OBTAINING THE BEST FROM REGULATION AND COMPETITION

Edited by
Michael A. Crew
Menahem Spiegel

OBTAINING THE BEST FROM REGULATION AND COMPETITION

Topics in Regulatory Economics and Policy Series

Michael A. Crew, Editor

Center for Research in Regulated Industries
Graduate School of Management, Rutgers University
Newark, New Jersey, U.S.A.

Previously published books in the series:

Crew, M.:
Regulation Under Increasing Competition

Crew, M.A. and Kleindorfer, P. R.:
Emerging Competition in Postal and Delivery Services

Cherry, B.A.:
The Crisis in Telecommunications Carrier Liability: Historical Regulatory Flaws and Recommended Reform

Loomis, D.G. and Taylor, L. D.:
The Future of the Telecommunications Industry: Forecasting and Demand Analysis

Alleman, J. and Noam, E.:
The New Investment Theory of Real Options and its Implications for Telecommunications Economics

Crew, M. and Kleindorfer, P. R:
Current Directions in Postal Reform

Faruqui, A. and Eakin, K.
Pricing in Competitive Electricity Markets

Lehman, D. E. and Weisman, D. L.
The Telecommunications Act of 1996: The "Costs" of Managed Competition

Crew, Michael A.
Expanding Competition in Regulated Industries

Crew, M. A. and Kleindorfer, P. R.:
Future Directions in Postal Reform

Loomis, D.G. and Taylor, L.D.
Forecasting the Internet: Understanding the Explosive Growth of Data

Crew, M. A. and Schuh, J. C.
Markets, Pricing, and Deregulation of Utilities

Crew, M.A. and Kleindorfer, P.R.
Postal and Delivery Services: Pricing, Productivity, Regulation and Strategy

Faruqui, A. and Eakin, K.
Electricity Pricing in Transition

Lehr, W. H. and Pupillo, L. M.
Cyber Policy and Economics in an Internet Age

Crew, M. A. and Kleindorfer, P. R.
Postal and Delivery Services: Delivering on Competition

Grace, M. F., Klein, R. W., Kleindorfer, P. R., and Murray, M. R.
Catastrophe Insurance: Consumer Demand, Markets and Regulation

Crew, M. A. and Kleindorfer, P. R.
Competitive Transformation of the Postal and Delivery Sector

OBTAINING THE BEST FROM REGULATION AND COMPETITION

edited by

Michael A. Crew
Center for Research in Regulated Industries
Rutgers Business School – Newark and New Brunswick
Rutgers University
Newark, New Jersey, U.S.A.

and

Menahem Spiegel
Center for Research in Regulated Industries
Rutgers Business School – Newark and New Brunswick
Rutgers University
Newark, New Jersey, U.S.A.

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 0-387-23196-X
Print ISBN: 1-4020-7662-2

Created in the United States of America

Visit Springer's eBookstore at: http://ebooks.kluweronline.com
and the Springer Global Website Online at: http://www.springeronline.com

CONTENTS

This page intentionally left blank

SPEAKERS AND DISCUSSANTS

Raj Addepalli, Manager - Staff ISO Team, New York State Department of Public Service

Coleman Bazelon, Vice President, Analysis Group

Timothy Brennan, Professor of Policy Sciences and Economics, University of Maryland Baltimore County

Roger Camacho, Assistant Corporate Rate Counsel, PSE&G

Pradip Chattopadhyay, Utility Analyst - Electric Division, New Hampshire Public Utilities Commission

Richard N. Clarke, Director of Economic Analysis, AT&T

Michael A. Crew, Professor of Economics and Director – Center for Research and Regulated Industries, Rutgers Business School, Rutgers University

Michael DeCesaris, Associate, Cornerstone Research

Jeanne M. Fox, President – New Jersey Board of Public Utilities

John Garvey, Economic Analyst - Office of the Chief Economist, New Jersey Board of Public Utilities

Fred Grygiel, Chief Economist, New Jersey Board of Public Utilities

Ralph Izzo, President and Chief Operating Officer, PSE&G

Paul R. Kleindorfer, Anheuser Busch Professor of Management Science and Economics and Co-Director of Center for Risk Management, Wharton School, University of Pennsylvania

Gregory Leonard, Manager, Cornerstone Research

Stephen Levinson, Independent Consultant

Colin Loxley, Manager – Process Standards and Development, PSE&G

Karl McDermott, Vice President, NERA

Richard Michelfelder, Assistant Professor of Finance, School of Business – Camden, Rutgers University

Gbadebo Oladosu, R&D Associate, Oak Ridge National Laboratory

Carl Peterson, Consultant, NERA

SPEAKERS AND DISCUSSANTS

J. Arnold Quinn, Economist, Office of Market Oversight & Investigation, Federal Energy Regulatory Commission

Mark Reeder, Chief of Regulatory Economics, New York State Department of Public Service

James D. Reitzes, Principal, The Brattle Group

Adam Z. Rose, Professor of Geography, The Pennsylvania State University

Michael Rothkopf, Professor of Operations Research, Rutgers Business School, RUTCOR, Rutgers University

David J. Salant, Co-CEO, Optimal Markets, Incorporated and Adjunct Senior Research Scholar, Columbia University

Derek Salvino, Associate, ICF Consulting, Inc.

Adam C. Schumacher, Associate, The Brattle Group

Richard Simnett, Chief Scientist, Telcordia Technologies (TM)

Menahem Spiegel, Associate Professor and Associate Director of the Center for Research and Regulated Industries, Rutgers Business School, Rutgers University

John Sumpter, Vice President - Regulatory, Pac-West Telecomm, Inc.

Steve Sunderhauf, Manager-Program Design and Evaluation, PEPCO

Nusha Wyner, Director, Energy Division, New Jersey Board of Public Utilities

J. Douglas Zona, Senior Advisor, Cornerstone Research

SPONSORS

PSE&G

AT&T

New Jersey Resources Corporation

Pac-West Telecomm, Inc.

NUI Corporation - Elizabethtown Gas Company

This page intentionally left blank

PREFACE AND ACKNOWLEDGEMENTS

This book is the result of two Research Seminars at the Center for Research in Regulated Industries, Rutgers—The State University of New Jersey on October 24, 2003, and May 7, 2004. Twenty six previous seminars in the same series resulted in *Problems in Public Utility Economics and Regulation* (Lexington Books, 1979); *Issues in Public Utility Economics and Regulation* (Lexington Books, 1980); *Regulatory Reform and Public Utilities* (Lexington Books, 1982); *Analyzing the Impact of Regulatory Change* (Lexington Books, 1985); *Regulating Utilities in an Era of Deregulation* (Macmillan Press, 1987); *Deregulation and Diversification of Utilities* (Kluwer Academic Publishers, 1989); *Competition and the Regulation of Utilities* (Kluwer Academic Publishers, 1991); *Economic Innovations in Public Utility Regulation* (Kluwer Academic Publishers, 1992); *Incentive Regulation for Public Utilities* (Kluwer Academic Publishers, 1994); *Pricing and Regulatory Innovations under Increasing Competition* (Kluwer Academic Publishers, 1996); *Regulation under Increasing Competition* (Kluwer Academic Publishers, 1999); *Expanding Competition in Regulated Industries* (Kluwer Academic Publishers, 2000); and Markets, Pricing, and Deregulation of Utilities (Kluwer Academic Publishers, 2002).

Like the previous Research Seminars, these seminars received financial support from leading utilities. The views expressed, of course, are those of the authors and do not necessarily reflect the views of the sponsoring companies: AT&T, FirstEnergy Corporation, NUI Corporation - Elizabethtown Gas Company, Pac-West Telecomm, Inc., and Public Service Electric and Gas Company.

Company managers freely gave their time and advice and, on several occasions, provided information about their industries. We especially thank David Blank, Richard Clarke, Frank Delany, Laurence Downes, John Graham, Mary Patricia Keefe, Amey Mesko, and Steve Levinson and John Sumpter. Ralph Izzo, President and Chief Operating Officer of PSE&G was the keynote speaker at the Seminar on October 24, 2003. Jeanne M. Fox, President of the New Jersey Board of Public Utilities, was the keynote speaker at the Seminar on May 7, 2004. The interest of sponsors in the program, which originated with the first Research Seminar in 1978, has continued ever since and has been a major factor in the success of the program.

Many thanks are owed to the distinguished speakers and discussants, listed on pages vi and vii, for their cooperation in making the seminars and this book possible. Most of them worked very hard in achieving deadlines,

without which the speedy publication of this book would have been impossible.

We would especially like to thank Jeremy T. Guenter, Senior Administrative Assistant, at the Center for Research in Regulated Industries for his excellent work in typesetting the book on schedule. This involved working with the authors, the publisher, and with us to insure that the numerous tasks associated with producing a book were performed. All of his duties concerning the production of the book he performed thoughtfully and effectively. This is the tenth book he has set in Microsoft Word and he is continuously making improvements in the process to the benefit of the Center, the authors and the publisher.

MICHAEL A. CREW
MENAHEM SPIEGEL

Chapter 1

Regulation and Competition as Complements*

Timothy J. Brennan
University of Maryland Baltimore County and Resources for the Future

1. INTRODUCTION

Reform and in many cases removal of economic regulation in the U.S. and around the world over the last quarter century is predicated on the perspective that regulation and competition are demand-side substitutes as means for setting prices and product characteristics in an economy.[1] For our purposes, we define competition as when prices, quantities, and quality decisions are left to private buyers and sellers, mediated by market

---

* Email: brennan@umbc.edu. Thanks for useful comments and suggestions go to Carlos Candelario, Michael Crew, Roger Comacho, Fred Grygiel, Paul Kleindorfer, Jonathan Lesser, Karl McDermott, Menahem Spiegel, and other participants at the Research Seminar on Public Utilities (Oct. 23, 2003) and the 23rd Annual Eastern Conference of the Advanced Workshop in Regulation and Competition (May 19, 2004), both sponsored by the Center for Research in Regulated Industries, Rutgers University.
The original idea for this paper came about in preparing Chapter 6 of the 1997 *Economic Report of the President,* "Refining the Role of the Government in the U.S. Market Economy," drafted by the author while he was on the staff of the Council of Economic Advisers. Opinions expressed here are my own and do not represent those of the CEA at the time, but I want to thank Cristian Santesteban, Joseph Stiglitz, Chad Stone, and David Sunding at that time for many useful discussions. Opinions and errors remain the sole responsibility of the author.

[1] We refer to these relationships in a "demand side" manner, as different means for achieving a goal demanded by the public—efficient production and allocation of a particular good or service or some dimension of it.

exchange. Regulation refers to the extent to which those decisions are dictated by government bodies rather than left to private parties.[2]

That regulation and competition are substitutes for making production and allocation decisions is undeniable. However, as the difficult policy experiences associated with "deregulating" major industries such as telecommunications and regulations have shown, increasing competition seems to bring with it more, not less, regulatory attention. We also have witnessed increasing use of market methods in regulatory contexts, as exemplified by emissions permit trading in the environmental arena. Thus, the conventional perspective of treating regulation and competition overemphasizes the "substitutability" between regulation and competition, at the expense of their occasional yet important "complementarity." Insisting that regulation and competition are themselves competitors, with one winning at the expense of the other, can lead to undue neglect of the ways in which careful and creative use of one of these can help attain the best from the other.[3]

A primary implication of the complementarity of regulation and competition follows from the definitional negativity of the cross-elasticity of demand. If regulation becomes more expensive, the demand for competition may fall, not rise. This comes out in partially deregulated sectors, where some markets are competitive while others remain regulated monopolies. Suppose regulation of the interface between competitive sectors (long distance service, electricity generation) and monopoly transport (local exchange service, the transmission grid) is more difficult than thought. Then, turning that potential competition into actual entry and rivalry that leads to lower prices and higher quality services may be *less* likely to take place. Arguments that such regulation is impossible may not lead to deregulation of the entire sector, but deregulation of none of it.

Viewing competition and regulation only as sides in a battle may contribute to having industrial and policies driven by ideological commitments (e.g., for or against "markets") rather than by careful theoretical and empirical assessments of the ways in which both might contribute to maximizing welfare or satisfying other social objectives. Some of the complementarities we review are legal mainstays in making competition more effective, often not thought of as regulation. Three such

---

[2] Markets and industrial sectors may be neither thoroughly competitive nor thoroughly regulated. Some dimensions of a product (e.g., safety attributes) may be regulated while others (e.g., price) left to competition. As emphasized in the discussion below, some markets within a sector may be regulated while others are competitive. Regulation along one dimension may complement competition in another.

[3] Price (1994) treats regulation and competition as complements in a study of the British gas market, but only in the sense that regulation was necessary to deal with market power following what, in her view, was flawed privatization.

complementarities are setting "prices" to be paid for breach of contract or tort liability, antitrust as regulation of market structure, and laws and rules regarding corporate governance and information disclosure.

An important nexus between regulation and competition is how judicious application of the former can help expand the role of the latter in industries where regulation was widespread and continues to be imposed in part of the sector. Three such industries where regulation has been taken a strong role in managing a transition to competition have been oil pipelines, telecommunications and electricity.[4] Pipelines were the subject of regulatory reform when vertical integration came to be seen as a means for monopolizing downstream oil markets (Shenefield, 1978). Telecommunications has been subject to a divestiture and rules regarding access to the local exchange to facilitate competition in long distance service (Brennan, 1987). More recently, the Telecommunications Act reflects a view (prescient or inaccurate) that regulation of interconnection between local telephone providers can permit competition in local markets, while preserving the network externalities that make telecommunications valuable.

Electricity has seen similar regulatory interventions to promote competition among electricity producers at the wholesale level (sales to distribution utilities) and in retail sales to end-users (Brennan, Palmer and Martinez ("BPM"), 2002). An important aspect of this has been the design and formation of intermediate institutions that would separate ownership of firms in competitive industries (electricity generation, retail marketing) from control of regulated bottleneck assets (transmission grids). Whether efforts to institute competition can preserve reliability and prevent the exercise of market power during peak demand periods remains an ongoing concern, coming to the public's attention following the Northeast blackout of August 14, 2003. Following that event, the *Washington Post* reported, "operators [were] forced to plead for cooperation from balky energy companies because they lacked the authority to require action to protect the grid" (Behr and Gray, 2003).

Potential complementarities between regulation and competition have some noteworthy implications. First, the complementarity can go in the

---

[4] At least two other potential examples that also may fit the complementarity model. One is separating railroad tracks from rail services, allowing different rail lines to use monopoly tracks at a regulated rate. These were imposed on Union Pacific and Southern Pacific as a condition for allowing them to merge. See Press Release, "Surface Transportation Board Issues Decision Resolving Trackage-Rights Dispute Arising From 'Union Pacific-Southern Pacific Merger," Sep. 3, 2003.

