ELECTRICITY MARKET DESIGN AND EFFICIENT PRICING

William W. Hogan

Mossavar-Rahmani Center for Business and Government
John F. Kennedy School of Government
Harvard University
Cambridge, Massachusetts 02138

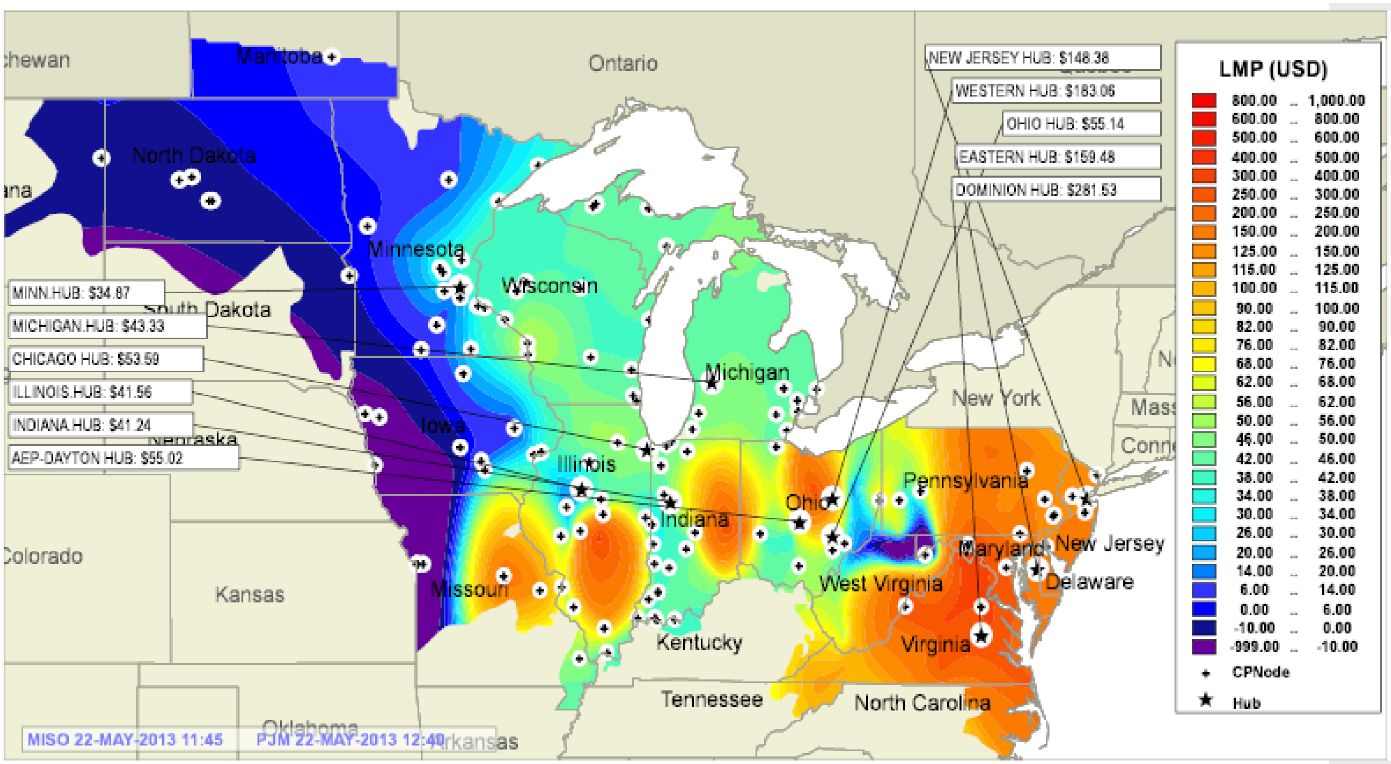
Harvard Electricity Policy Group

Cambridge, MA

October 2, 2014

NETWORK INTERACTIONS Locational Spot Prices

RTOs operate spot markets with locational prices. For example, PJM updates prices and dispatch every five minutes for over 10,000 locations. Locational spot prices for electricity exhibit substantial dynamic variability and persistent long-term average differences.



Missouri MPS -$71.25, Dominion Hub $281.53. May 22, 2013, 12:40pm.

From MISO-PJM Joint and Common Market, http://www.jointandcommon.com/

A promising direction is the FERC initiative to consider an array of issues affecting price formation.

"…the Commission believes there may be opportunities for RTOs/ISOs to improve the energy and ancillary service price formation process. (FERC Notice, June 23, 2014)

- **Use of uplift payments:** Use of uplift payments can undermine the market's ability to send actionable price signals.
- **Offer price mitigation and offer price caps:** All RTOs/ISOs have protocols that endeavor to identify resources with market power and ensure that such resources bid in a manner consistent with their marginal cost.
- **Scarcity and shortage pricing:** All RTOs/ISOs have tariff provisions governing operational actions (e.g., dispatching emergency demand response, voltage reductions, etc.) to manage operating reserves as they approach a reserve deficiency. These actions often are tied to administrative pricing rules designed to reflect degrees of scarcity in the energy and ancillary services markets. … To the extent that actions taken to avoid reserve deficiencies are not priced appropriately or not priced in a manner consistent with the prices set during a reserve deficiency, the price signals sent when the system is tight will not incent appropriate short and long-term actions by resources and loads.
- **Operator actions that affect prices:** … to the extent RTOs/ISOs regularly commit excess resources, such actions may artificially suppress energy and ancillary service prices or otherwise interfere with price formation."

All energy delivery takes place in the real-time market. Market participants will anticipate and make forward decisions based on expectations about real-time prices.

