Price Responsive Demand into PJM Wholesale Power Markets and System Operations", March 9, 2009, available at http://www.hks.harvard.edu/hepg/Papers/2009/Centolella%20%20Ott%20PJM%20PRD%2003092009.pdf.

D. York and Martin Kushler, "Exploring the Relationship between Demand Response and Energy Efficiency: A Review of Experience and Discussion of Key Issues", American Council for an Energy-Efficient Economy, Report Number U052, arch 2005.

R. Earle, S. Newell, A. Faruqui, Attila Hajos, Ryan Hledik, *Fostering Economic Demand Response in the Midwest ISO*, December 30, 2008, report prepared for the Midwest ISO.

A. Faruqui and L Wood, "Quantifying the Benefits of Dynamic Pricing in the Mass Market", Edison Electric Institute, January 2008.

L. Garver, "Effective Load-Carrying Capability of Generating Units," *IEEE Transactions on Power Apparatus and Systems* v. PAS-85, no. 8 (1966) 910-919.

IBM, "Ontario Smart Price Pilot Final Report", Ontario Energy Board, July 2007.

M. Jamali, "Generating Capacity Reliability Indices", MSc Thesis, University of Saskatchewan, 1979.

E. Kahn, "Effective Load Carrying Capability of Wind Generation: Initial Results with Public Data", *The Electricity Journal*, December 2004.

Federal Energy Regulatory Commission, 125 FERC ¶ 61,071, Order 719, Wholesale Competition in Regions with Organized Electric Markets, Final Rule, October 17, 2008.

Federal Energy Regulatory Commission, *Assessment of Demand Response and Advanced Metering*, Staff Report, December 2008.

R. Letzler, “Applying Psychology to Economic Policy Design: Using Incentive Preserving Rebates to Increase Acceptance of Critical Peak Electricity Pricing”, Center for the Study of Energy Markets, Working Paper 162, 2007.

North American Electric Reliability Corporation (NERC), *Data Collection for Demand-Side Management for Quantifying its Influence on Reliability*: *Results and Recommendations*, December 2007.

Summit Blue Consulting, “Evaluation of the 2005 Energy-Smart Pricing Plan”, Community Energy Cooperative, August 1, 2006.

Inset Box 1: Reliability Indices

Perhaps the most basic probabilistic reliability index is the loss of load probability (LOLP) index introduced by Calabresse in the 1940s.[1] By LOLP we mean the probability of an outage in a given hour. That is, the probability that in a given hour the amount of available generating capacity is less than load. The loss of load expectation (LOLE) as defined in the main text is the expected number of hours of outage in a given year. This is simply the sum of the LOLP for each hour over the number of hours in a year.[2]

The effective load carrying capacity (ELCC) reliability measure was developed in order to measure the reliability impacts of units.[3] ELCC is the amount of new load that can be added to a system after a new unit is added while keeping the same level of reliability as measured by LOLE.[4] So, ELCC measures the capacity contribution of a unit to a power system.

It is important to note that the ELCC is both dependent on the unit itself as well as the system to which it is added. For example, units that are relatively large compared to the system as a whole will have a lower ELCC than units whose capacity make up a smaller percentage of total system capacity other things being equal. The capacity contribution of thermal units is often thought of as being equal to their derated capacity where the derated capacity is the full unit capacity times one minus the forced outage rate.[5] The ELCC of a thermal unit, however, can be less than the unit's derated capacity. As a result, comparing the derated capacity of thermal units to each other is not a completely fair comparison of the contribution to reliability of each unit. More to the point for this discussion, comparing the ELCC of demand response (or other resources to the derated capacity of thermal units) will tend to make demand response appear to contribute relatively less to reliability than thermal units. An alternative that perhaps could provide more direct comparison for demand response would be to calculate the number of combustion turbine equivalents (ETCs). In other words, in order to achieve the same level of LOLE as the resource added, how many combustion turbines would have to be

---

[1] Our terminology differs a little bit from the original terminology in that LOLP originally meant the number of days per year of expected capacity shortages. More precisely speaking, this is an expectation rather than a probability, so we use the term LOLP, as others do, to refer to the probability of outage in a given hour.

[2] If for hour i we write the loss of load probability as $LOLP_i$, then LOLE can be written as $LOLE = \sum_{i=1}^{8760} LOLP_i$. And, $LOLP_i = \Pr(\sum C_j < L_i)$ where C_j is the random variable of capacity of generating unit *j* in hour *i* and L_i is the load in hour *i*.

[3] Garver (1966).

[4] If *C'* is the random variable of new capacity, then *ELCC* is the MW of new load added to each hour such that $LOLE = \sum \Pr(\sum C_j + C' < L_i + ELCC)$ where *LOLE* is the original loss of load expectation before adding the new unit *C'*.

[5] In other words, a thermal unit with capacity of 100 MW and a forced outage rate of 5% (or .05) is credited with providing 95 MW of reserves.

added to the system. A drawback of this measure is that it is even more system specific and dependent than ELCC because the appropriate combustion turbine to use as the equivalent will vary in size and forced out rate depending on the system.

It is sometimes objected that LOLP and LOLE are not good measures of system reliability because they are only measures of the probability of outage or the expected amount of time with outages. An outage of 100 MW over an hour has a greater reliability impact than an outage of 10 MW. Even if the 100 MW outage has half the probability of the 10 MW outage, the 100 MW outage should be given more weight than the 10 MW outage. While the LOLP and LOLE measures do not do this, an alternative measure, the expected unserved energy (EUE) weights the size of the outage with the probability of the outage to give a measure that takes this into account. Simply put, the EUE is the average number of MWh of outage one could expect to incur.

ELCC can then be redefined in terms of EUE rather than LOLE, so than the ELCC could be defined as the amount of load that can be added in each hour until the EUE is the same as before the resource was added to the system. In principle this method using EUE can result in a different answer for ELCC than when using LOLE. For the system we have studied, however, using EUE rather LOLE results in very little difference in the results, so we have only reported the results using ELCC as defined with LOLE. Moreover, LOLE and EUE are closely related in that LOLE is the marginal rate of increase in EUE.[6]

[6] That is, if the load increased by a small increment in every hour, the LOLE gives the resulting rate of increase in the EUE.