A second is gas pipelines (Price (1994) covers the British experience). In that sector, gas pipelines purchased gas at the well and sold it at the point of delivery, rather than charged a stand-alone tariff paid by shippers. The two are equivalent; the delivery charge is just the difference between the downstream sale price and the upstream purchase price.

reverse direction, with markets helping to attain regulatory goals. Second, as observed above, where regulation and competition are complements, an increase in the cost of the former will depress demand for the latter. If satisfactory regulation is very difficult to achieve, the competition it complements will also be difficult to attain. This can lead to failures of the market, industry, and regulatory structures instituted to bring about the intended economic benefits (McDermott and Peterson, 2004). Third, contexts where regulation and competition go together can create a rift between the "efficiency" justifications for competition and those based on "libertarian" values associate with the virtues of maximizing individual discretion and minimizing the role of the state in society. Efficiency and libertarianism often go together; but when regulation promotes efficiency by enabling competition, these values come into conflict.

We proceed as follows: Section 2 observes that regulation and competition may be assessed not only on efficiency grounds, but also on the distribution of benefits and the degree to which they preserve economic liberty. Section 3 sets out the arguments for why regulation is often viewed as an inferior substitute for competition in promoting efficiency. Section 4 sets out general principles for viewing regulation and competition as complements. Section 5 identifies regulatory aspects of the legal system—common law, antitrust, and corporate governance—that complement competition. Section 6 shows how regulation complements competition in partially deregulated sectors, looking at oil pipelines, telecommunications, and electricity. Section 7 concludes the paper by suggesting extensions of the analysis in additional policy contexts, how competition may complement regulation, conditions that promote complementarity, and implications for conflict between efficiency and libertarian norms.

2. ALTERNATIVE NORMS

Observing that regulation and competition can be complements does not refute the conventional generalities as to why regulation is often an inferior substitute for competition. Before looking at ways in which regulation and competition can be complements, it is useful to review how they are viewed as substitute institutions for the production and allocation of goods and services desired by consumers at prices close to cost, i.e., economic efficiency. The concepts of complementarity and substitutability, however, require at the outset an assessment of the different "product characteristics" on which policy makers and the public might choose how to blend regulation and competition as mechanisms for guiding the production and allocation of goods and services.

Efficiency is not the only criterion used to compare regulation and competition. Competition and regulation might also be judged by the distribution of benefits. Despite the normative weight economists place on efficiency, it is probably not the leading political determinant of whether we have regulation or competition. The allocation mechanism we get, and how it is implemented, is more likely driven by some complex combination of the distribution of influence among potential winners and losers, evaluated over the range of policy options (taxes, subsidies, trade and entry barriers) affecting the distribution of wealth. Whether an industry gets deregulated, for example, is more likely to depend on whether the net balance of political clout rests with those who would do better under competition, not whether society in the aggregate would do so (Stigler, 1971). That the distribution of net benefits influences policy hardly implies that distributive equity is that standard by which policies are judged.

A second criterion involves libertarian considerations, i.e., the extent to which mechanisms for producing and distributing goods and services maximize individual autonomy and minimize the role of the state. Competitive markets often are viewed as both promoting both economic efficiency and expanding individual liberty. Market failures, however, may break this association, when state regulatory mandates promote efficient outcomes while reducing freedom to set prices or outputs. As Sen (1970) pointed out, when one person's welfare may be harmed by another's actions, liberal principles—giving persons a range of alternatives over which they can make unfettered choices—can conflict with the Pareto principle—instituting A over B when all prefer A to B. Maximizing freedom need not maximize efficiency. That libertarian and efficiency considerations can differ creates a potential tension in advocating for the expansion of competition—a topic to which we return in the concluding section.

3. REGULATION AS AN INFERIOR SUBSTITUTE FOR COMPETITION

To put in context the possibility of complementarity between regulation and competition, it is useful to review how, on efficiency grounds, regulation comes to be viewed as an inferior substitute for competition. The list of reasons is familiar. Going back to Hayek (1945), economists have appreciated the fact that competitive markets have enormous informational advantages. Market prices reflect the marginal cost of production that buyers can compare against their individual marginal benefits to determine if purchases are worthwhile. At the same time, those prices signal to sellers the marginal value of goods that is compared against marginal cost so as to

make profit maximization compatible with economic efficiency. Markets achieve this without requiring that information on marginal costs and willingness to pay be transmitted to a central authority that would then make those decisions.

Along with better information, competition tends to trump regulation as a means for achieving economic efficiency because it "eliminates the middleman" of a public agent. Not only does adding that middleman introduce costs related to added information transfer and error. To the extent the agent has the authority to make production and allocation decisions, it can be expected to take its own interests into account and neglect those of the producers and consumers whose net economic benefits would be maximized. Under competition, those who reap the revenues and bear the costs make production decisions, and those who compare benefits to the price they have to pay make purchase decisions. Each principal has incentives that match efficient outcomes; bringing in the government introduces "incentive compatibility" problems (Laffont and Tirole, 1993).

For these reasons, a necessary condition for regulation is that competition fails to reach efficient outcomes because of a market failure, a situation in which "high transaction costs" prevent mutually beneficial exchanges (as in Coase, 1960). Of most interest here is when technological conditions—scale economies in production, network externalities—lead to there being few firms or only one firm in a market. (Antitrust laws, discussed below, are designed to prevent artificial limitations of competition, e.g., collusion, large mergers, or cornering markets in scarce inputs.)

In developed economies, for most products, conditions of perfect competition do not hold. There are often few suppliers of goods, or products are differentiated, giving some sellers a modicum of ability to set prices above marginal cost and limit output below efficient levels. However, the aforementioned informational and incentive problems with regulation relative to markets remain. These typically limit regulation to settings in which there is but one seller in a large homogenous market of sufficient value to consumers that unfettered monopoly would lead to adverse distributional consequences as well as inefficiency.

Even where there is monopoly, information and incentive problems make regulation problematic. If the rate of return is estimated to be too great, the regulated firm will have an incentive to bias its inputs toward capital (Averch and Johnson, 1962). Absent the ability to institute non-linear pricing, the second-best optimum will have prices above marginal cost, with price equal to average cost if the firm sells only one product to one relevant class of consumers (Baumol and Bradford, 1970). Regulators, no less self-interested than other economic agents, can be expected to respond to political influence, which in turn is likely to be disproportionately exercised by the regulated firm, rather than by the consumers regulation is nominally

intended to protect (Stigler, 1971). Regulation may serve as a way to redistribute wealth to specific constituencies in ways that would not be politically viable if instituted as explicit taxes and subsidies (Posner, 1971).

More recent analyses extended these criticisms. The main source for data on costs will often be the regulated firm, giving that firm an incentive to overstate costs in order to get the regulator to set higher prices (Baron and Myerson, 1982). Regulatory compliance leads to rigidity in both the technologies used to provide the regulated product and the structure of prices used to provide the services. A manifestation of the latter is rate averaging that creates cross-subsidies and divergences of prices from costs (Breyer, 1984). Regulated firms can have an incentive to enter unregulated businesses, in order to evade regulatory constraint by either non-price discrimination against unaffiliated competitors in downstream markets or misallocating costs of unregulated services to the regulated sector, generating anticompetitive cross-subsidies (Brennan, 1987). Even if the regulator can overcome these problems, setting price equal to cost squelches the incentive to minimize cost; prices may have to be set independently of costs to restore this incentive (Brennan, 1989). Moreover, regulation stifles entrepreneurial initiatives to come up with innovative ways to deliver services, imposing dynamic costs greater than those from static inefficiencies due to higher prices (Winston, 1993). These considerations are important in determining how threatening a monopoly should be before risking these regulatory costs.

4. REGULATION AND COMPETITION AS COMPLEMENTS

If all markets worked perfectly, we would not need regulation to either substitute for or complement competition. However, such is not the case, rendering the potential complementarity worthy of investigation. From here on out, we take as given the possibility that monopoly or other market failures can be sufficiently severe and persistent to warrant some form of regulation. That does not mean that regulation is a usually inferior but occasionally superior substitute for competition, but only a substitute.

As the ideas of "substitute" and "complement" imply in the normal contexts of goods and services rather than for allocation institutions, markets are interconnected, not independent. Choosing the better means for making production, pricing, and allocation decisions in one sector can improve the performance of the means chosen in other sectors. In most cases, competition in one market will make competition in another work better, and vice versa. The "theory of the second best" applies, in that reducing the

divergence from the optimum in one market would warrant reducing divergence from the optimum in others. However, when regulation may be the better mechanism in one sector, it can improve the prospects for competition in related sectors. Competition in one sector can also make regulation more effective in another sector. It is in this sense that regulation and competition can be complements.

The particular relationship between goods that we want to explore is when one good is, at least in a stylized way, a factor used in the production or distribution of the other. The ubiquitous input to competitive exchange is the "transaction cost" of the exchange itself. Common law—property rights enforcement, tort liability, contract resolution—is a form of government regulation designed to minimize transaction costs or replicate the low transaction cost outcome, complementing competition in most of its practical manifestations.

In other settings, the object of the regulation is the input to the competitive sector. The dominating endeavor in the deregulation of utilities has been to devise ways in which monopoly transport sectors can be restructured and regulated so that competition can work in newly opened markets using that transport. Three such examples are pipelines used to ship crude oil and petroleum products, local telephone networks used by long distance providers to originate and terminate calls, and the transmission grids used by generators to deliver electricity to distribution utilities and end users.

5. THE LEGAL SYSTEM AS A REGULATORY COMPLEMENT TO COMPETITION

The most pervasive way in which regulation complements competition is through the legal system's ability to facilitate exchange, encourage competition, and promote effective corporate governance. George Stigler (1981) once criticized regulatory economics for neglecting what he regarded as the most pervasive and fundamental regulatory structure in society—the legal system. As he put it, "If the economic theory of contracts, torts, and property, for example, are not part of the theory of regulation, I don't know where in economics this work belongs" (Stigler, 1981, p. 73). The ubiquity of the law hides its importance in the economic system. As Coase (1960) established, the legal system, most importantly unambiguous property rights, are necessary for effective competition.

Complementarity here arises because transacting is an input to exchange itself. Market analyses focus on the latter, but it is regulation through the law that ensures that transactions can be supplied at low cost, or their

outcomes replicated, so competition can flourish. When transaction costs are small, the state can define, interpret, and enforce property rights so meaningful exchange can take place. The fundamental practical definition of what it means to "buy" something is that the seller transfers to the buyer the right to call upon the police power of the state to threaten credibly to punish and thus deter theft.

On the other hand, if transaction costs are high, because property rights are ambiguous or exchange is difficult to arrange, the legal system can reproduce the outcome that would have ensued had costs been lower. When exchange takes place but agreements are incomplete, contract law allows courts to determine if breach occurred, supplying the terms that parties would have agreed to but for the costs of writing them. Courts can also determine the price to be paid in case of breach so as to encourage breach only when the costs of doing so are less than the benefits. When exchange at any level is impossible, e.g., in coping with unforeseen accidents, tort law sets prices in terms of liability judgments, in order to induce efficient care.

Antitrust enforcement is another form of regulation that complement competition, by setting the ground rules for conduct and structure that allow competition to determine prices, output, and product quality.[5] On the conduct side, the laws prohibit price fixing, market allocation, and, more problematically, conduct that makes it difficult for competitors to survive. Structural regulation in antitrust is manifested through prohibiting mergers that, in the language of the Clayton Act, may tend to inhibit competition, either by facilitating coordination (e.g., collusion) or enhancing unilateral market power (e.g., single firm dominance or a less competitive oligopoly) (Department of Justice and Federal Trade Commission, 1997).[6]

An additional way in which regulation complements competition in a complex capitalist economy is when financial and securities regulation improves the accuracy of information investors have regarding the value of their investments and how well corporate management acts to promote their interests. This concern leads to regulation of corporate governance and securities. The extent to which markets could supply accurate information and effective means of governance without such regulation remains a subject of debate, e.g., whether laws prohibiting insider trading promote market efficiency (Fishman and Hagerty, 1992). Despite this ongoing academic

---

[5] Guasch and Spiller (1999 at 287-300) discuss antitrust and regulation as if they were complements, but the substance of their discussion is only that regulation is not a complete or perfect substitute for antitrust enforcement.

[6] Karl McDemott observes that if antitrust is a form of regulation and, but for antitrust, firms would merge or collude to exercise market power, then "the only unregulated capitalist is an unregulated monopolist." On that score, not even the monopolist would be immune, as it would still have to deal with (and benefit from) financial regulation that protect its investors and property rights that protect its assets.

debate, recent notable bankruptcies and alleged corporate malfeasance by prominent entrants into deregulated sectors (Enron in electricity, WorldCom in telecommunications) suggest a continuing important complementarity between regulation in capital markets and competition downstream.[7]

6. PARTIAL DEREGULATION: MANAGING THE VERTICAL INTERFACE

The ways in which the law—common law, antitrust, securities—can complement markets rely on an abstract notion of "input" (transactions, rivalry, information) to a generic form of "output" (exchange, markets, investment). The most explicit and compelling concrete relationships in which regulation complements competition arise in industries where some vertical stages are open to competition while others remain monopolies. This category has come into policy prominence in industries where the impetus to replace regulation with competition, for the reasons listed above, hits limits imposed by scale economies or network externalities.

The sectors within these industries that have been difficult to regulate are related to a transport service used to distribute or move a competitive service between sellers and buyers. The three examples we review here—oil pipelines, local telecommunications, and the electricity grid—exemplify a relationship in which a regulated transportation service is an input to the downstream sale of a competitive product (oil, long distance telephony, wholesale electricity). The central initiatives in these industries took place about once per decade, with pipelines in the 1970s, telecommunications in the 1980s, and electricity in the 1990s, although no single decade contained the debates for any of them, reflecting the complexities involved. In many cases, certainly the last two, strident debates continue.

6.1 Oil Pipelines

Pipelines possess substantial scale economies for two reasons. A first contributor is the substantial fixed costs of acquiring rights-of-way and laying the lines over long distances. A second is that the material cost of the pipeline, based on the line's circumference, will be proportional to its radius. However, the carrying of the pipe will be proportional to the square of the radius. These together imply that average shipping costs, holding distance

[7] Fred Grygiel comments that regulation to prevent financial malfeasance may not be viewed as a complement by management in target firms, which will resist efforts to limit their independence in the name of promoting effective shareholder corporate governance and restoring faith in financial markets.

constant, would be inversely proportional to the radius of the pipe. That conclusion neglects the energy costs of moving the oil through the line, but to the extent that those costs are proportional to the quantity shipped, these scale economies would persist. In regions of the country with numerous oil fields, parallel pipelines have been constructed as more oil is discovered, but in others, a single large pipeline serves a specific market. One is the Trans Alaska Pipeline System (TAPS), running from the Prudhoe Bay oil fields on the North Slope to a terminal in Valdez. A second is the Colonial Pipeline, delivering petroleum products from the Louisiana to points in the southeastern U.S. and as far north as New Jersey.

