Do Centrally Committed Electricity Markets Provide Useful Price Signals?☆

Ramteen Sioshansi*, Ashlin Tignor*

Integrated Systems Engineering Department, The Ohio State University, Baker Systems Engineering Building, 1971 Neil Avenue, Columbus, OH 43210, United States

Abstract

Centrally committed markets rely on an independent system operator to determine the commitment and dispatch of generators. This is done by solving a unit commitment model, which is an $\mathcal{NP}$-hard mixed integer program that is rarely (if ever) solved to complete optimality. We demonstrate, using a case study based on the ISO New England system, that near-optimal solutions that are very close to one another in terms of overall system cost can yield very different generator surpluses and prices. We further demonstrate that peaking generators are more prone to surplus differences between near-optimal solutions and that transmission buses that are most prone to binding transmission constraints experience the greatest price fluctuations. Based on these findings, we discuss the potential benefits of a decentralized market design in providing more robust price signals.

Keywords: Unit commitment, market design, pricing

1. Introduction

Electricity service has undergone a major change in the past three decades, with a shift away from the traditional vertically integrated utility model toward competitive markets. These market restructuring efforts have been undertaken in the hopes of increasing system efficiency, reducing costs, spurring more economic investments, and improving the quality of service. Joskow and Schmalensee (1983) provide one of the first analyses of whether the successes in restructuring other industries could carry over to electric power systems.

The promise of this paradigm shift is that markets provide price signals to the different agents in the electricity supply chain. Generator profits can vary depending on factors such as operating costs and constraints, location within the transmission network, and the system load. Wholesale market prices and resulting generator profits are intended to provide market-based signals for what kind of generation should enter or leave the market to reach a long-run equilibrium. Similarly, Hogan (1992); Chao and Peck (1996); Bushnell and Stoft (1996, 1997) argue that locational marginal price (LMP) differences within a transmission network indicate the need for and can spur generation and transmission capacity expansion to relieve binding power flow constraints. Markets can also more efficiently ration electricity demand, for instance through the use of real-time retail prices that are based on wholesale prices. Borenstein (2002, 2005); Borenstein and Holland (2005) discuss the efficiency benefits of such a real-time pricing (RTP) scheme compared to the use of time-invariant retail pricing.

☆The authors would like to thank Richard O'Neill and Shmuel Oren for making us aware of these issues with pricing in centrally committed markets. Eugene Litvinov and Tongxin Zheng provided insights regarding modeling ISO New England's unit commitment problem as well as model data. Matthew Roberts, Armin Sorooshian, the editor, and three anonymous reviewers provided helpful comments and suggestions. The second author was supported by The Ohio State University College of Engineering Honors Committee. Any opinions and conclusions expressed in this paper are those of the authors only.

*Corresponding author

Email addresses: sioshansi.1@osu.edu (Ramteen Sioshansi), campbell.856@osu.edu (Ashlin Tignor)

An issue underlying these discussions is how prices should be determined and how much operational control should be vested in a central decision-maker. Power systems require tight coordination between different participants, due to non-convexities of and complementarities between agent actions. Generator operations are non-convex because of the discrete nature of each generator's online state—a generator can only be in an online or offline state at any particular time. Generator operations are further complicated by technical constraints, such as ramping limits, minimum and maximum loads, and minimum up- and down-times. Complementarities arise due to the just-in-time nature of electricity supply and the inability to direct flows within a transmission network. The just-in-time nature of power systems requires electricity demand and supply constantly be exactly balanced at all locations in the network. Any supply imbalance can result in voltage or frequency deviations, damage generators or customer electronics, and threaten system stability. Moreover, power flows within a transmission network are governed by Kirchhoff's laws and cannot be directed. Thus the amount of energy that can be injected or withdrawn at one node in the network can depend on injections and withdrawals elsewhere, since power flows in opposite directions along a transmission element cancel each other out. These two factors imply that the ability of one generator to feasibly adjust its output is intimately related to the actions of others. At a minimum, load-balance requires that at least one generator adjust its output in reaction to an adjustment by another. Power flows can, however, require coordination between multiple generators.

Under the traditional paradigm, this coordination is handled by the monopoly utility. The utility solves a unit commitment (UC) problem,[1] which ensures that the system load is satisfied at least cost while respecting power flow and generator operating constraints. Most restructured markets mimic this activity by granting an independent system operator (ISO) broad authority to make commitment and dispatch decisions. Under such a centrally committed market design the ISO collects bids from the market participants—generators indicate their operating costs and constraints while load-serving entities and large customers submit bids to purchase energy. The ISO uses these bids to determine the least-cost feasible commitment and dispatch of the system, and provides the market participants with schedules indicating the amount of energy they produce and consume. The ISO also determines LMPs, based on the dual variable of the locational load-balance constraints, to settle these transactions.

One of the issues that plagues the use of UC models is their computational complexity. Because of the non-convex nature of generator commitment decisions, UC problems are typically modeled as mixed-integer programs (MIP), which Guan et al. (2003) show to be $\mathcal{NP}$-hard. As such, UC problems are rarely, if ever, solved to complete optimality. Streiffert et al. (2005) explain that since PJM has a short time window after receiving bids from the market to determine the commitment and dispatch of the system, the model is solved for a fixed period of time and the best solution (in terms of cost) found within that time is used. They indicate that in almost all cases the UC model is able to find solutions that are within a fraction of a percent away from optimal.

The only effect of using such a near-optimal solution under the traditional monopoly paradigm would be that the utility's costs would slightly increase. In a competitive market, however, these different solutions can have economic consequences for generation assets that are no longer owned by a single entity. Johnson et al. (1997); Sioshansi et al. (2008a) use small test cases to demonstrate that near-optimal solutions that have negligible differences in overall system costs can yield vastly different generator surpluses. These surplus differences arise for two reasons. One is that the actual commitment and dispatch of a generator may differ between solutions. The second is that wholesale prices, which are determined endogenously in the UC model, can be affected by changes in generator commitments between solutions. Sioshansi et al. (2008a) further demonstrate that the magnitude of these effects do not necessarily decrease as the UC problem is solved closer to optimality.

A question that these analyses leave open, however, is whether these differences accumulate over time or if they 'cancel out' in the long-run. Johnson et al. (1997) examine surplus differences over a single week while Sioshansi et al. (2008a) consider one day, both showing that there can be significant surplus differences. If these differences are truly random in nature, however, then one might expect them to offset

[1] Muckstadt and Koenig (1977); Hobbs et al. (2001) provide background, formulation, and solution methodologies for UC problems.

in the long-run. This paper examines the economic implications of centralized unit commitment over a year using a case study based on the ISO New England system. By adjusting the stopping criteria of the solver used to optimize the UC model, we generate different near-optimal solutions that yield substantive annual generator surplus and price differences. We further demonstrate that peaking generators tend to see greater surplus variability between different near-optimal UC solutions, which can have important implications for long-term generation investment. We also examine the effect of demand response in a hypothetical case with RTP, showing that although hourly loads can vary substantively between near-optimal UC solutions, system demand over the year is relatively insensitive. We finally contrast centrally committed markets with decentralized designs that leave commitment decisions to individual generators, discussing pros and cons of the two designs. The remainder of this paper is organized as follows: section 2 describes the model, data, and case study used in our analysis; section 3 discusses our results; and section 4 concludes with our comparison of centralized and decentralized market designs.

