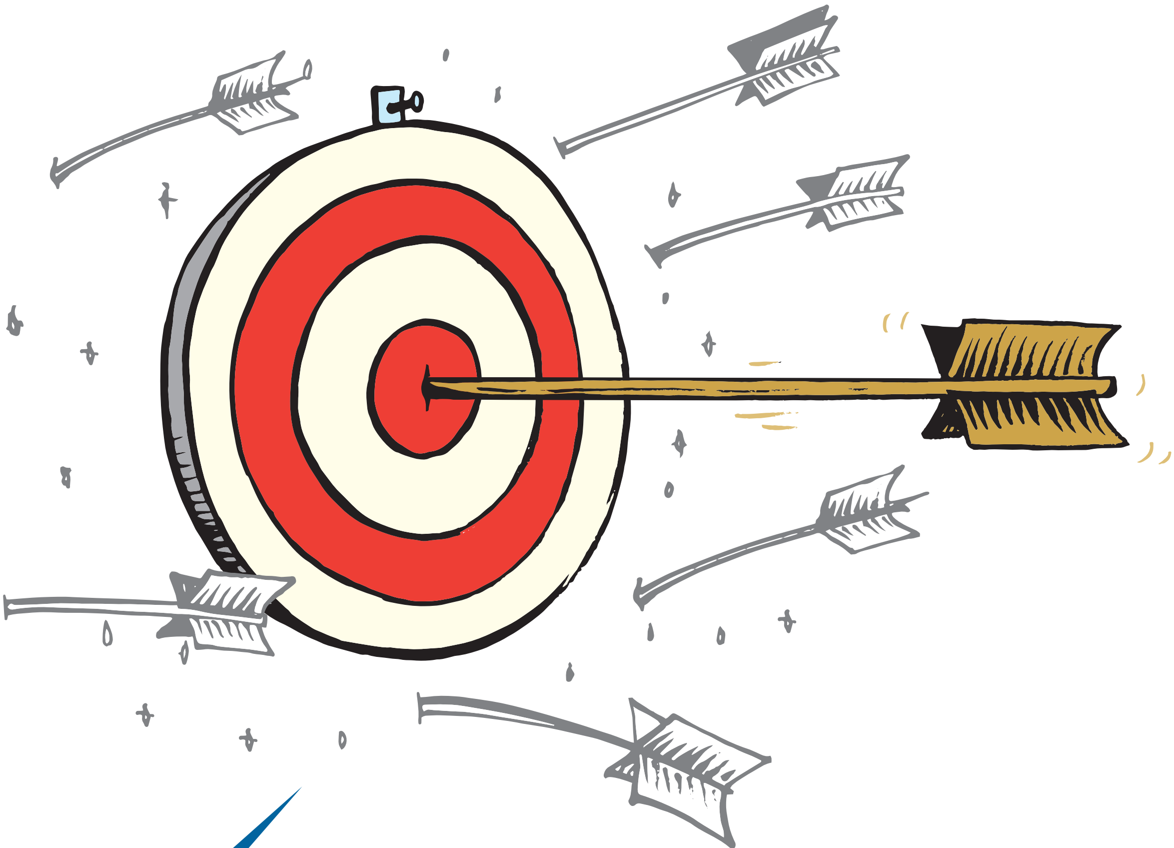
Hitting the Mark on Missing Money:

How to Ensure Reliability at Least Cost to Consumers

Author

Michael Hogan



Electronic copies of this paper and other RAP publications
can be found on our website at **www.raponline.org**.

To be added to our distribution list,
please send relevant contact information to
info@raponline.org.

September 2016

RAP®

Executive Summary

Getting the formation of prices in wholesale electricity markets right is key to ensuring reliability, delivering value for money, and empowering and protecting consumers. Yet many of the measures proposed to address what is known as the "missing money problem" instead create a new problem: misallocated money, overcompensating some resources and undercompensating others. The consequences of this misallocation put the business case for low-carbon power system innovation at risk, a particular concern at this time of transformation in the sector. This paper offers a brief refresher on how we should expect energy prices to form in a modern system, the ways in which they should be expected to shape critical investment decisions, and some of the ways energy price formation can go wrong. With this as a foundation, author Michael Hogan lays out a robust and sustainable approach to ensuring a reliable, low-carbon electric supply at the lowest reasonable cost.

The causes of "missing money" include failing to properly value the demand for balancing requirements, administrative measures (such as price caps) intended to rein in market power, and beneficial public policy measures whose design does not account for any price distortion effects. To tackle the problem effectively, regulators have three options. Top priority should be given to redressing the root causes of the missing money directly. Because this will take time, however, policymakers can reinforce their efforts by adopting administrative mechanisms that add missing money back into energy and balancing services markets. These two strategies, deployed in tandem, offer the best chance to ensure reliability at least cost. A capacity remuneration mechanism, which compensates investors in capacity resources outside the energy and balancing markets, is a third-best option. If resorted to, it should be designed as much as possible to recognize the higher relative value of more flexible resources; it should be accompanied by a thorough reform of the process for assessing the amount of capacity really needed to 'keep the lights on' in accordance with the established standard; and it should be a supplement to, rather than a substitute for, measures to improve the quality of energy price formation, with the ultimate objective that at some point in the future it will no longer be needed.

Table of Contents

Introduction . . . 3

Energy Price Formation and Its Role in Investment . . . 3

The Demand for Flexibility and the Role Energy Price Formation Plays . . . 6

Identifying Causes of Missing Money . . . 9

Failure to Properly Value the Demand for Balancing Requirements . . . 9

Mitigating Market Power in the Absence of Effective Competition and Demand Response . . . 9

Politics, Public Policy, and Prices . . . 11

Miscellaneous Sources of Price Distortion . . . 11

A Strategy for Fixing Missing Money Problems . . . 11

Redress the Root Causes of Missing Money . . . 12

Administrative Mechanisms for Energy and Balancing Services Market Pricing . . . 12

Administrative Mechanisms that Operate Outside of the Energy and Balancing Markets . . . 14

Which Option is Less 'Administrative'? . . . 14

Conclusions . . . 15

Acronyms

CRM	Capacity Remuneration Mechanism
ISO	Independent System Operator
kWh	Kilowatt-Hour
MWh	Megawatt-Hour
RES	Renewable Energy Systems
RTO	Regional Transmission Operator
VoLL	Value of Lost Load

Table of Figures

Figure 1: Most Capital-Intensive Industries Recover Fixed Costs in Markets Based on Unit Prices . . . 3

Figure 2: Illustrative Peak Period Supply/Demand Curves – Three Scenarios . . . 6

Figure 3: Weekly Average Energy Prices (2012) in Markets With and Without Out-Of-Market Capacity Remuneration Mechanisms . . . 7

Figure 4: Average Annual Wholesale Prices (2015) in Five U.S. ISO/RTO Markets . . . 8

Figure 5: Traditional Peak-Shaving vs. a Summer Week in a High-RES System . . . 8

Figure 6: Generating Capacity (GW) in the European Union . . . 10

Figure 7: Illustrative Example of Administrative Reserve Shortage Pricing Curve . . . 12

Figure 8: ERCOT Summer Loads & Resources . . . 13

Figure 9: PJM Capacity Demand Curve for Delivery Year 2016–2017 . . . 14

Table of Tables

Table 1: Representative Rank Order of Marginal Costs (Excluding Price Responsive Demand) . . . 5

Introduction

It has become commonplace to hear it said that wholesale electricity markets are plagued by a "missing money problem." These markets, often referred to as "energy-only" markets, have encountered challenges in practice. So long as these challenges persist it is possible that investors in resources needed to satisfy demand for a reliable supply may, in some cases, fail to earn what they would expect to earn were the market working as expected. This can in turn impede needed investment, a phenomenon often referred to as the missing money problem.

