POWER

PWP-033

Dynamic Efficiency and the Regulated Firm: Evidence from Interfirm Trade in Electricity Markets

Matthew W. White

September 1995

This paper is part of the working papers series of the Program on Workable Energy Regulation (POWER). POWER is a program of the University of California Energy Institute, a multicampus research unit of the University of California, located on the Berkeley campus.

University of California Energy Institute
2539 Channing Way
Berkeley, California 94720-5180
www.ucei.berkeley.edu/ucei

Dynamic Efficiency and the Regulated Firm: Evidence from Interfirm Trade in Electricity Markets

Matthew W. White*

Stanford University

September 1995

Abstract

This paper presents an empirical analysis of the value of a coordinated market exchange mechanism. I present a model of efficient trading mechanisms under uncertainty, and develop a measure of the value of an interfirm trade agreement in the context of sequential 'buy-versus-produce' decision-making by firms. The theory is applied to estimate the value of a formal trading institution in the California electricity market, where an interutility *power pool* has been proposed to restructure the electric power industry. I develop an empirical model of the optimal production and trading decisions for a firm in such a pool, and estimate state-contingent willingness-to-trade functions for each of the four major utilities in this market. With this information, I estimate the distribution of future costs that would obtain if an efficient exchange mechanism arbitraged away observed differences between willingness-to-buy and willingness-to-sell among the sample firms. The principal finding is that with the simple, relatively state-independent bilateral contracts observed in this market, the sample firms achieve within 4% of the theoretical minimum expected costs available with a complete state-contingent exchange mechanism. This difference represents an opportunity cost of approximately $250 million per year. I conclude with regulatory and managerial explanations for the absence of a more efficient state-contingent trading mechanism, and implications for deregulating electric power markets.

*Assistant Professor of Economics, Graduate School of Business, Stanford University, Stanford CA 94305-5015. E-mail: white@gsb.stanford.edu. Thanks to Severin Borenstein, Richard Gilbert, Alvin Klevorick, Daniel McFadden, Ariel Pakes, and John Rust for helpful suggestions and discussions, as well as seminar participants at Berkeley, Harvard, Michigan, Northwestern, Stanford, Wisconsin, and Yale. Financial support from the University of California Energy Institute is gratefully acknowledged.

1. Introduction

Despite the fundamental role of complete contingent contract markets in neoclassical equilibrium theory, economists are hard-pressed to identify a market organized as such. The absence of markets with this ideal organization of exchange is commonly attributed to the complexity of the set of possible future events and the costs of enumerating and contracting upon them. Such *transaction costs* then explain the decision of optimizing economic agents to forsake complete contingent contracts as a basis for exchange; in this sense, transaction costs are the opportunity cost of incomplete contracts. Beginning with Coase (1937), economists have made great strides in identifying the sources and implications of these transaction costs of exchange. Indeed, the analysis of economic institutions is at its core an attempt to determine an efficient organization of exchange—of goods, of information, of whatever—in the absence of complete contingent contracts. Nevertheless, we have made disappointingly little progress in quantifying the magnitude of transaction costs in empirical settings. The opportunity costs of incomplete contracts, measured in meaningful economic units, remains largely an open question.

This paper addresses this question in the context of interfirm trade. Interfirm trade agreements are contractual mechanisms designed to reduce the transaction costs of trade, enhancing the ability of firms to make efficient 'buy-versus-produce' decisions. In contrast to the predictions of standard cartel theory, however, the value of coordinating production behavior lies not in the ability to manipulate revenues, but rather in the ability to effect mutually beneficial reductions in costs. Examples include the payment clearing systems implemented among banks and financial institutions; passenger swaps among airlines, which mitigate the costs of overbooking; and the routing of rival firms' calls by major telecommunications carriers to reduce network congestion. These agreements represent formal *trade mechanisms:* A set of well-specified, state-contingent transaction and settlement rules governing future exchange. This paper addresses the value of complexity in an interfirm trading mechanism, where complexity is understood as the richness of the class of conditions identified ex ante by transacting parties. The objective is to quantify the opportunity cost when agents base exchange on simple, state-independent contracts, rather than specifying contingent contracts for numerous foreseeable but unlikely events affecting the value of subsequent exchange.

The immediate motivation for this paper stems from a curious institutional phenomenon in the organization of electric power markets. In most of the United States, these (wholesale) markets are decentralized: utilities buy and sell power from one another under negotiated bilateral agreements, in order to reduce the cost of providing power to their retail customers. But in some regions of the country, firms have voluntarily established coalitions with multilateral trading agreements. These coalitions, or *pools*, adopt sophisticated market-clearing mechanisms to implement efficient exchange among member firms. The role of the pool in clearing the market is essentially that of a state-contingent "price-caller," analogous to the classic Walrasian auctioneer in an Arrow-Debreu exchange economy.

These pooling agreements are a substitute for the more prevalent bilateral trading relationships among firms, presenting us with two different organizational forms of market exchange in the same industry. This heterogeneity is somewhat perplexing: If these pooling agreements are more efficient, why aren't they the ubiquitous form of market organization? Managers of firms with pooling agreements predictably state that these agreements make it easier for their firm to buy from others, as an alternative to more costly internal production. If such agreements are optimal solutions to exchange among some firms, however, why are they not for similarly situated others?

This issue is of pressing interest as a number of states, most notably California, are on the verge of substantially deregulating the electric power industry. A leading proposal calls for existing vertically-integrated electric utilities to divest control of generation and transmission, for the deregulation of the resulting new generation companies, and for the establishment of an independent, regional pooling institution to operate a spot market and to insure delivery by operating the utilities' transmission assets.[1] By creating a standard commodity market for electricity, such an institution could, in principle, function at least as well as the elaborate systems for multilateral arbitrage in the existing 'tight' power pools in the Northeast U.S. Both of these market institutions stand in sharp contrast to the current organization of interfirm trade in California (and most of the United States), where firms engage in exchange principally via relatively simple, long-term bilateral contracts. One of the key issues that has emerged in the (de)regulatory debate is the relative merits of these forms of market organization.

This begs the difficult question of *measuring* comparative organizational efficiency: How well do these multilateral trading institutions perform, relative to the decentralized, bilateral market? Toward an answer, this paper presents a model of interfirm trade, and an empirical framework for estimating the value of an interfirm pooling agreement. From an empirical perspective, the difficulty in estimating the value of a market institution like a power pool is that these institutions induce a structural change in the behavior of firms. Specifically, the trading process implemented among firms is an agreement to share the capacity for production (albeit with time-varying transfer prices), so that pooling through an interfirm trading mechanism effectively corresponds to a technological change in each firm's minimum

[1] See, e.g., Southern California Edison Company's "PoolCo" proposal (SCE (1994)) or Wisconsin Electric's vertical divestiture discussion (Abdoo (1995)).

cost function.[2]

The implication here is that classical econometric approaches to estimating firm-level production and cost functions (e.g., Christensen and Greene (1976, 1978), Fuss and McFadden (1978)) cannot identify the gains attributable to participation in a pooling agreement. In any sample of firms, the data reflect the historical self-selection of some firms into pooling agreements and others into their absence. Reduced-form estimates from such data cannot inform the counterfactual. To make such an inference, one needs a theory of how pooling agreements reallocate production among participants, and from this an empirical analysis of how production behavior would be modified by the resulting production possibilities set.

This paper pursues such an approach, providing a behaviorally-consistent empirical method that does not impose upon the data the assumption that an efficient mechanism already exists. Specifically, this paper develops an empirical model that allows us to determine a sequence of equilibrium production and exchange strategies for each firm, given the different trading opportunity sets (or *regimes*) that each firm might face. This empirical model is applied to assess the value of an efficient pooling mechanism in the California electricity market, and the extent to which existing trading behavior is consistent with a hypothesis of efficient trade.

Section 2 begins with a theory of efficient trading mechanisms under uncertainty. The objective is to develop a measure of the value of an interfirm pooling agreement in the context of sequential 'buy-versus-produce' decision-making by firms. I argue that the value of a such an agreement lies in the ability of participating firms to reduce (in a first-degree stochastic-dominance sense) their ex ante distribution of future production costs. Section 3 presents an empirical model of this stochastic optimization problem as a Markovian decision process. This structural approach to each firm's intertemporal production problem allows one to estimate the distribution of future production costs with and without an efficient pooling mechanism.

Section 4 applies this model to a set of interconnected electrical utilities in the California electricity market. Using data on existing power exchange contracts and detailed demand, cost, and technology data for each firm, I estimate production cost distributions under existing trading practices and construct willingness-to-buy and willingness-to-sell functions for each firm in this market. This process involves solving the firms' production problems and computing each firm's optimal decision rules, a set of state-contingent production and trading policies. With this information, I estimate the distribution of future costs that would obtain if a state-contingent exchange mechanism efficiently arbitraged away observed differences between willingness-to-buy and willingness-to-sell among the sample firms.

[2] Joskow and Schmalensee (1983) observe that pools in the electric power industry are, to varying degrees, attempts to capture economies of scale without the attendant costs of horizontal integration.

The principal finding is that the difference between these two distributions is both statistically and economically significant, and on average represents approximately 4% of the total costs of production for these firms. Section 5 presents these results in detail, and Section 6 explores regulatory and managerial explanations for why these firms might fail to capture these apparent gains from trade. In particular, I argue that there are transaction costs of capturing these efficiency gains stemming from the dynamic problem of *regulatory commitment:* The managers of these firms should reasonably expect that regulators will reduce their future revenues in response to activities undertaken to lower costs, and the regulators cannot commit to do otherwise. Hence, the firms "underinvest" in cost-saving activities, anticipating an expropriative *ratchet effect* in future regulatory proceedings.[3] I conclude in Section 7 with some implications for restructuring the organization of these markets.

2. Pooling Agreements: Theory

Consider a simple multi-period model of interfirm trade. In each period, each of n firms in a market realizes demand from consumers, and an attendant cost of production. Given (potential) revenue as a pre-determined function of consumer demand, each firm then considers how much to produce, buy, and sell. After executing trades with one another, the firms deliver the product to final consumers. We will assume that unserved demand is lost and that output must be sold in the period it is produced, but that current production costs may depend on past production decisions.

Formally, let $N = \{1, 2, \ldots, n\}$ denote the set of firms in our economy, indexed by i. At each time t, $t \in \mathbb{Z}$, firm i faces demand $y_i(t)$ where $y_i(t) \in Y \subset \mathbb{R}_+$. Each firm's production technology is characterized by a minimum cost function $c_i : Y \times U \to \mathbb{R}_+$, where $u_i(t) \in U$ is a set of time-varying, random cost components specific to firm i. We make the regularity assumptions that $U \subset \mathbb{R}^K$ for some $K < \infty$ and that for each $i \in N$, c_i is bounded (and Borel measurable) on $Y \times U$. We will take consumer demand to be exogenous; for our purposes, the determination of consumer prices is peripheral and is left outside the model.

The uncertainty in each firm's technology, and therefore in marginal costs, suggests a potential incentive for the firms to trade amongst themselves ex post and prior to delivery. In the presence of transaction costs (of any sort), however, the firms have an incentive to contract ex ante; that is, to agree to transaction and settlement rules governing execution of future exchange. We formalize such rules as a *state-contingent trade mechanism.*

[3] This "ratchet effect" is a subject of much attention in the incentive regulation and repeated agency literatures, although measures of its empirical significance remain to be established. See, e.g., Laffont and Tirole (1993) for a survey of the effect of commitment in dynamic agency problems.