Prior to the mid-1970s, oil pipelines were nominally regulated, but in practice regulation was nonexistent or ineffective (Spavins, 1979). Pipelines were generally owned by the oil companies that shipped oil or refined products through them. If the shippers own all of the oil upstream of the pipeline and lack market power at the termination point, they have no incentive to depress upstream values and no ability to increase oil prices downstream, rendering regulation unnecessary. However, if vertical integration is not complete, the pipeline owners have an incentive to raise fees to unintegrated shippers. If the output from the pipeline were substantial enough to set the downstream price, the shippers would have an incentive to reduce capacity of the pipeline in order to raise that price.

The Antitrust Division of the Department of Justice undertook a two-pronged initiative to counter pipeline market power. The first prong was to reform pipeline regulation where pipelines held market power so it followed a credible method that would lead to reasonable rates.[8] Toward this end, it joined rate proceedings at the Federal Energy Regulatory Commission involving TAPS and the Williams Pipeline Company. The second prong was structural reform to allow independent shippers to obtain access at reasonable rates. Going short of full divestiture, the Department recommended "competitive rules" by which shippers would have unilateral rights to expand capacity and obtain ownership interests in the line consonant with these reasonable rates (Shenefield, 1978).

A question debated at the time and still unresolved is the extent to which these "competitive rules" require underlying regulation or replace them. If latecomers make no contribution to initial construction costs, they obtain a "second mover advantage," with no firm bearing the initial cost of the facility. In the alternative, if latecomers have to become owners, the purchase price of an ownership share has to be regulated. As Shenefield (1978 at 208) observed, this merely replaces short-term per-unit regulation

[8] As noted above, many pipelines overlap, leading the Department of Justice to recommend deregulation for all crude oil pipelines except TAPS and a few major product pipelines (Untiet, 1987).

with longer-term capacity access regulation. In either case, whether the regulation is in the form or price, expansion rights, or some combination of the above, oil pipelines provided an initial example of how regulation at the transport stage can complement competition in related product markets.

6.2 Telecommunications

Following the 1956 settlement of the antitrust case filed against it in 1949 by the U.S. Department of Justice (DOJ), AT&T agreed to limit its operations to the communications sector. This was done to keep AT&T from leveraging its telephone monopoly into other markets, such as broadcasting or computing. During the 1950s and 1960s, the separation problem became more acute as markets within the telecommunications sector—long distance service, customer premises equipment (CPE), and "enhanced" information services—saw increasing entry, while the core local service market remained a regulated monopoly.

Solutions to the separation problem had different degrees of success in different sectors. Opening CPE markets was relatively successful, following court mandates for the Federal Communications Commission (FCC) to permit customers to use non-AT&T equipment on AT&T's network. After AT&T's efforts to block interconnection directly or through exorbitant tariffs were thwarted, the FCC came up with interconnection standards that permitted entry, competition, and declining prices of CPE (Brock, 1994 at ch. 6).

Long distance was less successful, in part because entry threatened not just AT&T's monopoly in that market but an inefficient system of taxing long distance service to hold down local rates (Brock, 1994 at ch. 8). Even after the courts forced AT&T to originate and terminate long distance traffic carried by independent microwave-based carriers, competition in this sector was slow to develop. FCC-mandated negotiations over the tax long distance entrants should pay were protracted. A related issue was that AT&T continued to give these entrants inferior connections that increased consumer inconvenience, billing uncertainty, and noisier lines.

The FCC continued to attempt to complement nascent competition in long distance and what was then called "enhanced services" or "information services," using telephone lines and the network to process and communicate higher speed digital communications, through behavioral regulations that allowed AT&T to continue to participate in both regulated and unregulated markets. Dissatisfaction with the results, and a belief that AT&T's discriminatory refusals to interconnect and "pricing without regard to cost" violated the antitrust laws, led to another DOJ lawsuit in 1974 (Brennan, 1987). The resulting divestiture ten years later put the regulated

monopolies in new companies that, with the exception of selling CPE, were not allowed to participate in competitive markets. The divestiture model left open the problem of how to set "access charges" in ways that maximized competitive potential in long distance while respecting the political imperatives behind the local service subsidy. However, this model ushered in a competitive era not only in long distance but also in enhanced services, CPE manufacturing, and central office switches.

For a variety of legal and political reasons, the "quarantine" established by the divestiture was not to last. From the standpoint of complementarity between regulation and competition, the most interesting development involved the possibility of competition within the local telephone exchange. Local telephone service had been a natural monopoly for two reasons—physical economies of scale in constructing local lines and switches, and "network externalities" created by the value of everyone being on the same system. Technological advances could make this market competitive if they both eliminated these physical scale economies and facilitated interconnection between customers of these different local carriers. The latter would, in effect, turn the "network externalities" into a public good by no longer requiring that they can be achieved only through a single firm owning the entire system.

The Telecommunications Act of 1996 and its subsequent implementation by the Federal Communications Commission may be viewed as Congress's attempt to institute a regulatory scheme that would accomplish two goals. A first is the divesting of the "network externalities" from the incumbent local carrier by forcing it to interconnect with competitors under nondiscriminatory terms and conditions overseen by state and federal regulators. A second was allowing new entrants to purchase "network elements" or the entire service at wholesale from incumbents if pieces of the local exchange, or the entire exchange, were not provided by competitors. An important adjunct was to set up conditions under which incumbent local exchanges would be allowed to provide long-distance service to their customers. The FCC's primary responsibilities have been to define the "network elements" and set their prices, and to decide when statutory conditions for allowing incumbent local carriers to provide long distance service in any given state are satisfied.

In theory, this regulatory structure would optimally select out those markets within the local exchange that are competitive from those that are not, and allow efficient pricing of the latter to facilitate competition in the former. Whether this scheme has been successful was and remains contentious (Crew, Kleindorfer, and Sumpter, 2004). The FCC's definitions of network elements and methods for setting prices have been to the

Supreme Court twice.[9] The pricing scheme has been criticized as being less than required to compensate incumbent carriers for the costs incurred in providing these services (Kahn, 2001).[10] Network element definition was the subject of considerable controversy in August 2003 when a majority of the FCC, not including its chairman, voted to let states continue to decide when switching should be offered by incumbents to their competitors; the D.C. Circuit in March 2004 overturned the FCC's decision.[11] The lesson seems to be that effective regulation is necessary to complement competition, but that designing and implementing such regulation is costly. If so, when regulation is difficult, it may be competition in complementary markets that suffers.

6.3 Electricity

Interest in introducing competition in electricity began with technological change that reduced scale economies in generation, allowing multiple suppliers vying to serve a particular set of customers (Joskow and Schmalensee, 1983). This realization was contemporaneous with initiatives under the 1978 Public Utility Regulatory Policy Act (PURPA) to open up utility networks to the carriage of electricity from cogenerators and renewable fuel users. Although motivated by conservation, the PURPA experience showed that the transmission system could deliver electricity from independent power producers.

In 1992, Congress passed the Energy Policy Act to open utility grids to all independent suppliers. The Federal Energy Regulatory Commission (FERC) issued its first implementation orders, Orders 888 and 889, in 1996, setting out principles for independent transmission system operation, nondiscriminatory access, information provision, and cost recovery.[12] These regulations were followed by FERC's Order 2000, setting out principles for

---

[9] AT&T vs. Iowa Utilities Board, 525 U.S. 366 (1999), Verizon Communications vs. FCC, 535 U.S. 467 (2002).

[10] One controversy is whether a standard that suggests that network element prices should include a premium for historical cost would warrant a larger than otherwise discount for competitors who want to purchase local service at wholesale for resale as part of an integrated telecommunications package.

[11] Federal Communications Commission, "Report and Order on Remand and Further Notice of Proposed Rulemaking," In the Matter Of Review of the Section 251 Unbundling Obligations of Incumbent Local Exchange Carriers," CC Docket No. 01-338, Aug. 21, 2003; U.S. Telecom Association v. Federal Communications Commission, 359 F.3d. 554 (D.C. Cir. 2004).

[12] Federal Energy Regulatory Commission, "Promoting Wholesale Competition Through Open Access Non-discriminatory Transmission Services by Public Utilities; Recovery of Stranded Costs by Public Utilities and Transmitting Utilities," Order No. 888 (Apr. 24, 1996); "Open Access Same-Time Information System (formerly Real-Time Information Networks) and Standards of Conduct," Order No. 889 (Apr. 24, 1996).

"regional transmission organizations" (RTOs) that would manage transmission grids over multistate areas.[13] While all of this was going on at the federal level, many states were making moves to open retail electricity markets—direct sales to end users—under their authority, up until the California electricity market crisis of 2000-01 (BPM, 2002 at chs. 3-5).

Corporate, political, and technical realities combine to make the effort to open electricity markets at least as and perhaps more complex than similar efforts in telecommunications. The most fundamental problems arise out of the physics of the transmission system and electricity itself. Because storing electricity is expensive, production has to be kept continually equal to consumption. A system will be more reliable the more it is interconnected, where surplus power in one region can make up for shortfalls in others. Since electricity takes all paths from its origin to its destination—routing costs are prohibitive as well—grids thus become regional monopolies, despite being owned by particular utilities and falling to some extent within the purview of individual states.

Interconnectedness of the grid has another important byproduct. If one electricity supplier fails to meet the consumption of its customers, not only will its customers be blacked out, but so too will all others. The industry will require an ongoing central authority—private grid operator or public regulator—to ensure that the grid is kept stable. Whether the reach of this central authority needs to so large to keep meaningful competition from succeeding is perhaps the core policy question facing the electricity sector (BPM, 2002 at 194-97).

If competition remains feasible, the interconnectedness of the regional grid may require additional regulation to handle the "blackout externality," i.e., the fact that a load imbalance at one point in the system can cause blackouts elsewhere. Such externalities can justify policies to require "real time meters," conservation programs to limit consumption at peak demand periods, and capacity reserve requirements (Brennan, 1998, 2003, 2004). Facilitating competition consistent with reliability will also require that operational and regulatory authority over transmissions systems be regional, and perhaps international, to match the regions over which grids operate (BPM, 2002 at ch. 12).

Even if acceptable reliability can be achieved through such measures, the problem of setting prices remains. Ideally, prices for transmission would be negligible in low demand periods when grids have excess capacity, and rise to reflect the opportunity cost of capacity when lines are congested. Since congestion can vary along different segments of the grid, many have argued that prices should be "nodal," with separate rates for separate routes (Hogan,

---

[13] Federal Energy Regulatory Commission, "Regional Transmission Organizations," Order No. 2000 (Dec. 20, 1999).

1992). Allowing rates to rise with congestion, however, gives the holder of such rights the incentive to reduce transmission capacity. FERC's longstanding interest in independent system operation and regional transmission organizations also reflects a concern that utilities controlling a grid may not provide equal access to unaffiliated power producers (BPM, 2002 at 76-80).[14]

For all of these reasons, effective regulation of substantial components of the electricity sector, and exercised over wide regions, remains a necessary complement to competition in wholesale and retail electricity markets. How that should best be carried out remain would be contentious and difficult even without traumatizing events such as the California crisis of 2000-01 and the Northeast blackout on August 14, 2003.

7. EXTENSIONS AND IMPLICATIONS

Complementarity between regulation and competition is not limited to common law and sector-specific regulation of natural monopolies within industries. Information used by consumers to evaluate products may be asymmetrically distributed or otherwise costly to provide and convey in a credible manner, leading to the dissolution of markets for those products. Regulation that assures consumers of minimal safety or quality standards or facilitates product comparisons can facilitate entry and competition that might not otherwise take place.

Complementarity can go in the reverse direction, with markets used to assist regulation. When regulation appears necessary to mitigate a market failure, competition may help attain the most efficient results. A key difficulty with information is getting good information on the costs of supplying the regulated service. In that sense, cost information is an input to regulation, just as transactions are an input to market exchange. Competition can complement such regulation by taking advantage of the market's ability to reveal costs.

The leading example of using competition to complement environmental regulation is trading of emissions permits (Schmalensee *et. al.,* 1998).

---

[14] Another potential boundary line between regulated and competitive sectors in electricity is between local distribution, a regulated monopoly, and retail selling of electricity. Full separation here would require that traditional utilities would no longer sell electricity to industrial, commercial, and residential users. It remains less than clear whether residential users have sufficient interest in choosing energy suppliers to make retail competition for that sector worthwhile. Some predict that it will not happen, with the only competition being that in which the state regulator chooses the default supplier for consumers who do not bother to pick a competitive power company. (Thanks to Roger Comacho for suggesting this point.)

Competitive permit markets supply information as to whom can comply with regulatory requirements at least cost, improving regulatory performance and making it possible to achieve reduced concentrations of airborne pollutants.[15] Markets also can help find the least cost providers of services mandated by regulation. Examples include auctioning telecommunications spectrum for specific uses (Coase, 1959) or competing to be the provider of universal or default services (Salant, 2004).

The complementarity of regulation and competition in partially regulated industries will turn on the degree to which the implementation of the former in monopoly sectors can be insulated from vertically related competitive markets. The lessons from telecommunications and electricity are that insulation will be most successful when firms that provide regulated services are kept out of competitive markets, through a divestiture if the separate services had been provided by a vertically integrated firm. With a clean break, incentives for discrimination and the ability to cross-subsidize disappear. Without a clean break, regulators, the regulated incumbent, unaffiliated competitors, and consumers engage in what seem to be eternal legislative, administrative, and legal struggles over how competition is implemented. The best case for complementarity will thus be when the regulated and competitive services are complements on the demand side, but not the supply side. Supply-side complementarity will typically imply scope economies that make it costly, perhaps prohibitively so, to separate regulated and competitive sectors sufficiently to allow competition in vertically related markets to thrive.[16]

Finally, the conventional linkage between economic efficiency and political libertarianism breaks down if regulation and competition are complements. Advocates of competition frequently claim that markets both deliver goods efficiently to consumers and minimize the scope of state control over private conduct. However, the U.S. experience with oil pipelines, telecommunications, and electricity indicates that regulation—of a firm's price, with whom it deals, and the services it can provide—may be required to promote competition in other markets. In some settings, e.g., partially deregulated sectors, market advocates may have to decide between

---

[15] The net benefits of an emissions permit-trading scheme, relative to command-and-control alternatives, depend on the degree to which emissions from one source have the same harms as emissions from another. Differences in effects across sources tend to reduce the relative advantages of permit trading (Oates, Portney and McGartland, 1989).