> **Many of the measures proposed to replace missing money create a new problem: misallocated money, overcompensating some resources and undercompensating others.**

That said, there is no single approach to meeting the demand for reliability. Experts agree that a growing share of variable renewable resources increases the value of flexibility elsewhere in the system,[1] value that can only be seen clearly in prices reflecting real-time conditions in the wholesale electricity market. Yet many of the measures proposed to replace missing money operate outside of that market, on different time scales and using different parameters. They dilute and thus subvert the unique role energy prices can and should play in shaping investment in the resources needed to satisfy demand for reliable supply *at the lowest reasonable cost.* As a result, in trying to replace missing money they create a new problem: *misallocated money,* i.e., overcompensating some resources and undercompensating others.

The consequences of misallocation extend far beyond simple questions of fairness between groups of investors. It can create structural incentives to invest in a mix of resources ill-suited to the underlying needs of the system, particularly a low-carbon power system. It can obscure the true value of energy storage and flexible demand as supply becomes less controllable. As a result, the business case for innovation can be seriously compromised and consumers can face significantly higher costs for reliability. This risk is particularly acute now, given the fundamental transformation under way in the sector.

"Keeping the lights on" is about more than just investment in generating capacity; it is also about delivering value for money. It is about empowering and protecting consumers. Getting the formation of prices in wholesale electricity markets right remains a key to tying these pieces together. This paper offers a brief refresher on how we should expect energy prices to form in a modern power system, the ways in which they should be expected to shape critical investment decisions, and some of the ways energy price formation can go wrong. With this as a foundation, the paper lays out a robust and sustainable approach to ensuring a reliable, low-carbon electric supply at the lowest reasonable cost.

Energy Price Formation and Its Role in Investment

With the transition of wholesale electricity from a natural monopoly to a naturally competitive industry, the competitive wholesale electricity markets introduced in parts of Europe, North America, and Australia/New Zealand during the past 30 years were conceived of as true markets in which the wholesale price of electricity

Figure 1

Most Capital-Intensive Industries Recover Fixed Costs in Markets Based on Unit Prices

1 For a comprehensive reference, see: International Energy Agency. (2014). *The Power of Transformation – Wind, Sun and the Economics of Flexible Power Systems.*

is the price at which the quantity of supply willing to sell matches the quantity of demand willing to buy (the "market clearing price"). The current and expected price of electricity was meant to be the principal basis for decisions regarding investment, production, and distribution.[2] That is, wholesale electricity markets were, in principle, meant to be no different from any other commodity market (see Box 1).

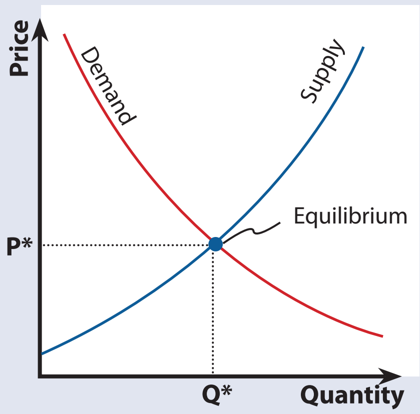
In reality, few commodity markets live up to the theoretical ideal, which is why various forms of regulatory and administrative intervention can be appropriate. The more important the commodity, the more important it is that the integrity of the market be reinforced by judicious interventions. Electricity is both an especially important commodity and one that has historically exhibited its own particular set of challenges.

First, electricity is more difficult and more expensive to store than most commodities. As a result, electricity markets are more susceptible to being manipulated by withholding production. This puts a high premium on reducing market concentration, and it means the competitiveness of the market must be monitored and enforced as close to real time as possible. In many cases, the response to these challenges has been to impose direct and indirect price controls in an attempt to mitigate possible market power abuses. This both undermines legitimate price formation and simply postpones the necessary work of ensuring competitive markets, because no market, however constituted, can function without effective competition.

Second, even when wholesale prices are allowed to reflect actual market conditions, demand for electricity has tended to be "inelastic" – relatively unresponsive to higher costs during shortages or to greater opportunities during surpluses. Although some fraction of load does place a very high value on uninterrupted service, the inelasticity of demand is more generally attributable to the common practice of non–time-varying retail pricing. This practice

Box 1

How Commodity Markets Work



When supply is tight relative to demand, high-cost sellers can clear the market, raising the clearing price. During *acute shortages,* marginal sellers may even be able to bid prices above their short-run production costs. In a normal commodity market, this "pricing power" is limited in the first instance by competition among high-cost suppliers, substitutes, and the threat of competitive entry, and ultimately by buyers' ability to choose not to buy if the price exceeds the value they place on the commodity. Conversely, when demand is low relative to supply, the lowest-cost suppliers can price others out of the market, lowering the clearing price. In *acute surpluses* buyers may even be able to bid prices below the short-run production costs of the lowest-cost sellers. However, in a properly functioning market, suppliers rely on the fact that extreme downward pressure on prices is counterbalanced by opportunistic buyers who can usefully increase their purchases as prices decline.

In a healthy commodity market, a seller or group of sellers should not be in a position to create and then benefit from artificial short-ages, nor should a buyer or group of buyers be in a position to create and benefit from artificial surpluses.[3] Buyers and sellers in such markets typically seek to manage their exposures to risks associ-ated with supply shortages and surpluses by entering into mutually beneficial contracts, either directly or through trading exchanges. Such undertakings typically constitute the principal support for investment.

2 See: Joskow, P. (2008). Lessons Learned from Electricity Market Liberalization. *The Energy Journal.* Special Issue. "The overriding reform goal has been to…ensure that an appropriate share of [societal] benefits are conveyed to consumers through prices that reflect the efficient economic cost of supplying electricity and service quality attributes that reflect consumer valuation." (p 11).

3 When sellers are in a position to create and benefit from artificial shortages, it is referred to as a *monopoly* or, in the case of a group of sellers, an *oligopoly.* When buyers are in a position to create and benefit from artificial surpluses, it is referred to as a *monopsony* or *oligopsony.*

arose, in turn, from the combination of monopoly retail franchises and the historical impracticality of serving any but the largest individual customers selectively based on their willingness to pay. The result has been that virtually all loads in a given area are served for the same price in shortage hours and in surplus hours, until none of them are served.[4] Consequently, the demand impacts of fluctuating wholesale prices have played out only in longer time frames, if at all.

The fact that demand has been relatively inelastic does not mean consumers place an unlimited value on reliability. The value of continuous service (often referred to as the "value of lost load" or VoLL) varies widely depending on the energy service in question, from near zero (for example, when charging an electric vehicle at 2 AM) to tens of thousands of euros per megawatt-hour (MWh) (say, at a hospital). Cost-effective, convenient options for consumers to act on that range of values are expanding rapidly, but for now we continue to rely principally on standards set by public officials that impute a single reliability value for all loads. These standards typically impute a value of tens or even hundreds of thousands of euros per MWh that consumers would be willing to pay to avoid an interruption of service.

System operators apply that value by acting, in effect, as the buyer and seller of last resort, procuring the various types of balancing resources needed to ensure the system meets the public reliability standard. They are procured from the same pool of resources used to meet the demand for energy. Prices should reflect the full marginal cost of all actions required to balance supply and demand. When the demand for energy begins to eat into what is needed in reserve to meet the reliability standard, the true marginal cost of a kilowatt-hour (kWh) of energy includes the cost system operators should be willing to incur to reserve resources, or should charge to release resources to meet the rising demand for energy.

In this way the wholesale energy market is meant to reflect not only the short-run marginal cost of energy sold in the energy market, but also the marginal cost of all actions required to meet the demand for reliable energy, including the cost of activating controllable demand-side options[5] (see Table 1). When supply margins are tight, the demand for energy and balancing resources can drive marginal costs well above the variable cost of the last kWh sold in the forward market. Multiple studies support this finding, even in a system with very high penetration of variable renewables.[6] This in turn creates the economic window for the growing suite of opportunities for consumers to play their role in balancing supply and demand.