- **Real-Time Prices:** In a market where participants have discretion, the most important prices are those in real-time. "Despite the fact that quantities traded in the balancing markets are generally small, the prevailing balancing prices, or real-time prices, may have a strong impact on prices in the wholesale electricity markets. … No generator would want to sell on the wholesale market at a price lower than the expected real-time price, and no consumer would want to buy on the wholesale market at a price higher than the expected real-time price. As a consequence, any distortions in the real-time prices may filter through to the wholesale electricity prices."[1]
- **Day-Ahead Prices:** Commitment decisions made day-ahead will be affected by the design of day-ahead pricing rules, but the energy component of day-ahead prices will be dominated by expectations about real-time prices.
- **Forward Prices:** Forward prices will look ahead to the real-time and day-ahead markets. Although forward prices are developed in advance, the last prices in real-time will drive the system.
- **Getting the Prices Right:** The last should be first. The most important focus should be on the models for real-time prices. Only after everything that can be done has been done, would it make sense to focus on out-of-market payments and forward market rules.

[1] Cervigni, G., & Perekhodtsev, D. (2013). Wholesale Electricity Markets. In P. Rinci & G. Cervigni (Eds.), *The Economics of Electricity Markets : Theory and Policy*. Edward Elgar, p. 53.

The purpose of *ex post* or dispatch-based pricing is to determine "prices consistent with the actual usage by applying the marginal tests of economic dispatch."[2] Examples include:

- **Ex Post LMP:** Utilize the actual dispatch to simplify the model for calculating consistent locational prices.
- **Scarcity Pricing and the Operating Reserve Demand Curve:** Price the scarcity of operating reserves and stimulate demand participation.
- **Demand Response:** Incorporate demand response in the pricing model to reflect scarcity conditions and avoid price reversals.
- **Reliability Unit Commitment:** Recognize reliability constraints in the dispatch and the pricing model.
- **Voltage Support:** Recognize operator actions for difficult to model problems in the economic dispatch by incorporating constraint approximations in the dispatch.
- **Extend Locational Marginal Pricing (ELMP):** Incorporate the effects of unit commitment, block loaded units, and other lumpy decisions in prices to minimize the related uplift charges.

[2] See W. Hogan, "Electricity Market Design and Efficient Pricing: Applications for New England and Beyond," June 24, 2014, at www.whogan.com.

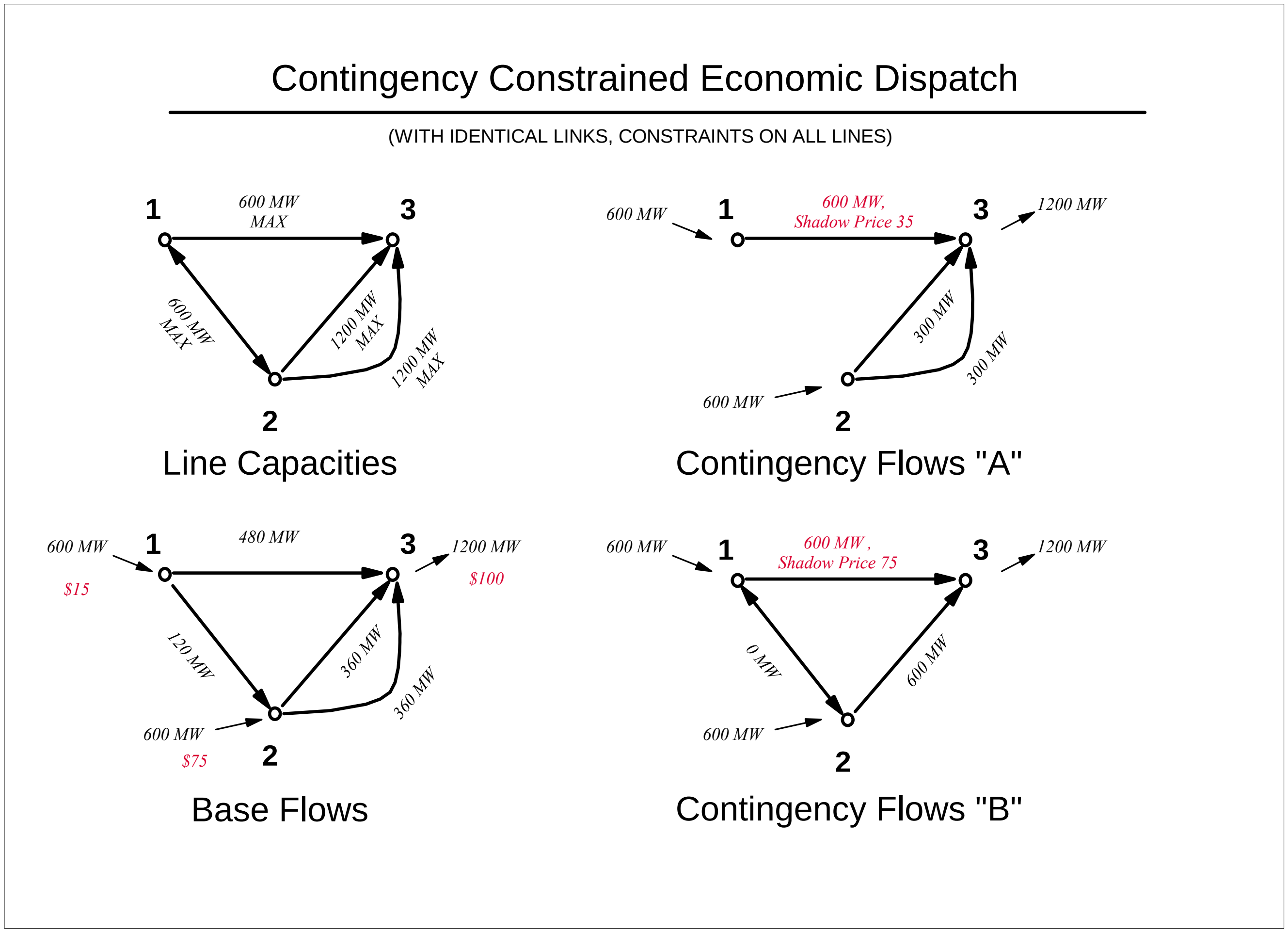
The basic security-constrained, economic dispatch formulation provides the foundation and the framework for real-time and day-ahead electricity spot market pricing.