Definition 1: A *trade mechanism*, $M = \{g, r\}$, is a trading rule $g : (Y \times U)^n \rightarrow Y^n$ and a transfer policy $r : Y^n \times (Y \times U)^n \rightarrow \mathbb{R}^n$, contingent upon each participating firm's demand and cost realizations.[4]

We can view the implementation of a trading mechanism as a "pre-market" phase in which, at the close of some period t, (i) each player chooses whether to participate in a production "pool," a set $Q \subseteq N$ to be governed by some trade mechanism, and (ii) the players that decide to participate then choose a trade mechanism. Normalizing this "pre-market" time period to 0, participating firms produce according to the mechanism rather than their individual demand in each period $t > 0$. In (a perfect Bayesian) equilibrium, the ex ante allocation of revenue in the pool must make participation individually rational for each joining firm; moreover, the decision not to participate by any firm must be optimal given the mechanism subsequently chosen by the group.

In order for a mechanism to be feasible, we require a trade mechanism to satisfy a balanced-budget condition: it must clear the market in each period, and it cannot violate the total revenue endowment of the pool.

Definition 2: A mechanism M achieves a *balanced-budget* if for each t, almost surely

(i) $\sum_{i \in Q} g_i(y(t), u(t)) - y_i(t) = 0$ (*market clearing condition*)

(ii) $\sum_{i \in Q} r_i(g; y(t), u(t)) = 0$ (*no subsidy condition*)

where $y(t) \in Y^n$ denotes the n-vector of demand, and $u(t) \in U^n$ the set of nK disturbances at time t.

In the absence of a state-contingent trade mechanism, we allow that an individual firm i may engage in exchange with others, provided that i's resulting production $y_i^*(t)$ is feasible ($y_i^*(t) \in Y$ and $\{y_i^*(t)\}_{i \in N}$ clear the market) and that the corresponding net benefit afforded each firm, $w_i(t)$, is non-negative. We will refer to the market output vector $y^*(t)$ as the *status quo*.

We can now consider the pool participation decision for an individual firm more precisely. The firm's problem is to choose whether to participate in the mechanism. If the mechanism chosen specifies production $g_i(t)$ and transfer $r_i(t)$ for firm i in state $(y(t), u(t))$, then the ex post value of participation for any period t is $v_i(t) - w_i(t)$, where $v_i(t)$ is the transfer

[4] The assumption of a time-invariant mechanism is without loss of generality, and will be relaxed in Section 3.

received less the incremental costs of production under the mechanism:

$$v_i(t) = r_i(g(t); y(t), u(t)) - [c_i(g_i(t), u_i(t)) - c_i(y_i(t), u_i(t))]. \tag{1}$$

Assuming risk neutrality, a firm would agree to participate in a mechanism over a horizon T if

$$E\left[\sum_{t=1}^{T} \beta^t (v_i(t) - w_i(t))\right] \geq 0 \tag{2}$$

where β is the firm's per period discount factor and where the expectation is with respect to the distribution of the random sequence $\{u(t)\}$.

The firms' (collective) problem is to choose jointly an optimal mechanism to allocate production and the resulting gains from trade. If $Q \subseteq N$ firms participate, their cooperative problem is to maximize $E\left[\sum_{i \in Q} \sum_t \beta^t v_i(t)\right]$, subject to the balanced-budget constraint. The value V of a trade agreement over a horizon of duration T among $Q \subseteq N$ is the maximum (ex ante) gains from trade sustainable as an equilibrium:

$$V_T(Q) \equiv \max_{g \in \mathcal{G}, r \in \mathcal{R}} E\left[\sum_{t=1}^{T} \beta^t \sum_{i \in Q} [c_i(y_i^*(t), u_i(t)) - c_i(g_i(t), u_i(t))]\right] \tag{3}$$

subject to: (i) $\quad \sum_{i \in Q} g_i(y(t), u(t)) - y_i(t) = 0 \qquad t = 1, 2, \ldots, T,\ a.s.$

(ii) $\quad \sum_{i \in Q} r_i(g(t); y(t), u(t)) = 0 \qquad t = 1, 2, \ldots, T,\ a.s.$

(iii) $\quad E\left[\sum_t \beta^t (v_i(t) - w_i(t))\right] \geq 0 \qquad \forall\, i \in Q$

where $\mathcal{G}$ denotes the space of Y^n-valued Borel functions on $(Y \times U)^n$, $\mathcal{R}$ defined similarly (with respect to r), and where the expectation is taken with respect to the random sequences $\{u_i(t)\}$, $i \in Q$. Constraints (i) and (ii) are the market clearing and no-subsidy conditions for a balanced mechanism; constraint (iii) is the participation constraint for each firm. In an equilibrium, (iii) must hold with the opposite inequality for each $i \notin Q$; the ex ante allocation of surplus must make participation in the agreement individually rational for each joining firm, and the decision not to participate by any firm must be optimal given the mechanism subsequently chosen by the group.[5]

[5]An equilibrium trading rule achieving $V_T(N)$ is not necessarily an efficient trading rule, however. A trading rule $g^e \in \mathcal{G}$ is *efficient for* Q if it maximizes the objective in (3) subject only to $\sum_{i \in Q} g_i^e(y(t), u(t)) - y_i(t) = 0\ \forall\, t$; that is, if g^e realizes the maximum possible gains from trade among Q. Letting $W_T(g^e, Q)$

Because the value of future exchange is uncertain, the value function $V_T(Q)$ is a population moment. Define the random variable:

$$s(y^*, u) = \sum_{t=1}^{T} \beta^t \sum_{i \in N} c_i(y_i^*(t), u_i(t)) \tag{4}$$

and let $\mathcal{S} \equiv \{s(y,u) : (y,u) \in (Y \times U)^{nT}\}$. If g^* is a solution to program (3), then the total cost sums $s(g^*, u)$ and $s(y^*, u)$ represent draws from the distribution of total pool production costs (in present value terms) with and without a trade mechanism. Denoting these distributions by Ψ and Φ, respectively, the value of a trade mechanism is the extent to which Ψ "improves" upon Φ, where improvement means $\int_{\mathcal{S}} \Psi - \Phi$, that is, $V_T(Q)$. Since the set of feasible mechanisms includes the *trivial mechanism*, $r_i = 0$ and $g_i = y_i^*$, we have $\Psi \geq \Phi$: An optimal trading mechanism is a first-degree stochastically dominating shift of the distribution of total pool production costs from Φ to Ψ.

We can interpret Φ and Ψ as a decomposition of the likelihood that the law of one price fails to obtain under the status quo. Intuitively, a difference between Φ and Ψ reflects the likelihood that a state of the world will be realized such that there are residual opportunities for trade in this market. Such opportunities arise when there are differences in marginal costs among the firms under the status quo exchange outcome.[6] If the status quo is a competitive equilibrium, then an optimal mechanism can at best replicate it; in this case, $\Phi = \Psi$. If these two distributions differ, then we may infer that existing trading practices fail to exhaust the opportunities for Pareto improving exchange.

Moreover, the distance between these two distributions reflects the economic significance of this failure, in the sense of measuring how far existing trading practices are from the "no arbitrage" condition of a competitive equilibrium. Specifically, the optimal value function $V_T(N)$ defined in (3) is the expectation of a distance metric on the space of sequences of demand and marginal cost functions. The values realized by this random welfare metric correspond to the ex post gains from trade under the mechanism, relative to the status quo. The distribution Φ describes the firms' likely total costs if they operate under the existing exchange agreements, and Ψ reflects the reduction in total costs achievable with adoption of an optimal pooling agreement.

---

denote this optimal value, a trading rule is *efficient* if it is efficient for N, since (by a simple opportunity set argument) $W_T(g^e, N) = \sup_{Q \subseteq N} W_T(g^e, Q)$. In theory, it is possible that the cost functions $\{c_i\}_{i \in N}$ imply that no balanced, efficient mechanism simultaneously satisfies all n participation constraints (3, *iii*) for each state. This arises because of *coalitional instability* if the value $V_T(Q)$ for any $Q \subset N$ is "too large" relative to $V_T(N)$ (cf. Moulin (1988), Thm. 4.1); in other terms, the superadditive game (3) can have an *empty core*. Whether such instability problems are an issue here is an open empirical question.

[6] This marginal cost condition is not entirely correct, for the usual reasons: marginal costs may differ in equilibrium if the cost functions have discontinuities, and boundary solutions may be optimal if the cost functions are sufficiently divergent (i.e., a shutdown outcome).

Estimation of the two distributions, Φ and Ψ, and the population moment $V_T(N)$, is the primary empirical objective of this paper; they characterize the value of an optimal trading mechanism. The basic approach to estimating the unknown parameters Φ and Ψ is to use time-series demand and cost data to estimate $\sum_{i \in N} c_i(y_i^*(t), u_i(t))$ and $\sum_{i \in N} c_i(g_i^*(t), u_i(t))$, where g^* is an optimal trading rule in the sense of (3), and then base inference on simulated empirical distributions of the discounted payoff sums $s(y^*, u)$ and $s(g^*, u)$. The largest step of this approach is to estimate the cost functional $c_i(y_i(t), u_i(t))$ for each firm $i \in N$, for which we employ a structural model of an individual firm's intertemporal production problem.

3. Stochastic Production Costs: Inference and Estimation

Knowledge of the cost functions can be equivalently viewed as knowledge of willingness-to-trade functions for each firm in the market. In the short run, a firm's willingness-to-trade is determined by its marginal cost function: A firm will prefer to buy from another producer—as a substitute for production within the firm—whenever the price is less than the firm's marginal cost. Conceptually, we wish to know whether there are differences between willingness-to-buy and willingness-to-sell among the firms at the observed market quantities; and if so, what would be gained if these differences were arbitraged away.

To provide an answer, we want to determine willingness-to-trade functions for each firm in the market, given the set of existing bilateral exchange contracts. Two factors make this a difficult problem, however. The first is that in this market, as in many others, a firm's willingness-to-trade is inherently stochastic. In electricity markets, this uncertainty arises primarily because firms' marginal costs are subject to random shocks. These shocks (called *forced outages*) arise from recurrent machine failures, and can significantly affect the marginal cost of producing power by forcing substitution to higher cost resources or execution of options to purchase power from others. This latter consequence suggests the second problem: future realizations of a firm's marginal cost function typically depend upon previous production decisions. In general, electric power producers face both intertemporal contractual and physical factor constraints (e.g., water stocks for hydroelectric production and purchase contracts structured as options). Thus, *current* period decisions can affect *future* marginal costs.

Given this stochastic environment, an appropriate basis for inference is to determine the *likelihood* of differences in marginal costs among the firms (at each point in some time interval). Ideally, we would like to base inference on time-series data describing actual marginal cost sequences realized by the firms. From such data we could directly determine whether there are marginal cost differences; and, given a suitable probability model of the economic process generating the data, we could assess the statistical significance of any differences. Unfortunately, we do not observe the actual marginal cost sequences realized by the sample firms. But with available data and some theory, we can determine the probability of generating any particular marginal cost sequence. The first basis for inference is then whether this distribution places much mass in regions where the firms have divergent marginal costs, and little mass where the marginal costs are similar. If so, we have evidence that there are gains to be realized from arbitrage. The second step is to determine how much is likely to be gained; that is, to estimate the corresponding discounted total cost distributions Φ and Ψ.

An Empirical Model

Toward this end, this section considers a structural model of each firm's sequential 'buy-versus-produce' decision making problem. Essentially, each firm seeks to minimize the current and future costs of meeting (exogenous) consumer demand, given (i) a portfolio of generating assets with different costs and operating constraints, and (ii) a menu of contracts for power from other producers, with varying prices, terms, and delivery conditions. Defining the status quo trading regime as the set of existing bilateral exchange contracts, each firm seeks a sequential least-cost production and exchange strategy, or *policy*. The optimal sequence of production and trading rules is given by the solution to an intertemporal optimization problem under uncertainty.

Formally, the production problem facing an electrical utility, given its capital stock, is a *stochastic scheduling problem.* For a firm $i \in N$, we suppose that this capital stock consists of K_i distinct production units, either owned by the firm or contracted for from others.[7] Each unit $k = 1, 2, \ldots, K_i$ is characterized by:

(i) A set of output levels, $X_{ik}(t) \subset \mathbb{R}^{K_i}$, which may depend on (calendar) time; and

(ii) A cost function, $m_{ik} : X_{ik}(t) \to \mathbb{R}_+$.