Table 1

Representative Rank Order of Marginal Costs

(Excluding Price Responsive Demand)

System Resource	Full Marginal Cost (€/MWh)
Generation capacity	20-250
Imports	20-1,000
Secondary (operating) reserves	250-5,000
Emergency generation	500
Primary (regulation) reserves	500-9,000
30-minute responsive back-up	1,400
30-minute controllable demand response	2,400
10-minute controllable demand response	2,600
10-minute responsive back-up	3,700
Emergency load-shedding	9,000

Adapted from: Pfeifenberger, J., Spees, K., Carden, K., & Wintermantel, N. (2013). Resource Adequacy Requirements. The Brattle Group; and Newell, S., et al. (2014). Estimating the Economically Optimal Reserve Margin in ERCOT. The Brattle Group.

Figure 2 illustrates demand and market price impacts under three different scenarios during a period of tight supply margins in a hypothetical system (resulting from either peak demand or from the unavailability of a significant amount of generation). In Scenario a, the demand for balancing reserves is not reflected in market clearing prices, the marginal costs of "emergency"

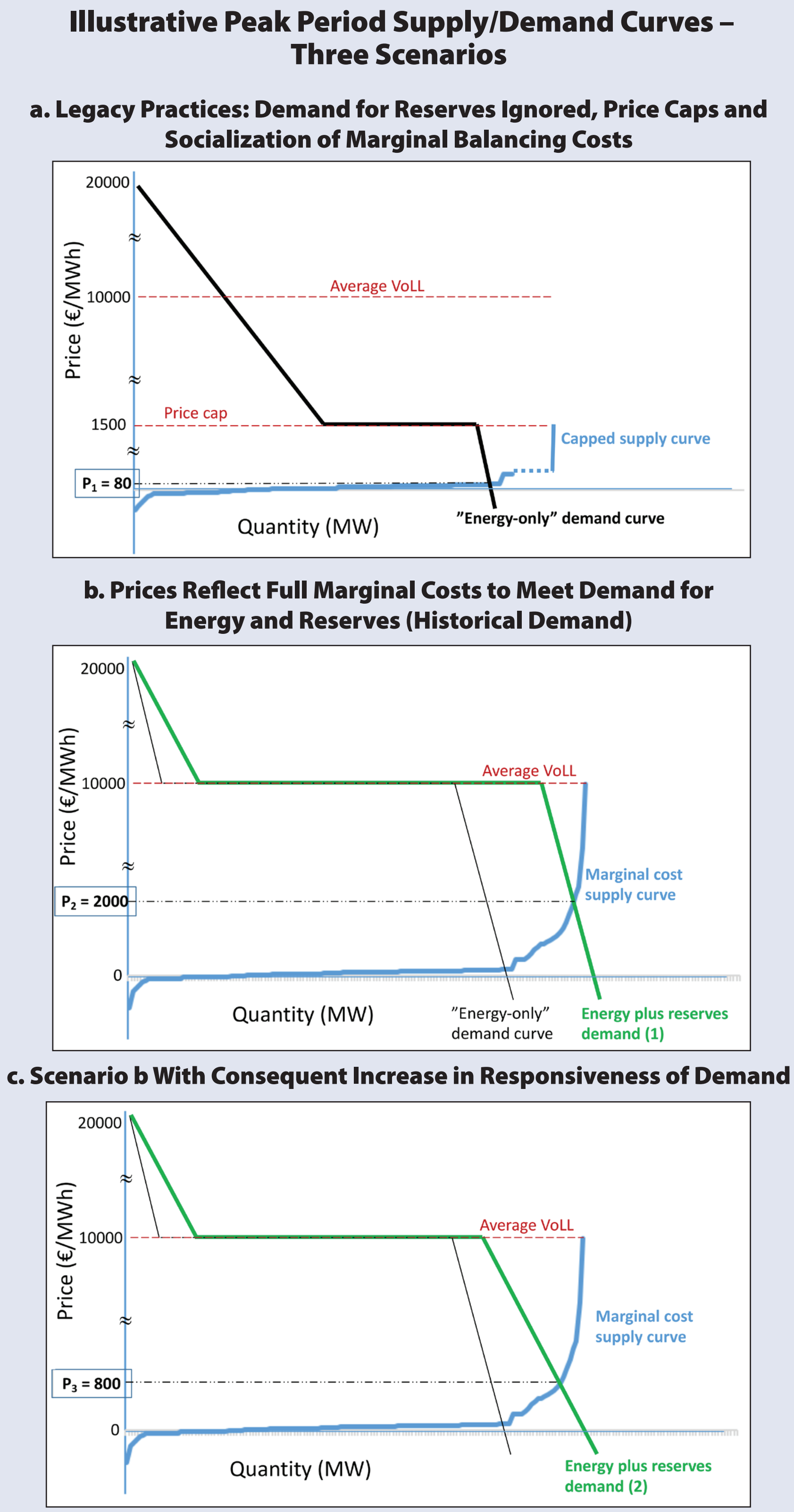
4 It is this characteristic, referred to as "non-excludability," that has led to the treatment of electricity as a "public good." Technology is rapidly eroding the non-excludability of electricity.

5 In essence this describes the difference between a market based simply on "economic dispatch" (as commonly misconstrued) and the actual design basis of "security-constrained economic dispatch."

6 See: McKinsey & Co. et al, for ECF (2011). *Power Perspectives 2030: On the Road to a Decarbonized Power Sector in Europe.* See also: Energy Analyses. (2016). *Price Formation in Power Markets Dominated by Low OPEX Technologies.* This unpublished study for DONG found no basis to expect wholesale price collapse in high-RES markets where demand for both energy and balancing reserves is fully reflected in prices.

resources available to the system operator are socialized or ignored, and the price of supply is capped below the average value of lost load. In Scenario b, the marginal costs of all balancing actions are reflected in the supply curve, the price cap has been lifted to the average value of lost load, and the demand curve now reflects the full demand for both energy and reserves. The true marginal cost of meeting historical demand for reliable energy under typical peak-period conditions is now reflected in the clearing price. Scenario c illustrates the moderating impact of investments in greater demand responsiveness when prices are allowed to reflect the full cost of "keeping the lights on," investments that now have a visible business case.

Figure 2a, b, and c



Adapted from: Hogan, W. (2005). On An 'Energy Only' Electricity Market Design for Resource Adequacy. Cambridge, MA: Harvard University.

In practice, most wholesale electricity markets have yet to exhibit the intended pricing behavior (Scenario c). This is largely attributable to two issues: the imposition of various forms of price controls – a well-recognized issue – and a failure to update balancing market practices so that the full marginal cost of balancing services is properly reflected in energy market clearing prices – a much less well-recognized issue.[7] These issues are addressed later.

The Demand for Flexibility and the Role Energy Price Formation Plays

The previous section addressed the question of how energy market price formation is meant to drive investment in the *quantity* of capacity resources needed to meet established reliability standards. But when money goes missing or is diverted from energy price formation, what else is lost, other than a quantum of remuneration? The answer, in a word, is flexibility.

If there is one thing about a decarbonized power sector on which nearly everyone agrees, it is that it will need to become more flexible. What that means in practice is that as the

7 Historical technical and economic barriers to responsive demand are falling rapidly. The key remaining barriers are in retail tariff design, including adoption of time-varying retail pricing, a discussion of which is beyond the scope of this paper.

share of variable sources of energy grows, there will be growing value in the ability of other components of the power system – demand, transmission, distribution, more controllable generating resources – to respond in a timely manner both when there is a lot of energy available at a given time or in a given place, and also when there is only a limited amount of energy available.

By far, the most reliable way to determine what kind of flexibility is needed, when, and what it's worth is by revealing the value of energy during periods and in locations where it's plentiful, and during periods and in locations where it's dear. That differential is, by definition, the value of flexibility. Ignoring energy price distortion or trying to replace money missing from the energy market without restoring the variability in the value of energy from one hour to the next or between one transmission-constrained area and another, creates a different kind of missing money – money missing from the remuneration of investment in more flexible resources.

By far, the most reliable way to determine what kind of flexibility is needed, when, and what it's worth is by revealing the value of energy during periods and in locations where it's plentiful, and during periods and in locations where it's dear. That differential is, by definition, the value of flexibility.