[16] In the railroad industry, effective regulation of a monopoly infrastructure (railroad tracks) could complement a competitive market (in running trains). Citing both efficiencies in safety and operations from coordinating track and train operations, and the possibility of economies of density in running trains themselves, Pittman (2003) expresses skepticism that such separation would be effective.

competition in one market and regulation in others, rather than assuming that competition and freedom from state regulation always fit hand-in-hand.

Attaining the best from regulation and competition requires that we declare an end to the war between the two. Treating policy choices as a simplistic, ideologically driven battle between "regulation" and "competition" may do justice neither to those institutions nor to the public that they serve.

REFERENCES

Averch, Harry and Leland Johnson. 1962. "Behavior of the Firm Under Regulatory Constraint." *American Economic Review* 52: 1052-69.

Baron, David and Roger Myerson. 1982. "Regulating a Monopolist with Unknown Costs." *Econometrica* 50: 911-30.

Baumol, William and David Bradford. 1970. "Optimal Departures from Marginal Cost Pricing." *American Economic Review* 60: 265-83.

Behr, Peter and Steven Gray. 2003. "Grid Operators Spotted Overloads, But Ind. Controllers Couldn't Force Power Companies to Cut Output." *Washington Post.* Sep. 5, 2003: E1, E10.

Brennan, Timothy. 1987. "Why Regulated Firms Should Be Kept Out Of Unregulated Markets: Understanding the Divestiture in *U.S. v. AT&T." Antitrust Bulletin* 32: 741-93.

Brennan, Timothy. 1989. "Regulating by 'Capping' *Prices." Journal of Regulatory Economics* 1: 133-47.

Brennan, Timothy. 1998. "Demand-Side Management Programs Under Retail Electricity Competition," Resources for the Future Discussion Paper 99-02 (1998), available at http://www.rff.org/disc_papers/PDF_files/9902.pdf.

Brennan, Timothy. 2003. "Electricity Capacity Requirements: Who Pays?" *Electricity Journal* 16(8): 11-22.

Brennan, Timothy. 2004. "Market Failures in Real-Time Metering: A Theoretical Look." *Journal of Regulatory Economics* 16 (8).

Brennan, Timothy, Karen Palmer and Salvador Martinez. 2002. *Alternating Currents: Electricity Markets and Public Policy.* Washington: Resources for the Future.

Breyer, Stephen. 1984. *Regulation and Its Reform.* Cambridge, MA: Harvard University Press).

Brock, Gerald. 1994. *Telecommunication Policy for the Information Age: From Monopoly to Competition.* Cambridge, MA: Harvard University Press.

Coase, Ronald. 1959. "The Federal Communications Commission." *Journal of Law and Economics* 2: 1-40.

Coase, Ronald. 1960. "The Problem of Social Cost." *Journal of Law and Economics* 3: 1 -44.

Crew, Michael, Paul Kleindorfer, and John Sumpter. 2004. "Bringing Competition to Telecommunications by Divesting the RBOCs." In *Obtaining the Bestfrom Regulation and Competition,* edited by M.A. Crew and M. Spiegel. Boston, MA: Kluwer Academic Publishers.

Department of Justice and Federal Trade Commission. 1997. *Horizontal Merger Guidelines.* Washington: U.S. Department of Justice.

Fishman, Michael and Kathleen Hagerty. 1992. "Insider Trading and the Efficiency of Stock Prices." *RAND Journal of Economics* 23: 106-22.

Guasch, J. Luis and Pablo Spiller. 1999. *Managing the Regulatory Process: Design, Concepts, Issues, and the Latin America and Caribbean Story.* Washington: The World Bank.

Hayek, Freidrich. 1945. "The Uses of Knowledge in Society." *American Economic Review* 35: 519-30.

Hogan, William. 1992. "Contract Networks for Electric Power Transmission", *Journal of Regulatory Economics* 4: 211-42.

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

Kahn, Alfred. 2001. "Telecom Deregulation: The Abominable TELRIC-BS." Washington: The Manhattan Institute. Available at www.manhattan-institute.org/html/kahn.htm.

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

McDermott, Karl and Carl Peterson. 2004. "The Anatomy of Institutional and Organizational Failure." In *Obtaining the Bestfrom Regulation and Competition,* edited by M.A. Crew and M. Spiegel. Boston, MA: Kluwer Academic Publishers.

Oates, Wallace, Paul Portney, and Albert McGartland, "The Net Benefits of Incentive-Based Regulation: A Case Study of Environmental Standard Setting." *American Economic Review* 79: 1233-42.

Peltzman, Sam. 1976. "Toward a More General Theory of Regulation." *Journal of Law and Economics* 14: 109-48.

Pittman, Russell. 2003. "Vertical Restructuring (or Not) of the Infrastructure Sectors of Transition Economies." *Journal of Industry, Competition and Trade* 3: 5-26.

Posner, Richard. 1971. "Taxation by Regulation." *Bell Journal of Economics and Management Science* 2: 22-50.

Price, Catherine. 1994. "Gas Regulation and Competition: Substitutes or Complements?" In *Privatization and Economic Performance,* edited by M. Bishop, J. Kay, and C. Mayer. Oxford: Oxford University Press: 137-61.

Salant, David. 2004. "Multi-Lot Auctions: Application to Regulatory Restructuring." In *Obtaining the Bestfrom Regulation and Competition,* edited by M.A. Crew and M. Spiegel. Boston, MA: Kluwer Academic Publishers.

Schmalensee, Richard, Paul Joskow, Denny Ellerman, Juan Pablo Montero, and Elizabeth Bailey. 1998. "An Interim Evaluation of Sulfur Dioxide Emissions Trading." *Journal of Economic Perspectives* 12: 53-68.

Sen, Amartya K. 1970. "Conflicts and Dilemmas." in *Collective Choice and Social Welfare.* San Francisco: Holden Day: 78-86.

Shenefield, John. 1978. Testimony Before the Subcommittee on Antitrust and Monopoly of the Committee on the Judiciary, United States Senate, Concerning Oil Company Ownership of Pipelines, reprinted in *Oil Pipelines and Public Policy: Analysis of Proposalsfor Industry Reform and Reorganization,* edited by E. Mitchell. Washington: American Enterprise Institute: 191-215.

Spavins, Thomas. 1979. "The Regulation of Oil Pipelines." In *Oil Pipelines and Public Policy: Analysis of Proposals for Industry Reform and Reorganization,* edited by E. Mitchell. Washington: American Enterprise Institute: 77-105.

Stigler, George. 1971. "The Theory of Economic Regulation." *Bell Journal of Economics and Management Science* 2:3-21.

Stigler, George. 1981. "Comment on Joskow and Noll." In *Studies in Public Regulation,* edited by G. Fromm. Cambridge, MA: MIT Press: 73-77.

Untiet, Charles. 1987. "The Economics of Oil Pipeline Deregulation: A Review and Extension of the DOJ Report." Antitrust Division Economic Analysis Group Discussion Paper EAG 87-3. Washington: U.S. Department of Justice.

Winston, Clifford. 1993. "Economic Deregulation: Days of Reckoning for Microeconomists." *Journal of Economic Literature* 31: 1263-89.

Chapter 2

Bringing Competition to Telecommunications by Divesting the RBOCs*

Michael A. Crew, Paul R. Kleindorfer, and John Sumpter
Rutgers University, University of Pennsylvania, and PacWest Telecommunications

1. INTRODUCTION

The road to competition in telecommunications has been a long one and, despite the major technological advances in microelectronics, wireless and optical fiber, the industry is only partially competitive. In this paper we argue that the main barriers to competition are the Regional Bell Operating Companies (RBOCs) as they control the bottleneck of access to the local wireline network. The Telecommunications Act of 1996 (the 96 Act) attempted to change this by allowing RBOCs into long-distance, provided they opened up their networks to competitors. This has proved to be very difficult to do because of the nature of the local networks and the problems of interconnections to them. These problems have meant that the competitors known as "Competitive Local Exchange Carriers" (CLECs) have not been able to compete on equal terms with the RBOCs. As the RBOCs were the gatekeepers, the CLECs were always concerned that absent regulation, the RBOCs would gouge them with high prices and that even with regulation the

* We would like to thank Ralph Ahn for programming assistance. David Mandy and Dennis Weisman were generous in reading our paper very carefully and provided us with some critical but useful comments, for which we thank them. This paper was also presented at the CRRI's 16th Annual Western Conference, June 25-27, 2003, San Diego, California; we would like to thank the participants for their helpful comments not least the discussant, Steve Levinson, and Tim Brennan.

RBOCs would sabotage their operations. In this paper we revisit this problem, which has been of considerable interest in the literature and in regulatory proceedings. In particular, we argue that the bottleneck, sabotage and monopoly issues are such that divestiture by the RBOCs of their local networks, albeit a very drastic step, currently is the most promising approach to making the industry more competitive relative to the main alternatives under consideration, namely, the *status quo,* the creation of fully separated subsidiaries (but wholly owned by the RBOCs) and to lifting all regulation of RBOCs, namely *laissez-faire.*

The paper proceeds in section 2 by stating the problem. It goes beyond the background and summary of existing work on sabotage and economies foregone and discusses the welfare tradeoffs. Section 3 provides a simple model of sabotage that extends a model developed by Weisman and Kang (2001), henceforth, WK. The principal result is that divestiture is welfare enhancing absent major losses in economies of scope. Section 4 sets out our proposal and examines some issues of practical implementation. Section 5 is by way of summary and implications. An appendix provides a proof of the major proposition on divestiture.

2. BACKGROUND AND STATEMENT OF THE PROBLEM

Following the Bell System Divestiture in 1984 (the 1984 Divestiture), which separated AT&T from its local operating companies, the industry has undergone major change, including some considerable technological advances and new forms of competition especially from wireless. The long distance market became intensely competitive as a result of the 1984 divestiture, which was its intent. Long distance competition resulted from equal access and balloting. We should note here that, in this context, equal access provides effectively identical use of the local network by Long Distance Carriers to connect to their customers. Such equal access has not taken place for local exchange competition. The 1984 Divestiture was not designed to result in local competition. That was left to the 1996 Act. However, the networks owned and operated by the RBOCs, although they have been substantially upgraded and have benefited from technological advances, have remained a bottleneck, a monopoly in effect. RBOCs assert that significant competition exists. It is true that competition exists from wireless. However, there is minimal competition from other wireline technologies. The competition from cable that appears to be the main plank of C. Michael Armstrong's vision for a vertically integrated competitor to the RBOCs' fixed networks has not yet materialized Despite powerful pleas

by the RBOCs and others, for example, Danner and Wilk (2003), we find it hard to characterize the RBOCs' networks otherwise than a bottleneck, a monopoly. They have similar properties to electricity distribution networks. Just as the RBOCs face competition from wireless networks, so do electricity distribution networks face competition from gas distribution networks in the energy market. In the case of electricity, gas and local telecommunications regulation of price, terms and conditions of service exists primarily because of significant monopoly power. In the case of wireless there are several operators and competition appears to be vigorous, almost certainly too vigorous for the operators' tastes. Moreover, wireless operators in addition to providing some competition with local access provide more competition for long distance service. Indeed, the competition from wireless is such that the distinction between local and long distance has now blurred. Arguments by the RBOCs to the effect that their networks should no longer be regulated have been considered carefully by regulators and generally but not universally rejected. While the recent FCC "Triennial Review" decision appears to have changed the regulatory landscape regarding CLEC access to the RBOC's local network, it did not deregulate the market for local telephone service.

We are proceeding on the basis that the RBOCs have a monopoly in their local networks and will therefore be subject to regulation, essentially the situation that has prevailed since 1996 but has much longer roots. We therefore see RBOCs as fully vertically integrated suppliers of telecommunications services who are obliged by regulatory authority to provide access to their networks to CLECs. This situation confers huge advantages on the RBOCs relative to the other carriers that rely on access to the local network. This advantage is *per se* a major problem when it comes to competition. If one firm has great advantages relative to another, then the situation is the basic structure of a natural monopoly, implying great difficulties when a competitive solution is sought. These advantages drive our argument for divestiture of the RBOCs' local networks in that divestiture puts all players on equal terms but with possible loss of economies of scope. If, in addition to these advantages, the RBOCs are also in a position to sabotage, discriminate against or otherwise disadvantage rivals then the case for divestiture is even stronger.

For some time there has been a concern in practice and in the regulatory economics literature as to whether vertically integrated providers (VIPs) like the RBOCs have an incentive to discriminate. For example, Economides (1988), Mandy (2000) and Weisman (1995), WK have studied this situation at some length. Mandy provides a summary and analysis of the state of the debate including a critical review of the assumptions employed by the

participants of the debate. WK, (p125) summarize the results of their analysis as follows.

> *Discrimination always arises in equilibrium when the vertically integrated provider (VIP) is no less efficient than its rivals in the downstream market, but it does not always arise when the VIP is less efficient than its rivals.* Numerical simulations that parameterize the regulator's ability to monitor discrimination in the case of long-distance telephone service in the U.S. reveal that *pronounced efficiency differentials are required for the incentive to discriminate not to arise in equilibrium.*[1] [Emphasis added]

Based on this, raising rivals' costs through discrimination or sabotage clearly cannot be rejected. Reiffen and Ward (2003, p39-40) argue that "...well-established economic principles indicate that a regulated monopolist with an affiliate in an unregulated business may have an incentive to deny the affiliate's competitors access to an 'essential' input, or more generally, degrade the quality of service of the input supplied to the competitors." CLECs have always argued that the RBOCs have discriminated against them putting them at a severe disadvantage. Crandall (2001) has argued that RBOCs have received only a very small number of complaints. Mini (2001) argued that the RBOCs prior to 1996 were more cooperative to CLECs than was the old GTE.[2] Similarly, Reiffen and Ward (2003) is supportive of the hypothesis that RBOCs discriminate. RBOCs, while arguing that they do not discriminate, have provided evidence to the effect that they treat CLECs' orders for service differently from their own internally generated orders. Indeed, one of their augments is that if they were forced to form separate subsidiaries, let alone divest their networks, that they would face a dramatic increase in their costs arising from a change in ordering system.[3] Thus, there seems to be reasonable grounds to suspect that RBOCs treat CLECs differently at least with respect to processing their orders relative to

---

[1] As Dennis Weisman has noted, this assumes no quality of service penalties and parity provisions that may temper the incentive for sabotage.