Let $B(d)$ define the benefits of bid-in load (d) and $C(g)$ the cost of generation (g) offers. Incorporate other relevant variables such as unit commitment decisions in the control variables in u . The net load at each location is defined as the vector $y = d - g$. Aggregate losses are $L(y,u)$. Finally the transmission constraints appear in the vector function $K(y,u)$. With these definitions, we treat the underlying security-constrained economic dispatch problem as

$$\begin{aligned} &\underset{d\in D, g\in G, u\in U}{Max} B(d) - C(g) \\ &s.t. \\ &\quad d - g = y, \\ &\quad L(y,u) + \iota^t y = 0, \\ &\quad K(y,u) \le 0. \end{aligned}$$

This is a complicated problem with a large number of variables and constraints. With thousands of locations and thousands of transmission lines, the complete statement of the problem can run into millions of variables and millions of constraints. Fortunately, system operators are familiar with this model and have workable methods using a blend of optimization tools and operator judgment to approximate an economic dispatch solution.

Security-constrained economic dispatch models include a large number of security constraints.



The complicated economic dispatch model lends itself to a relatively simple pricing model.

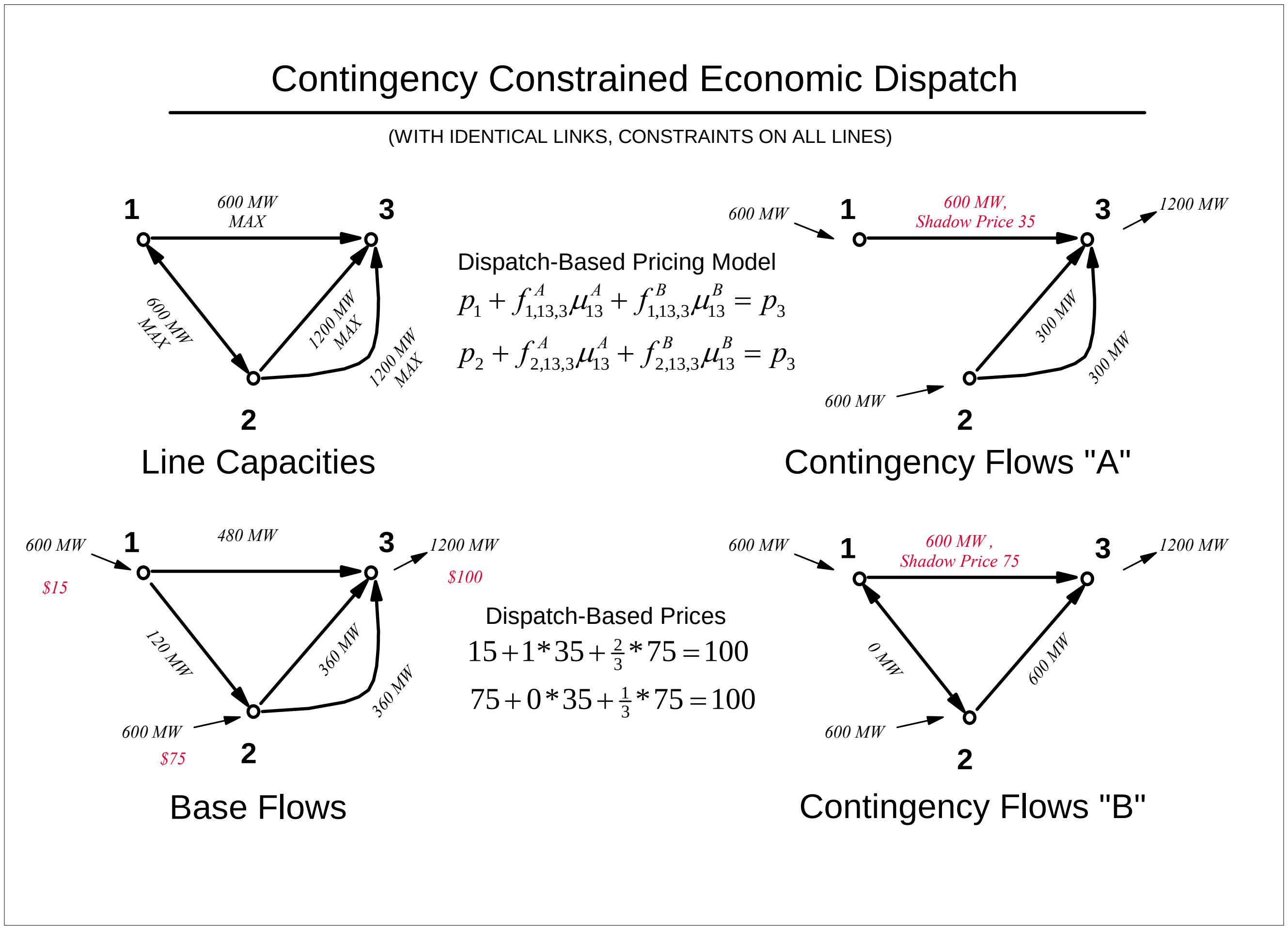
One of the earliest implementations of the dispatched-based approach was to extract the implied LMP values from the solution to the linear approximation of the full dispatch problem, where the linear approximation is based on the actual dispatch with binding transmission constraints in K^*. The set of binding constraints is not known before the dispatch is determined. Given the linear approximation of the binding constraints, the ex post LMP model would be:

$$\begin{aligned}
&\underset{d\in D, g\in G, u\in U}{Max} B(d) - C(g) \\
&s.t. \\
&\quad d - g = y, \\
&\quad L(y,u) + \iota^t y = 0, \\
&\quad \nabla K^*\left(y^*, u^*\right)\left(y - y^*, u - u^*\right) \le 0.
\end{aligned}$$

Given the dispatch, the information needed to formulate this problem is both relatively simple and readily available. The critical elements would be the "shift factors" that define the derivatives of the binding constraints K^*, which is a small subset of the full list of possible constraints in K.

The resulting dispatch-based prices are easy to compute and as good as the approximation to the economic dispatch.

Dispatch-based pricing utilizes the actual dispatch to produce a much simplified pricing model.



Dispatch-based pricing can incorporate pricing approximations that reflect explicit and implicit constraints.

- **Linear Pricing Model:** Under minimal conditions, the exact pricing model employs the linear approximation of the constraint derivatives.
- **Implicit Constraints:** Operator judgment deals with implicit constraints that may be difficult to specify or very non-linear. But a linear approximation would suffice for the pricing model.
- **Approximation Quality:** Developing a good approximation itself requires a judgment call. But an imperfect approximation would almost always be better than ignoring the constraint when determining prices.
- **Enhanced Dispatch Models:** Approximations of important pricing effects may be included in dispatch models. An example is in scarcity pricing and operating reserves.

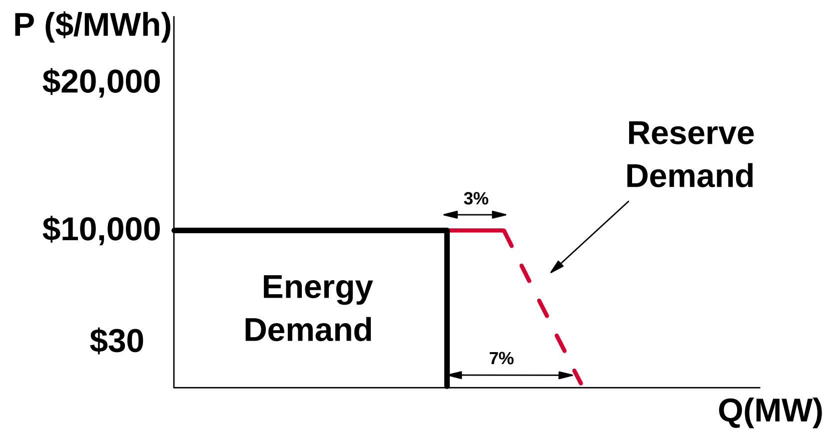
ELECTRICITY MARKET Scarcity Pricing and Volatility

Inadequate scarcity pricing dampens real-time price volatility, and has a material impact on incentives for innovation. Fixed rates, including pre-determined time-of-use rates, dampen volatility. Levelized rates and socialized costs eliminate volatility. Accurate scarcity prices would capture the marginal welfare effects of consumption and generation. Assuming cost recovery on average, incomplete scarcity pricing implies various forms of inefficiency.

- **Energy Efficiency and Distributed Generation.** With levelized rates, passive energy efficiency changes such as insulation are efficient only for customers with the average load profile. Customer load profiles are heterogeneous, so there is too little or too much incentive for most. For distributed generation and active load management, such as turning down air conditioning when away from home, sees too little incentive when it is needed most during high periods of (implicit) scarcity prices.
- **Load Management.** Changing the load profile to arbitrage price differences over time depends on exploiting price volatility. Suppressing and socializing scarcity prices dampens incentives for load management.
    - **Load Shifting.** Cycling equipment of moving consumption to "off-peak" hours receives too little incentive.
    - **PHEV/EV.** Managing the charging cycle for electric vehicles will affect the economics of both cars and the electricity system. Inadequate scarcity pricing and rate smoothing dampen incentives and raise costs.
    - **Batteries.** The principal benefit of batteries, from high tech flow batteries to low tech ceramic bricks, is profit from price arbitrage. Smooth prices undo the incentives for battery deployment.

Operating reserve demand curve would reflect capacity scarcity.

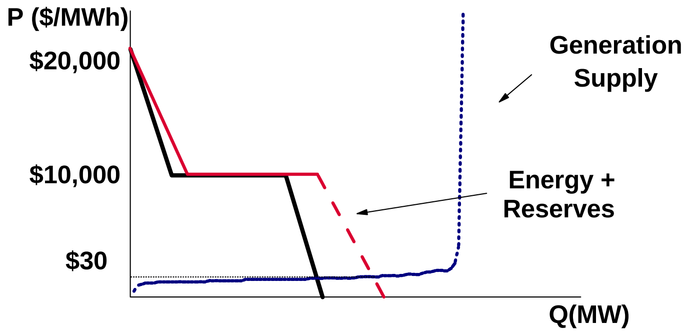
Illustrative Reserve Demand



There is a minimum level of operating reserve (e.g., 3%) to protect against system-wide failure. Above the minimum reserve, reductions below a nominal reserve target (e.g., 7%) are price senstive.