Figure 3 charts weekly average wholesale energy prices in six regional markets. The price swings evident in the markets on the left make visible the value of investment in flexible resources in a way that the markets on the right have struggled to replicate. Furthermore, a wider range of allowable short-term prices can lead to lower, not higher average energy prices to consumers, even before adding back the cost impact of the capacity remuneration mechanisms, as strongly suggested by Figure 4 (which compares three of the markets presented in Figure 3). Although a number of factors account for differences in average prices in these markets, the end result is consistent with the conclusion that markets that reflect the full marginal cost of reliability during periods of resource shortage and surplus, and therefore the full value of resource flexibility, can also deliver comparable reliability at lower cost to consumers.

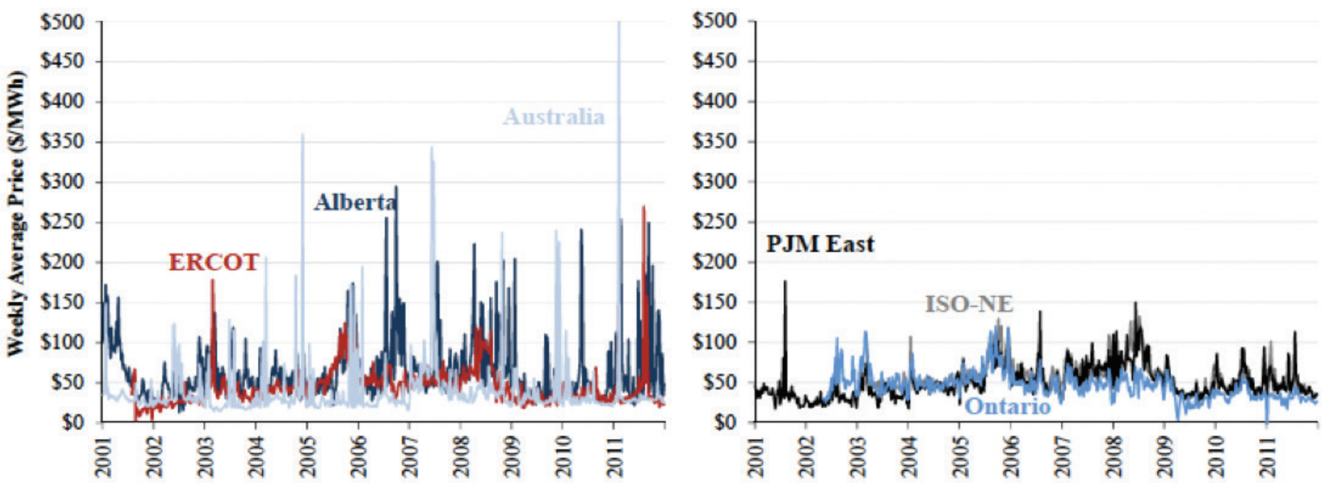
Valuing flexibility was useful but hardly mission-critical when systems were dominated by largely controllable production sources, but that is set to change dramatically. As the share of less controllable renewable supply continues to grow, it will become increasingly challenging and unnecessarily costly to meet established expectations for reliable supply without a significant increase in system flexibility. At the same time, there are different forms of flexibility and different strategies for providing them. Some options can be quite expensive and may not turn out to be cost-effective. Without visibility to the value of investment in different kinds of flexibility, there is little incentive to innovate more cost-effective options for providing them.

Unlike capacity, flexibility is not an easily quantifiable product. It is relatively straightforward to identify a

Figure 3

Weekly Average Energy Price (2012) in Markets

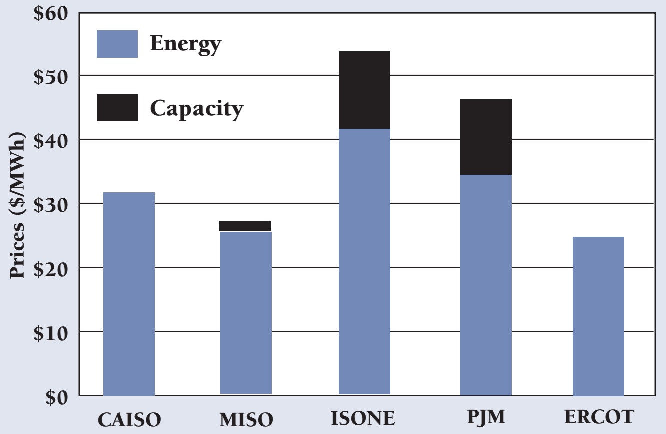
With (left) and Without (right) Out-Of-Market Capacity Remuneration Mechanisms



Source: Ventyx & Australia Electricity Market Operator data, figures reproduced from Newell, S. et al, (2012). ERCOT Investment Incentives and Resource Adequacy. The Brattle Group.

Figure 4

Average Annual Wholesale Prices (2015) in Five U.S. ISO/RTO Markets



Source: Northbridge Group

MW of generating capacity or an equivalent amount of interruptible load. It is a little more complicated but nonetheless feasible, even with variable renewable capacity, to determine a resource's "capacity value" (or "firm capacity") by assessing the probability it will be available when most needed. Even as the power system is decarbonized, the specification for a unit of firm capacity will remain unchanged. But what is a unit of flexibility?

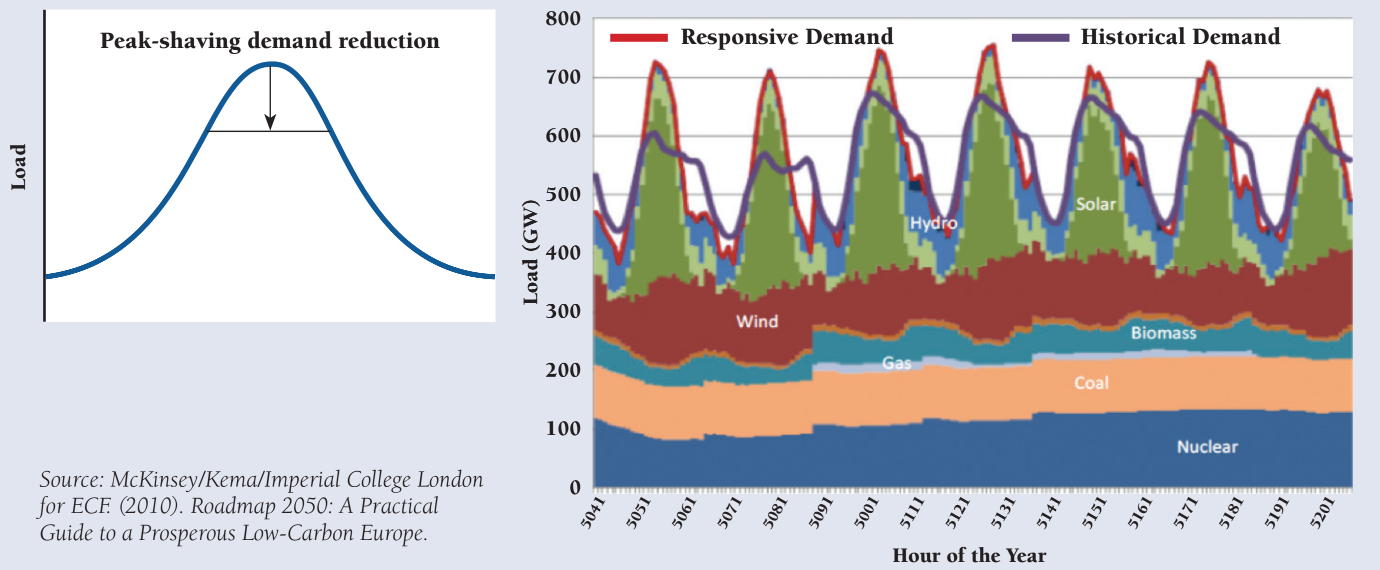
As the share of variable resources grows, temporal and spatial patterns of variability in the value of energy will change over time. Exactly how they will change will depend on which pathways are chosen and how the technologies evolve. If one chooses to establish a separate remuneration mechanism for "flexibility," what exactly should it be designed to remunerate? How does one ensure value for money in committing to pay years in advance for a given amount of a particular kind of resource capability? An alternative is to subdivide capacity remuneration mechanisms (CRMs) based on resource capabilities, yet in addition to the challenge of determining how to subdivide them, proponents of CRMs consistently maintain, with good reason, that doing so to any significant degree would reduce liquidity and create an unacceptable risk of market power abuse. We are still struggling, a decade on, with how to ensure value for money in the design of CRMs for a product as comparatively simple and static as firm capacity. The challenge of doing so for a product as varied and mutable as flexibility presents a far greater challenge.