A firm's *production decision* at time t, $x_i(t)$, is a set of feasible output levels for each unit. Associated with each decision is a current period production cost,

$$\mu_i(x_i(t), y_i(t)) \equiv \sum_k m_{ik}(x_{ik}(t)) + \eta \cdot \max\{0, [y_i(t) - \textstyle\sum_k x_{ik}(t)]\} \tag{5}$$

where $y_i(t)$ is the demand for electricity at time t. The constant $\eta > 0$ reflects the penalty price of excess demand when the utility fails to satisfy its (legal) requirement to serve all

[7]The capital stock of each firm is treated as fixed in this analysis since the lags involved in adding physical capital are long relative to the time interval T we consider.

consumer demand at given prices.[8]

The firm seeks to schedule output levels from each unit to minimize the present value of production costs over some suitable horizon. The difficulty facing the firm is that future production costs are uncertain, and may depend on current period production decisions. This randomness is manifest in the *state variables* $\{y_i(t), u_i(t)\}$, some of which are observable to the firm at t but unobservable prior to t. In particular, the set of feasible output decisions at t is constrained by a stochastic *capacity state*, $u_i(t)$, which describes the availability of each unit of capital k for production in period t.

In order to characterize the behavior of these random technology shocks, we posit a simple model of the firm's production technology as follows. Electric power is produced with two basic technologies: hydro-electric facilities and thermal-electric plants. For the former, given an existing stock $u_{ik}(t-1)$ of hydro-electric energy (i.e., water) and the previous output level $x_{ik}(t-1)$, next period's capacity state is given by a simple stock adjustment model:

$$u_{ik}(t) = u_{ik}(t-1) - x_{ik}(t-1) + \omega_{ik}(t) \tag{6}$$

provided that the stock $u_{ik}(t)$ is non-negative and obeys an upper storage bound. The term $\omega_{ik}(t)$ accounts for stock changes due to exogenous factors, such as inflows and evaporation, and production $x_{ik}(t-1)$ may be positive or negative (i.e., water can be pumped uphill as a means of storage). Thus, we may describe the *feasible set* of output decisions, $X_{ik}(u_{ik}(t), t)$, by

$$X_{ik}(u_i(t), t) = \{x_{ik}(t) \in X_{ik}(t) : x_{ik}(t) \leq u_{ik}(t)\}. \tag{7}$$

In contrast, thermal-electric units are subject to random generator failures characterized by an *outage state*, $\omega_{ik}(t)$:

$$\omega_{ik}(t) = \begin{cases} 1 & \text{if unit } k \text{ is available at } t \\ 0 & \text{if otherwise} \end{cases} \tag{8}$$

We assume that the likelihood of an outage state at $t+1$ depends on the outage state at t but is independent of the outage history prior to t. The evolution of this Markov process proceeds independently of the output decisions $\{x_{ik}(t)\}$ according to the state transition

[8]The penalty price η reflects peak-load incremental costs, to induce firms to make appropriate capacity expansion decisions.

probabilities

$$p_{ik}(a,b) = P_{\omega_{ik}}(\omega_{ik}(t) = b \,|\, \omega_{ik}(t-1) = a), \qquad a, b \in \{0,1\} \tag{9}$$

The probability $P_{\omega_{ik}}$ characterizes thermal unit k in terms of its *failure rate*, $p_{ik}(1,0)$, and its *repair rate*, $p_{ik}(0,1)$. Given an outage state $\omega_{ik}(t)$, the capacity state $u_{ik}(t)$ for a thermal unit is the capacity available at time t:

$$u_{ik}(t) = \omega_{ik}(t) \cdot \sup X_{ik}(t) \tag{10}$$

Thus, as with hydro-electric facilities, the set of feasible output decisions for unit k is described by (7).

The cost m_{ik} of any particular unit may reflect the firm's cost of production from that unit, or the price the firm pays for output from a unit (whether specifically identified or not) owned by another firm. Thus (5) allows for the purchase of electricity from others as a substitute for production using the firm's assets, incorporating the cost of electricity purchased under a contract with other producers in the regime without a pool. The firm's problem is to determine an optimal schedule of production from its own units and purchases from others to minimize the current and future costs of meeting demand.

Properly posed as a dynamic problem, the firm seeks a strategy (or *policy*), $\pi_i = \{\pi_{it}(y_i(t), u_i(t)),\ t = 1, 2, \ldots, T\}$, to minimize the intertemporal cost functional

$$J_{\pi_i} = E\left[\sum_{t=1}^{T} \beta^t \mu_i(\pi_{it}(y_i(t), u_i(t)), y_i(t))\right] \tag{11}$$

where the functions $x_i(t) = \pi_{it}(y_i(t), u_i(t))$ are state-contingent *decision rules* taking values in the feasible set $X_i(u_i(t), t)$, and the expectation is taken with respect to the state process $\{u_i(t)\}$. The *optimal value function* is the minimand:

$$J^*_{\pi_i} = \inf_{\pi_i \in \Pi_i} J_{\pi_i} \tag{12}$$

where the *admissible set* Π_i is the collection of all production policies with feasible decision rules for each t.

Since the feasible set $X_i(u_i(t), t)$ is known at t but random prior to t, clearly an optimal production policy involves a sequential decision-making process on the part of the firm. This intuition is made precise by Bellman's Optimality Principle, which states that if $\{\pi^*_{it}(y_i(t), u_i(t)),\ t = 1, 2, \ldots, T\}$ is an optimal policy beginning at $t = 1$, then the truncated policy $\{\pi^*_{it}(y_i(t), u_i(t)),\ t = s, s+1, \ldots, T\}$ is optimal for the continuation process beginning at any time s, $1 < s \leq T$. More formally, Bellman's Principle states that the optimal value $J^*_{\pi_i}$ defined by (11) and (12) is given by J_0 via the recursive functional

$$J_t(y_i(t), u_i(t)) = \inf_{x_i(t) \in X_i(u_i(t),t)} \{\mu_i(x_i(t), y_i(t)) + \beta E\left[J_{t+1}(y_i(t+1), u_i(t+1))\right]\} \tag{13}$$

where $J_{T+1} \equiv 0$ and where the future value $E[J_{t+1}(\cdot)]$, $1 \le t < T$, is defined by

$$E\left[J_{t+1}(y_i(t+1), u_i(t+1))\right] = \int_{\{u_i\}} J_{t+1}(y_i(t+1), u_i(t+1)\, P_{u_i}(du_i | u_i(t), x_i(t)) \tag{14}$$

Thus, given the conditional probabilities in (9) as primitives in the firm's production problem, the firm uses backward induction via the recursion (13) to compute a sequence of optimal state-contingent decision rules $\pi^*_{it}(y_i(t), u_i(t))$, $t = T, T-1, \ldots, 1$. The decision rules are then sequentially applied as each state $(y_i(t), u_i(t))$ is realized, implying a sequence of actual production costs $\{\mu_i(\pi^*_{it}, y_i(t)),\ t = 1, 2, \ldots, T\}$.

From the cost function $\mu_i(\pi^*_{it}, y_i(t))$, it is straightforward to determine the marginal cost at time t from the price of the unit or contract on the margin at $y_i(t)$. Thus, given an (exogenous) demand series, the sequence of optimal decision rules provides us with a mapping from the distribution of the model primitives (the ω's) to the distribution of marginal cost sequences. This map embodies the behavioral assumption that each firm does the best it can to minimize its costs, given the contractual relations in place with its other suppliers. The optimal decision rules allow us to determine the likelihood that, even with least-cost production and exchange strategies, there are marginal cost differences among the firms under status quo exchange outcomes.

Value of a Pool

To link this empirical model with the value of a pooling agreement developed in Section 2, observe that:

$$c_i(y^*_i(t), u_i(t)) \equiv \mu_i(\pi^*_{it}, y_i(t)) \tag{15}$$

since by construction, $\mu_i(\pi^*_{it}, y_i(t))$ is the minimum current period cost of satisfying demand $y_i(t)$ in state $u_i(t)$. Thus for a given realization of the state process $\{y_i(t), u_i(t)\}$ for each $i \in N$, the series $\{\sum_{i \in N} \mu_i(\pi^*_{it}(y_i(t), u_i(t)), y_i(t))\}_{t=1}^T$ constitutes the sequence of total production costs for the set N of firms without a pool, and from (4),

$$s(y^*, u) = \sum_{t=1}^{T} \beta^t \sum_{i \in N} \mu_i(\pi^*_{it}(y_i(t), u_i(t)), y_i(t)) \tag{16}$$

constitutes a random draw from the distribution Φ with

$$E\left[s(y^*, u)\right] = \sum_{i \in N} J^*_{\pi_i} \tag{17}$$

Equation (17) provides the first equation relating the structural formulation of the firm's production problem to the value of a pooling mechanism defined in Section 2. Similarly, the firms' expected costs in a pooled regime, $E[s(g^*, u)]$, can be expressed in terms of an optimal value function by observing that the total costs under an optimal state-contingent trading rule, g^*, are the optimal value of the firms' collective minimization of $\sum_{i \in N} J^*_{\pi_i}$; in this case, the firms' choice set is the set of schedules with time t decision rules on the union of the firms' time t feasible productions sets, and with the individual demands in (5) replaced by the pooled demand $y_N(t) = \sum_{i \in N} y_i(t)$ in the aggregate current period cost function. Thus, if $J^*_{\pi_N}$ denotes this optimal value, we have $E[s(g^*, u)] = J^*_{\pi_N}$ and so:

$$V_T(N) = \sum_{i \in N} J^*_{\pi_i} - J^*_{\pi_N} \tag{18}$$

Equation (18) expresses the value function $V_T(N)$ we wish to determine in terms of the estimable quantities $J^*_{\pi_i}$ and $J^*_{\pi_N}$. Specifically, given the cost m_{ik} of production of i's kth unit for all i, k, and estimated transition probabilities $\hat{P}_\omega$, we compute $\{\hat{\pi}^*_{it}(y_i(t), u_i(t))\}$ and $\hat{J}^*_{\pi_i}$ for each $i \in N$ via the recursion (13), and the pooled optimal decision rules $\{\hat{\pi}^*_{Nt}(y(t), u(t))\}$ and optimal value $\hat{J}^*_{\pi_N}$ via the analogous recursion on the pool's opportunity set. The difference is an estimator of the expected gains from trade under an optimal pooling agreement.

Estimation of Φ and Ψ

Now if we observed the actual state process $\{u_i(t)\}$ for each $i \in N$, we could use the estimated decision rules $\{\hat{\pi}^*_{it}\}$ and $\{\hat{\pi}^*_{Nt}\}$ to estimate the *actual* production costs that would have been incurred during the sample period under the pool and status quo regimes, respectively. Unfortunately, we do not have data on the actual state realizations, so the best we can do is to estimate the distributions generating these production costs. That is, the true production costs that would have been incurred under a mechanism g^* are an unobserved random sequence, and $s(g^*, u)$ is an unobserved random draw from Ψ. What we can achieve is to describe the probability that the actual costs under each regime are equal to some particular value, that is, to estimate the probability that $s(y^*, u)$ and $s(g^*, u)$ lie in a given interval on $\mathcal{S}$.

While the model above supplies a representation for $V_T(N)$ via (18), no such expressions are readily available for Φ and Ψ. Minimization of (11) implies a measurable transformation from P_ω to Φ and Ψ exists via (10) and (15), but this involves the map $(y_i(t), u_i(t)) \mapsto x_i(t)$

defined by a decision rule π_{it}^* for which we have no analytic, closed-form representation. Thus, one has to look to alternate, numerical-based estimation schemes. In particular, the estimated decision rules $\{\hat{\pi}_{it}^*\}$ and $\{\hat{\pi}_{Nt}^*\}$ and transition probabilities $\hat{P}_\omega$ form the basis for a *simulation estimate* of the unknown distributions Φ and Ψ.