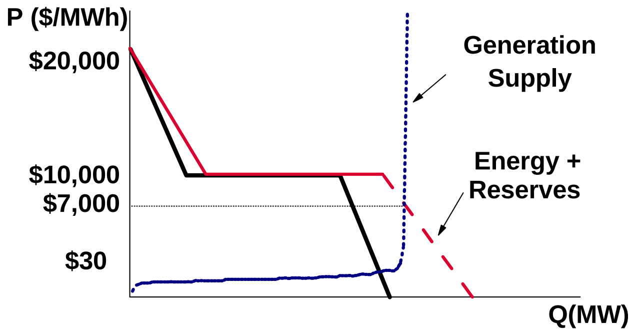
Market clearing addresses the "missing money."

Normal "Energy Only" Market Clearing



When demand is low and capacity available, reserves hit nominal targets at a low price.

Scarcity "Energy Only" Market Clearing



When demand is high and reserve reductions apply, there is a high price.

What are the relevant first principles that could guide better scarcity pricing? There are many ideas that would be included under the general framework of economic dispatch. A suggestive list for operating reserve pricing would include:

- Connecting to the value of loss load and other emergency actions.
- Including a representation of the uncertainty of net load changes and the loss of load probability.
- Integrating minimum contingency reserve requirements.
- Maintaining consistency between energy and reserve prices.
- Coordinating day-ahead and real-time settlements.
- Co-optimization of reserves and energy.
- Providing a consistent representation of any locational differences in valuing reserves.

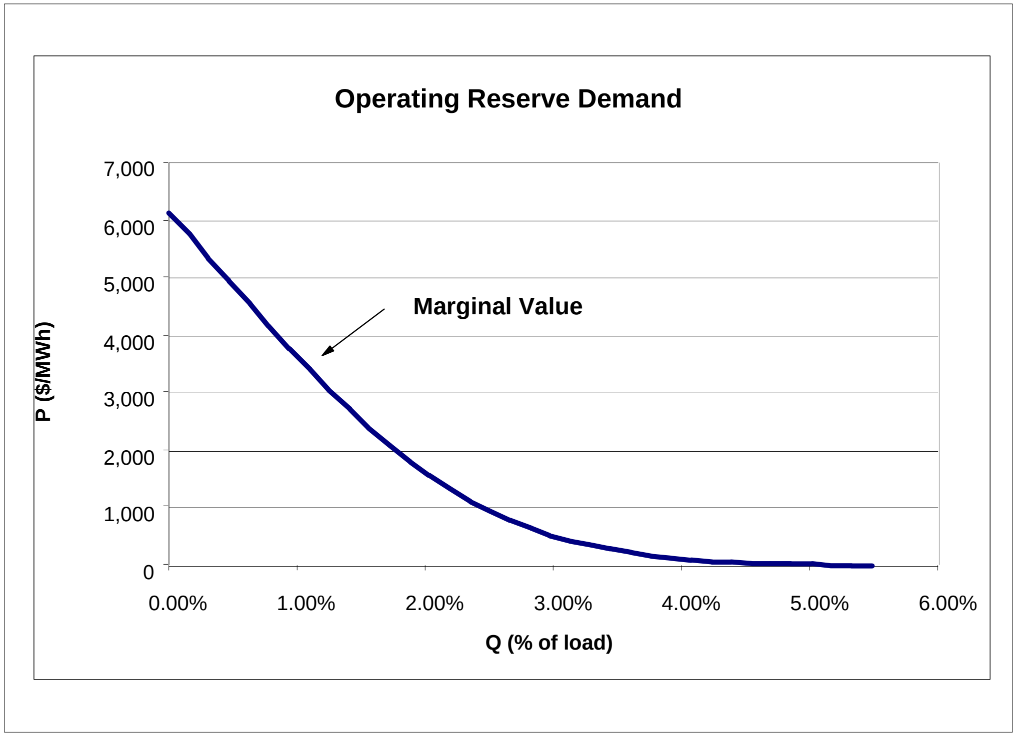
The most general principle would be to provide a pricing framework that incorporates reasonable prices for actions that the system operator may take to provide a security constrained economic dispatch. "As a general principle, competitive and efficient market prices should be consistent with the marginal cost of the marginal action taken to satisfy the market's demand." (IMM, ERCOT 2012 State of the Market Report, p. 82)

Operating reserve demand curve (ORDC) is a complement to energy demand for electricity. The probabilistic demand for operating reserves reflects the cost and probability of lost load.[3]

Example Assumptions

Expected Load (MW)	34000
Std Dev %	1.50%
Expected Outage %	0.45%
Std Dev %	0.45%
Expected Total (MW)	153
Std Dev (MW)	532.46
VOLL ($/MWh)	10000

Under the simplifying assumptions, if the dispersion of the LOLP distribution is proportional to the expected load, the operating reserve demand is proportional to the expected load.