2. Model and Data

2.1. Model

UC problems are typically modeled as MIPs. Binary variables track the discrete online/offline state of each generator during each time period, while continuous variables track the amount of energy and reserves provided by each generator. The model is formulated to maximize social welfare, which is the difference between the integral (up to the amount of load served in each time period) of the inverse demand function and total generation cost. The model includes load balance, power flow, reserve, and generator operating constraints. Throughout our analysis we assume that the UC problem is modeled at hourly timesteps. While we base the formulation of our UC problem on the model used by ISO New England, our model does not exactly capture ISO New England operations. Thus our results should be viewed as illustrative of the effect of relying on near-optimal UC solutions in making operational decisions as opposed to indicating actual revenues earned by particular generators in the ISO New England system.

We first define the following parameters used in our UC model:

T: set of hours in planning horizon;

L: set of transmission elements;

N: set of transmission buses;

G: set of generators;

$G(n)$: set of generators that are located at bus n;

f_l: flow limit on transmission element l;

$\gamma_{l,n}$: power transfer distribution factor between bus n and transmission element l;

$c_g^v(\cdot)$: variable cost function of generator g;

c_g^n: spinning no-load cost of generator g;

c_g^s: startup cost of generator g;

$\kappa_{g,t}^-$: minimum generating capacity of generator g if it is online during hour t;

$\kappa_{g,t}^+$: maximum generating capacity of generator g in hour t;

ρ_g^-: maximum amount generator g can decrease its generation between consecutive hours;

ρ_g^+: maximum amount generator g can increase its generation between consecutive hours;

τ_g^+: minimum number of hours generator g must remain online when it is started up;

τ_g^-: minimum number of hours generator g must remain offline when it is shutdown;

$\mu_{g,t}$: binary parameter that equals 1 if generator g is must-run in hour t, equals 0 otherwise; and

$p_{n,t}(\cdot)$: inverse demand function for energy at bus n in hour t.

We also define the following decision variables:

$q_{g,t}$: energy produced by generator g in hour t;

$r_{g,t}^{10}$: ten-minute operating reserves provided by generator g in hour t;

$r_{g,t}^{30}$: thirty-minute operating reserves provided by generator g in hour t;

$u_{g,t}$: binary variable that equals 1 if generator g is online in hour t, equals 0 otherwise;

$s_{g,t}$: binary variable that equals 1 if generator g is started up in hour t, equals 0 otherwise;

$h_{g,t}$: binary variable that equals 1 if generator g is shutdown in hour t, equals 0 otherwise;

$\xi_{g,t}^1$: binary variable that equals 1 if generator g is the largest contingency in hour t, equals 0 otherwise;

$\xi_{g,t}^2$: binary variable that equals 1 if generator g is the second-largest contingency in hour t, equals 0 otherwise;

ϕ_t^1: size of largest contingency in hour t;

ϕ_t^2: size of second-largest contingency in hour t;

$d_{n,t}$: energy used by consumers at node n in hour t; and

$e_{n,t}$: net energy exported from node n in hour t.

The model formulation is given by:

$$\max \quad \sum_{t \in T} \left(\sum_{n \in N} \int_0^{d_{n,t}} p_{n,t}(x)dx - \sum_{g \in G} \left(c_g^v(q_{g,t}) + c_g^n \cdot u_{g,t} + c_g^s \cdot s_{g,t} \right) \right); \tag{1}$$

$$\text{s.t.} \quad \sum_{g \in G(n)} q_{g,t} = d_{n,t} + e_{n,t}, \qquad \forall\, n \in N, t \in T; \tag{2}$$

$$\sum_{n \in N} e_{n,t} = 0, \qquad \forall\, t \in T; \tag{3}$$

$$\phi_t^1 \geq q_{g,t}, \qquad \forall\, g \in G, t \in T; \tag{4}$$

$$\phi_t^1 - q_{g,t} \leq (1 - \xi_{g,t}^1) \cdot \max_{g \in G}\{\kappa_{g,t}^+\}, \qquad \forall\, g \in G, t \in T; \tag{5}$$

$$\sum_{g \in G} \xi_{g,t}^1 = 1, \qquad \forall\, t \in T; \tag{6}$$

$$\phi_t^2 \geq q_{g,t} - \xi_{g,t}^1 \cdot \kappa_{g,t}^+, \qquad \forall\, g \in G, t \in T; \tag{7}$$

$$\phi_t^2 - q_{g,t} \leq (1 - \xi_{g,t}^2) \cdot \max_{g \in G}\{\kappa_{g,t}^+\}, \qquad \forall\, g \in G, t \in T; \tag{8}$$

$$\sum_{g \in G} \xi_{g,t}^2 = 1, \qquad \forall\, t \in T; \tag{9}$$

$$\sum_{g \in G}(q_{g,t} + r_{g,t}^{10}) - \phi_t^1 \geq \sum_{n \in N} d_{n,t}, \qquad \forall\, t \in T; \tag{10}$$

$$\sum_{g \in G} (q_{g,t} + r^{10}_{g,t} + r^{30}_{g,t}) - \phi^1_t - \phi^2_t \geq \sum_{n \in N} d_{n,t}, \qquad \forall\ t \in T; \tag{11}$$

$$\kappa^-_{g,t} \cdot u_{g,t} \leq q_{g,t}, \qquad \forall\ g \in G, t \in T; \tag{12}$$

$$q_{g,t} + r^{10}_{g,t} + r^{30}_{g,t} \leq \kappa^+_{g,t} \cdot u_{g,t}, \qquad \forall\ g \in G, t \in T; \tag{13}$$

$$-\rho^-_g \leq q_{g,t} - q_{g,t-1} \leq \rho^+_g, \qquad \forall\ g \in G, t \in T; \tag{14}$$

$$r^{10}_{g,t} \leq (1 - \xi^1_{g,t}) \cdot \rho^+_g / 6, \qquad \forall\ g \in G, t \in T; \tag{15}$$

$$r^{10}_{g,t} + r^{30}_{g,t} \leq (1 - \xi^1_{g,t} - \xi^2_{g,t}) \cdot \rho^+_g / 2, \qquad \forall\ g \in G, t \in T; \tag{16}$$

$$u_{g,t} - u_{g,t-1} = s_{g,t} - h_{g,t}, \qquad \forall\ g \in G, t \in T; \tag{17}$$

$$\sum_{\iota = t - \tau^-_g}^{t} h_{g,\iota} \leq 1 - u_{g,t}, \qquad \forall\ g \in G, t \in T; \tag{18}$$

$$\sum_{\iota = t - \tau^+_g}^{t} s_{g,\iota} \leq u_{g,t}, \qquad \forall\ g \in G, t \in T; \tag{19}$$

$$u_{g,t} \geq \mu_{g,t}, \qquad \forall\ g \in G, t \in T; \tag{20}$$

$$-f_l \leq \sum_{n \in N} \gamma_{l,n} \cdot e_{n,t} \leq f_l, \qquad \forall\ l \in L, t \in T; \tag{21}$$

$$r^{10}_{g,t}, r^{30}_{g,t} \geq 0, \qquad \forall\ g \in G, t \in T; \tag{22}$$

$$d_{n,t} \geq 0, \text{ and} \qquad \forall\ n \in N, t \in T; \tag{23}$$

$$u_{g,t}, s_{g,t}, h_{g,t}, \xi^1_{g,t}, \xi^2_{g,t} \in \{0, 1\}, \qquad \forall\ g \in G, t \in T. \tag{24}$$