Specifically, let $\tilde{\omega}(t)$ be a sequence of simulated state disturbances: a T-fold sequence drawn from the product measure $\hat{P}_\omega$, given an initial state $\omega(0)$. Application of the decision rules $\{\hat{\pi}_{it}^*\}$ and $\{\hat{\pi}_{Nt}^*\}$, $i \in N$, and the state transition equations (6) and (10) provide a sequence of simulated production decisions $\tilde{x}_N(t) = \hat{\pi}_{Nt}^*(y(t), \tilde{u}(t))$ and $\tilde{x}_i(t) = \hat{\pi}_{it}^*(y_i(t), \tilde{u}_i(t))$, $i \in N$. Given unit costs $\{m_{ik},\ \forall\ i,k\}$, application of (5) and (15) (the latter modified for the pooled case in the obvious manner) provides a simulation estimate of the total cost sequences $\{\sum_{i\in N} c_i(g_i^*(t), u_i(t)),\ \sum_{i\in N} c_i(y_i^*(t), u_i(t))\}_{t=1}^T$ with and without a trade mechanism.

With R independent replications of these bivariate sequences, we can construct estimators $\tilde{\Phi}$ and $\tilde{\Psi}$ which will approximate $\hat{\Phi}$ and $\hat{\Psi}$ (under the model) uniformly on $\mathcal{S}$. In particular, if we let $\tilde{s}^r(y^*, u) = \sum_t \beta^t \sum_{i\in N} c_i(y_i^*(t), \tilde{u}_i^r(t))$ and $\tilde{s}^r(g^*, u) = \sum_t \beta^t \sum_{i\in N} c_i(g_i^*(t), \tilde{u}_i^r(t))$ denote the rth replication simulation estimate of $s(y^*, u)$ and $s(g^*, u)$, respectively, then for R large we may use the empirical distribution functions

$$\tilde{\Phi}(s) = \frac{1}{R}\sum_{r=1}^{R} \mathbf{1}(\tilde{s}^r(y^*, u) \leq s) \quad \text{and} \quad \tilde{\Psi}(s) = \frac{1}{R}\sum_{r=1}^{R} \mathbf{1}(\tilde{s}^r(g^*, u) \leq s) \tag{19}$$

as good approximations to $\hat{\Phi}$ and $\hat{\Psi}$.[9]

4. Applying the Empirical Model

Data

The model is applied to estimate the value of a pool among the four major electrical utilities in California over a 744 hour period during July 1992. The data for this study include the hourly demand for electricity from each firm, generating unit production cost and operating characteristics data for each firm during this period, and the terms and prices for the power purchase contracts held by the sample firms during this period (with generators both within and outside California). The data for this study were provided by the California Energy Commission and are described in detail in the Data Appendix.

The sample firms are the Pacific Gas and Electric Company (PG&E), the Southern California Edison Corporation (SCE), the San Diego Gas and Electric Company (SDG&E),

[9]The indicated uniform convergence is a trivial application of the Glivenko-Cantelli Theorem (cf. Billingsley (1987), Thm. 20.6).

Table 1
Characteristics of Sample Firms

	PG&E	SCE	SDG&E	LADWP
Demand (1,000 kw)				
Peak	18,373	15,690	2,886	4,750
Average	12,733	10,583	1,965	3,058
Minimum	8,004	6,107	1,159	1,729
Capacity (1,000 kw)				
Hydro				
Fixed	2,514	514	0	71
Variable	3,654	1,919	0	1,648
Thermal				
Fixed	3,481	7,728	854	3,148
Variable	5,315	5,964	1,561	1,925
Contracts with Non-Utility Generators				
Fixed	3,384	4,384	182	1
Variable	0	0	0	0
Contracts with Other Utilities				
Fixed	450	1,314	0	129
Variable	1,825	2,266	1,694	786
Totals	20,623	24,089	4,291	7,708

Notes: Capacity figures are for peak demand periods. "Fixed" is the minimum output level within a unit class or the must-take level for contracts. "Variable" is the residual on-peak capacity within a unit class.

and the Los Angeles Department of Water and Power (LADWP).[10] The first three are investor-owned utilities and the latter is the largest municipal electric utility in the U.S. Each of these utilities is vertically integrated into electricity generation, transmission, and distribution, and has extensive interconnections with other utilities both within and outside of California. The sample firms have non-overlapping service territories and together account for over 95% of all retail electricity sales in California.

Table 1 presents some demand and production unit statistics for the sample firms, and the scatterplot in Figure 1 shows the average energy cost and capacity for each production unit or power supply contract. Although the demand statistics in Table 1 are given for

[10] At the risk of committing gross neglect of institutional detail, for expositional convenience I will continue to refer to the four utilities as "firms" despite LADWP's public agency status.

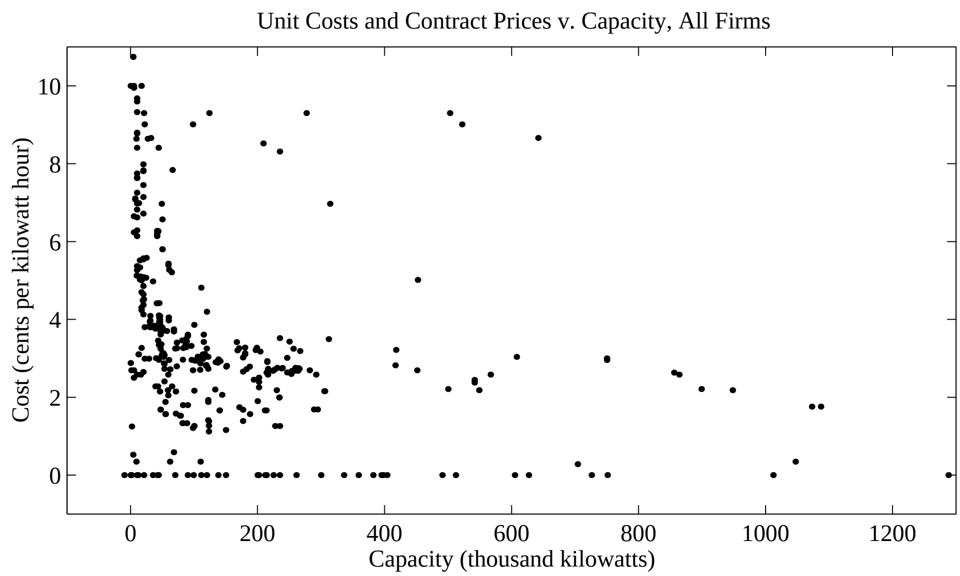


Figure 1: Average costs for units and contracts, all firms

the July 1992 sample period, they are indicative of the relative sales volume of the utilities on an annual basis. The capacity statistics in Table 1 present the aggregate generation capacity of each utility by type, as well as the capacity available from other utility and non-utility suppliers. It is assumed that the existing short and long-term contracts to which one or more of the sample firms is party, summarized in Table 1 and Figure 1, represent the opportunities for trade in the status quo regime. It is also assumed that in the pool regime, existing contracts with suppliers outside the pool remain unchanged over the sample period (although the output may be reallocated to other members within the pool).

Implementation and Computational Issues

Application of the model of Sections 2 and 3 to the data requires solution of the firms' optimal decision rules. Unfortunately, it is rare that Markovian decision processes such as the firm's optimal scheduling problem can be solved for an explicit analytical representation of an agent's optimal policies. Thus, we must resort to numerical solution of the firm's (and the pool's) production and exchange problem.

Solution of this problem presents a substantial computational task, for several reasons. Foremost is the fact that instead of optimizing for a fixed sequence of actions $\{x_1, x_2, \ldots, x_T\}$, one needs to optimize for a complete sequence of functions $\{\pi_1, \pi_2, \ldots, \pi_T\}$ so as to allow the optimal action x_t to vary as a best response to the current state in each period; i.e., to compute state-contingent strategies. Second, these state-contingent strategies will typically differ in each period t. In addition to the truncation effect of a finite horizon, many of the firms' costs and scheduling constraints depend explicitly on calendar time (e.g., time-of-day prices and delivery conditions for power purchased from others). Thus, the conventional Markov decision process assumption of a stationary admissible set is inappropriate. Third, while the state space for thermal-electric units is the discrete set of extremal points of a 2^{K_i}-dimensional unit hypercube, the state space for hydro-electric units is an interval bounded by the minimum and maximum storage levels; the feasible output sets for both technologies are intervals as well. Technically, these continuous state and action spaces require optima selected from infinite dimensional function spaces. To approximate such optima numerically we must discretize the continuous spaces, but even at a relatively coarse approximation the computational task involved in the optimization in (14) is formidable.

To minimize these computational difficulties, we exploit a number of properties of the model to simplify solution of (13) and, more importantly, to simplify the continuation value term (14). The first of these properties govern the unit outage process:

(P1) *The outage state of a thermal unit evolves independently of other units.*

(P2) *The outage state of a thermal unit evolves independently of past thermal unit actions.*

In addition to (P1) and (P2), we make the following specific distributional assumption regarding the evolution of the outage state process:

(A1) *The time to failure and time to repair for each thermal unit are independently geometrically distributed.*

Geometric arrival times are the discrete-time analog of the common exponential arrival time assumption for continuous-time waiting processes. In particular, it implies that state transitions are a *memoryless process*: the hazard rate for the transition from one outage state to another is a constant function of time. This has the useful property that the conditional transition probabilities can be readily derived from the long-run outage frequency (as an estimate of the stationary outage probability) and the expected time to repair a unit after an outage.

Given the curse of dimensionality arising in large stochastic optimization problems, the model is made computationally feasible with the following simplifying assumptions:

(A2) *The state of a hydro-electric unit is pre-determined by the past hydro-electric actions; that is,* ω_{ik} *for a hydro-electric unit is non-stochastic.*

(A3) *Within a unit, the cost of a unit of output is constant over the variable range of production.*

(A4) *Within classes of homogeneous units, available capacity is constant at its expected value under the stationary outage distribution.*

Assumption (A2) allows us to reduce our computational burden substantially, by making the evolution of the hydro-electric elements of the state vector $u_i(t)$ a predictable function of current period information. Assumption (A2) is reasonable if the uncertainty regarding future exogenous stock changes is small, e.g., if evaporation and inflow patterns are sufficiently well-understood as to be accurately predicted over the model horizon.

Assumption (A3) makes possible an explicit pre-ordering of the optimal scheduling sequence of variable-production thermal units. The data indicate maximum and minimum capacities for each production unit; in some cases, these values are equal, and when they are not the firm is allowed to choose any output level in the interval between the minimum and the maximum. The number of thermal units owned by a utility ranges from a dozen for the smallest utility in the sample, SDG&E, to more than thirty for the largest firm, PG&E. Many of these units are relatively homogeneous (within technology and vintage classes) and, more importantly, have similar costs of production. Assumption (A4) exploits these properties, and is tantamount to assuming that a weak law of large numbers eliminates the uncertainty of outages on the total capacity of a given class. This is clearly a strong assumption for large, "base-load" thermal units, the outage of which can induce significant period-to-period variation in production costs. It is more appropriate for modeling the behavior of the numerous small units used to provide peak-period capacity, as well as for power from multiple independent generators having similar contract terms with the utility.