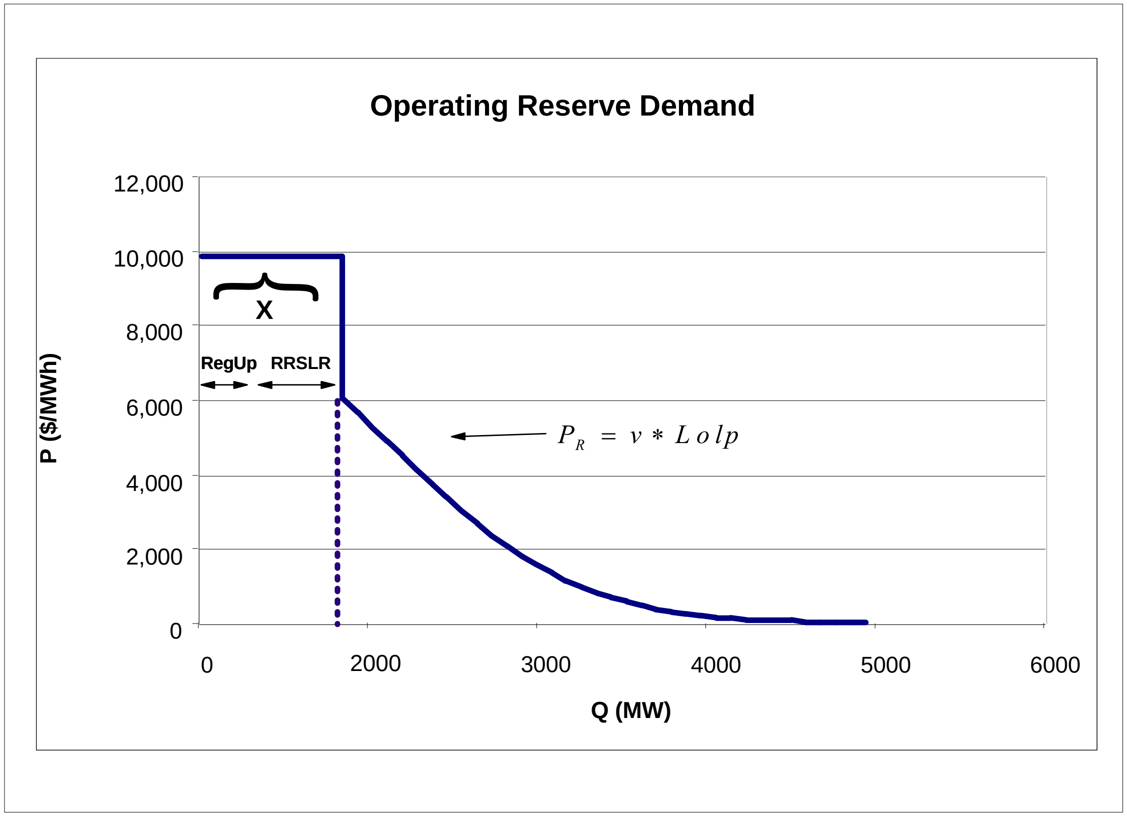
[3] "For each cleared Operating Reserve level less than the Market-Wide Operating Reserve Requirement, the Market-Wide Operating Reserve Demand Curve price shall be equal to the product of (i) the Value of Lost Load ("VOLL") and (ii) the estimated conditional probability of a loss of load given that a single forced Resource outage of 100 MW or greater will occur at the cleared Market-Wide Operating Reserve level for which the price is being determined. … The VOLL shall be equal to $3,500 per MWh." MISO, FERC Electric Tariff, Volume No. 1, Schedule 28, January 22, 2009, Sheet 2226.

The deterministic approach to security constrained economic dispatch includes lower bounds on the required reserve to ensure that for a set of monitored contingencies (e.g., an n-1 standard) there is sufficient operating reserve to maintain the system for an emergency period.

Suppose that the maximum generation outage contingency quantity is $r_{Min}\left(d^0, g^0, u\right)$. Then we would have the constraint:

$$r \geq r_{Min}\left(d^0, g^0, u\right) = X.$$

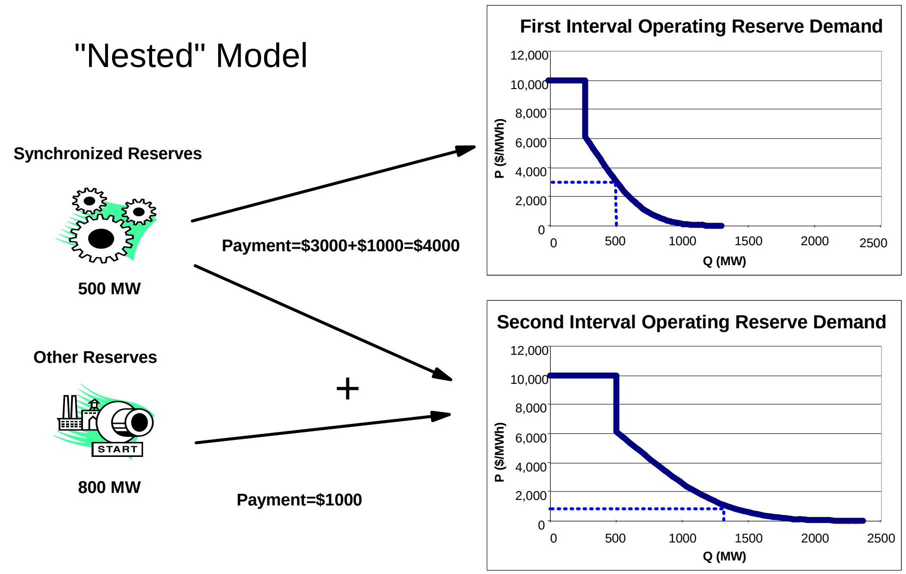
In effect, the contingency constraint provides a vertical demand curve that adds horizontally to the probabilistic operating reserve demand curve.



If the security minimum will always be maintained over the monitored period, the marginal price at r=0 applies. If the outage shocks allow excursions below the security minimum during the period, the reserve price starts at the security minimum.