Objective function (1) maximizes social welfare. In the cases without RTP the demand is not price-responsive, and as such the integral term is fixed. Thus in these cases the objective is equivalent to minimizing generation cost. In the price-responsive cases we assume that the inverse demand functions are non-increasing step functions, meaning that the integral term is a concave piecewise-linear function of $d_{n,t}$. The variable cost functions, $c^v_g(q_{g,t})$, are convex piecewise-linear functions of their arguments. Thus the objective function of the MIP is linear in the decision variables.

Constraint set (2) enforces hourly nodal load-balance constraints. They ensure that total generation at each node exactly equals the sum of demand at the node and net exports out of the node. Constraint set (3) ensures that there are no losses in the transmission network by forcing the net exports from all of the nodes to sum to zero. Constraint sets (4) through (11) ensure that the hourly reserve requirements are met. The model includes two contingency reserve products: 10- and 30-minutes reserves. The amount of 10-minute reserves required in each hour is determined by the largest contingency, which is the generator scheduled to produce the most energy in each hour. The amount of 30-minute reserves required is determined by the second-largest contingency, which is the generator scheduled to produce the second-most energy in each hour. Constraint set (4) defines the amount of 10-minute reserves required in each hour in this way, by forcing ϕ^1_t to equal $\max_{g \in G} q_{g,t}$. Constraint set (5) determines which generator is the largest contingency in each hour—if a generator is not the largest contingency the left-hand side of the inequality is strictly positive, which forces $\xi^1_{g,t}$ to equal zero. Constraint set (6) ensures that exactly one generator is the largest contingency in each hour. Constraint set (7) similarly defines the second-largest contingency. The $\xi^1_{g,t} \cdot \kappa^+_{g,t}$ term on the right-hand side of the inequality ensures that the largest contingency is not also counted as the second-largest. Constraint sets (8) and (9) identify which generator is the second-largest contingency. Constraint set (10) ensures that the sum of scheduled generation and 10-minute reserves less the generation of the largest contingency is sufficient to meet the total scheduled demand. Constraint set (11) enforces a similar restriction for 30-minute reserves.

Constraint sets (12) and (13) enforce the minimum and maximum generation requirements for each generator when it is online. They also constrain energy and reserves provided by a generator that is offline to equal zero. Constraint set (14) enforces ramping limits on how much energy production can be increased or decreased between consecutive hours. Constraint sets (15) and (16) limit the amount of 10- and 30-minute reserves that each generator can provide based on one-sixth and half of the hourly ramping limit. The $\xi^1_{g,t}$ and $\xi^2_{g,t}$ terms on the right-hand sides of these inequalities also prevent the two largest contingencies providing reserves. Constraint set (17) defines generator startups and shutdowns in terms of changes in the value of the $u_{g,t}$ state variables. Constraint sets (18) and (19) enforce the minimum down- and up-time constraints when generators are shutdown or started up. Constraint set (20) enforces the generator must-run requirements.

Constraint set (21) enforces the power flow limits on the transmission elements. Our model uses a linearized dc-approximation of the power flows. Constraint sets (22) and (23) enforce variable non-negativity and constraint set (24) forces the generator state and contingency definition variables to be binary.

2.2. Data

We study the ISO New England system using historical data from the year 2005. The generator and transmission data used in our analysis are based on data obtained from ISO New England for the single day (14 February, 2005) studied by Sioshansi et al. (2008a,b). This data set includes 276 dispatchable generators, six transmission elements, eight load zones, and transmission buses corresponding to each generator modeled. The ISO New England data also include virtual and transaction bids, although we exclude these from our analysis and only model the commitment and dispatch of physical assets. Generators do have the flexibility to adjust the cost and constraint parameters used in the centralized UC, in some cases on a daily basis (some parameters can be adjusted less frequently). We use these parameters from the single day in our analysis due to data availability. Thus our results do not capture any effect of generators adjusting these bids on the resulting commitment and dispatch. Moreover, it is likely that our dataset does not include the full set of dispatchable generators in the system, since some generators may have been offline in February due to planned outages and are thus not included in the model data. Indeed, there are five days of the year during which the system cannot be feasibly dispatched due to binding generator ramping constraints, which is possibly due to the exclusion of such units from our dataset. We include slack variables, which have high objective function penalties, in the load balance constraints to allow the model to find a feasible solution for these days while accounting for intertemporal commitment and dispatch dynamics. We exclude these days, on which the load cannot be feasibly served, from our surplus, price, and demand analysis since LMPs cannot be properly computed due to the arbitrarily high cost coefficients on the slack variables.

In the cases without RTP we use actual load data for the year, obtained from ISO New England, in our analysis. In the cases with RTP we generate inverse demand functions, which are approximated as non-increasing step functions with 100 equal-width segments. Following Borenstein et al. (1997); Sioshansi (2010), the hourly demand functions are calibrated to go through the point defined by the historical hourly load and the retail price of electricity. Retail electricity price data obtained from the United States Department of Energy's Energy Information Administration for each state in the ISO New England service territory are used (each of the demand zones falls within a single state). We only model own-price demand elasticities and consider cases in which the elasticity ranges between -0.10 and -0.30. These values are consistent with short-run elasticities reported by King and Chatterjee (2003).

2.3. Cases Studied

The UC model is formulated using `AMPL` 12.1 and solved using the branch and cut algorithm in `CPLEX` 12.1. All of the default `AMPL` and `CPLEX` settings are used, except for the termination criteria. The default termination criteria used by `CPLEX` is a relative optimality gap[2] of less than 10^{-4}. We consider cases in which the termination criteria is a relative optimality gap of less than 10^{-2}, 10^{-4}, 10^{-6} and 0. The first three cases yield near-optimal solutions, all of which are within 1% of the MIP-optimum, while the last case yields a MIP-optimal solution.

[2]The relative optimality gap is defined as the absolute optimality gap (the difference between the best upper- and lower-bounds on the optimal objective function value) divided by the best lower-bound.

Mimicking actual ISO operations, the UC problem is solved over the year in a rolling fashion one day at a time. Following Sioshansi (2010), we solve for each day's unit commitment and dispatch using a 48-hour optimization horizon. The use of a 48-hour optimization horizon ensures that sufficient generating capacity is left online at the end of each day to meet the following day's loads and minimizes unnecessary generator cycling.