Demand response was traditionally a one-dimensional source of flexibility (load reduction), typically limited in when and how often it could be used. New opportunities are multiplying rapidly for demand to become more responsive, for instance in the controlled charging of electric vehicles and in grid-integrated water and space heating. The value of that sort of demand response does not lie in reducing demand a few times a year when demand is reaching the limits of the system. It lies in the ability to *decrease or increase* demand as frequently as every day, sometimes for only a matter of seconds, and at other times shifting electricity deliveries by hours. As illustrated in Figure 5, flexible demand will be increasingly valuable but will need to be increasingly multidimensional.

New sources of flexibility (both supply- and demand-

Figure 5

Demand Flexibility: Traditional (left) vs. a Typical Summer Week in a High-RES System



Source: McKinsey/Kema/Imperial College London for ECF. (2010). Roadmap 2050: A Practical Guide to a Prosperous Low-Carbon Europe.

side) are expected to expand rapidly and to be viable at a wide range of price points. Investors and innovators will need to see a business case for developing these new, more varied strategies for increasing system flexibility. Regulators and consumers will need to see that they are getting value for money. Resources good at providing some kinds of flexibility tend to be not so good at providing others. The dynamism of effective energy pricing represents the most transparent means of signaling what kind of flexibility is needed, revealing what it is actually worth to consumers and bringing forward the most cost-effective responses. Regardless of whether we pursue a low-cost pathway to a low-carbon power sector will depend on how well we cultivate or replicate that functionality.

Identifying Causes of Missing Money

There are a number of ways money for investment in resources needed to meet the demand for reliability at the lowest reasonable cost can genuinely "go missing" from energy prices. The most common causes are summarized below.[8] The key phrase, of course, is 'needed to meet the demand for reliability' (see box 2).

Failure to Properly Value the Demand for Balancing Requirements

One common problem is a failure to update the traditional approaches used by system operators to procure and charge for reserves and other balancing services. Historical practices effectively subsidized the cost of energy during shortage periods by overcharging for energy during non-shortage periods. This was harmless enough when the industry was a regulated or state-owned monopoly and supply was highly controllable, but it undermines price formation in a market environment and becomes especially problematic as supply becomes less controllable. In a market environment, the real-time value of these services should factor into, although not necessarily set, the market clearing price for energy (as demand flexibility increases, the market can clear at a price below the level implied by the public reliability standard). This should be complemented by reform of the processes by which these services are best procured, for instance through regular short-term auctions, incorporation of non-traditional sources such as demand response, and reduction of minimum bid sizes.

Mitigating Market Power in the Absence of Effective Competition and Demand Response

Where there are concerns about whether the market is sufficiently competitive to prevent abuses, the risk for missing money can also arise from administrative measures intended to correct for or prevent market actors from taking advantage of a dominant market position. Such measures most commonly take the form of caps limiting market prices, although prices can be limited in less obvious ways as well, such as keeping a "strategic reserve" that can enter the market at artificially low prices.[9] In practice, the setting of these caps can be fairly arbitrary, up to several times the level of the highest marginal generation cost but only a small fraction of what the value of energy can be in hours when the market is tight.[10]

It would seem simple enough to recommend that price caps be removed, but in some cases the concerns about the level of competition or the potential for market manipulation may be justified. Markets cannot function without the check of effective competition, with the latitude to form effective energy prices being a principal casualty. Ensuring competition is a non-negotiable prerequisite for the market in general, much less for proper energy price formation. And although the system operator's role as buyer/seller of last resort sets an upper limit on shortage pricing, full confidence in unfettered energy market prices is unlikely until demand has developed a more dynamic capacity to clear the market based on the range of values consumers actually place on continuous service, rather than at an administratively set average.

The history and technical characteristics of the modern electricity industry make ensuring competition challenging, although quite feasible, as has been demonstrated in multiple power markets around the world. It is not something that is easily left to traditional competition authorities who lack the specialized technical

8 See: Pope, S. L. (2014, October). *Price Formation in ISOs & RTOs: Principles & Improvements;* Hogan, W. W. (2014, June). *Electricity Market Design and Efficient Pricing: Applications for New England and Beyond.*

9 It is possible to design a true strategic reserve, one that does not unduly interfere with market price formation.

10 A different sort of "cap" – price floors to limit or prevent negative prices – also causes money to go missing, but specifically from flexibility investments. Distorting surplus prices has the same effect as distorting scarcity prices in obscuring the value of investments in flexibility.

expertise required to assess electricity market behavior. On the other hand, only the more sophisticated, independent, and well-resourced energy regulatory authorities have the requisite capacity. There are good examples of the evolution of effective institutional frameworks and processes for monitoring, reporting on, and acting to ensure fair competition, but in too many markets this has yet to be accomplished.[11]

Unless and until consumers and public authorities have good reason to be confident in the effectiveness of both supply-side market competition and demand-side flexibility, energy price formation can be expected to be compromised by measures to mitigate the risk for market power abuse, creating a risk for missing money. It

Box 2

Low Prices ≠ Missing Money

Public discussions of an esoteric topic like electricity market design can easily obscure the fact that, as in any commodity market, low prices and poor returns on investment usually have a very simple explanation: surplus capacity. In a market oversupplied with capacity, there is virtually always more than enough capacity to meet the combined demand for energy and balancing reserves and "scarcity value" is virtually nonexistent. In this case low prices and poor returns are not evidence of missing money, they are an accurate reflection of the marginal value of capacity. Too often what is claimed to be "missing money" is in reality the difference between what an asset is actually worth and what its owners would like it to be worth.

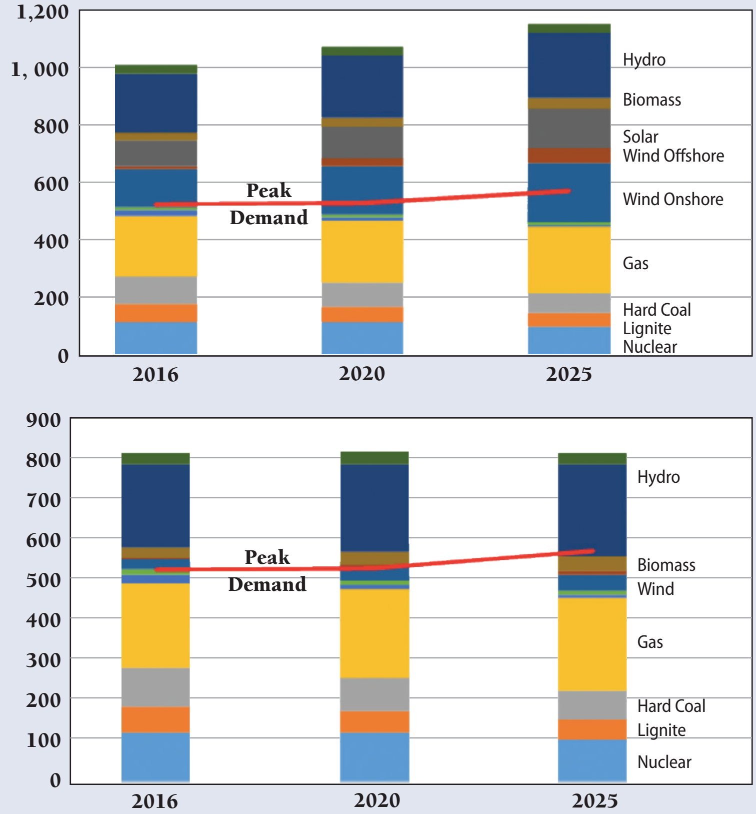
As Figure 6 illustrates for Europe, mature power markets in Europe and North America have indeed built up significant surplus capacity over the past decade or more. This is attributable to several factors, including the Great Recession, increased end-use efficiency, and policy-driven investment in renewables, compounded by various local efforts to protect incumbent industries. Continued improvements in efficiency, declining energy intensity of economic activity, and plans to continue support for investment in low-carbon supply mean the surplus is not expected to disappear anytime soon without concerted efforts to reduce redundant capacity, even after factoring in a conservative discount for the lower reliability "capacity credit" planners are able to attribute to intermittent resources. (In practice this credit varies by location but can be quite a bit higher than depicted here.)