Multiple types of operating reserves exist according to response time. A nested model divides the period into consecutive intervals. Reserve schedules set before the period. Uncertainty revealed after the start of the period. Faster responding reserves modeled as available for subsequent intervals. The operating reserve demand curves apply to intervals and the payments to generators include the sum of the prices for the available intervals.

Multiple Operating Reserve Demand Types (Intervals)



Nested model with two intervals, decisions made before uncertainty revealed.

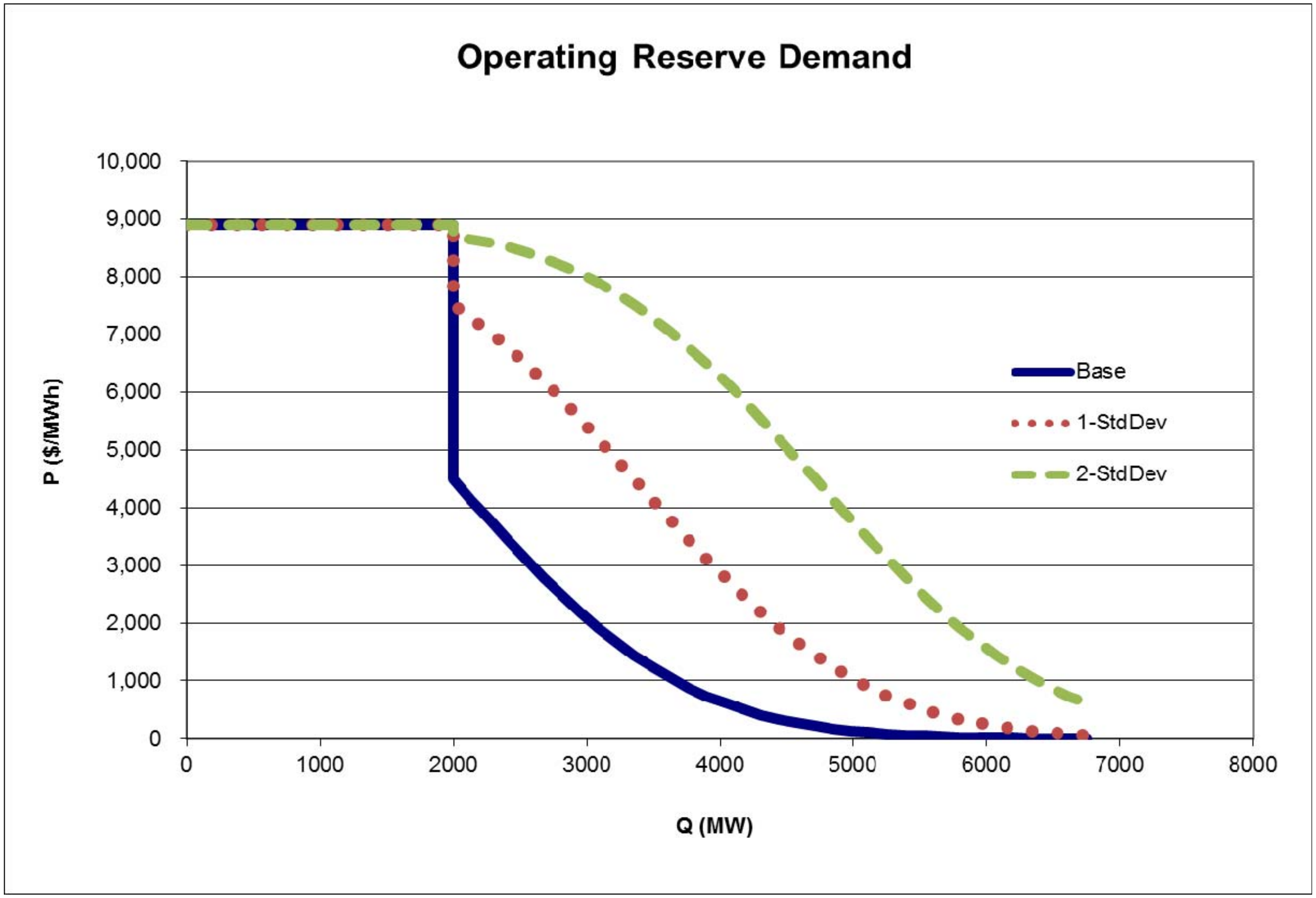
An augmented ORDC would impose conservative assumptions on the basic model. The intent would be to provide both a reliability margin of safety, an associated increase in total operating reserves, and energy payments to address the missing money problem. The three principal parameters of the ORDC are the value of lost load (VOLL), the minimum contingency level (X), and the loss of load probability (LOLP).

- **VOLL.** The VOLL price applies when conditions require involuntary load curtailment. It is important that this price be paid to generation and charged to remaining load. Hence, an upper bound on a conservative VOLL would be the maximum price we were willing to charge in the face of load curtailment. It may be better to err in the direction of a higher VOLL, but this may not be enough to address the reliability goal and provide the missing money.

- **X.** The minimum contingency level is more directly connected to reliability. However, if the minimum contingency threshold is set too high, we would produce periods when VOLL prices were being imposed but no non-market interventions were needed. Regulators would have to defend applying the VOLL when it was not required.

- **LOLP.** The short-term load and generation changes that give rise to the LOLP summarize a complex process. The models applied employ certain assumptions about the accuracy of the system approximations and the ability to avoid problems like human error typically found in events that threaten the stability of the system. A conservative approach to reliability is already part of the motivation for the use of contingency constraints to define secure operations. However, it would be consistent to extend this reliability motivation to a conservative estimation of the LOLP. This would avoid the conflicts that arise with too high a VOLL or too high an X.