With these assumptions (A1-A4), it is possible to simplify considerably the evaluation of the continuation value, (14), to

$$E\left[J_{t+1}(y_i(t+1), u_i(t+1))\right] = \sum_{\{u_i\}} J_{t+1}(y_i(t+1), u_i(t+1)) \prod_{k=1}^{K} P_{u_{ik}}(u_{ik}(t+1)|u_{ik}(t), x_{ik}(t)) \tag{20}$$

and furthermore assumption (A2) implies that

$$P_{u_{ik}}(u_{ik}(t+1)|u_{ik}(t), x_{ik}(t)) = \begin{cases} P_{u_{ik}}(u_{ik}(t+1)|u_{ik}(t)), & k \in TH_i \\ \mathbf{1}(u_{ik}(t+1) = u_{ik}(t) - x_{ik}(t)), & k \in HY_i \end{cases} \tag{21}$$

where TH_i and HY_i are firm i's set of thermal-electric and hydro-electric units, respectively. With (20) and (21), it is feasible to solve Bellman's equation (13) for the optimal value function by recursive iteration and discrete Monte Carlo integration.[11] Under (A3), it is feasible to determine optimal thermal decisions in a continuous space because of the complete ordering property. However, to implement (20) we still must treat both the hydroelectric unit state and action spaces as a discrete grid of points, an approximation to the underlying continuous spaces.

Finally, because available data include the demand state sequences realized by each firm, the model can be solved using the actual demand sequences realized by the firms. More precisely, we assume that:

(A5) $y_i(t)$ = *the empirical demand process for firm* i.

Assumption (A5) is a strong condition, as it states that $y_i(t)$ is a *previsible* process. It implies that each decision maker (correctly) behaves as if there is no uncertainty in the path of future demand. This assumption is driven by computational feasibility requirements, as it reduces the state space of demand to a single path through Y^T.[12]

A consequence of (A5) is that the estimated value of an optimal pooling mechanism must be interpreted as conditional on this particular demand series. Under (A5) we are computing the likelihood (under P_ω) of marginal cost differences among the firms at the demand quantities observed for July 1992. It is important to note that while the resulting estimates can indicate the gains that might have been achieved during this period, they may not necessarily reflect the gains for, say, July of a prior or subsequent year.

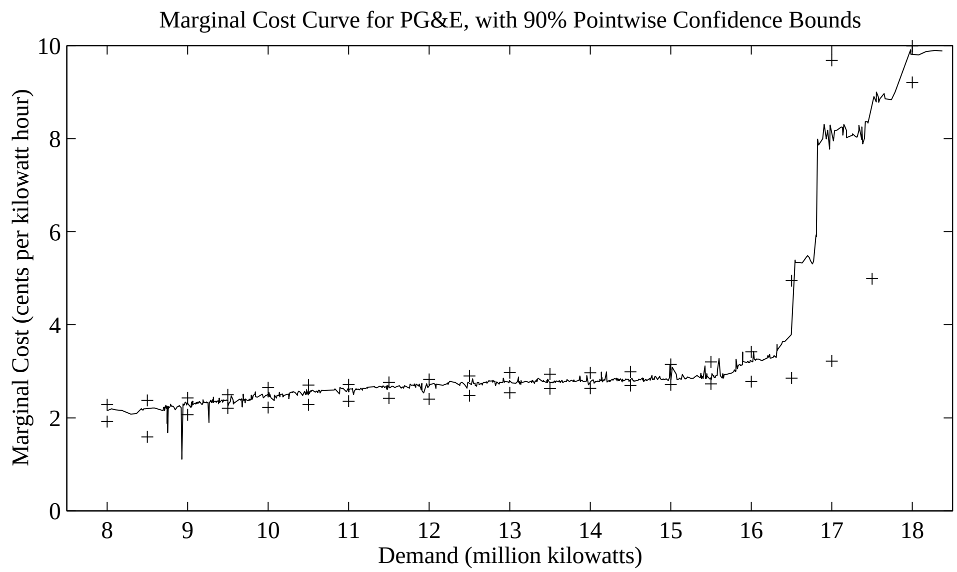
5. Numerical Results

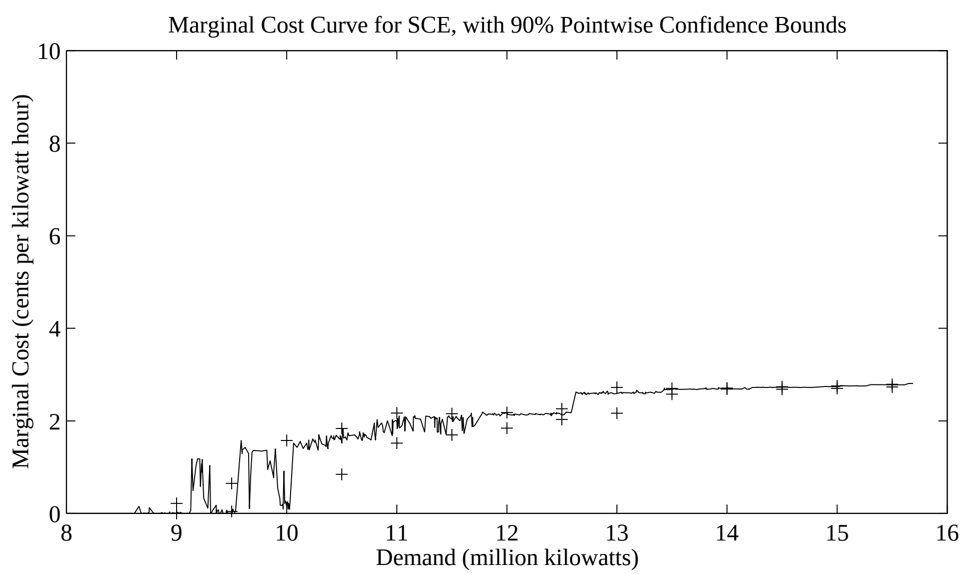
Marginal Cost Functions

The results indicate that in most states of the world, there are measurable differences in marginal costs among the sample firms. These differences are observed under the assumption that the firms' operators seek to do the best they can to minimize costs, given the existing set of bilateral contracts as their trading opportunity set. The differences in marginal costs among the firms are readily apparent from the estimates of each firm's marginal cost function in Figures 2 through 5.

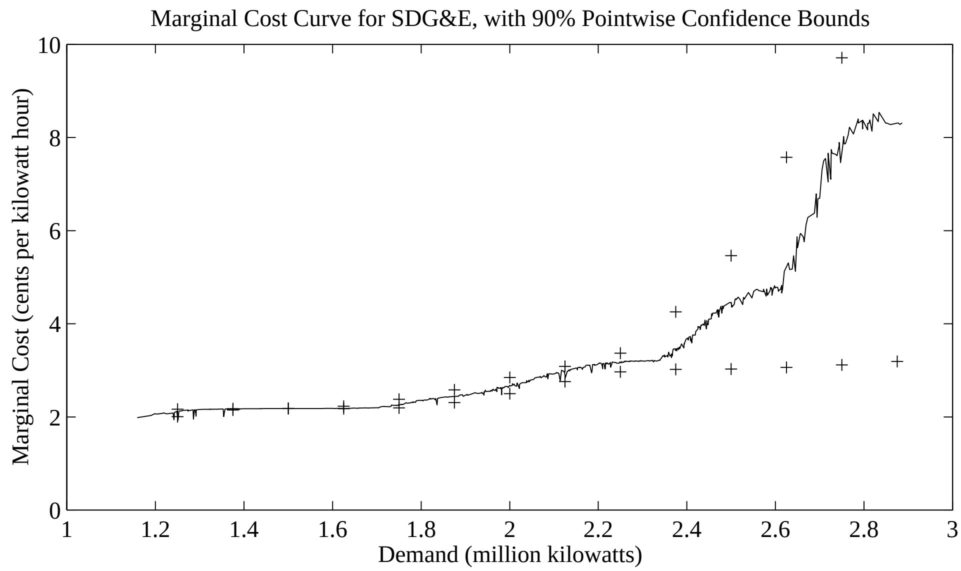
[11] Estimation of the pooled firms' decision rules required approximately 80 hours on a Sparc 10/42.

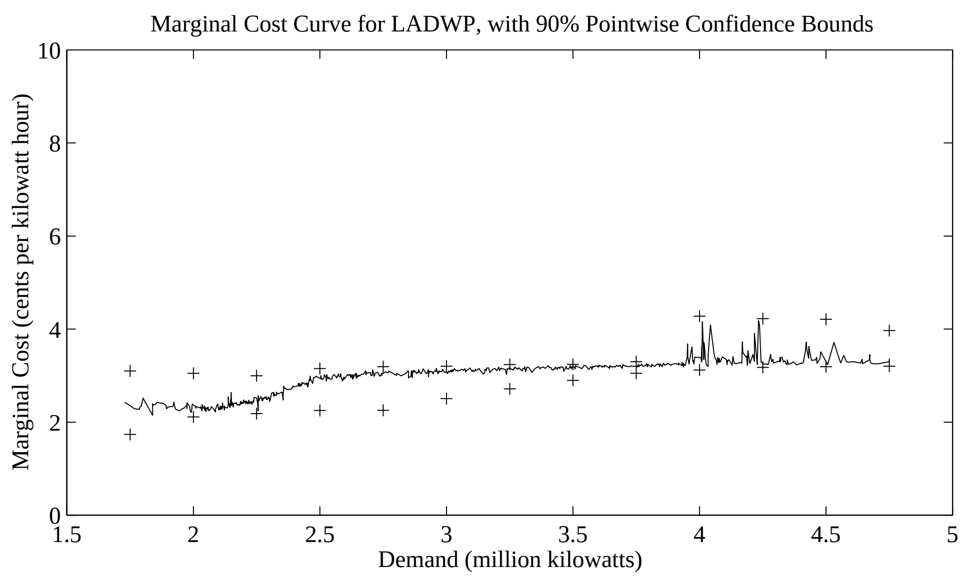
[12] The high-order serial correlation of demand at hourly intervals both renders a Markovian demand model effectively infeasible (the relevant "states" entail long histories), and mitigates the nominally heroic assumption of previsibility.





Figures 2 and 3: Marginal Cost Curves for PG&E and SCE





Figures 4 and 5: Marginal Cost Curves for SDG&E and LADWP

Construction

Given each firm's state-contingent production cost function, the simulated state sequences can be readily used to construct marginal cost curves for each firm. For a given state in a given period, the marginal cost function of any firm is simply a monotone step function based on an ordering from the intertemporally optimal schedule of all available production units and contract options. The expected marginal cost function estimates in Figures 2–5 are the likelihood-weighted averages of these state-contingent marginal cost functions. Because of the approximate correspondence between time and demand, these marginal costs curves appropriately account for the optimal intertemporal use (i.e., the shadow price) of low-cost limited-factor resources, particularly hydroelectricity.

Each marginal cost estimate is constructed by applying the optimal decision rules, given the empirical demand process, to $R = 200$ independent collections of (firm and unit-specific) simulated outage sequences.[13] The decision rules then imply a minimal total cost and corresponding marginal cost sequence for each replication. A weighted-mean marginal cost estimate for each of a pre-determined set of demand values (the observed demand data) in the domain of Figures 2–5 is then computed by averaging the R marginal cost values. The solid line in each figure is the natural interpolant of these 744 weighted averages (otherwise known as the 'connect-the-dots' estimator). The pointwise confidence bands (indicated by the '+' marks in each figure) are determined by the 5th and 95th percentiles of the empirical distribution of marginal costs at each demand quantity.

Before addressing the economic implications of Figures 2–5, there are some statistical issues relating to the marginal cost function estimates that merit discussion. First, because the terms of many of the bilateral contracts among the utilities depend explicitly on calendar time, the distribution of the state-contingent marginal cost functions for each firm does not, in general, generate a stationary (function-valued) process. Specifically, the estimated marginal cost functions in Figures 2–5 are a mixture of an (ostensibly) stationary distribution on a 'on-peak' class of marginal cost functions, and a stationary distribution on an 'off-peak' class of marginal cost functions. The dependence of contract terms on calendar time is largely manifest through this binary on-peak/off-peak condition, suggesting that the process may be appropriately viewed as a mixture with a (near degenerate) time-varying mixing probability. Unfortunately, this distinction is not as clean as one might like; the demand frequency distributions for each subperiod have overlapping support, particularly conditional on nearing a switching time. This phenomenon explains in large part the substantial volatility in certain regions of the marginal cost estimates, and the occasionally

[13] For each of the R replications, an initial condition is drawn at random using the long-run historical outage frequencies in the data (as an approximation to the stationary outage distribution).

large local non-monotonicities. For example, the spikes in the 4–4.5 million kilowatt region for LADWP arise because adjacent demand realizations occur in different time classes.