We assume that the ISO settles transactions using uniform prices generated by the UC model. Energy is priced using LMPs, which are determined by the dual variable on the hourly nodal load-balance constraints (constraint set (2) in the model). The 10- and 30-minute reserves are similarly priced based on the dual variable on the hourly reserve requirement constraints (constraint sets (10) and (11) in the model). These dual variables are found by fixing the optimized values of the binary variables and solving the resulting linear program (LP). If we let $\eta_{n,t}$ denote the hourly LMPs and ζ_t^{10} and ζ_t^{30} the hourly marginal prices for 10- and 30-minute reserves, respectively, then the total energy and reserve revenues earned by generator g are given by:

$$R_g = \sum_{t \in T} \left(\eta_{n,t} \cdot q_{g,t} + \zeta_t^{10} \cdot r_{g,t}^{10} + \zeta_t^{30} \cdot r_{g,t}^{30} \right), \tag{25}$$

where n is defined such that $g \in G(n)$. We can also define the as-bid surplus of each generator as the difference between the energy and reserve revenues and the costs incurred by the generator on the basis of the cost bids submitted to the ISO:

$$S_g = R_g - \sum_{t \in T} \left(c_g^v(q_{g,t}) + c_g^n \cdot u_{g,t} + c_g^s \cdot s_{g,t} \right). \tag{26}$$

It is important to stress that S_g does not necessarily equal the actual realized profit of the generator, since the cost data in the generator's bids could be misstated.

An issue with using a linear pricing scheme in a market with non-convexities is that the payments can be economically confiscatory. Gomory and Baumol (1960); Shapley and Shubik (1971); Wolsey (1981); Leonard (1983); Williams (1989); Scarf (1990, 1994); Crema (1995); Williams (1996); Bikhchandani and Mamer (1997) examine the confiscatory nature of linear payments in general markets with non-convexities and O'Neill et al. (2005) consider the specific case of centrally committed electricity markets. Almost all ISOs that operate a centrally committed market, including ISO New England, overcome this confiscation by means of make-whole payments. These are supplemental payments that are given on a daily basis to ensure that no generator has a negative as-bid surplus on any day. If we let $S_{g,d}$ denote the as-bid surplus of generator g on day d, then its make-whole payment on day d is:

$$M_{g,d} = \max\{0, -S_{g,d}\}. \tag{27}$$

The make-whole payment is defined such that if the generator has a negative as-bid surplus, then the payment exactly compensates it for this surplus shortfall. Otherwise, the make-whole payment is zero and the generator earns any inframarginal rents from energy and reserve sales. Our surplus analysis assumes that such a make-whole provision is in place.

3. Results

3.1. Effect of Near-Optimal UC Solutions with Fixed Loads

We first analyze the economic effect of using near-optimal UC solutions if system loads are fixed and do not respond to real-time price signals.

3.1.1. Generator Surplus Differences

In the case with fixed loads there are 25 generators that are not committed or dispatched at any point during the year. Moreover, there are an additional 80 generators that are committed and dispatched identically between the four different sets of UC solutions. Table 1 provides summary statistics of the increase in annual generator surpluses between near- and MIP-optimal solutions (including these 105 generators).

Positive values indicate a surplus increase with the near-optimal solution. The last row of the table provides the coefficient of variation, which is a unitless metric of the dispersion of the surplus increases. The table demonstrates that the findings of Johnson et al. (1997); Sioshansi et al. (2008a) hold over the long term. In some cases generators can lose all of or more than triple the surplus that they would earn in a year if a near- as opposed to MIP-optimal solution is used to operate the system. The table also shows that the surplus differences do not necessarily decrease as the UC problem is solved closer to optimality. Indeed, decreasing the optimality gap stopping criteria from 10^{-2} to 10^{-4} causes the magnitude of the average and maximum surplus differences to increase. The last three columns of table 1 show that generators can have a marginal decrease in surplus per MWh generated. This occurs when a generator is dispatched to generate less energy and provide more reserves, yielding a surplus increase (due to reserve payments) and a generation decrease.

Table 1: Summary statistics of the increase in annual generator surpluses between near- and MIP-optimal solutions with fixed system loads.

	[$]			[%]			[$/MWh]		
	10^{-2}	10^{-4}	10^{-6}	10^{-2}	10^{-4}	10^{-6}	10^{-2}	10^{-4}	10^{-6}
Average	-27,733	43,035	63	11	25	5	146	-27	-1
Minimum	-2,095,886	-780,826	-1,087,571	-100	-100	-100	-2,044	-9,923	-628
Maximum	711,213	5,807,417	1,099,788	1,783	903	522	15,240	5,397	383
cv	6.60	8.48	1,628.78	14.00	4.55	10.34	8.93	44.59	145.69

Table 2 provides further details on the annual earnings of 13 generators that see relatively large surplus changes among the different UC solutions. The first four columns show the annual surplus earned by the generators. The next three show the surplus difference between each near- and the MIP-optimal solution. The last two columns show the surplus range, which is the difference between the highest and lowest surplus earned among the four UC solutions. This range is given in absolute dollars and as a percentage of the MIP-optimal surplus. The table further demonstrates that the surplus earned by individual generators can be highly sensitive to different UC solutions. Generator 285, for example, sees a close to $7 million swing in its annual surplus between the different solutions, which is 81% of its MIP-optimal surplus. The table also shows that the magnitude of surplus differences do not necessarily decrease with smaller optimality gaps.

Table 2: Annual surplus of a select group of generators with fixed system loads.

Unit	Surplus [$ 000]				Surplus Increase [$ 000]			Surplus Range	
ID	10^{-2}	10^{-4}	10^{-6}	0	10^{-2}	10^{-4}	10^{-6}	[$ 000]	[%]
070	24,421	25,010	25,084	25,426	-1,006	-417	-342	1,006	4
088	3,544	4,017	3,709	3,560	-16	457	149	474	13
090	5,746	5,852	5,907	6,043	-297	-191	-136	297	5
091	6,646	6,737	6,810	6,946	-300	-209	-136	300	4
123	3,658	3,775	3,724	3,580	79	195	145	195	5
133	8,238	8,217	8,141	7,934	304	284	208	304	4
224	1,199	1,552	1,232	1,202	-4	350	30	353	29
235	5,489	4,868	4,774	4,778	711	91	-3	715	15
261	485	202	29	26	460	177	3	460	1,783
274	204	137	84	84	120	53	0	120	143
285	7,327	13,884	9,177	8,077	-750	5,807	1,100	6,557	81
327	3,251	3,351	3,199	3,111	140	240	88	240	8
328	2,532	2,676	2,466	2,440	92	237	27	237	10

Figure 1 shows the cumulative surplus increase of seven generators between solutions with an optimality gap of 10^{-4} and 0. The figure shows that while some generators, such as units 088, 133, 224, 327, and 328, see gradual surplus differences, others, such as 070 and 091, can see very drastic differences on a single day.