[2] Recall that the old GTE was allowed to enter the long distance market essentially unrestricted while the RBOCs were not. Since 1996 the RBOC situation has been more similar to that of the old GTE. It may not be unreasonable to infer that their behavior might change bringing the "new" RBOCs behavior more closely aligned to that of the old GTE.

[3] For instance, in a current proceeding in California, SBC argues that dire consequences will result if it is forced to form a fully separate subsidiary and is forced to use non-integrated ordering systems. SBC claims that "such integrated processes are more cost efficient than...contracting with a separate network company for services," (Palmer, Reilly and Cartee 2003, page 7).

internally generated orders and that the differences between the two systems are significant as admitted by SBC. The issue is whether this different treatment constitutes sabotage or raises CLECs' costs. This is potentially an important issue as there are considerable deadweight losses associated with sabotage. By contrast the literature has concentrated on the incentives to discriminate while ignoring the welfare economic effects. We attempt, partially at least, to remedy this with our model in section 3.

3. A BASIC MODEL OF SABOTAGE AND WELFARE

We employ the basic Cournot model of WK, but add features that are essential for a welfare analysis. We focus only on the duopoly case in which there is a vertically integrated incumbent V and a single entrant E. V provides both an essential access good, priced at a regulated price *w* to E, as well as a complementary good, bundled with access. We assume some economies of scope across the bundled stages of production for V in that the unit cost of providing the access good is assumed to be $c + e$ for E and c for V, where $e \geq 0$.

We define welfare in the usual manner, incorporating the additional costs of regulatory oversight.

$$W(w,\delta) = \int_0^Q P(q)dq - P(Q)Q + \Pi_V + \Pi_E - C_R(\delta) \qquad (1)$$

where the inverse demand curve P(Q) is given by

$$P(Q) = A - BQ \qquad (2)$$

The regulatory monitoring function $C_R(\delta)$ is some convex, increasing function of regulatory precision $1/\delta$, where δ is defined as "...the maximum percentage distortion in the independent rival's complementary cost that [V] can affect without certain detection by the regulator."(WK, p129)

The VIP's profits Π_V are given by

$$\Pi_V = (w - c - e)q_E + (P(Q) - c - s_V)q_V - C_V(s_E - s_{0E}) \qquad (3)$$

where $C_V(s)$ is the cost to V of effecting discriminatory access policies to E, with

s_V = V's unit cost of the complementary good

s_E = unit cost of the complementary good for the entrant E after actions by V that may increase this unit cost

s_{0E} = initial unit cost of the complementary good for the entrant E (or the cost before distortion by V)

Thus, $s_E - s_{0E} \geq 0$ is the contrived unit cost increase resulting from V's possible sabotage or other type of discriminatory access policies. Note that these cost increases affect E, but they also require expenditure by V of $C_V(s_E - s_{0E})$ to effect.

The entrant's profits Π_E are given by

$$\Pi_E = (P(Q) - w - s_E)q_E \tag{4}$$

The following constraints are imposed on the magnitude of V's effect on E's cost of the complementary good:

$$s_{0E} \leq s_E \leq (1+\delta)s_{0E} \tag{5}$$

As in WK, we can easily compute the asymmetric Cournot equilibrium outcome, which is identical to that of WK, for the case n = 2, i.e.

$$q_V = \frac{1}{3B}[A - 2(c + s_V) + w + s_E] \tag{6}$$

$$q_E = \frac{1}{3B}[A - 2(w + s_E) + c + s_V] \tag{7}$$

Total demand and price are given by:

$$Q = q_V + q_E = \frac{1}{3B}[2A - c - s_V - w - s_E] \tag{8}$$

$$P(Q) = A - BQ = \frac{1}{3}[A + c + s_V + w + s_E] \tag{9}$$

Given the above, there are several cases that could be computed. These include solutions to the following problems:

First-Best, Welfare-Optimal Solution: Solve for all decision variables $(w, \delta, q_V, q_E, s_E)$ to maximize Welfare in (1).

Cournot-constrained Welfare-Optimal Solution: Solve for the decision variables (w, δ, s_E), with q_V and q_E determined by the Cournot solution, so as to maximize W. Think of this solution as the outcome resulting from nearly perfect regulation (the Regulator can set both δ and s_E, the latter subject to (5)), where the outcome in the market is known to everyone to be determined by Cournot-Nash interactions between V and E.

Cournot-constrained, Profit-maximizing Welfare-Optimal Solution: Solve for the decision variables (w, δ) with q_V, q_E determined by the Cournot solution and s_E determined by V so as to maximize V's profits at the Cournot solution. This is the outcome of a boundedly rational regulator who must expend resources to monitor discriminatory behavior (as embodied in the cost function $C_R(\delta)$) and who knows that the outcome in the market will be determined by V and E as Cournot, with V rationally expecting this outcome and setting s_E so as to maximize V's profits at the resulting Cournot outcome.

To obtain the Cournot-constrained, Profit-maximizing Welfare-Optimal Solution, we need to solve for the profit-maximizing s_E, given a Cournot-Nash outcome. Substituting the Cournot-Nash equilibrium solutions given above into V's profit function, we obtain:

$$\Pi_{VC} = \left(\frac{w-c-e}{3B}\right)\left[A - 2(w+s_E) + c + s_V\right] + \left(\frac{P(Q)-c-s_V}{3B}\right)\left[A - 2(c+s_V) + w + s_E\right] - C_V(s_E - s_{0E}) \tag{10}$$

where $P(Q)$ is given as the Cournot-Nash price in (9) above. Substituting for $P(Q)$ from (9), it is easily seen that (10) is strictly concave in s_E. Thus, the optimal solution to maximizing Π_{VC} in (10) subject to the linear constraint set (5) as characterized by the first order necessary conditions, is found. To make the intuition clear in what follows, let us consider the following functional specifications for C_V and C_R:

$$C_V(s) = as^2, \text{ for some } a > 0; \qquad C_R(\delta) = \frac{b}{0.1+\delta}, \text{ for some } b > 0 \tag{11}$$

C_R is a decreasing function of δ. Thus, as δ increases, regulatory precision $(1/\delta)$ decreases. With C_V given in (11), a bit of algebra on the first

order necessary conditions for maximizing (10) yields the following profit-maximizing solution s^*_E to maximizing Π_{VC} in (10) subject to (5):

$$s^*_E = Min[Max(\hat{s}_E, s_{0E}), (1+\delta)s_{0E}] \tag{12}$$

where δ is regulatory precision, as set by the regulator, and where

$$\hat{s}_E = \frac{2[A - 2(c + s_V) + w] - 6[w - c - e] + 18aBs_{0E}}{18aB - 2} \tag{13}$$

where we assume that $9aB > 1$, so that the denominator in (13) is always positive. We note that when $w \leq c + e$, (13) implies that $\hat{s}_E > s_{0E}$ whenever $q_V(s_{0V}, s_{0E}) > 0$. Thus, at the Cournot solution, sabotage is optimal for V under very general conditions.

We can now summarize the basic results for the three problems examined here.

P1. First-Best, Welfare-Optimal Solution: Solve for all decision variables $(w, \delta, q_V, q_E, s_E)$ to maximize Welfare. The solution here is easily shown to satisfy the following conditions:

1. Regulatory precision δ should be set to the least costly level possible i.e. to minimize $C_R(\delta)$ in (11);
2. No contrived cost inflation should occur: $s_E = s_{0E}$;
3. The least costly producer between V and E (which will be V if and only if $c + s_V \leq c + e + s_{0E}$) should produce total industry output;
4. Access price should be $w = 0$, and end-to-end price should be set to the marginal cost of the least costly producer.

P2. Cournot-constrained Welfare-Optimal Solution: Solve for the decision variables (w, δ, s_E), with q_V and q_E determined by the Cournot solution (6)-(7), so as to maximize W. The solution here is also easy to obtain and satisfies the following conditions:

1. Regulatory precision δ should be set to the least costly level possible;
2. No contrived cost inflation should occur: $s_E = s_{0E}$;
3. Access price should be:

$$w^* = Max[2c + 6e + 5s_{0E} - 4s_V - A, 0] \tag{14}$$

Here we would expect some loss in welfare relative to First Best because Cournot competition leads to non-zero profits and non-marginal cost pricing,

but still we would expect to see no wasted effort in distorting costs, and this is true.

P3. Cournot-constrained, Profit-maximizing Welfare-Optimal Solution: The Regulator sets the decision variables *(w,* δ*)* with q_V, q_E then determined by the Cournot solution and s_E determined by V according to (12)-(13), i.e. so as to maximize V's profits at the Cournot solution. In this case, as expected from WK, V does find it optimal to drive up its rival's cost. A tradeoff ensues for both V and the Regulator, in that it is costly for V to drive up E's costs, and it is costly for the Regulator to monitor V so as to mitigate the contrived cost increases inflicted on E. Interestingly, the Regulator uses the access price w to partially correct for the excess profits generated at Cournot equilibrium for both V and *E.* By decreasing access price, especially in the case in which V is more efficient than E, the Regulator can, in effect, drive E's costs down through decreasing access price, and can move the resulting Cournot-Nash equilibrium towards more efficient industry-wide pricing and output. In the process, some of V's profits are, of course, sacrificed as *w* is set below cost *(c + e)* to provide E access.[4]

Table 1 below shows the base case values for the numerical examples and Table 2 shows the solutions to problems P1, P2 and P3 above. We provide two solutions to P2 and P3. The first of these P2a-P3a does not restrict the range of the access price w; the second P2b-P3b restricts w to be no lower than V's cost of providing the access good to E *c + e.* Interestingly, the former case provides higher welfare under both the assumptions of P2 and P3. The reason is simple. The Regulator can increase the Cournot industry output by decreasing *w* (see (8)), sacrificing V's profits in the interest of increasing industry output and decreasing price. If the Regulator is able to control industry output through adjustment of the price instrument "w", then it is efficient to do so. Indeed, even if V's profits are constrained to be non-negative (see case P3a below), a price of $w = 0$ can be the most efficient non-negative price when the Regulator anticipates that Cournot competition between V and E will determine the ultimate market price of the bundled good.

Table 1: Parameters for Base Case

A	B	Sv	C	e	a	b	s0E
1000	*5*	200	*100*	*20*	*0.4*	200	*180*

[4] The idea of setting the price of access below cost is not original to us; for example, Mandy (2001) and Panzar and Sibley (1989).

Table 2: Illustrating Base-Case Results

	Case P1a	Case P2a	Case P2b	Case P3a	Case P3b
w	120.00	0.00	120.00	0.00	120.00
δ	1.000	1.000	1.000	0.015	0.020
s_E	180.00	180.00	180.00	182.73	183.62
Π_V	0.00	-44.44	10888.89	61.50	10982.78
Π_E	0.00	19635.56	10888.89	19408.45	10664.99
q_E	0.00	63.00	47.00	62.00	46.00
q_V	140.00	39.00	47.00	39	47.00
P(Q)	300.00	493.33	533.33	494.24	534.54
W	48818.18	45080.4	43373.74	43312.11	41647.81

Note that δ in our examples is constrained to be no greater than 1, so that $\delta = 1$ is the most relaxed, i.e. least costly, regulatory oversight of potential discriminatory behavior. For example, P1 a is the First-Best Solution, while P2a is the welfare-optimal solution when the Regulator is constrained to Cournot outcomes after setting *w,* δ and s_E. P2b is the same as P2a, except here the Regulator requires V to sell access at no lower than marginal cost *(c + e)*. P3a and P3b are the (more realistic) outcomes associated with allowing V to raise E's costs, subject to the endogenous regulatory constraint (5), with δ set by the Regulator to maximize welfare, given the anticipated discrimination of V and the cost of controlling it. Comparing P1 with P2b and P3b, we see that welfare, output and the magnitude of sabotage $(s_E - s_{0E})$ move in the expected direction, following the results recorded above.

Now let us consider a final approach to the problem above, that of divestiture. In this case, a separate firm, denoted D (for divested firm) takes over the assets of V associated with producing the access good, leaving the former firm V with only assets for the upstream or complementary good, which competes directly with E. The profits of the divested access firm, Π_D, are defined as[5]:

$$\Pi_D = (w - c - e)(q_E + q_V) - C_D(s_V - s_{0V}) - C_D(s_E - s_{0E}) \qquad (15)$$

[5] Note that we that V no longer has a cost advantage relative to E in integrating with its former parent division, now the divested firm D, in the sense that the provision of access carries a constant marginal cost for both V and E of *c + e*. We maintain the cost disadvantage "e", now applying to both V and E, in order to do a welfare comparison with the undivested case.

where $C_D(s)$ enjoys the same properties as C_V (e.g., is of the form (11) with $s = s_E - s_{0E}$), and where the entrant E and the now upstream divested division of V make profits (compare with (3) and (4))[6]:

$$\Pi_{VD} = (P(Q) - w - s_V)q_V \tag{16}$$

$$\Pi_{ED} = (P(Q) - w - s_E)q_E \tag{17}$$

Assuming as before that V and E compete in Cournot-Nash fashion, we obtain from (16)-(17) the following Cournot outcomes:

$$q_{VD} = \frac{1}{3B}[A - 2s_V - w + s_E] \tag{18}$$

$$q_{ED} = \frac{1}{3B}[A - 2s_E - w + s_V] \tag{19}$$

Total demand and price at the Cournot equilibrium are then given by:

$$Q_D = q_{VD} + q_{ED} = \frac{1}{3B}[2A - 2w - s_V - s_E] \tag{20}$$

$$P(Q) = A - BQ = \frac{1}{3}[A + 2w + s_V + s_E] \tag{21}$$

These results lead to the following problem:

P4. Cournot-constrained, Profit-maximizing Divested Solution: The Regulator sets the decision variables (w, δ) to optimize welfare (see (23) below) with q_V, q_E then determined by the Cournot solution (18)-(21), where D chooses s_V and s_E to maximize (15) subject to:

$$s_{0F} \le s_F \le (1 + \delta)s_{0F}, \quad F \in \{E, V\} \tag{22}$$

---

[6] Note that we assume that V is no longer in a position to raise E's costs and therefore no longer incurs any costs itself in this regard.

where welfare is defined as:

$$W_D(w,\delta) = \int_0^{Q_D} P(q)dq - P(Q)Q_D + \Pi_{VD} + \Pi_{ED} + \Pi_D - C_R(\delta) \qquad (23)$$

Table 3 below provides results for the base case of Table 1 for the Divested model. For comparison purposes, we restrict attention to the case in which $w \geq c + e$ is required, so that the access good is not subsidized. Thus, the column "Case P3, e = 20" repeats the results of Table 2 above for Case P3b). The columns "Case P4, e = n" reflect the results of maximizing (23) subject to (22) and the constraint $w \geq c + e$. Comparing P3 with P4, we see the expected result that welfare is increased when moving to divestiture. Of course, if the access cost difference "e" for divested and undivested access were high enough, then divestiture would not be welfare enhancing.[7] This implicit assumption and argument of our paper is that e is not sufficiently large to undermine the welfare benefits of our divestiture proposal.