Randomization errors also play a part in explaining the volatility evident in the figures. In particular, the low amplitude noise that pervades most of the estimates would likely be dampened by increasing the number of replications above the 200 performed here. Alternatively, one might use a functional estimator to smooth out the noise and develop a tighter estimate of the average marginal cost function. Unfortunately, the economic structure underlying these data suggests that as operators make incremental changes in output, marginal costs are likely to exhibit kinks and discontinuities as new units are started up or shut down. Thus, it is important that any estimator of the marginal cost function be consistent for discontinuous functions on Y; this attribute motivates the choice of the interpolant estimator used in the figures.

There are also some idiosyncratic sources of volatility in the estimates, a striking example of which is the volatility of SCE's marginal costs in the 9 to 10 million kilowatt (kw) range. This particular phenomenon is attributable to the high outage rates of San Onofre, a large nuclear generating station operating in SCE's fixed output region. The 9–10 million kw region contains a stochastic 'cliff' where marginal cost leaps between 0 and approximately 1.4¢ per kilowatt hour; this cliff represents where SCE moves from a demand quantity below the aggregate short-run minimum output level to the first unit capable of variable output production.

Discussion

The most important implication of Figures 2–5 is that the four firms have quite different marginal costs. While SCE has the highest average cost per unit output (see Table 2), its average production of 10.5 million kw lies in a range where its marginal costs are below 2¢ per kilowatt-hour (kwh), lower than any of the other firms. PG&E, however, on average produces 12.7 million kw at a marginal cost of 2.7¢ per kwh, with a normal daily range from just over 2¢ per kwh (off-peak) to 3¢ per kwh (on-peak). Moreover, PG&E and SDG&E occasionally operate in a range where their marginal cost exceeds 8¢ per kwh, more than twice the maximum observed for SCE and LADWP.

The basic source of these marginal cost differences is that these four utilities have chosen fairly different firm-level production technologies. Although these (and most other) utilities generate electricity with the same basic commercial generation technologies, the historical investment decisions made by each firm have lead to significantly different aggregate production functions. Whether it is optimal for the firms to have chosen such different firm-level technologies is an open question beyond the scope of these data; to some extent, the technology differences reflect differences in regional factor endowments. Nevertheless, given

Table 2
Estimated Production Costs, Status Quo and Pool Regimes

	Total Production Costs[a] (Million $)		Production Costs per Kwh[a] (Cents/Kwh)	
	Mean	S.E.	Mean	S.E.
Status Quo				
PG&E	212.60	2.05	2.24	0.02
SCE	243.67	0.43	3.09	0.01
SDG&E	37.72	0.13	2.58	0.01
LADWP	51.64	0.35	2.26	0.02
Sum	545.63	2.06	2.59	0.01
Pool Regime	523.21	1.79	2.48	0.01
Difference (Status Quo − Pool)	22.42	2.67	0.11	0.01

[a] $T = 744$ hours, $R = 200$ replications.

these capital investments as sunk costs, the firms clearly have realized technologies which afford significant opportunities for trade. These results are in contrast to the position of (at least one of) the sample firms regarding the efficiency of existing trading mechanisms. In response to investigations by the California Public Utilities Commission, PG&E has stated (bracketed terms added):

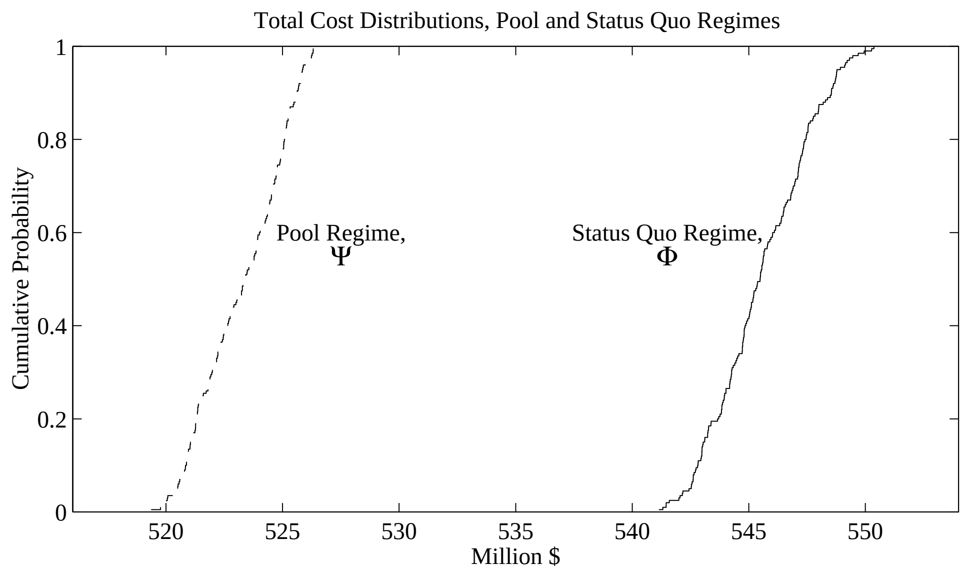
> [Existing agreements] already provide an efficient mechanism for capturing and sharing the benefits which arise from the small difference in second-by-second [marginal] dispatch costs between California's utilities. These cost differentials are extremely slight since, at virtually all times, these companies have identical resources at the margin.[14]

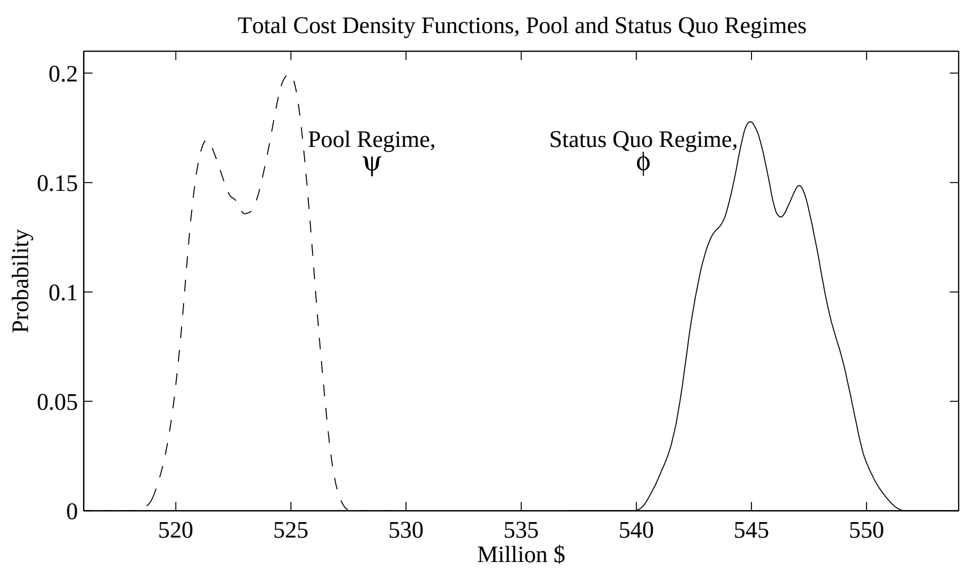
Value of a Pooling Agreement

Given that we observe marginal cost differences among the sample firms, the question of immediate interest is: How much would be gained if an optimal pooling mechanism arbitraged away these differences among the firms' marginal costs?

Table 2 presents the estimated total production costs for the four firms in both the status quo and the pool regimes. Because of the discontinuities in the firms' marginal cost functions, the value of a pooling mechanism is not determined from the incremental

[14] Pacific Gas and Electric Company (1994, p. 8).





Figures 6a and 6b: Estimated Total Cost Distributions

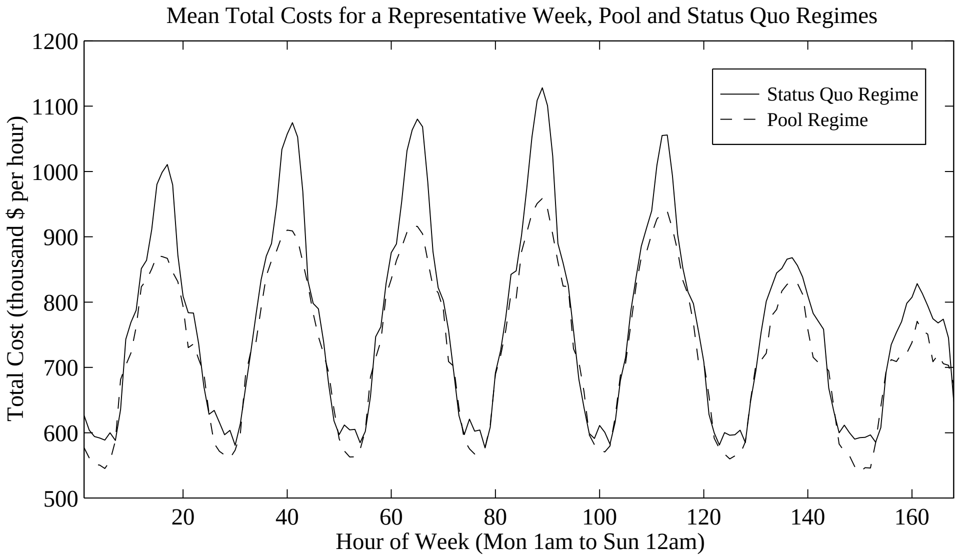


Figure 7: Total Costs for a Representative Week

quantities that equalize marginal costs. Rather, the value function (18) is used to determine the expected gains after computing the approximate optimal decision rules (under (A1-A5)) for each firm individually, and then for the set of all firms operating as a pool. Given the estimated state-contingent costs for each firm in each regime, it is straightforward to estimate Φ and Ψ: Application of the decision rules to the simulated state sequences provides R total cost values for each regime, the empirical distributions of which provide estimates of Φ and Ψ.

Figures 6a and 6b present the estimates of the distributions Φ and Ψ. The empirical distribution functions of total costs during the sample period $\hat{\Phi}$ and $\hat{\Psi}$ are plotted in Figure 6a, and estimates of the corresponding densities are shown in Figure 6b. The data points for each are the total costs over the 744 hour sample period resulting from application of the optimal decision rules to $R = 200$ simulated state sequences. Summary statistics for each distribution are given in Table 2. The difference between sample means, $\hat{V}_T(N)$, is \$22.42 million, with a sample standard deviation of \$2.67 million. This figure represents approximately 4% of the firms' total costs in the non-pooled regime, on average. The first-order stochastic dominance property of g^* is readily apparent in Figure 6a; moreover, the infimum of the support of $\hat{\Phi}$ exceeds the supremum of the support of $\hat{\Psi}$. The density function estimates in Figure 6b are constructed with a biweight kernel smoother under the assumption that the true data generating distribution is Gaussian (separate smooths were performed for each density in Figure 6b).[15] The bimodality in the distributions of total costs is principally attributable to bimodality in the underlying distribution of demand.

Despite its computational cost, a particularly appealing attribute of estimating the structural model developed in Section 3 is the ability to examine the intertemporal distribution of the gains from trade. Figure 7 presents the total production costs of the firms in the time domain in both regimes, for a representative week within the 744 hour sample period. These curves are the pointwise (with respect to time) averages of the R total cost sequences in each of the two regimes. Figure 7 indicates that the largest gains from a more efficient trading mechanism arise from exchange at the times of peak demand (approximately 3–6 pm during weekdays), and to a lesser extent at the troughs when some firms (such as SCE) find themselves operating below their short-run fixed output level.