This is why claims of missing money cannot be evaluated without an independent, transparent, and comprehensive framework for assessing the quantity and quality of available resources against a clear set of criteria for what is needed. This is referred to as a "resource adequacy" or "generation adequacy" assessment.

Figure 6

Generating Capacity (GW) in the European Union

(above) All Capacity; (below) Capacity with Zero Credit for Solar, 20% Credit for Wind



Source: ENTSO-e 2016 System Outlook & Adequacy Forecast

11 See, e.g., Keay-Bright, S. (2016). *Can We Trust Electricity Prices?* Brussels: The Regulatory Assistance Project.

must be noted that recent experience suggests[12] that the same dynamic comes into play with CRMs, with real and imagined concerns about competition leading to efforts to limit and undermine capacity market price formation.

Politics, Public Policy, and Prices

Some beneficial public policy measures can have the unintended consequence of distorting energy prices if not accounted for in the design of the policy. One example is public support for investment in new resources in a market that is already fully supplied – or oversupplied – by existing resources. This can include policies to promote renewables, nuclear, or other favored classes of resources. Although such measures can be well justified, for instance in the case of new zero-carbon resource investment and particularly energy efficiency,[13] there are consequences for energy price formation that must be anticipated and redressed. Without countervailing measures – such as an early retirement program[14] – this can force a market into overcapacity and depress market prices. Another (and often related) example is the decision by public officials to support the continued viability of resources that would otherwise be forced out of the market. This includes efforts to protect local generators and measures to subsidize generation that relies on domestic fuel, particularly coal.

These sorts of policy measures are seldom intended to distort energy prices, but they can do so nonetheless. Price formation can also be distorted by policies that have the unintended effect of supporting a suboptimal mix of resources, such as public support for inflexible baseload plants when the market is – or should be – favoring investment in more flexible resources.

Miscellaneous Sources of Price Distortion

Much of the money genuinely missing from energy market prices can be traced to a range of flaws in price formation, large and small, owing to various legacy market rules and vestigial practices carried over from the days of vertical monopolies, closed local systems, and administered cost-of-service pricing.[15] Examples include the exclusion of the cost of calling upon voluntary interruptible loads from the setting of market prices; separately compensating inflexible resources for the costs of responding to dispatch instructions; or failing to allow prices to reflect where there are local surpluses or shortages caused by transmission congestion.

In most cases these issues are known or can be readily identified. Remediation – fixing the problem that gives rise to distorted energy prices in the first place – is largely a matter of willingness and competence. Until this can be accomplished, there is a risk of missing money that may warrant some form of administrative mechanism. However, it must be recognized that remediation of this sort of price distortion is entirely feasible. Given the risk that is always present with administrative mechanisms – that they will lead to suboptimal outcomes compared to what would be achieved with effective price formation – such mechanisms should be supplanted by direct remediation as soon as practicable.

A Strategy for Fixing Missing Money Problems

There are, broadly speaking, three different ways to attack legitimate missing money problems:

1) Redress the root causes of missing money;
2) Implement administrative investment remuneration mechanisms that inject missing money back into energy and balancing services market prices; and
3) Implement administrative investment remuneration mechanisms that compensate investors in capacity resources outside of the energy and balancing services markets.

These need not be mutually exclusive, but which approaches are chosen, and the order in which they are prioritized, may have important consequences.

12 See admonishments to state regulators and legislators in, for example: PJM. (2016, May). *Resource Investment in Competitive Markets.* PJM White Paper; ISO New England. (2015, October). *The Importance of a Performance-Based Capacity Market to Ensure Reliability as the Grid Adapts to a Renewable Energy Future.* ISO New England White Paper.

13 Cost-effective investments in end-use efficiency are constrained by a number of well-documented market barriers. Where efficiency is less expensive or more valuable than generation, policy support for efficiency that reduces generator margins reinforces rather than undermines both market fundamentals and security of supply.

14 See: Buck, M., Hogan, M., Redl, C. (2015). *The Market Design Initiative and Path Dependency.* Regulatory Assistance Project (Brussels) and Agora Energiewende (Berlin).

15 See: Pope, S. L. (2014, October). *Price Formation in ISOs & RTOs: Principles & Improvements;* Hogan, W. W. (2014, June). *Electricity Market Design and Efficient Pricing: Applications for New England and Beyond.*

Redress the Root Causes of Missing Money

The reality is that the conditions necessary for strong price formation are the conditions necessary for the power market to do what power markets are meant to do – deliver reliable electricity *at the lowest reasonable cost*. There can often be valid reasons to choose to elevate some other desirable outcome over the goal of lowest cost. But when weighing options for ensuring reliable electricity at least cost, the first best option, when it is available, is to improve energy price formation, especially as we move to a low-carbon power system.

The most commonly cited causes of missing money have been discussed previously, along with ways in which each of them could be directly redressed. Options range from relatively straightforward measures like modernizing the rules and procedures for procuring and pricing balancing services, to more challenging measures like setting gate closures much closer to real time or implementing locational pricing that reflects actual congestion boundaries, to building a continuous market monitoring and enforcement framework consistent with established best practice, robust enough to provide the confidence needed to relax and eventually remove the various forms of price control that have been and are being adopted.

A related issue in many markets is that deliberate choices to depart from pure market outcomes may create imbalances in the market that call for countervailing measures by a central authority. An example would be overcapacity created by support for investment in targeted resources in an already oversupplied market (see Box 2). An appropriate administrative response may be targeted assistance for the early retirement of resources legitimately stranded as a result.[16] How one intervenes to deal with these issues will determine how closely the results come to a least-cost reliability solution.

An appropriate administrative response may be targeted assistance for the early retirement of resources legitimately stranded as a result. How one intervenes to deal with these issues will determine how closely the results come to a least-cost reliability solution.

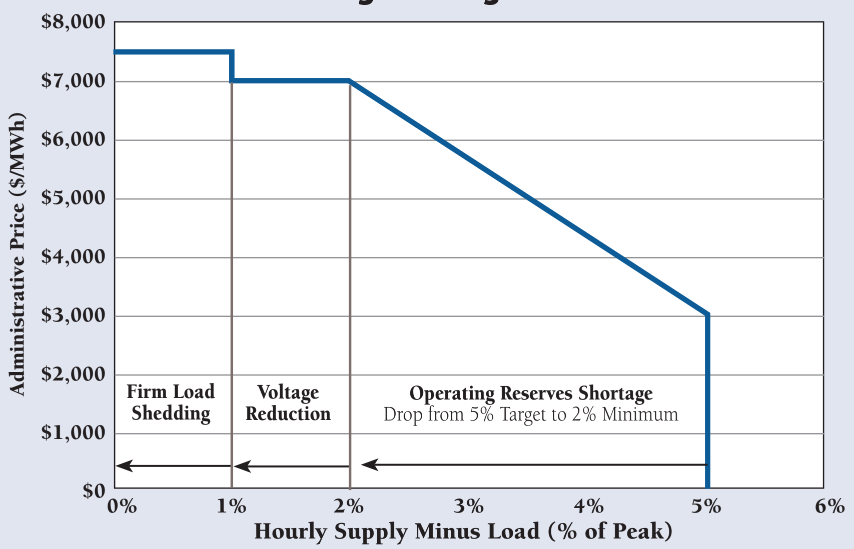
Administrative Mechanisms for Energy and Balancing Services Market Pricing

Although priority should be given to improving energy price formation wherever possible, in most markets this remains at best a work in progress. The practical reality may be that some form of administrative mechanism may be deemed necessary at least on a provisional basis to ensure that the right amount of the right kind of investment is fairly compensated.