For instance, generator 070 sees more than a \$234,000 surplus decrease between the solutions with optimality gaps of 10^{-4} and 0 on 16 September. This is its largest single-day surplus decrease and represents about 56% of the total surplus loss that it experiences over the year. The unit is dispatched nearly identically between the two solutions—it produces 11,693 MWh of energy on this day in the solution with an optimality gap 10^{-4} as opposed to 11,805 MWh in the MIP-optimal solution. The surplus difference is because the LMP that it receives in hour 15 is \$502/MWh in the MIP-optimal solution as opposed to only \$60/MWh in the near-optimal solution. These LMP differences are due, in turn, to a different set of generators with different fixed and variable generation costs being committed and marginal in the two sets of solutions—the MIP-optimal solution commits a generator with lower fixed and higher variable costs in hour 15. While there are LMP differences in the other hours of the day, these are smaller and average about \$2/MWh in magnitude and only yield a \$9,000 surplus reduction.

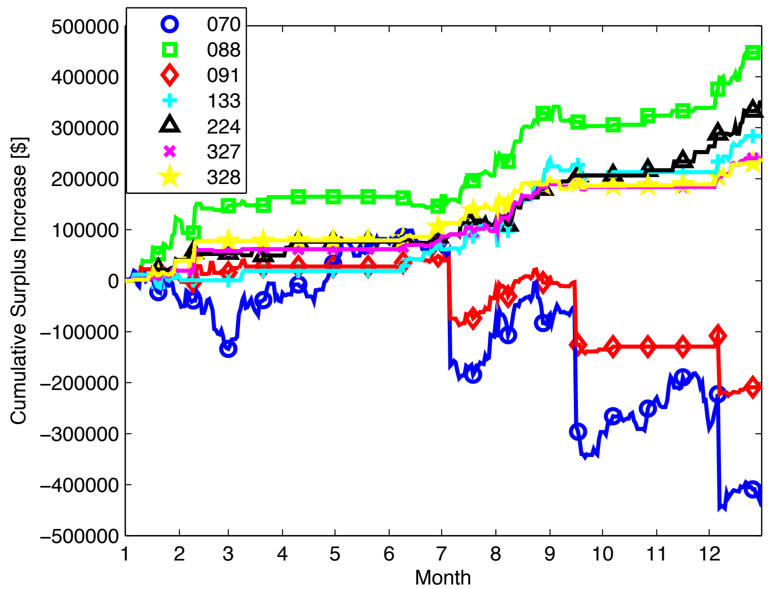

Figure 1: Cumulative surplus increase of seven generators between UC solutions with an optimality gap of 10^{-4} and 0 with fixed system loads.

3.1.2. Effect of Generator Characteristics on Surplus Differences

Figure 1 also shows that most of the surplus differences for this set of generators occur in the second half of the year, particularly in the summer. Table 3 shows the standard deviation of the total daily generator surplus differences between each near- and the MIP-optimal solution during each quarter of the year. The table shows that all of the near-optimal solutions yield the greatest surplus variability during the summer quarter. This suggests that the high system loads during the summer contributes to the surplus differences. The commitment of peaking generators, which tend to have high operational flexibility (*e.g.* high ramp rates, low minimum generating loads, and low minimum up-times), low startup and no-load costs, and high variable generation costs, can be sensitive on high-load days. On such days the ISO often chooses between running a peaking generator for an hour or two to meet the load peak, or starting up a less flexible generator with higher fixed operating costs earlier in the day to ramp-up to meet the peak. Since peaking generators may only run during a handful of hours of a year, the ISO choosing the less flexible generator on a single high-load day can drastically affect the peaking generator's annual surplus.

We use a regression model to test for such a relationship between surplus variability and generator characteristics. The model has the form:

$$y_g = X_g^\top \beta + \epsilon_g, \tag{28}$$

Table 3: Standard deviation of total daily generator surplus differences between near- and MIP-optimal solutions during each quarter of the year with fixed system loads.

	Optimality Gap		
Quarter	10^{-2}	10^{-4}	10^{-6}
1	348,313	309,985	310,921
2	318,191	245,806	226,677
3	569,281	473,439	367,816
4	452,033	327,336	279,332

where y_g is the standard deviation of the total annual surplus earned by generator g between the four sets of UC solutions, X_g is a column vector of explanatory variables, and ϵ_g is an error term. The explanatory variables included in the model are the minimum up-time, maximum ramp-up rate, spinning no-load cost, and minimum and maximum generating capacity of each generator. The model is estimated using ordinary least squares (OLS), and visual inspection and correlation analysis of the residuals do not reveal endogeneity in the model. Table 4 shows the OLS estimates of the model, with starred values denoting estimates that are significant at the 99% level. The table shows that generators with greater operating flexibility, *i.e.* lower minimum generating capacities and up-times, and higher ramping rates, tend to see higher surplus variability. Generators with lower spinning no-load costs see high surplus variability as well. Surplus variability also increases with the maximum capacity of a generator, since larger generators tend to earn more surplus, which gives higher absolute differences between the solutions.

Table 4: OLS estimates of regression model (28). Regression R^2 is 0.6376. Starred values are significant at the 99% level.

Variable	Coefficient
Constant	-16,386
Minimum up-time	-4,149 *
Maximum ramp-up	65 *
Spinning no-load cost	-81 *
Minimum generating capacity	-1,013 *
Maximum generating capacity	1,502 *

3.1.3. Price Differences

The UC model endogenously generates a set of hourly LMPs for energy and hourly prices for contingency reserves. The model also generates a set of hourly slack bus prices for energy. The slack bus is a fictitious bus added to the transmission network to model power flows using the power transfer distribution factors. The slack bus price is defined as the dual variable on constraint (3) in the LP relaxation of the UC model given when the binary variables are fixed to their optimized values. This dual variable can be viewed as giving the marginal price of energy in each hour, without pricing transmission congestion. Indeed, LMPs can be defined as the sum of the slack bus price and a congestion cost. Figure 2 is a histogram of differences in the slack bus price between each near- and the MIP-optimal solution. The figure shows that these price differences tend to be quite small on average. There is, for instance, zero price difference between solutions with an optimality gap of 10^{-2}, 10^{-4}, and 10^{-6} and the MIP-optimal solution in 2,978, 3,355, and 3,568 hours of the year, respectively. There are, however, hours with extremely large price differences of up to \$130/MWh in magnitude.