A conservative assumption addressed at reliability would be to increase the estimate of the loss of load probability. A shift of one standard deviation would have a material impact on the estimated scarcity prices. The choice would depend on the margin of safety beyond the economic base.



Improved pricing through an explicit operating reserve demand curve raises a number of issues.

Demand Response: Better pricing implemented through the operating reserve demand curve would provide an important signal and incentive for flexible demand participation in spot markets.

Price Spikes: A higher price would be part of the solution. Furthermore, the contribution to the "missing money" from better pricing would involve many more hours and smaller price increases.

Practical Implementation: NYISO, ISONE, MISO and PJM implementations dispose of any argument that it would be impractical to implement an operating reserve demand curve. The only issues are the level of the appropriate price and the preferred model of locational reserves.

Operating Procedures: Implementing an operating reserve demand curve does not require changing the practices of system operators. Reserve and energy prices would be determined simultaneously treating decisions by the operators as being consistent with the adopted operating reserve demand curve.

Multiple Reserves: The demand curve would include different kinds of operating reserves, from spinning reserves to standby reserves.

Reliability: Market operating incentives would be better aligned with reliability requirements.

Market Power: Better pricing would remove ambiguity from analyses of high prices and distinguish (inefficient) economic withholding through high offers from (efficient) scarcity pricing derived from the operating reserve demand curve.

Hedging: Day-ahead and longer term forward markets can reflect expected scarcity costs, and price in the risk.

Increased Costs: The higher average energy costs from use of an operating reserve demand curve do not automatically translate into higher costs for customers. In the aggregate, there is an argument that costs would be lower.

William W. Hogan is the Raymond Plank Professor of Global Energy Policy, John F. Kennedy School of Government, Harvard University. This paper draws on research for the Harvard Electricity Policy Group and for the Harvard-Japan Project on Energy and the Environment. This paper was supported by Vitol Inc. The author is or has been a consultant on electric market reform and transmission issues for Allegheny Electric Global Market, American Electric Power, American National Power, Aquila, Atlantic Wind Connection, Australian Gas Light Company, Avista Corporation, Avista Utilities, Avista Energy, Barclays Bank PLC, Brazil Power Exchange Administrator (ASMAE), British National Grid Company, California Independent Energy Producers Association, California Independent System Operator, California Suppliers Group, Calpine Corporation, CAM Energy, Canadian Imperial Bank of Commerce, Centerpoint Energy, Central Maine Power Company, Chubu Electric Power Company, Citigroup, City Power Marketing LLC, Cobalt Capital Management LLC, Comision Reguladora De Energia (CRE, Mexico), Commonwealth Edison Company, COMPETE Coalition, Conectiv, Constellation Energy, Constellation Energy Commodities Group, Constellation Power Source, Coral Power, Credit First Suisse Boston, DC Energy, Detroit Edison Company, Deutsche Bank, Deutsche Bank Energy Trading LLC, Duquesne Light Company, Dyon LLC, Dynegy, Edison Electric Institute, Edison Mission Energy, Electricity Corporation of New Zealand, Electric Power Supply Association, El Paso Electric, Energy Endeavors LP, Exelon, Financial Marketers Coalition, FTI Consulting, GenOn Energy, GPU Inc. (and the Supporting Companies of PJM), GPU PowerNet Pty Ltd., GDF SUEZ Energy Resources NA, Great Bay Energy LLC, GWF Energy, Independent Energy Producers Assn, ISO New England, Koch Energy Trading, Inc., JP Morgan, LECG LLC, Luz del Sur, Maine Public Advocate, Maine Public Utilities Commission, Merrill Lynch, Midwest ISO, Mirant Corporation, MIT Grid Study, Monterey Enterprises LLC, MPS Merchant Services, Inc. (f/k/a Aquila Power Corporation), JP Morgan Ventures Energy Corp., Morgan Stanley Capital Group, National Independent Energy Producers, New England Power Company, New York Independent System Operator, New York Power Pool, New York Utilities Collaborative, Niagara Mohawk Corporation, NRG Energy, Inc., Ontario Attorney General, Ontario IMO, Ontario Ministries of Energy and Infrastructure, Pepco, Pinpoint Power, PJM Office of Interconnection, PJM Power Provider (P3) Group, Powerex Corp., Powhatan Energy Fund LLC, PPL Corporation, PPL Montana LLC, PPL EnergyPlus LLC, Public Service Company of Colorado, Public Service Electric & Gas Company, Public Service New Mexico, PSEG Companies, Red Wolf Energy Trading, Reliant Energy, Rhode Island Public Utilities Commission, Round Rock Energy LP, San Diego Gas & Electric Company, Secretaría de Energía (SENER, Mexico), Sempra Energy, SESCO LLC, Shell Energy North America (U.S.) L.P., SPP, Texas Genco, Texas Utilities Co, Tokyo Electric Power Company, Toronto Dominion Bank, Transalta, TransAlta Energy Marketing (California), TransAlta Energy Marketing (U.S.) Inc., Transcanada, TransCanada Energy LTD., TransÉnergie, Transpower of New Zealand, Tucson Electric Power, Twin Cities Power LLC, Vitol Inc., Westbrook Power, Western Power Trading Forum, Williams Energy Group, Wisconsin Electric Power Company, and XO Energy. The views presented here are not necessarily attributable to any of those mentioned, and any remaining errors are solely the responsibility of the author. (Related papers can be found on the web at www.whogan.com).