Once the decision is made to adopt an administrative mechanism, preference should be given to mechanisms that directly address energy and balancing market pricing. The most widely recognized version is sometimes referred to as "administrative reserve shortage pricing."[17] This approach leverages the central administrative role the system operator currently plays in virtually all energy markets acting on behalf of consumers'

Figure 7

Illustrative Example of Administrative Reserve Shortage Pricing Curve

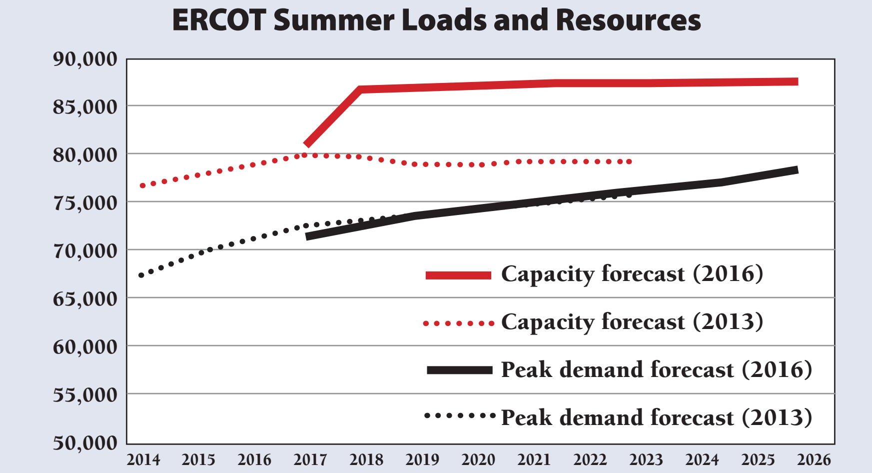


Source: Brattle Group/Astrape Consulting

16 In some instances, assets are stranded as a result of government action not reasonably foreseeable by management, and it may be fair for the public to bear some of the cost. In others, it is the result of poor management decisions, and shareholders should bear the costs. Often it is a mix of both, and sorting it out can be a complex affair. We take no position on that question here, but rather focus only on the need for retirement.

17 In perhaps the most well-known example, the ERCOT market in Texas, this measure is referred to as an "Operating Reserve Demand Curve."

Figure 8



Source: ERCOT (Capacity, Demand, and Reserves Reports, May 2013 and May 2016)

interest in reliable service. As system operators look to position required balancing resources for the next day or the next few hours, this approach sets the price they are willing to pay for reserves. To do so they use a demand curve that tracks the extent to which reserves available in future balancing intervals are expected to fall short of what is needed[18] (see Figure 7). If market prices do not already reflect the system operator's demand for balancing resources, they are adjusted up (or down) to a level that is administratively determined to do so.

By ensuring a price signal that reflects the full cost of meeting the demand for energy and reliability during any given dispatch interval, this mechanism mimics price formation in a fully functional energy market. It affords market participants the opportunity and the incentive to respond before shortages become acute. As the risk for higher and more volatile prices rises in proportion to the need for new investment, it spurs the growth of commercial risk management activities, such as bilateral long-term contracts, which in turn provide a business case for needed investment. In giving greater visibility to temporal swings in the value of producing energy it reveals more efficiently than alternative out-of-market measures the value of investing in resources (including "smart demand" technologies) better suited to responding to frequent swings in the availability of variable generation.

Versions of this approach have been adopted in a number of jurisdictions. In Europe, a version was introduced in the Great Britain market as part of the Electricity Balancing Significant Code Review, with the hope that it will eventually make redundant the recently adopted capacity market. In May 2014 the ERCOT market in Texas implemented a reserve shortage pricing mechanism as its principal administrative mechanism, so far with good results (see Figure 8). It was recently adopted by the PJM market operator alongside their existing capacity auction as part of an overall effort to reform market price formation. An early version (the "Capacity Payment") was introduced in the late 1980s in the England & Wales Pool. Implemented in conjunction with a robust market monitoring regime, an administrative reserve shortage pricing mechanism also affords system operators a tool for mitigating market power abuse when shortages develop.

Administrative Mechanisms That Operate Outside of the Energy and Balancing Markets

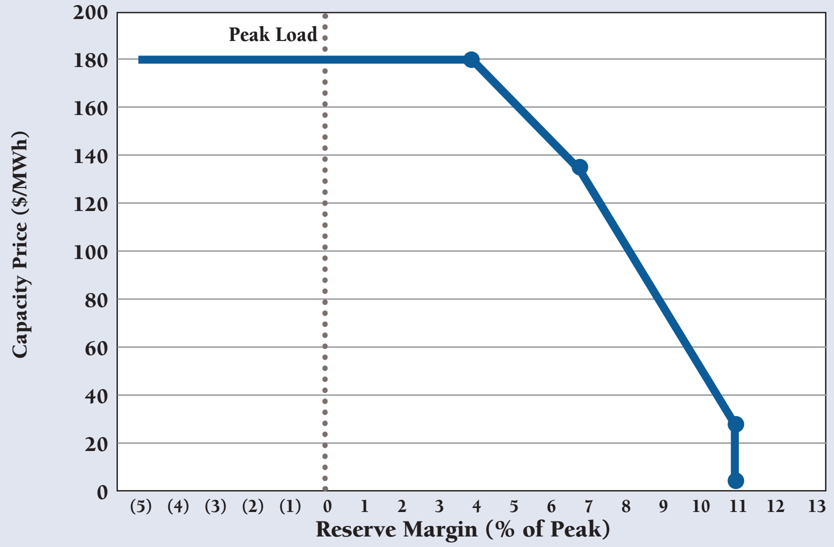
A different approach to remunerating investors, one that is particularly favored by generators, is to establish a separate revenue mechanism that operates outside of energy and balancing markets. This out-of-market approach, in its various forms, is the CRM (see p8).[19] In a CRM, a central authority (usually the system operator) attempts to gauge how much money will be missing from energy markets months or years into the future to pay for the investment costs of the capacity assumed to be needed to meet the reliability standard. The CRM arranges guaranteed payments, either directly from system operators or through obligations placed on retail suppliers, to providers of that quantity of dependable capacity, ideally over and above whatever those providers would expect to earn in the energy and balancing services markets. Rather than directly addressing energy prices, a CRM extracts some components of the energy value chain – investment and other fixed costs – and treats them as a discrete product called "capacity."

18 For an extensive description of the concept, see: Hogan, W. W. (2014, 27 February). *Electricity Scarcity Pricing and Resource Adequacy.* Harvard Energy Policy Group; Electricity Scarcity Pricing Through Operating Reserves. (2013). *Economics of Energy & Environmental Policy,* vol. 2, issue 2.

19 For a description of the most common types of CRM, see: Agency for the Cooperation of Energy Regulators. (2013, July). *Capacity Remuneration Mechanisms and the Internal Market for Electricity,* Appendix A.

Figure 9

PJM Capacity Demand Curve for Delivery Year 2016–2017



In perhaps its most well-known form – a forward capacity market or auction – a CRM operates in a similar fashion to administrative reserve shortage pricing (see Figures 7 and 9). The CRM establishes the price consumers should be willing to pay for capacity for a given period based on a demand curve created by the system operator. The key difference is that it does so months or a few years into the future rather than hours or a day in advance. The quantities and prices of capacity offered are then stacked to create a supply curve, and any and all capacity that clears is paid the clearing price for that period.[20] To the extent the quantity of capacity offered falls short, the clearing price would be adjusted to the price set by the curve.

Despite early hopes for greater simplicity and transparency, most CRMs that have been in operation for extended periods of time have encountered frequent political controversy and creeping complexity.