Figure 3 is a histogram of differences in the maximum congestion cost between each near- and the MIP-optimal solution. We define the maximum congestion cost in hour t as:

$$\max_{n \in N} \eta_{n,t} - \min_{n \in N} \eta_{n,t}, \tag{29}$$

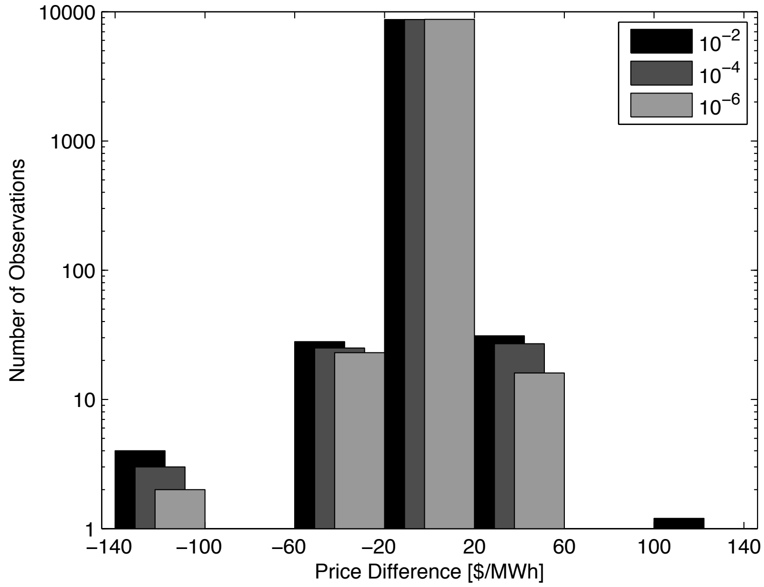

Figure 2: Histogram of difference in price of energy at slack bus between near- and MIP-optimal solutions with fixed system loads.

which is the difference between the highest and lowest LMP in the hour. Comparing figures 2 and 3 shows that congestion costs are considerably more sensitive to different UC solutions than the slack bus price, suggesting that there can be locational differences in the effect of near-optimal UC solutions on prices. Among the eight load zones, the Connecticut zone tends to have the highest LMPs—indicating that this zone lacks low-cost generation and that binding transmission constraints can limit energy imports. This zone also sees the most extreme congestion cost differences between the near- and MIP-optimal solutions. The other zones see congestion cost differences of -$33/MWh to $32/MWh between the near- and MIP-optimal solutions, whereas these differences range between -$436/MWh and $428/MWh in the Connecticut zone. The standard deviation of the hourly congestion differences between the near- and MIP-optimal solutions is $9/MWh in the Connecticut zone, as opposed to $3/MWh at the other zones. Taken together, these results show that customers and generators in the more congested Connecticut zone can expect greater energy price variance due to the use of near-optimal UC solutions.

3.2. Effect of Near-Optimal UC Solutions with Price-Responsive Loads

An economic rationale behind RTP is that electricity demand will better reflect consumer preferences since it responds to real-time marginal cost signals. Our results in section 3.1.3 show that LMP differences between near- and MIP-optimal UC solutions can be high, potentially affecting the efficiency of this rationing. When demand elasticity is incorporated into the model, differences in the total annual load between the near- and MIP-optimal solutions are relatively small at less than 0.25%. Moreover, the magnitude of these differences decreases as the UC problem is solved closer to optimality. Nevertheless, the system load in individual hours can vary more significantly between near- and MIP-optimal solutions, indicating that demand in a given hour does not optimally react to price signals.

Figure 4 is a scatter plot of the difference in hourly system loads between near- and MIP-optimal solutions and demonstrates this. The horizontal axis shows the hourly difference in load between MIP-optimal solutions with demand elasticities of -0.1 and zero, as a percentage of the inelastic demand. This axis reflects the MIP-optimal amount of demand rationing. The vertical axis shows the hourly difference in load between each near-optimal solution with a demand elasticity of -0.1 and the MIP-optimal solution with a demand elasticity of zero, also as a percentage of the inelastic demand. Thus points that do not lie on the 45° line indicate hours in which demand is not fully efficiently rationed if the UC problem is not solved to

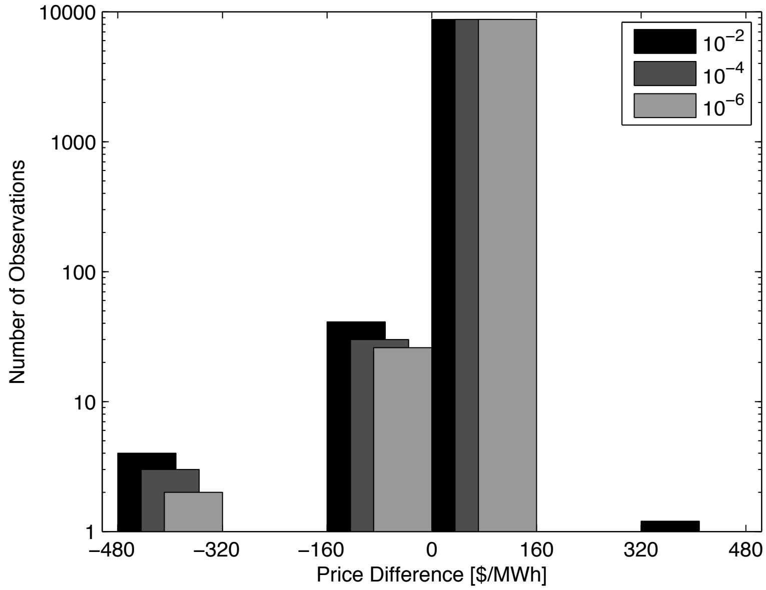

Figure 3: Histogram of difference in maximum congestion cost between near- and MIP-optimal solutions with fixed system loads.

complete optimality. Figure 4 shows that solving the UC problem closer to optimality tends to reduce these hourly demand differences, since the scatter plot for the solutions with optimality gaps of 10^{-4} and 10^{-6} are more tightly clustered around the 45° line. The figure also shows that loads with an optimality gap of 10^{-2} tend to be less than the MIP-optimum, especially during hours in which the MIP-optimal demand increases relative to the fixed-load case.

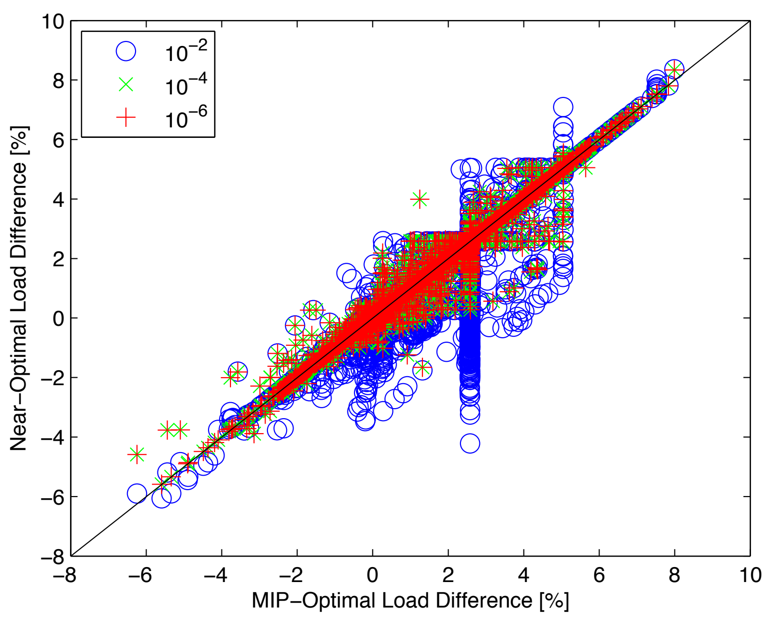

Figure 4: Scatter plot of difference in hourly system load under MIP-optimal solution against difference in system load under near-optimal solutions with a demand elasticity of -0.1.