These numerical results illustrate the following general conclusions for this case.

1. Subject to the breakeven constraint $\Pi_D \geq 0$, the welfare-optimal access price, given Cournot competition between V and E, is $w^* = c + e$.

2. From (15), we see that as long as access price w is set no lower than c + e, so that D can earn non-negative profits, the divested firm D will no longer find it worthwhile to increase either E or V's costs (i.e., the left-hand inequality in (22) holds for both E and V at optimum), as this would simply depress E or V's output and therewith D's profits. This occurs because D has no downstream operations and therefore has no downstream profits at stake.

3. Assuming that D and its predecessor vertically integrated V are both required to make non-negative profits on access, i.e. when $w \geq c + e$ is imposed, welfare is increased in the divested case relative to the undivested case for e sufficiently small.[8]

---

[7] For the base case data of Table 1, the critical value of e* = 28.61. For values less than e* divestiture is welfare enhancing, and for values greater than e* divestiture decreases welfare.

[8] The proof requires the mild additional requirement that the vertically integrated firm V produces non-zero output at the Cournot equilibrium when no cost distortion occurs, that is when $s_F = s_{0F}$, for F = V, E in (6).

Table 3: Illustrating the Results of Divestiture on Welfare (Subject to the Constraint that $w \geq c + e$)

	Case P3 e=0	Case P3 e=20	Case P4 e=0	Case P4 e=20
w	100.00	120.00	100.00	120.00
δ	0.018	0.020	1.00	1.00
s_E	183.30	183.62	180.00	180.00
s_V	200.00	200.00	200.00	200.00
Π_V	10361.13	10982.78	10275.56	9680.00
Π_E	11952.89	10664.99	12168.89	11520.00
q_E	49.00	46.00	49.33	48.00
q_V	46.00	47.00	45.33	44.00
P(Q)	527.77	534.54	526.67	540.00
Π_D	NA	NA	0.00	0.00
W	42924.24	41647.81	44667.07	42178.18

Concerning the last point, a proof of this assertion is included in the Appendix. One first shows the intuitive fact that, subject to the constraint $w \geq c + e$, the optimal solution for both P3 and P4 obtains at $w = c + e$, i.e. there is no rationale for the regulator to increase access prices beyond marginal cost in this model. Next, one shows directly from (1) and (23) that, at $w = c + e$, welfare under divestiture is greater than under vertical integration for e sufficiently small. The noted claim results.[9]

The model and the simulations of this section, along with the discussion of section 1 indicate that there are potentially serious problems with a structure involving a VIP in the midst of a group of competitors. Cases P1-P3 provide a number of insights on the difficulties of attaining the benefits of competition in the case where there is a monopoly VIP. Some of the solutions are going to be difficult to achieve in practice. While, as shown in Table 2 (P2a-P3a), it may be possible to improve efficiency by forcing down the VIP's access price below cost, this is not likely to be acceptable and has not proved so. Given, the advantages conferred upon the VIP, it seems that *laissez faire* would lead to one supplier, like the old days of the integrated Bell System. The alternative is to promote competition in those parts of the telecommunications value chain where it seems to be viable and ensure competitors' access to the monopoly network on equal terms. Thus, to make the telecommunications competitive, some minimal monopoly may have to

[9] Using the values for the simulation as reported in Tables 1-3, $e < 28.61$ would result in Divestiture being preferred.

be accepted with regulation still playing a major role in the future of telecommunications.

4. NATURE OF THE DIVESTITURE

The 1984 Divestiture was executed through a court approved Plan of Reorganization (POR). The separation of monopoly "local" facilities from competitive "long distance" facilities was consummated through the POR. The 1984 Divestiture, along with the subsequent equal access and balloting, led to a vigorous competitive long distance market. This contention is supported by the FCC's non-dominance finding for AT&T, and similar findings by state commissions. The 1984 Divestiture achieved its goal.

As expected, the 1984 Divestiture did not lead to effective competition in the local market. Since the 1984 Divestiture the seven RBOCs and GTE have consolidated into four ILECs. SBC absorbed Pacific Telesis and Ameritech as well as little SNET, which was not an RBOC. Verizon absorbed NYNEX, Bell Atlantic and GTE. BellSouth and Qwest remain essentially unchanged by consolidation with other ILECs. All of the RBOCs, except Qwest, have major interests in wireless. Indeed, this is the only area in which they effectively compete against one another. While the 1984 Divestiture was not designed to deliver a competitive local market, the 96 Act was expressly intended to achieve that result.

The 96 Act immediately gave the RBOCs the opportunity to compete for long distance customers outside their regions but they did not do so. This, in itself, is testimony to the power of the VIP. The VIP, no doubt, understood the dangers of opening the door to freewheeling competition to other powerful companies, namely, fellow RBOCs. In addition, they were presumably aware of the difficulties of competing where they did not own the local network. Under our proposal each of these monopolies would be split into two independently owned companies, a wholesale network company (NetCo) and a retail company (Retail Company). The NETCO would be a wholesale only entity providing services only to other carriers. These Carriers Carriers (CCs) would be regulated and provide no retail services. Their only customers would be retail telephone carriers. While the details require considerable attention we sketch below how the industry would now be organized including the nature of the regulation.

The argument for divestiture of the RBOCs can be understood in the context of the question "should airlines own airports?" In the case of the airline industry, the monopoly element is the airport with its concourses, gates, air traffic control and runways (the literal "last mile"). The competitive airline carriers invest in their planes, ticket ordering systems and retail marketing but they obtain shared equal access to the runways and air

traffic control system from third parties. No one to our knowledge has proposed allowing an airline to own all (or any) commercial airports.[10] In the case of the RBOCs under our proposal, the network would be divested from the retail sales, marketing, billing and customer service.

The policy goal would be to obtain the benefits of competition and efficient use of bottleneck facilities. The RBOCs currently use two sets of OSSs (operating Support Systems) to process service requests. First are the internal OSSs that support the vertically integrated operations of the RBOC retail sales organization. Second is the "wholesale" OSSs established to process CLEC service requirements. Under our proposal, the RBOC would choose which set of OSSs would be used uniformly after divestiture. After divestiture, the NETCO would use the same OSSs to serve all CLECs, including the new Retail Company formed in the divestiture.

The Retail Company would (in general) retain the administrative buildings, CRIS billing system (end-user billing), retail sales and marketing organizations and systems, as well as wireless operations and facilities. NETCO would retain all central office and outside plant facilities and buildings (except for the wireless assets noted above), CABs billing system, and wholesale marketing support. There are two choices regarding "long distance" facilities. Such facilities can either be left with the NETCO, or separated using the same rules as the 1984 Divestiture POR. That choice can be made by the agency supervising the divestiture. Our preference is to leave the long distance facilities with the Retail Company so that the NETCO could provide wholesale LD capabilities to its carrier customers. Since much of the existing RBOC long distance service is provided through resale of other long distance carriers' wholesale services, this is likely to be a smaller issue than during the 1984 Divestiture.

The NETCO would remain initially regulated. As CLECs (including the new Retail Company) make new network investments, state PUCs should evaluate NETCO's market power. As network elements are duplicated and lose their status as bottleneck facilities, the degree of regulation should be reduced. The PUC could choose either classic Rate Base/Rate of Return regulation or price-cap regulation or some combination including earnings sharing. NETCO would retain the interconnection and unbundling obligations of the former RBOCs. NETCO would gain a new enthusiasm for selling UNEs to carriers, since that would become its principal source of revenue. CLECs would be less concerned about UNE pricing since all would be clearly and transparently treated the same.

---

[10] Even with this ownership safeguard, competition is not guaranteed in the airline industry. Certain hubs are, indeed, dominated by one carrier.

The regulation of the Retail Company after Divesture would be quite different from that of the NETCO. One possibility would be not to regulate it in any way, as it no longer controlled bottleneck facilities. However, given that it would initially have a market share of over eighty per cent there may be some concern about market dominance. Some transitory oversight regulation may therefore be in order. This could be similar to the manner in which AT&T was regulated after the 1984 Divestiture. Since the Retail Company would no longer control bottle-neck facilities, its form and degree of regulation should be reduced over a period of time that is expected to be shorter than the 10 years it took AT&T to be declared non-dominant by state and federal regulators. The Retail Company would have no statutory resale obligations. However, we expect that as the Retail Company comes to understand that it has no market power (assuming that is the case), it would actively seek resale opportunities.

Divestiture of the RBOCs should lead to benefits similar to the divestiture of the Bell System. Whichever OSSs are selected by NETCO, they will be applied to all interconnecting carriers on the same non-discriminatory basis, which is not the current case. Two OSS systems mechanically allow for a set of discriminatory schemes limited only by imagination. The opportunity for discrimination begins with design differences and can be a simple as staffing choices (i.e., quality of staffs). If there is an incentive to discriminate, then the existence of two systems provides the means to implement such discrimination. In practical experience, harm can be imposed via limits on volume, time to process, unexplained rejects, and errors in processing (for example, installing a digital loop w/o testing prior to turn-over). Because the opportunity for discrimination is limited by imagination, this list of examples is illustrative, not comprehensive.

The effective difference between the RBOCs' internal OSS and the systems designed for CLECs are documented in the filings made by CLECs. The significance of those differences are documented in SBC's filing with the California PUC opposing separation. In that filing, SBC admitted that its integrated system was more efficient than the OSS made available to CLECs. The fact that there are two separate systems makes it impossible for both systems to be equally effective and efficient. When all retail carriers face the same costs and processing quality to process an order and provide service, concerns regarding discriminatory treatment, sabotage and price-squeeze will be ameliorated or eliminated. Prior to divestiture, each RBOC has superior access to the network than does CLECs. That condition is eliminated in divestiture.

The competitive pressure of multiple retail carriers having the same access to the network will lead to innovation and lower costs. NETCO's

"largest" customer will initially be the Retail Company, but all vendors in any market have a largest customer. NETCO will find incentive to serve all of its customers well. The relationship between NETCO and its customers will be external and arm's length. Regulators will have an easier time ensuring non-discriminatory treatment. The NETCO will have the creative ideas of multiple customers as a resource (all of the CLECs), rather than the ideas ofjust one retailer (the centralized planning of the RBOC). [11]

5. CONCLUDING DISCUSSION

The debate on the structure and regulation of telecommunications will certainly continue. Our aim in this paper has been to propose a solution that is at the same time radical but has also been tried and tested with the 1984 Divestiture. The RBOCs, as they are currently structured, are increasingly like the vertically integrated Bell System. The major difference is that, in contrast to the old Bell System, they have lawful local competitors who are at a disadvantage because they have to use the RBOCs' bottleneck facilities. Our initial attempts at modeling and simulation have indicated that maintaining the vertically integrated structure is problematical. One possible approach is for the incumbent to sell access to the entrants at below cost. This is obviously something that the incumbent will oppose and it will encourage sabotage. This is the situation that the incumbents currently claim they face while categorically denying practicing sabotage.

By contrast divestiture, based on our analysis, is likely to be efficiency enhancing and may lead to real competition in that all the competitors compete on equal terms. Bringing about the "New Divestiture" may be difficult. It took years for the Government's case against AT&T to be resolved with the 1984 Divestiture. There is currently no litigation of equivalent magnitude against the RBOCs. What is going on is more akin to trench warfare. The RBOCs and the CLECs have entrenched positions. To the extent that state regulatory commissions reduce the payments for UNEs, or UNEPs or require the formation of fully separate subsidiaries the impact will become more unfavorable on the RBOCs, who may then decide that Divestiture makes sense or is the lesser of two evils. Their management and their shareholders might then find the idea of a NETCO and a Retail Company more attractive as it would be more difficult for regulators to set NETCO prices in a non-compensatory manner and the Retail Company

[11] As Dennis Weisman remarked to us, at least historically, divestitures can increase the market value of the companies divested. The governments attempt to punish John D. Rockefeller succeeded only in making him a much wealthier man!

formed from the former RBOC would still be the dominant player with its large market share as well as significant assets in wireless and broadband. While notionally the Retail Company would compete on equal terms in practice it would have numerous advantages over the others. If the RBOCs were to recognize these benefits relative to the *status quo* they would conceivably even initiate Divestiture.

For now we may be voices crying in the wilderness. None of the major players, like AT&T, have embraced our proposal. Other like Reiffen and Ward (2003) and Faulhaber (2003), while showing concern for the problem of sabotage by the RBOCs, have held back from arguing for the New Divestiture. Faulhaber seems to believe that it would not provide sufficient benefits and that technological change will eventually resolve most of the problems. Reiffen and Ward are concerned about the loss of scope economies. We have serious doubts that they are significant. Indeed, to us it seems likely that capital markets will do a better job than management in allocating resources and the history of management excesses in over-expansion by American business supports this view. If Reiffen and Ward are correct and there are significant scope economies then even absent sabotage the ultimate solution will be for the RBOCs to become monopolists again like the old AT&T. The RBOCs would counter by arguing that their resulting monopoly would not be like that of the old AT&T. They would contend that significant competition would exist from cable and from wireless. Given the trend to consolidate wireless, which may include RBOC takeover of wireless companies thereby reducing the competition in wireless. What could occur is a duopoly - with the RBOCs dominating wireline and wireless and with cable companies offering telephony. The direction the duopoly would take is not clear. Cable companies might attempt to acquire wireless companies so that they would be able to compete with the RBOCs in offering a full range of broadband, wireline and wireless.

The choices may be a duopoly in the form suggested (with likely increasing dominance on the part of RBOCs, or collusion) or a divested wires-only wholesale carrier providing a competitive check on the cable entities. In the latter case there would be the possibility for many competitors to buy NETCO's services including cable operators where they did not have cable operations. Clearly we have a preference for the latter as we expect the potential for significant and widespread competition in telecommunications to be limited in the other cases. It may therefore be better to limit the monopoly to a divested carriers' carrier as we propose than to allow the monopoly power of the monopolists to continue to grow as the competitors are whittled away by sabotage.

APPENDIX

The purpose of this appendix is to demonstrate the validity of the assertion made in the text concerning P4, which we formulate in the Proposition below. We first note two lemmas, the proof of which is straightforward.

Lemma 1: The solution to maximizing W_D in (23) at the Cournot solution (18)-(21) subject to $w \geq c + e$ entails $s_{0F} = s_F$, for $F \in \{E, V\}$, so that no sabotage is optimal in the divested problem subject to a breakeven constraint.