System operators continue to experiment with different approaches to ensuring that capacity paid by a CRM actually performs when most needed. In response to widespread failures during an early 2014 event in the eastern U.S. markets known as the Polar Vortex, system operators there have implemented reforms to their CRMs that dramatically increase the penalties that capacity providers will face for failing to perform when called. These reforms place most, all, or even more than all CRM payments at risk for failure to respond, with virtually no exceptions allowed. They tend to favor more flexible resources, those able to avoid low-priced hours preceding and following shortage events, over resources that must operate through those hours to be sure they can perform during the shortage hours. Because CRM payments are typically fixed for a relatively short period – six months to a year at a time – they are in any case likely to be significant only during periods when capacity resources are growing scarce. As a result, directionally these reforms move the investment proposition under a CRM back toward what it would be in a market relying on effective energy market price formation.

Despite early hopes for greater simplicity and transparency, most CRMs that have been in operation for extended periods of time have encountered frequent political controversy and creeping complexity. One reason for this is the difficulty in striking an equitable and cost-effective balance between the benefits CRMs are meant to offer investors – a greater level of certainty around at least a portion of their revenues – and the long-term investment risks that CRMs shift back to consumers. CRMs were meant to reduce political risks to investors by flattening energy market price volatility and spreading the cost of reliability over all hours. Over time it has become apparent that politicians and other stakeholders are as likely to seek to interfere with CRM prices that "spike" to reflect the need for new investment. This is proving even more true as CRMs are reformed to improve their effectiveness.

Which Option is Less 'Administrative'?

It is sometimes claimed that CRMs are more market-driven than administrative reserve shortage pricing because they propose to pay only the auction clearing price, rather than adjusting the price to an administratively set curve. This argument fails for several reasons:

- Forecasts of the amount and type of capacity resources the market will need hours or a day in advance are likely to be much closer to actual market conditions than forecasts months or even

20 A "strategic reserve," sometimes categorized as a CRM, pays only designated resources to provide capacity, ideally only when the market fails to clear and thus those resources are prohibited from competing in the energy and balancing services markets. In practice, some strategic reserves are simply another form of price control.

years into the future. Indeed, experts have noted a consistent pattern of over-procurement among system operators administering forward capacity markets.

- Under existing reserve shortage pricing mechanisms the price is adjusted to the curve, if needed, only when available balancing requirements fall below the target. All CRMs work in a similar fashion, most notably auction-based CRMs – as long as the quantity offered exceeds what is specified, the price is set by bids, but in the event the quantity offered falls short, the price is administratively adjusted to the curve. (And as with reserve shortage pricing, the CRM demand curve constitutes a price ceiling.)
- Finally, CRMs are designed to procure an undifferentiated product: the capacity to produce energy when needed. But different capacity portfolios can differ dramatically in cost-effectiveness, particularly in a low-carbon system. Attempts to differentiate CRMs along various dimensions, including operational flexibility, raise legitimate concerns about liquidity and product design and are strongly opposed by CRM proponents.

Conclusions

Price formation and missing money are complex topics about which volumes have been written, but we have tried to cover the essential issues in this paper. They go to the heart of how electricity markets, where they have been adopted, are expected to accomplish their central function – *to deliver reliable electricity at the lowest reasonable cost.*

Claims that remuneration for needed investment is missing from energy markets and can only be recouped via a parallel forward capacity mechanism can be grouped into three themes:

- "Energy pricing was *never meant to* drive the investment needed to ensure security of supply, so we have to pay directly for capacity." That turns out not to be true.
- "For X reasons, energy prices in practice *can't* do the job alone and require a parallel form of payment for capacity." The need for administrative measures to compensate for flaws in energy market price formation may yet be with us for some time, but there are administrative options to improve energy prices themselves that can be effective and that offer advantages that would be difficult to replicate with a capacity mechanism. They should be used in preference to mechanisms that pay for capacity outside of the energy market.
- "In reality we *won't* do what is needed to enable proper price formation, and in any case we *won't* tolerate such prices, so we have to pay directly for capacity." Even if this argument is made based on an accurate understanding of shortage pricing – and it usually is not – it ignores the roles that aggressive market monitoring and the growing empowerment of consumers and demand management can play in offsetting concerns about energy market pricing. It also ignores that similar risks have emerged with CRMs. It is, in effect, an argument for re-regulation/re-nationalization. If that is what we're really about, we should be honest about it and commit to doing a proper job of it.

This doesn't mean an energy market cannot tolerate measures, such as emissions restrictions or targeted support for investment in renewables, that seek to accomplish goals other than the lowest reasonable cost for reliable electricity. Virtually all important commodity markets have learned to adapt to similar kinds of interventions. It does mean that energy markets need to be able to seek the lowest cost responses to the effects of such measures if they are to serve their intended purpose. In an ideal world they would do so by forming prices fully reflecting the ever-shifting balance between supply and demand. In reality markets are never perfect and require constant, judicious administrative intervention.

Electricity markets are especially challenging in this regard, in part because society has a lower tolerance for interruptions in the supply of electricity than, for instance, in the supply of tomatoes, and in part because most consumers are only now beginning to acquire the practical capacity to decide in acceptable ways whether and how much to buy at a given price. This calls both for sustained, detailed administrative efforts to improve price formation, but also for more active administrative measures to address the gap that remains when prices are set below what would be needed to drive needed investment – to replace missing money.

We have explored a range of options that have been developed to accomplish both objectives – to improve the quality of energy prices and to replace money that nonetheless goes missing from what is required to deliver reliable electricity at the lowest reasonable cost. We have looked at two kinds of mechanisms – those that inject money back into energy and balancing services prices close to real time (for example, administrative reserve shortage pricing) and those that divert revenues into a separate, parallel mechanism to set fixed payments for capacity months or years into the future (so-called CRMs).

CRMs should be considered a third-best option for ensuring reliability at the lowest reasonable cost. CRMs are very poor at differentiating among capacity resources on the basis of their flexibility, and they institutionalize the diversion of money from the energy and balancing services prices that are better suited to the job. At a time when most studies point to the value of moving from a conventional generating portfolio dominated by baseload to a smaller but equally reliable one dominated by flexible mid-merit plants, the fact that CRMs tend to place the same value on all firm capacity – and incent existing capacity to remain on the system regardless of its capabilities – is especially problematic.

If it becomes necessary to resort to a CRM, it should be designed to the extent practicable to recognize the relative values of more flexible versus less flexible resources, it should be accompanied by a thorough reform of the process for assessing the amount of capacity really needed to 'keep the lights on' in accordance with the established standard and it should be a supplement to, rather than a substitute for, measures to improve the quality of energy price formation, with the ultimate objective that at some point in the future it will no longer be needed.

Rather than defaulting to the third-best option, top priority should be given first to the pursuit of better energy price formation. Because this is expected to be a challenging and time-consuming project, it should be reinforced by administrative measures that address energy and balancing services prices directly. Some combination of these first two options – improved price formation and administrative shortage pricing mechanisms – should be given an opportunity to succeed before resorting to more desperate measures, especially where a proper resource adequacy assessment would indicate that the need for more capacity is years in the future, as is the case in most of Europe and North America. This approach can both address the problem of missing money and ensure that the risks and rewards for investment are not misallocated. Especially in the midst of the current transformation, it offers the best chance of ensuring that consumers enjoy the clean, reliable service they demand at the lowest reasonable cost.

The Regulatory Assistance Project (RAP)® is an independent, non-partisan, non-governmental organization dedicated to accelerating the transition to a clean, reliable, and efficient energy future.
We help energy and air quality regulators and NGOs navigate the complexities of power sector policy, regulation, and markets and develop innovative and practical solutions designed to meet local conditions.
We focus on the world's four largest power markets: China, Europe, India, and the United States.
Visit our website at **www.raponline.org** to learn more about our work.

Rue de la Science 23
B – 1040 Brussels
Belgium
Tel: +32 2 894 9300
www.raponline.org