Introducing demand elasticity tends to reduce price differences between near- and MIP-optimal UC solutions. This is shown in table 5 which provides summary statistics of slack bus price and maximum congestion cost differences between near- and MIP-optimal solutions for the different elasticity cases modeled. The table shows that both the range and standard deviation of the price differences are significantly reduced as a result of demand elasticity. Furthermore, the magnitude of the price differences are decreasing in the elasticity of demand.

Table 5: Summary statistics of slack bus price and maximum congestion cost differences [$/MWh] between near- and MIP-optimal solutions.

	Demand Elasticity			
	0	-0.1	-0.2	-0.3
Slack Bus Price				
Minimum	-130	-11	-12	-11
Maximum	128	26	20	17
Standard Deviation	4.08	1.29	1.12	1.12
Maximum Congestion Cost				
Minimum	-436	-40	-29	-20
Maximum	428	86	37	37
Standard Deviation	8.45	1.74	0.82	0.65

4. Conclusions

This paper expands upon the analysis of Johnson et al. (1997); Sioshansi et al. (2008a) and examines price and cumulative generator surplus differences over a one-year period between UC solutions that are solved to near-optimality. We demonstrate that the surplus differences arising from the use of a near-optimal UC solution do not necessarily zero out in the long-run, and that individual generator profits can be significantly increased or decreased in a seemingly arbitrary fashion. This is because commitment decisions can be persistent and a generator that is committed on one day can be more likely to be online the following day. This persistence is due to generator cycling costs and intertemporal operating constraints. We further show that solving the UC problem closer to optimality will not generally reduce the magnitude of these differences.

Using a regression model we demonstrate that generator operating and cost characteristics that are common to peaking and highly flexible generators tend to increase surplus variability between different near-optimal solutions. This finding suggests potential short- and long-run market inefficiencies. Flexible generators may, in the short-run, be incentivized to misstate cost and operating characteristics to the ISO in order to reduce profit variability and risk. This could reduce the efficiency of the centralized commitment, since the ISO would be optimizing the system using incorrect generator information. In the long-run, the greater surplus variability may deter investment in peaking and flexible generation, due to the inherent market revenue risk. This longer-term investment issue is especially important with higher penetrations of non-dispatchable renewable resources, which increase the need for system flexibility.

We also show that market prices can vary considerably between near- and MIP-optimal solutions. Importantly, we find that the more congested Connecticut zone sees greater price and congestion cost variability. This implies that loads and generators in congested regions of a power system are subjected to greater price risk than those elsewhere. Moreover, this price variability can raise the same issues related to long-term investments that are spurred by price signals. Namely, the variability in congestion costs can make a merchant transmission investment riskier and may not provide adequate signals for generation to be built within the zone. While the total annual system load does not vary considerably when demand response is introduced, we show that the demand in individual hours can deviate from the MIP-optimum, implying inefficient rationing.

Although our model and case study are representative of how centrally committed markets are operated and settled, we make some simplifying assumptions that potentially understate the effect of relying on near-optimal solutions. One is that UC models can include more complex modeling of generator states. This can include startup costs that depend on how long a generator has been offline, more complex state variables that better represent combined cycle units, or interdependencies between cascaded hydroelectric plants. Including such constraints can result in the UC model being more computationally complex, increasing the candidate set of near-optimal solutions and the resulting price and surplus variability. A second simplification is that we only consider a linearized dc load flow model that is directly embedded in the UC problem. ISOs typically only model dc load flows in their UC problems, since ac load flows would yield a non-linear MIP with a non-convex continuous relaxation, which is intractable. Instead, the solution generated by the UC problem is tested is an ac load flow model to ensure that the resulting flows are truly feasible. If not, the flow limits in the UC model are adjusted and the model resolved. The ISO iterates between the two models until arriving at a feasible solution. Because this iteration process can be time-consuming, ISOs may use conservative flow limits in the UC problem to reduce the required computational time. When the system is highly congested, these more conservative limits may yield a solution that is considerably different from what using the actual flow limits from the ac model would result in. Since we only model dc load flows, the MIP-optimal solution that we benchmark against is unlikely to be the truly optimal solution. Moreover, even if an ISO could solve a UC model with dc load flows to optimality, the same type of issues could arise due to the linearization of the power flows.

Taken together, these issues call into question the strength of price signals generated within a centrally committed market. Although the prices reflect the marginal cost of energy for a given commitment and approximation of the transmission network, the commitment and prices generated by the UC model are quite sensitive to the particular approximation and near-optimal solution used. Importantly, our findings suggest that a centrally committed market that is not solved to optimality will produce highly variable scarcity pricing signals. High LMPs at a particular load zone or bus are intended to signal the need for more transmission of generation at particular parts of the network. Similarly, a peaking generator being committed and setting a high marginal price indicates the need for more generation capacity. Our results show that these types of price signals are the ones that are most variable and sensitive to the near-optimal solution used, calling the value of the price signals into question.

While solving a UC problem that accounts for ac load flows to complete optimality will eliminate these pricing issues, this is not a practical solution. Alternatively, a decentralized market design could surmount these problems raised by centralized commitment. Such designs typically rely on generators to individually make unit commitment decisions. Generators then trade energy, reserve, capacity, and other products among themselves and with load-serving and other entities, either bilaterally or through a spot market. Indeed, Wilson (1997); Elmaghraby and Oren (1999) suggest that unit commitment decisions would be more efficiently made under such a decentralized design. As such, the restructured California and Texas markets were originally based on such a design (Sweeney (2006); Adib and Zarnikau (2006) detail the original designs of the California and Texas markets) and the Australian National Electricity Market is a hybrid between a centralized and decentralized market (Moran (2006) describes the Australian market). The original California and Texas market designs called for generators and load-serving entities (LSEs) to bilaterally trade energy and capacity products, and the ISO to solve a load flow problem with incremental and decremental generation offers solicited from market participants to ensure feasibility of the power flows. Although ac load flow models raise computational issues, they are easier to solve than a UC problem and the ISO would be more likely to find an optimal solution, eliminating the pricing issues inherent in centralized commitment. Moreover, proponents claim that decentralized commitment allows for greater flexibility than a regimented centralized commitment with fixed day- and hour-ahead market windows. For instance, a decentralized design may be able to better accommodate variable renewables, as generators and LSEs can adjust sale and purchase positions as more updated resource forecasts become available. Centralized markets may not be nimble enough to make such accommodations, since transactions are centrally cleared by the ISO.

Despite these benefits, decentralized commitment suffers from loss of coordination among generators. Indeed, Ruff (1994); Hogan (1994, 1995); Hunt (2002) support centrally committed markets because they claim that it is most efficient to have the ISO, which has the best information about the power system as a whole, make commitment and dispatch decisions. Sioshansi et al. (2008b, 2010) use a numerical example based on the same one-day ISO New England dataset used here and stylized examples to compare the productive efficiency and settlement cost of markets that rely on centralized and decentralized commitment. They demonstrate that even with perfect competition, the loss of coordination in commitment decisions due to a decentralized design can decrease efficiency by 4% and increase settlement costs by 85%. Sioshansi and Nicholson (2011) further use a stylized symmetric duopoly model to examine the effects of relaxing the perfect competition assumption, showing that the decentralized design can be more costly than the centralized, depending on the Nash equilibrium that the generators follow. Given these pros and cons between the two market designs, it is difficult to definitively state that one design is superior to another.