6. Discussion and Implications

Taken together, these results clearly suggest that there exist gains from trade which are not realized by existing trading practices. Even if each firm's operators do the best they can under the existing bilateral agreements, the findings above indicate that they are unlikely to achieve the gains from trade available under an optimal pooling agreement. Of course, the paradigm of the pool here essentially is a surrogate for a (more) complete contingent contract among the market participants. Power pools, both in theory and in practice, are an institutional vehicle for specifying state-contingent trading behavior for numerous foreseeable but unlikely events affecting the value of exchange. Thus, what we ought to conclude from these results is that the firms in this market have chosen not to sign a rich enough class of agreements to capture all the available gains from trade.

From an economist's perspective, these foregone gains from trade define the transaction costs of adopting a more complex trading mechanism in this market. It is difficult, however, to ascribe this opportunity cost to bounded rationality or limited "cognitive competence" on the part of the market participants.[16] The existence of elaborate pooling systems adopted by the utilities in New York and New England indicate that such mechanisms are clearly feasible solutions to the problem of the complexity of optimal exchange in electrical networks. Rather, the transaction costs of pooling in the market studied here likely reflect

[15]The biweight kernel is $K(s) = \frac{15}{16}(1 - s^2)^2 \cdot \mathbf{1}(|s| < 1)$, with asymptotically (MISE-sense) optimal bandwidth $h \sim 2.778\sigma R^{-1/5}$ under Gaussianity. As a check on this latter assumption, separate (two-sided) Kolmogorov-Smirnov tests of the hypotheses that $\hat{\Phi}$ and $\hat{\Psi}$ are Gaussian rejects the null for the latter and fails to reject for the former, at the 5% significance level (approximate finite-sample critical values for a K-S test of Gaussianity with unspecified mean and variance are given in Bickel and Doksum (1977, p. 381)). The density estimates in Figure 6b are intentionally undersmoothed by $\hat{h}/3$.

[16]Cf. Williamson (1989, p. 139).

factors that rationally lead the firms' decision makers not to adopt a pooling mechanism. The outstanding question before us is then: If there exist residual opportunities for trade in this market, why have the firms not voluntarily adopted a pool? I offer two arguments.

Regulatory Appropriation. This explanatory hypothesis is that the regulatory process dilutes the firms' incentives to reduce costs. The structure of electrical utility regulation in California, as in most other states, is a mixture of rate-of-return and cost-of-service regulation. Utilities are compensated by regulatory price-setting to cover their operating costs plus a fair rate of return on prudently invested capital. Pooling, at least over the time periods addressed by this study, is a factor which reduces the firms' operating costs—so if operating costs fall by 4%, the revenues required to maintain the appropriate rate of return on capital would be reduced commensurately. Moreover, if there are Averch-Johnson effects, over a longer horizon pooling institutions would allow the firms to build less new capital, exacerbating this incentive not to pool.

Of course, the objection can be made that there is regulatory lag in price setting, and that this provides an opportunity for the regulated firm to retain the benefits of cost reducing activities. Joskow (1974) and others have argued that the financial significance of such regulatory lag can be substantial, providing a considerable incentive to increase the efficiency of the regulated firm's operations. Nevertheless, it is doubtful that a move to a regional power pool in California or any other state would escape careful regulatory scrutiny. Establishment of a major power pool is a significant change in the organizational form of electric power production, and, if the experiences of utilities in other regions are any guide, would require the explicit regulatory approval of both the Federal Energy Regulatory Commission (FERC) and the state utility commission. To justify such a change, the utilities would almost certainly have to estimate and describe in considerable detail the expected savings from the pool. It is probably naive to expect that regulators would refrain from appropriating some, if not most, of the potential savings. In fact, when the Florida Power Pool trading mechanism was established in the early 1980's, the Florida Utilities Commission reviewed the savings from the power pool and declared that 80% of the subsequent savings from the pool should be remitted to ratepayers.

Of course, even if the firms were allowed to retain only a fraction of the net savings from a pooling mechanism, in principle they still have a positive incentive to capture this remaining payoff. Thus arises the question of whether there are *latent costs* of pooling: unobserved or unobservable costs of pooling which the firms expect would not be accounted for in the regulatory commission's calculation of the pool's net benefits. In particular, perhaps the costs of securing regulatory approval, negotiating agreements with other participants, restructuring long-standing operational practices, or other internal "adjustment costs," are sufficient to offset the firms' limited expected retained earnings from investing in a multilateral pooling mechanism.

Owner-Manager Separation. A second and complimentary explanatory hypothesis is that although a pooling agreement would be in the interests of profit-minded shareholders, managers may dislike a pool because of agency problems within the firm. In the case of (nominally) non-profit LADWP, such agency problems take the form of separation of utility managers' interests from those of civic leaders, who might prefer to reduce the utility's operating budget and allocate limited public resources to other city services.

While there are many potential sources of owner-manager agency problems, a particularly potent one in this context is that pooling reduces the scope of managerial control. Efficient power pooling institutions unambiguously lessen managerial control over a firm's assets. Numerous studies have implicated this loss of managerial sovereignty as an important source of resistance to pooling agreements. For example, interviews with utility representatives conducted for the California Energy Commission in 1981 "consistently revealed that, while utilities see benefits from pooling, they wish to see them achieved at the lowest level of centralization possible in order to preserve as much independence as possible. They recognized that pooling inevitably leads to loss of control by individual utilities. Many of the utilities expressed strong reservations about centralized dispatch and felt that most of the benefits could be obtained with more informal cooperation" (CEC (1981, p. 57)). Such perspectives strongly suggest that managers perceive the prerogative of managerial control to be sufficiently jeopardized by multilateral pooling agreements that they prefer to forego the potential benefits to shareholders.

Transportation Costs: A Caveat

In the electric power industry, as in most markets, there are transportation costs of trade. On the margin, these costs take the form of *energy losses* as electricity is transported over the transmission network. As incremental energy losses due to increased trade have not been explicitly accounted for in the estimation above, transportation costs present an important potential caveat to the estimated value of a pool.

As a rough indication of the magnitude of these costs, studies suggest that 1 to 3% percent of generation is typically lost before local distribution to consumers, and therefore must be made up on the margin.[17] In the pooled case, the estimated average marginal cost of production is 2.21¢ per kwh. Using the upper loss rate figure, to obviate $22.4 million in gross savings the corresponding average quantity traded among the pool would have to exceed 45 million kwh in each hour of the sample period. This volume exceeds the aggregate *peak* demand for all of the utilities in the sample. A more appropriate average

[17] FERC (1989, p. 60) calculations based on a 345kv line of 100 miles.

trading volume is about 10% of this figure, which reduces the value of trade to approximately \$20.2 million per month.

A variant of the transportation cost argument is that the firms may not be able to execute trades because they lack sufficient transmission capacity among them. Congestion on interconnecting lines would then foreclose the opportunity to realize potential gains from pooling. This is not an implausible argument in many regions of the Western U.S., but each of the four firms examined here has developed extensive interconnections with other utilities both within and outside of California. Moreover, if the pool allocates the use of this transmission capacity efficiently, then only the marginal, or lowest-value transactions, should be foreclosed if congestion limits the ability to execute trades during peak periods.

Most importantly, interutility transmission capacity is ultimately an endogenous variable. If over longer horizons the gains from hour-to-hour electricity trade within a pool are comparable to those identified here, then one should address the effect of transportation costs on trading opportunities in the context of the cost of additional transmission capacity. An extensive review reported by the FERC (1989, p. 46 ff.) estimates the average cost of new transmission capacity at 0.1¢ to 0.4¢ per kilowatt-hour, based on actual construction cost data for 274 lines installed during the mid-1980s.[18] For comparison, the average pairwise marginal cost difference among the sample firms in the status quo regime is approximately 0.9¢ per kilowatt-hour.

In summary, it is evident that transportation costs will erode some of the gross benefits of a pooling agreement. Clearly, the extent to which these costs limit the opportunities for trade must be refined before taking allegations of unappropriated gains from trade into the formal regulatory process. Nevertheless, the admittedly crude transportation cost estimates given here strongly suggest that they cannot offset the estimated gross savings of \$22.4 million per month.

7. Conclusion

Since the California Public Utilities Commission instituted a review of options to restructure and partially deregulate the electric power industry in that state in April 1994 (CPUC (1994)), substantial debate has ensued over what form a restructured electricity market should take. Both the Commission and various participants in the restructuring effort, including two utilities, have suggested that a power pool is an appropriate institutional

[18] Average capital costs for 100 and 400 mile transmission lines at 500kv are 0.08 and 0.32¢/kwh, respectively, assuming 50% loading, 16% cost of capital, and a 30-year facility life. PG&E has argued before the FERC that the average cost of a 300 thousand kw expansion of its main north-south corridor (probably the most congested interconnection in this market) would be approximately 0.2¢/kwh, assuming full utilization (FERC (1989, p. 58)).

mechanism to facilitate exchange in a competitive electric power market. Such pools have been successfully implemented elsewhere, notably the Northeastern U.S. and the United Kingdom.

Opinions have been sharply divided on the matter of the value of a pool in markets like California. On one hand, there is the actual experience of pools operating elsewhere. The New England Power Pool, which coordinates production among utilities in all six New England states, reports net savings of $76 million in 1991.[19] The New York Power Pool, which was designed to coordinate production and exchange among the eight major utilities in that state, reports savings averaging $17.8 million per month for the period between 1980 and 1987.[20,21] On the other hand, there are important players in the current restructuring effort who have argued vigorously that the value of a pool in California is negligible. In testimony filed before the state utility commission in July 1994, Pacific Gas and Electric argued:

> [Existing] mechanisms capture materially all of the economic benefits available from existing resources at the wholesale level. Because of the high level of efficiency and coordination which already exists, imposing a government-mandated power pool would not provide additional efficiency benefits for consumers.[22]

In an attempt to shed some light on this debate, this paper presents a structural model of agents' decisions to produce and exchange power through an interfirm pooling mechanism. Pools provide an institution to mitigate the transaction costs of trade, allowing participants to equalize their marginal costs of production in a stochastic environment. In application to data for the four major electrical utilities in California, I find that the economic benefit of a power pool among these firms is approximately four percent of the firms' operating costs for a 744 hour sample period during July 1992. At the risk of the usual hazards of statistical extrapolation, this corresponds to an annual total savings on the order of $250 million per year (1992$). These findings suggest that claims that there is little to be gained from regional power pooling agreements may be appropriately viewed with skepticism.

I conclude that while there appear to be savings associated with a more sophisticated electricity trading institution in California, these savings are unlikely to be achieved as a

---

[19] NEPOOL (1991, p. 11).

[20] NYPP (1990, Table 6–4). A restructuring of the New York Power Pool governing principles in 1987 led to significant incentives for parties to contract outside the pool. Data before and after that time are not strictly comparable.

[21] These estimates likely overestimate the actual savings attributable to the NYPP and NEPOOL, as the pools make conservative estimates of the volume of trade that would occur in their absence. Additionally, for comparison note that (each of) these pools serves a market approximately half the size of the California electricity market.

[22] Pacific Gas and Electric Company (1994, p. 9).

voluntary exercise by the existing firms under the existing regulatory regime. The adoption of a pooling agreement can be viewed as a regulatory game in which there are two equilibria. In the “good” equilibrium, firms adopt pooling agreements and regulators subsequently refrain from appropriating the benefits of the pool for consumers. In the “bad” equilibrium, however, firms refuse to adopt a pool because they perceive that regulators will not allow the firms to retain the resulting benefits for their shareholders. Thus, we observe heterogeneity in the adoption of pooling institutions, where the key explanatory variable is the utilities’ expectations of how regulators will treat their gains from trade. In states with a historically strict regulatory climate, such as California, utilities may believe that they are unlikely to retain the benefits resulting from a regional power pooling agreement. Thus, we are left in the unfortunate position where the inability of regulators to commit not to appropriate the benefits of cost-reducing activities by the regulated firm keeps these activities from being undertaken.