As noted in the text, Lemma 1 follows immediately by noting that if D drives up the cost of either E or V, it will simply result in decreased output at the Cournot equilibrium (18)-(19), lowering D's profits as given in (15).

Lemma 2: Assume that $q_V(s_{0V}, s_{0E}) > 0$ when $e = 0$ and $w = c$ (i.e., the Cournot solution (5) entails positive output for V when $e = 0$, $w = c$, and $s_F = s_{0F}$, for $F \in \{E, V\}$). Then the solution to maximizing W_D in (23) at the Cournot solution (18)-(21) subject to $w \geq c$ exceeds that of maximizing W_V in (1) at the Cournot solution (6)-(9) subject to $w \geq c$. Moreover, under the same conditions ($e = 0$ and $w = c$), $s^*_{EV} > s_{0E}$, sabotage occurs at the Cournot-Constrained Welfare Optimum (problem P3).

The proof of Lemma 2 follows directly from noting first that (when e = 0) maximizing either W_D or W_V subject to $w \geq c$ obtains at $w = c$. Using this together with (1) and (23) then yields, after some algebra, the desired result.

The following Proposition notes our main result concerning sabotage.

Proposition: Assume that $q_V(s_{0V}, s_{0E}) > 0$ when $e = 0$ and $w = c$. Then, for e > 0 sufficiently small, the solution to maximizing W_D in (23) at the Cournot solution (18)-(21) subject to $w \geq c+e$ exceeds that of maximizing W_V in (1) at the Cournot solution (6)-(9) subject to $w \geq c+e$.

The proof of the proposition proceeds as follows. We first note that at $w = c$ and $e = 0$, any feasible $(\delta_V, q_V, q_E, s_V = s_{0V}, s_E)$ satisfying (5)-(7) and (12)-(13) also satisfies (18)-(19) and (22). Thus, at optimum, we must have $W_D \geq W_V$. Actually, however, strict concavity of W_D in w and δ and the linearity of the constraint set (18)-(22) implies that strict inequality must obtain, i.e. $W_D > W_V$, if the solutions to maximizing W_D and W_V (at $w = c$, $e = 0$) are not identical. Noting the above Lemmas 1 and 2, sabotage occurs under regime V and does not occur under regime D, so that, in fact, the solutions are not identical under the noted conditions. We see therefore that $W_D > W_V$ when $e = 0$. But since the optimal solution is unique (under both regimes), we know that the optimal solution is continuous in e. Clearly, therefore there exists a neighborhood around $e = 0$ for which $W_D > W_V$ continues to obtain, as asserted.

REFERENCES

Robert W. Crandall. 2002. "An Assessment of the Competitive Local Exchange Carriers Five Years After the Passage of the Telecommunications Act." Washington, D.C.: Criterion Economics.

Danner, Carl R. and G. Mitchell Wilk, 2003. "The Next State in Local Telephone Regulatory Reform," In *Markets, Pricing and Deregulation of Utilities,* edited by Michael A. Crew and Joseph C. Schuh. Boston, MA: Kluwer Academic Publishers.

Economides, Nicholas.1988. "The Incentive for Non-Price Discrimination by an Input Monopolist." *Journal of Industrial Organization* 16(no. 3, May): 272-284.

Faulhaber, Gerald R. 2003. "Policy-Induced Competition: the Telecommunications Experiments," *Information Economics and Policy,* 15(no. 1, March): 73-97.

Mandy, David M. 2000. "Killing the Goose That May have Laid the Golden Egg: Only the Data Know Whether Sabotage Pays." *Journal of Regulatory Economics* 17(no. 2, March): 157-172.

Mandy, David M. 2001. "Price And Vertical Control Policies For A Vertically Integrated Upstream Monopolist When Sabotage Is Costly." Working Paper, Department of Economics University of Missouri, Columbia, MO.

Mini, Frederico. 2001. "The Role of Incentives for Opening Monopoly Markets: Comparing GTE and BOC Competition with Local Entrants." *Journal of Industrial Economics* 43(no. 3, September): 379-424.

Panzar, John and David S. Sibley. 1989. "Optimal Two-Part Tariffs for Inputs: The Case of Imperfect Competition." *Journal of Public Economics* 40: 237-249.

Palmer, William, Brian Reilly and Ruth Ann Cartee. 2003. *Analysis of the Cost Impacts of Forced Structural Separation and Divestiture.* Evanston, II: LECG.

Reiffen, David and Michael R. Ward. 2002. "Recent Empirical Evidence on Discrimination by Regulated Firms." *Journal of Network Economics*1(no. 1, March): 39-53.

Weisman, Dennis. 1995. "The Incentive to Discriminate by a Vertically Integrated Firm: the Case of RBOC Entry into Inter-LATA Long-Distance." *Journal of Regulatory Economics* 8(no. 3, November): 249-266.

Weisman, Dennis and Jaesung Kang, 2001. "Incentives for Discrimination when Upstream Monopolists Participate in Downstream Markets." *Journal of Regulatory Economics* 20(no. 2, September): 125-139.

Chapter 3

Multi-Lot Auctions

Application to Regulatory Restructuring

David Salant
Optimal Markets, Incorporated and Columbia University

1. INTRODUCTION

The details of an auction design and process can matter a great deal, and at times, in counter-intuitive ways. This is especially the case in multi-unit, multi-lot auctions for a few close substitutes. These types of auctions are becoming very common in the electricity sector. Multi-lot and/or multi-unit auctions are now commonly used for transmission rights, capacity entitlements and default service procurement. Examples include the New Jersey auction for Basic Generation Service (NJ BGS), the Regional Transmission Organizations PJM and the California Independent System Operator (CAISO) auctions financial transmission rights (FTRs), and the Texas capacity entitlements auctions. These are all cases involving multiple lots of a few types of products. The aim of this paper is to examine how fine detailed decisions about division of what is to be auctioned into lots and across time can affect the outcome of an auction. Of particular concern is how auction design can affect the likelihood of an inefficient matching of buyers and sellers. This type of inefficiency can occur in multi-lot and multi-unit auctions but not in single lot auctions. This paper explains the impact of auction design on actual electricity auctions.

This paper seeks to explain some of the theory of multi-unit, multi-lot auctions and to illustrate the impact of design on the outcome in general and to recent electricity procurement auctions in particular. As has been well-documented, auctions are now being used in lieu of administrative

proceedings for selling spectrum rights. An enormous amount has been written in both the academic and popular literature about the design of auctions for selling spectrum rights.

One of the more notable aspects of the spectrum auction experience is the fact that game theorists involvement in developing a novel approach for auctioning spectrum rights, the simultaneous, multiple-round, ascending auction (SMR). In particular, Preston McAfee as consultant to Airtouch, and Paul Milgrom and Robert Wilson advising Pacific Telesis developed the SMR auction mechanism.[1] This too has been well documented. The SMR approach has been adopted for several dozen spectrum auctions in the U.S. and elsewhere,[2] and in the energy sector for energy contracts and entitlements.[3]

Most auctions, especially in the energy sector, whether regulated or not, are still conducted using traditional sealed-bid approaches, or on-line, Yankee or English auctions.[4] Most sealed-bid procurements are multi-attribute; this means that an evaluation of winners requires a subjective assessment of relative advantages of the alternative combinations of offers. As is discussed in more detail below, sealed-bid approaches have both advantages and limitations.

The most common auction formats for selling multiple units of a single good, the English auction and variations thereof, seem especially poorly suited for situations in which a load is to be divided among multiple suppliers[5]. Each item is auctioned individually in a Standard English auction – the auctioneer will start by announcing a low price and gradually raise it as long as there is a bidder willing to accept the price. When identical lots are sold in a sequence of Standard English auctions, there is no guarantee that the prices of all lots will be the same or even that the highest valuation bidders will win the lots. An alternative multi-unit auction format, sometimes referred to as a Yankee auction, has bidding remains open for a specific period of time. At the end of the bidding window, the units are

---

1 Kwerel and Rosston (2000), McAfee and McMillan (1996) and Cramton (1997) provide discussion of the auction rule-making process.

2 SMR auctions for spectrum rights have been conducted in Australia, Austria, Canada, Germany, Guatemala, Italy, the Netherlands, Nigeria, Singapore, Switzerland, Taiwan, and the United Kingdom.

3 SMR auctions and variations have been conducted in Alberta, Canada, France, New Jersey, and Texas, and for selling pipeline capacity in Austria and Germany.

4 Yankee auctions are used here to refer to a multi-object variation of an English auction. See Easley and Tenorio (1999) for a description of *one* version of a Yankee auction.

5 Sequential English auctions give rise to declining price anomalies and afternoon effects. See Ashenfelter (1989) or Beggs and Graddy (1997) and Yankee auctions suffer deadline effects. Both can affect the efficiency of the outcome.

awarded to the highest bidders. Yankee auctions suffer from "deadline effects": bidders all try to wait until the very end to submit their bids so that they will not get topped by a small amount. A bidder can never be sure it will both beat the deadline and not give a rival a chance to respond. This means that Yankee auctions, like sequential English auctions, have high risks of inefficient allocations. In contrast, in most variations of SMR auctions, there is much less risk of this type.

The issue of appropriate auction design for multi-lot auctions has been analyzed extensively elsewhere. The simultaneous multiple round and clock auction formats were developed to ensure that auction prices of substitutes are in proportion to value differences and also allow bidders to make simultaneous offers for complementary lots. A clock auction is the most basic multi-lot auction format. A multi-unit clock (forward) auction is one in which bidders indicate how much they want to purchase at the starting price, and as the auction manager raises price, each bidder can maintain or reduce its demand. The auction ends when the total demand falls to the level of supply. In a reverse clock auction to sell, the auction manager raises, rather than reduces, prices from one round to the next until supply reaches the amount needed.

The focus of this paper is much less on the choice of auction design, that is, decisions about what type of auction, English, Dutch, Yankee, Japanese[6], clock, package bidding, etc. Rather, once having settled on an overall auction approach, such as a clock auction or a simultaneous multiple round, many other decisions that can greatly affect the outcome: how what needs to be auctioned is divided into lots and into different auctions, as well as the provisions in the auction for bidders to arbitrage price differences so as to ensure price differentials for lots are comparable to value or cost differentials. This is mostly NOT the case, even in simultaneous auctions in the energy sector.

This paper first reviews recent experience in electricity auctions. The paper discusses the status of default service procurement in some detail. Only New Jersey has adopted an open, multi-round auction approach for such procurements. Transmission rights auctions and energy capacity entitlement and asset auctions are also briefly discussed.

Section 3 provides a brief theoretical analysis of how the division of what is to be auctioned can affect the outcome. This analysis shows that prices can vary enormously, even for virtually identical objects. However, the fact

---

[6] A Japanese auction is one in which all bidders must indicate whether they are still in or not as the auctioneer raises price. The auction ends when there is only one bidder remaining in.

that prices vary so much need not have significant adverse effects on the economic efficiency of the outcome.

Section 4 provides some discussion of experience in some previous auctions. The discussion is intended to be suggestive rather than comprehensive of how the details of auction design can affect the outcome.

2. ELECTRICITY AUCTIONS

2.1 Electricity auction experience

This section is intended to recap experience using various forms of auctions in the electricity sector.

Restructuring over the past several years had led to introduction of markets, including auctions, in lieu of regulatory processes for dispatching and allocating energy resources. While there is significant trading in various forms of exchanges, my focus is on auctions and not other types of markets. The key distinction between auctions and other types of markets is that auctions are one (or a few) to many transactions. Exchanges and other types of competitive markets are many-to-many – there are many buyers and many sellers.

Experience in electricity auctions is still somewhat limited. Many different auction types have been used and are in use for electricity. A large fraction of the auctions and tenders used in the sector have consisted of one-shot, sealed-bids, usually requests for proposals (RFPs). As noted above, the RFPs are most often multi-attribute, although at times price is the sole determinate of the winners. Among the many examples of the sealed-bid auctions are the PJM, ISO-NE and NYISO transmission rights auctions. The NYISO capacity rights auctions are also all sealed-bid auctions. Maryland is currently conducting a default service procurement which is a sealed-bid auction (see http://rfp.bge.com/, http://www.pepcoholdings.com/rfp, and http://www.alleghenypower.com/Marketing/energyprocure.asp).

Some electricity auctions have required bidders to submit supply or demand functions. The California Power Exchange required bids to be piece-wise linear curves. When bids are required be supply functions, it is also often the case that the additional limitation to step functions is imposed. This is the true, for example, in the CAISO day-ahead market and in the NYISO capacity auctions.

Other than the NJ BGS auction (see www.bgs-auction.com) and the ill-fated California Power Exchange, all default service bidding processes have relied on different forms of sealed bid bidding or requests for proposals (RFPs). Utilities procuring electricity have, most commonly, sought multi-

attribute bids, with the result that the selection of winners is based on some sort of subjective evaluation. Often the set of bids that a utility deems to comprise the low cost set of offers is based the utility's forecasts of loads, which it then may use to help evaluate competing offers. This is necessarily a subjective process.

In many instances, the SMR auction and clock auction formats are well suited to *solve* the multi-lot procurement problem that is more commonly managed through sealed-bid RFPs. The SMR and the simpler clock auctions were developed specifically for multi-lot auctions of substitutes.[7] Milgrom (2004) has shown that with straightforward bidding, SMR auctions will achieve efficient allocations. Of course, bidders may face incentives to bid strategically and not straightforwardly. However, in many cases, the incentives to deviate from straightforward bidding can be limited (see Loxley and Salant (2004) and (McAdams (2001)) SMR auctions and its variants have had not been extensively used in the sector. Some versions of SMR auctions have been used for transmission rights in California (CAISO) and the Northeast (PJM, NYISO and ISO-NE, which also conduct capacity auctions), for energy contracts and entitlements (Texas, Alberta Power Purchase Arrangements, French Virtual Power Plants), and for default service procurement (New Jersey BGS).