References

Adib, P., Zarnikau, J., 2006. Texas: The most robust competitive market in north america. In: Sioshansi, F. P., Pfaffenberger, W. (Eds.), Electricity Market Reform: An International Experience. Elsevier.

Bikhchandani, S., Mamer, J. W., June 1997. Competitive equilibrium in an exchange economy with indivisibilities. Journal of Economic Theory 74, 385–413.

Borenstein, S., January 2002. The trouble with electricity markets: Understanding california's restructuring disaster. The Journal of Economic Perspectives 16, 191–211.

Borenstein, S., 2005. The long-run efficiency of real-time electricity pricing. The Energy Journal 26, 93–116.

Borenstein, S., Bushnell, J. B., Knittel, C. R., 1997. A Cournot-Nash equilibrium analysis of the New Jersey electricity market. Appendix A of Review of General Public Utilities Restructuring Petition, Final Report, New Jersey Board of Public Utilities, docket Number EA97060396.

Borenstein, S., Holland, S. P., Autumn 2005. On the efficiency of competitive electricity markets with time-invariant retail prices. The RAND Journal of Economics 36, 469–493.

Bushnell, J. B., Stoft, S. E., July 1996. Electric grid investment under a contract network regime. Journal of Regulatory Economics 10, 61–79.

Bushnell, J. B., Stoft, S. E., March 1997. Improving private incentives for electric grid investment. Resource and Energy Economics 19, 85–108.

Chao, H.-P., Peck, S., July 1996. A market mechanism for electric power transmission. Journal of Regulatory Economics 10, 25–59.

Crema, A., 21 September 1995. Average shadow price in a mixed integer linear programming problem. European Journal of Operational Research 85, 625–635.

Elmaghraby, W., Oren, S. S., October 1999. Efficiency of multi-unit electricity auctions. The Energy Journal 20, 89–115.

Gomory, R. E., Baumol, W. J., July 1960. Integer programming and pricing. Econometrica 28, 521–550.

Guan, X., Zhai, Q., Papalexopoulos, A., 13-17 July 2003. Optimization based methods for unit commitment: Lagrangian relaxation versus general mixed integer programming. In: Power Engineering Society General Meeting. Institute of Electrical and Electronics Engineers, Toronto, Ontario, Canada.

Hobbs, B. F., Rothkopf, M. H., O'Neill, R. P., Chao, H.-P. (Eds.), 2001. The Next Generation of Electric Power Unit Commitment Models. Kluwer, Norwell, Massachusetts.

Hogan, W. W., September 1992. Contract networks for electric power transmission. Journal of Regulatory Economics 4, 211–242.

Hogan, W. W., 1994. An Efficient Bilateral Market Needs a Pool. Testimony Before the California Public Utilities Commission.

Hogan, W. W., March 1995. Coordination for competition in an electricity market, working Paper.

Hunt, S., 2002. Making Competition Work in Electricity. John Wiley & Sons, New York, New York.

Johnson, R. B., Oren, S. S., Svoboda, A. J., March 1997. Equity and efficiency of unit commitment in competitive electricity markets. Utilities Policy 6, 9–19.

Joskow, P. L., Schmalensee, R., 1983. Markets for Power An Analysis of Electric Utility Deregulation. The MIT Press, Cambridge, Massachusetts.

King, C. S., Chatterjee, S., July 2003. Predicting California demand response. Public Utilities Fortnightly, 27–32.

Leonard, H. B., June 1983. Elicitation of honest preferences for the assignment of individuals to positions. Journal of Political Economy 91, 461–479.

Moran, A., 2006. The electricity industry in australia: Problems along the way to a national electricity market. In: Sioshansi, F. P., Pfaffenberger, W. (Eds.), Electricity Market Reform: An International Experience. Elsevier.

Muckstadt, J. A., Koenig, S. A., May-June 1977. An application of lagrangian relaxation to scheduling in power-generation systems. Operations Research 25, 387–403.

O'Neill, R. P., Sotkiewicz, P. M., Hobbs, B. F., Rothkopf, M. H., Stewart, W. R., 1 July 2005. Efficient market-clearing prices in markets with nonconvexities. European Journal of Operational Research 164, 269–285.

Ruff, L. E., June 1994. Stop wheeling and start dealing: Resolving the transmission dilemma. The Electricity Journal 7, 24–43.

Scarf, H. E., May-June 1990. Mathematical programming and economic theory. Operations Research 38, 377–385.

Scarf, H. E., Autumn 1994. The allocation of resources in the presence of indivisibilities. The Journal of Economic Perspectives 8, 111–128.

Shapley, L. S., Shubik, M., December 1971. The assignment game I: The core. International Journal of Game Theory 1, 111–130.

Sioshansi, R., April 2010. Evaluating the impacts of real-time pricing on the cost and value of wind generation. IEEE Transactions on Power Systems 25, 741–748.

Sioshansi, R., Nicholson, E., August 2011. Towards equilibrium offers in unit commitment auctions with nonconvex costs. Journal of Regulatory Economics 40, 41–61.

Sioshansi, R., O'Neill, R. P., Oren, S. S., May 2008a. Economic consequences of alternative solution methods for centralized unit commitment in day-ahead electricity markets. IEEE Transactions on Power Systems 23, 344–352.

Sioshansi, R., Oren, S., O'Neill, R., December 2010. Three-part auctions versus self-commitment in day-ahead electricity markets. Utilities Policy 18, 165–173.

Sioshansi, R., Oren, S. S., O'Neill, R. P., 2008b. The cost of anarchy in self-commitment based electricity markets. In: Sioshansi, F. P. (Ed.), Competitive Electricity Markets: Design, Implementation & Performance. Elsevier.

Streiffert, D., Philbrick, R., Ott, A., 12-16 June 2005. A mixed integer programming solution for market clearing and reliability analysis. In: Power Engineering Society General Meeting, 2005, IEEE. San Francisco, CA.

Sweeney, J. L., 2006. California electricity restructuring, the crisis, and its aftermath. In: Sioshansi, F. P., Pfaffenberger, W. (Eds.), Electricity Market Reform: An International Experience. Elsevier.

Williams, A. C., May 1989. Marginal values in mixed integer linear programming. Mathematical Programming 44, 67–75.

Williams, H. P., August 1996. Duality in mathematics and linear and integer programming. Journal of Optimization Theory and Applications 90, 257–278.

Wilson, R. B., 14 March 1997. Activity rules for a power exchange. In: Proceedings of the 2nd Annual POWER Conference of the University of California Energy Institute. Berkeley, California.

Wolsey, L. A., November 1981. Integer programming duality: Price functions and sensitivity analysis. Mathematical Programming 20, 173–195.