What this suggests is that the deregulatory proposals to create a regional power pool through vertical divestiture of the existing electric utilities may be necessary to resolve this problem. Absent the distortionary process of regulatory oversight, electric power producers will have considerable incentive to capture any gains from trade available in a deregulated wholesale power market. While the downstream activities of electricity transmission and distribution will likely require a regulatory presence for the indefinite future, there are clearly gains to be realized from a market structure that provides agents unmitigated incentives to arbitrage price differences among producers. Moreover, there are likely to be additional gains from market competition as new entrants seek to compete with utilities’ existing generating stock. The analysis of what market institutions would best facilitate such competition remains an important subject for future research.

Data Appendix

The data for this study are drawn principally from the California Energy Commission, 1992 Electricity Supply Planning Assumptions Report (ESPAR) Volumes I–III, and the California Energy Commission, 1992 Electricity Report Elfin Data Sets (EDS). Supplemental cost data were obtained from the U.S. Energy Information Administration publications listed in the references (DOE (1994a,b)). Demand data were provided by each of the utilities in this study via the Demand Forecasting Office of the California Energy Commission. Details regarding specific data items and how they are incorporated into the stochastic programming model for computation of optimal production schedules are described below.

Demand: Demand data are the hourly system loads for each utility's control area, monitored at the power control metering point for each utility. The demand data include system transmission and distribution losses actually incurred. Demand data also include the effects of DSM programs in place.

Thermal Unit Capacities: Thermal unit capacities for utility-owned units are drawn from the generating unit data sections of EDS, and ESPAR Table III–1A. Modeling is based on unit (as opposed to plant) level data, and at two block levels for units with minimum and variable capacity blocks. Minimum commitment block levels are used for all committed units. Units subject to must-run requirements due to voltage support or other loading requirements are committed and dispatched to at least minimum block level.

Fuel Prices: Utility-specific natural gas and distillate fuel prices are drawn from ESPAR Vol. III Section IV. Fuel prices vary monthly and represent the total cost of gas as delivered per mmbtu. Unit-specific nuclear fuel prices are drawn from the nuclear generation unit sections of EDS. For utility-owned thermal units, the price of a unit of output is determined by the fuel price times a thermal unit efficiency factor, plus a variable operations and maintenance component (see below).

Thermal Unit Efficiencies: Unit-specific heat rates by capacity block are drawn from the generating unit data sections of EDS, and used to calculate prices for output by capacity level.

Thermal Unit Forced Outage Rates: Historical unit-specific five-year average forced outage rates (FORs) and maintenance outage rates (MORs) are drawn from the generating unit data sections of EDS, and are described more generally in ESPAR Vol. III p. AIII–64 ff. Conditional failure rates are derived from the long-run FORs under the assumption that the mean time to repair after a forced outage is 48 hours, using assumptions (P1, P2, A1) in the text.

Variable O&M: Variable O&M costs are drawn from the generating unit data sections of EDS. These data provide variable O&M costs as a (unit-specific) constant per megawatt hour of generation. Data for utility-owned units outside the utility's control area include the cost of off-system losses. Nuclear O&M costs per kwh are calculated from the 1992 O&M cost data in DOE (1994a).

Unit Commitment: Since the sample period is July, all slow-start units not on long-term standby reserve are assumed to be committed and dispatched to at least minimum block level.

Spinning Reserve Margins: All four firms have a stated spinning reserve margin of 7%. The model accounts for spinning reserve requirements by basing dispatch requirements on demand (including losses) plus the spinning reserve margin requirement.

Hydro Availability: Capacity and energy data are drawn from the hydro supply data sections of EDS. Three classes of hydro resources are incorporated into the model. Run-of-river hydro is dispatched as a fixed, zero-price resource, at the capacity levels given in the data. Pondage hydro minimum and maximum per-period capacities are calculated net of minimum flow requirements (which are aggregated into the run-of-river capacity) and are treated as a fixed energy resource to be optimized. Monthly hydro availability data in Gwh are given in EDS, based on average hydro year conditions for California. Pumped-storage hydro generating capacity, pumping capacity, storage capacity, and inflow for PG&E Helms and LADWP Castaic units are drawn from EDS. Generation is at a zero price but pumping cost is the marginal cost at the time the energy is pumped, times the unit efficiency factor (marginal cost may be zero if demand is below minimum output of committed units). Hydro units have zero forced outage rates.

Imports: Firm power imports are incorporated individually as specific production units, including contractual time-of-day minimum and maximum capacity terms, including losses. Additional shape factors drawn from the Imports sections of EDS are used for contracts with capacity factors specified at less than 100% (energy limited contracts). Detailed price and delivery terms are described in ESPAR Volume III Tables II–2A–G. Seasonal diversity exchanges are specified similarly but with a zero import price and a must-run requirement during contractual export periods. Data on non-firm economy energy purchases from the Pacific Northwest and Desert Southwest regions are drawn from ESPAR Volume I. These prices and quantities represent the CEC's estimated quantities available and transferrable to California utilities by month, net of the transfer limits on the Pacific intertie and the Desert Southwest tie lines.

Self-Generation: Self-generation capacity data is drawn from EDS and ESPAR Vol. III Table II–1A–V. Self-generation capacity offsets demand at a zero price.

Qualifying Facilities: QF capacity is aggregated within technology-specific types with identical contract pricing terms, by host utility. QFs with Standard Offer #1 (SO1) series contracts are priced at "short-run avoided cost" ("SROC"). SROC is calculated endogenously whenever demand exceeds minimum load and is assumed to be equal to the lowest-cost available variable-production block during minimum load periods. QFs with Standard Offer #4 (SO4) contracts receive either a fixed price per unit output or a price based on a contractually specified thermal efficiency and the utility's cost of fuel (depending on the contract pricing option). Data on prices or efficiencies (as appropriate) are drawn from the QF sections of EDS. All QF capacities are derated by unit-specific forced outage rates (i.e., expected capacity levels are used) and treated as a non-random supply resource. Unit-specific FORs and MORs are used where available in EDS; generic technology-type plant-performance factors are used elsewhere, drawn from EDS and ESPAR Vol. III Table III–5F. Certain jurisdictional QFs are dispatchable by the host utility (during certain periods) and are treated as such; the balance of SO1 and SO4 contract holders are must-run base-load resources. Renewable QF (wind, solar) capacities are treated as fixed at the dependable capacity levels in the QF Renewables sections of EDS.

References

Abdoo, Richard A. 1995. Wisconsin Electric's View of a More Competitive Industry. *Public Utilities Fortnightly* 133(4): 17–19.

Acton, Jan Paul, and Stanley M. Besen. 1987. Assessing the Effects of Bulk Power Rate Regulation: Results from a Market Experiment. *Applied Economics* 19(5): 663–85.

Bertsekas, Dimitri P. 1987. *Dynamic Programming: Deterministic and Stochastic Models.* Englewood Cliffs, NJ: Prentice–Hall.

Bickel, Peter J., and Kjell A. Doksum. 1977. *Mathematical Statistics.* Englewood Cliffs, NJ: Prentice-Hall.

Billingsley, Patrick. 1987. *Probability and Measure,* 2nd Ed. New York: John Wiley and Sons.

California Energy Commission. 1980. *Costs and Benefits of Increased Power Pooling in California.* Systems Control, Inc. (May).

California Energy Commission. 1981. *Institutional Factors Affecting Increased Power Pooling and Utility Integration in the Western United States.* California Energy Commission, P300–81–001 (March).

California Energy Commission. 1992a. *Electricity Report.* Sacramento, CA: California Energy Commission.

California Energy Commission. 1992b. *Electricity Report: Data Appendices.* Sacramento, CA: California Energy Commission.

California Public Utilities Commission. 1994. *Order Instituting Rulemaking and Investigation on the Commission's Proposed Policies Governing Restructuring California's Electric Services Industry and Reforming Regulation.* Docket #R.94–04–031 and I.94–04–032 (April 20).

Christensen, Laurits R., and William H. Greene. 1976. Economies of Scale in U.S. Electric Power Generation. *Journal of Political Economy* 84: 655–676.

Christensen, Laurits R., and William H. Greene. 1978. An Econometric Assessment of Cost Savings from Coordination in U.S. Electric Power Generation. *Land Economics* 54(2): 139–155.

Coase, Ronald H. 1937. The Nature of the Firm. *Economica* 4: 386–405.

Electrical World. 1991. *Directory of Electric Utilities.* New York: McGraw-Hill Publishing Co.

Federal Energy Regulatory Commission. 1980. *Power Pooling in the U.S.* Washington D.C.: Federal Energy Regulatory Commission, Office of Electric Power Regulation, FERC–0049.

Federal Energy Regulatory Commission. 1989. *Transmission Task Force Report* (October).

Fuss, Melvyn and Daniel McFadden, eds. 1978. *Production Economics: A Dual Approach to Theory and Applications.* Amsterdam: North-Holland.

Joskow, Paul L. 1974. Inflation and Environmental Concern: Structural Change in the Process of Electrical Utility Regulation. *Journal of Law and Economics* 17: 291–327.

Joskow, Paul L., and Richard Schmalensee. 1983. *Markets for Power.* Cambridge, MA: MIT Press.

Keelin, Thomas W., et al. 1991. *Western Systems Power Pool Assessment, Final Report.* Menlo Park, CA: Strategic Decisions Group.

Laffont, Jean-Jacques, and Jean Tirole. 1993. *A Theory of Incentives in Procurement and Regulation.* Cambridge, MA: MIT Press.

Moulin, Herve. 1988. *Axioms of Cooperative Decision Making.* Cambridge: Cambridge University Press.

New England Power Pool. 1991. *Annual Report.* Holyoke, MA: New England Power Pool.

New York Power Pool. 1990. *Operating Report.* Schenectady, NY: New York Power Pool.

New York State Public Service Commission. 1991. *Report on the Management and Operations Audit of the New York Power Pool.* Albany, NY: New York State Department of Public Service.

Pacific Gas and Electric Company. 1994. *Comments on Competitive Wholesale Electric Markets and Market Institutions in a Restructured Electric Industry.* San Francisco, CA (July 26).

Rao, C. R. 1973. *Linear Statistical Inference and Its Applications.* New York: John Wiley.

Ross, Sheldon M. 1983. *Stochastic Processes.* New York: John Wiley.

Rust, John. 1987. Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica* 55(5): 999–1033.

Rust, John. 1993. Structural Estimation of Markov Decision Processes. In R. Engle and D. McFadden, eds., *Handbook of Econometrics,* Volume 4. Amsterdam: North Holland.

Schmalensee, R., and Benett W. Golub. 1984. Estimating Effective Concentration in Deregulated Wholesale Electricity Markets. *Rand Journal of Economics* 15(1): 12–26.

Sharkey, William W. 1979. Existence of a Core When There are Increasing Returns to Scale. *Econometrica* 47(4): 869–76.

Silverman, Brian W. 1986. *Density Estimation.* London: Chapman and Hall.

Southern California Edison Company. 1994. *Response of Southern California Edison (U 338–E) to Order Instituting Rulemaking and Order Instituting Investigation Dated April 20, 1994.* Rosemead, CA (June 8).

Stokey, Nancy, and Robert E. Lucas. 1989. *Recursive Methods in Economic Dynamics.* Cambridge, MA: Harvard University Press.

U.S. Department of Energy. 1994a. *Financial Statistics of Selected Investor-Owned Electric Utilities 1992.* Washington, DC: U.S. Department of Energy, Energy Information Administration DOE/EIA–0437/1.

U.S. Department of Energy. 1994b. *Financial Statistics of Selected Publicly Owned Electric Utilities 1992.* Washington, DC: U.S. Department of Energy, Energy Information Administration DOE/EIA–0437/2.

Williamson, Oliver E. 1989. Transaction Cost Economics. In R. Schmalensee and R. Willig, eds., *Handbook of Industrial Organization,* Volume 1. Amsterdam: North-Holland.