2.2 Default service procurement

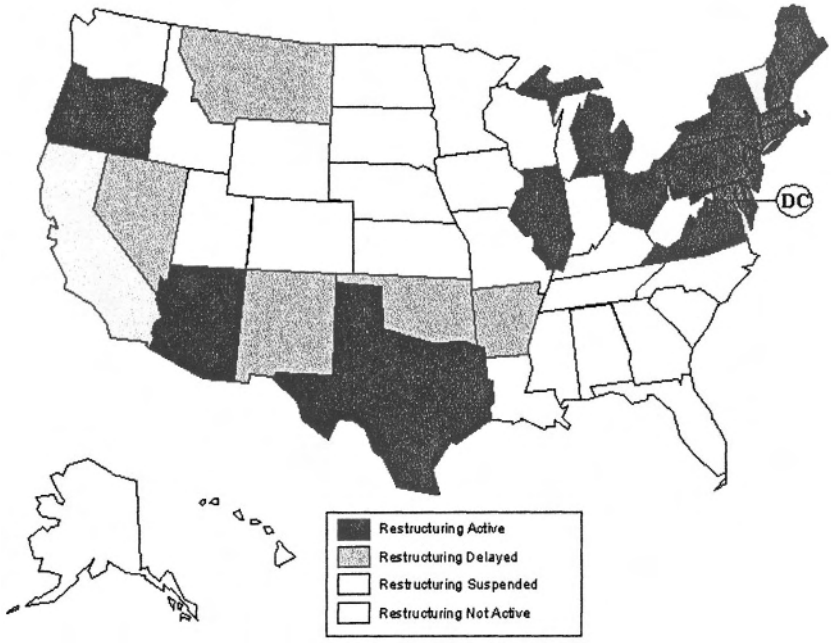
Over the past few years customer choice has been introduced in an increasing number of the U.S. states and is now in effect in many, but not all, states. The following map shows the status of retail competition in each of the 50 U.S. states.[8]

The figure details the status of each state's restructuring process in the United States. The States shaded black is going through the process or have gone through the process of permitting retail competition. The dark gray states are those in which there is a significant delay that occurred in the restructuring process. California is the only light gray state because it is the only state where the restructuring process is suspended. Finally the white states are those that are not currently pursuing restructuring.

---

[7] See Cramton (1997) for a description of SMR auction rules, and Loxley and Salant (2003) of the Simultaneous Descending Clock Auction. In a multi-round, multi-lot auction, bidding occurs in rounds. After each round, bidders are informed about the standing best (high bids in a forward auction, low bids in a reverse, procurement, auction) and are provided an opportunity to improve their offers. Activity rules require each bidder's offer in one round to, in some critical way, improve on its previous round offer. The auction ends when no bidder is willing to improve its offer.

[8] From http://www.eia.doe.gov/cneaf/electricity/page/fact_sheets/facts.html.



Restructuring in virtually every state in which is occurring leaves consumers with the option, but not usually the requirement, to choose a non-utility supplier for electricity. There are several models of retail choice. Under none of those models has the migration from the regulated default service to competitive offerings been rapid and complete. The load requirements of the pool of default service customers remain very uncertain.

Table 1 shows migration in the U.S. states that have introduced retail choice. It shows significant variation in customer switching rates, from a low of less than 1/10 of 1% in New Jersey to an excess of 15% in Ohio as measured by the percentage of customers who have switched. In some other jurisdictions migration from default service has been less than 1% whereas in others it has exceeded 30%. The migration of larger customers has tended to exceed that of smaller ones, so the percentage of load migrating will be higher than the percentage of customers. However, in many jurisdictions even the migration of large customers remains quite low. While the fraction of consumers migrating from the regulated default service can vary, it has remained well below 50% in virtually every jurisdiction and tends to be no more than 10%.

Table 1: 2003 Customer Participation in Retail Choice -2003

State	Population	Customers of Competitive Suppliers	Total Customers Eligible for Retail Choice	Migration Rate	Competitive Residential Load	Competitive Commercial Load	Both C&I Load	Competitive Industrial Load	Total Competitive Load
Arizona	5,456,453	N/a	n/a	n/a	0.00%	0		0.00%	0.00%
California	35,116,033	72,422	10,580,906	0.68%	0.80%	1.4*/13.9#		35.70%	13.30%
Connecticut	3,460,503	N/a	n/a	n/a	1.30%		0.20%		1.20%
D.C.	570,898	25,115	197,359	12.73%	11.10%		59.60%		49.40%
Delaware	807,385	N/a	n/a	n/a	n/a		n/a		n/a
Illinois	12,600,620	27,896	687,980	4.06%	0.00%	26.90%		50.10%	25.80%
Maine	1,294,464	8,713			1.60%	30.40%		72.10%	34.80%
Maryland	5,458,137	74,870	2,074,243	3.61%	3.90%			29.10%	17.20%
Massachusetts	6,427,801	84,532	2,544,495	3.32%	2.20%	11.4*/17.4#		43.50%	21.60%
Michigan	10,050,446	N/a	n/a	n/a	n/a	n/a		n/a	7.30%
New Jersey	8,590,300	2,573	3,651,148	0.07%	n/a		n/a		1.80%
New York	19,157,532	388,308	7,279,618	5.33%	5.50%		26.20%		18.90%
Ohio	11,421,267	747,951	4,681,053	15.98%	13.90%	15.20%		11.70%	12.90%
Oregon	3,521,515	36,503	1,213,858	3.01%	n/a		0.00%		0.00%
Pennsylvania	12,335,091	278,429	n/a	n/a	5.60%	10.70%		11.00%	8.70%
Rhode Island	1,069,725	2,132	468,015	0.46%	n/a	n/a		n/a	12.90%
Texas	21,779,893	469,106	n/a	n/a	4.80%	27.20%		81.10%	38.20%
Virginia	7,293,542	2,584	1,300,763	0.20%	n/a	n/a	n/a	n/a	n/a

#Large Commercial; *Small Commercial; n/a = Not Available; n/o = Not Open

The exact dates at which retail choice started differ across states. In most of the states listed above, retail choice has been offered for several years. In some states, migration has increased over time, but in others, retail choice has never attracted many consumers. In California, retail access was introduced in 1998 and suspended in 2001.

Differences in default service rates have been responsible, in part, for the differences in customer migration. The models differ in other ways between the states. In many states, such as California, default service rates were regulated under legislation that settled some other matters related to competitive restructuring. In a number of states, the default service rates have been so low that there has been virtually no switching during the period in which switching was permitted. Ohio has a number of large distribution companies serving the different parts of the state. In one part of Ohio, migration has exceeded 15%, whereas in other parts of the state served by other distribution companies, migration has been less than 1%. These differences can, in part, be explained by a flexible approach to setting default service rates, allowing each company to have somewhat different rates, and by other factors such as differences in what it costs competitive suppliers to serve the different regions. Many other details of competitive restructuring differ from state to state.

What Table 1 shows is that despite restructuring, utilities still retain a very large share of the load, and an even larger share of the customers. In all these states, the distribution companies have remained in the retailing business. However, in virtually every state, restructuring has resulting in separation of generation and distribution, at least at an operational and managerial level, if not total divestiture. This has left the utilities with the obligation to procure electricity for the very large fraction of customers, and through market-based mechanisms, such as auctions.

Moreover, the situations differ across states, and even across regions within a state. One of the key differences is the extent to which the distribution companies have divested or otherwise separated their generating assets. Some states, such as California, have required electric distribution companies to divest most or all of their generating facilities. In other places the utilities have been required to divest some control, in the form of contracts or entitlements (for example, Texas, and the Canadian Province of Alberta). In yet other jurisdictions the utilities had options, but had to structurally separate generation from distribution (e.g. New Jersey). In New Jersey, one of the three main utilities, PSE&G, retained ownership of its distribution and the other two divested theirs.

2.3 Default Service Procurement Procedures

As was mentioned above, two states, New Jersey and California have introduced open bidding processes for the utilities to purchase electricity on behalf of their default service customers.[9] The process used in New Jersey has been more successful than that used in California. This subsection describes some of the differences, without trying to quantitatively attribute differences in outcomes to specific differentiating factors. The process in California was largely abandoned as a result of persistent price spikes and recurring shortages requiring emergency measures, such as rolling blackouts. This paper is not intended to review the causes of the California energy crisis (see Borenstein, Bushnell and Wolak (2002), Joskow and Kahn (2002), and Wolak (2003)).

Among the differences between New Jersey and California was that fact that in California utilities, for a variety of reasons, relied largely on a day-ahead market, the CALPX, for electricity procurement. In the New Jersey Basic Generation Services (BGS) auctions, default service procurement costs are determined in auctions for longer term, 10 month to 34 month, contracts. This did not mean that there is no day-ahead market in New Jersey nor is concentration of ownership of physical resources lower.

Besides the fact that the utilities made most of their purchases in longer-term markets in New Jersey and in shorter-term markets in California, there were other significant differences in the procurement processes. Most significantly, the auctions in the two states were conducted using different bidding rules. The CALPX auction required bids to be piecewise linear supply schedules, and was a one-shot, sealed-bid process. The rationale for use of supply schedules was to allow bidders, in a single round bidding process, to state different prices depending on the amount supplied. This cannot be accomplished in a one-shot process in which bidders name only one price and/or one quantity. In contrast, the New Jersey BGS auction is a clock auction. Bidders make initial quantity offers at starting prices specified by the auction manager. As the auction manager lowers prices bidders can maintain or reduce their supplies. Prices stop ticking down when supply offered for each type of tranche falls to amount required.[10]

---

[9] Maryland recently introduced a sealed-bid, multi-stage, auction for default service procurement.

[10] This is a simplified description of the BGS auction rules. As there were different types of tranches, bidders could switch across types after each price tick, but no bidder could increase the total number of tranches offered. Loxley and Salant (2004) contain a more complete description.

As described in Loxley and Salant (2004), physical concentration was, if anything, higher in New Jersey than in California.[11] The integration of New Jersey, Pennsylvania, Delaware and Maryland energy markets, through the PJM operation of the transmission grid, though, resulted in an effectively less concentrated market in New Jersey than in California. Although PJM operates a day-ahead market, the combination of the PJM trading rules, the BGS auction rules, and supply/demand balances have apparently kept the New Jersey utilities from having to pay high prices to maintain supplies. Since trading was suspended in the CALPX, the California utilities have been negotiating purchases in more conventional RFP processes. Current plans for subsequent purchases do not include any reintroduction of market or auction processes.

The Maryland utilities are procuring energy requirements through a series of sealed-bid auctions for portions of the load. The load requirements are being divided into categories. In each auction, several tranches from each of the utilities will be put up for bid at a single time. Bidders can make offers for any number of tranches in each category. The Maryland RFP requires bidders to guess as to which types of tranches at which time might prove the best values. The tranches are for a variety of terms, of up to three years.[12] Section 3 reviews the design of this RFP in more detail.

Utilities in most other states have relied on processes that more closely resemble a negotiation rather than open bidding. One question is as to why there is continued reliance on negotiation rather than auctions or other market mechanisms. Auctions are a form of market process in which competition between bidders determines prices and allocations. The winning bidders are those who are willing to offer the best (lowest) prices. This will tend to achieve the efficient outcomes which are usually one of the main goals of such procurement processes.

Despite the apparent success of the BGS auctions in New Jersey (see Loxley and Salant (2004)) and the potential benefits of competitive bidding, this model has not been copied elsewhere. In part, this may be due to the failure of the CALPX which suggest that auctions pose significant risks. These risks are one reason for reluctance to introduce auctions, despite potential advantages. Because many of the fine details can matter, it can be difficult for policy-makers to gain confidence from previous experience when the situation they face differs in some, possibly very important, details.

---

[11] See Loxley and Salant (2004) and McAdams (2001) for a discussion of mechanisms to limit the impact of concentration.

[12] See http://www.pepcoholdings.com/rfp/announce_102903.html or http://rfp.bge.com/. The schedule and the bid structure from that RFP are described in an appendix.

Regulatory restructuring in any one state will differ from the restructuring in other states in some details. The following describes some ways, beyond those already mentioned, in which restructuring varies across states. One obvious difference between states is the concentration of energy resource holdings. More concentrated are these holdings the more difficult it can be to introduce competition for the provision of default service electricity requirements. A supplier with a large share of the capacity being offered in an auction will have a larger incentive to withhold supply than one having a small share of the available capacity. This follows because the reduction of supply needed to increase price by a given magnitude will be a larger fraction of the small firm's supply. Therefore such a firm would recoup less from the withheld supply, than would a large firm.

The ability of any procurement auction to achieve competitive outcomes will be affected by concentration among suppliers. The potential for different firms to compete to supply the default service needs of a utility's customers will depend crucially on transmission markets. An owner of generation capacity can compete to serve load effectively only if it can cost effectively acquire transmission rights to serve that region. The more transparent and efficient is the organization and operation of transmission markets into a region, the more resources that will be able to compete to serve that region. Even within a state, it may be difficult for generators in one part of the state to compete for load in other parts of the state. This appears to be partly the case in Ohio where there are two separate regional transmission organizations (RTOs), PJM West and the Midwest ISO. Seams between RTOs, and even within RTOs can segment the market creating pockets of market power and limiting the resources that can provide competition for serving the default service load in any one area.

Differences in transmission rights allocation procedures and markets enhance or reduce the competition in an auction. The following example, derived from the New Jersey experience, is intended to illustrate how transmission rights allocations which integrate two or more markets can affect competition in an auction. More specifically, the allocation of transmission rights in New Jersey was coordinated to increase integration between Pennsylvania and New Jersey markets (see Loxley and Salant (2004)).

Transmission market organization can affect concentration of resource holdings that can effectively compete to serve a region. If, for example, there are two similar regions each having loads in similar proportions to native generation, and there is no transmission capacity connecting the two regions, or transmission tariffs are high, then the generators in each region will not be able to compete effectively for load in the other. A transmission line opening up to connect the two regions can effectively integrate the two

markets and increase the number of firms that can compete load in the each region. Each purchaser can then choose from twice as many suppliers. If more regions were to be integrated in such a fashion, the competition in each region would increase further.

The impact of transmission allocation procedures on concentration and competition to supply energy services in any one region will depend on all the market rules. Whether and what type of link there is between concentration, and auction and market prices would be necessarily dependent on the specific auction rules. For most auction formats, there is no formulaic link between any measure of concentration, such the Hirschman Herfindahl Index, and the auction outcome, although, as was noted above, suppliers with larger market shares can have greater incentives to withhold supply in a procurement auction.

Some auction designs can enhance the impact of concentration and concentration can directly affect price. The optimal auction design can be affected by concentration. When there are many, many small suppliers, the ability of one supplier to affect the auction outcome may be small independent of the auction format. In contrast, a highly concentrated market can require measures to mitigate the impact on bidder withholding.[13]

Integration of the two or more regional markets does *not* require transmission facilities with capacity equal to the load in each market. A system operator can arrange dispatch independent of location, and levy charges for congestion only when load imbalances require it. A part of the auction design process is to account for differences in concentration and to introduce measures to ensure appropriate allocations even where competition is very limited. The allocation of transmission rights can facilitate or deter competition between generators across regions. The larger the market, the less effect any one party is likely to have on price. Decisions about complementary transmission rights can be a primary determinant of how many independent entities can bid, and therefore can also be a determinant about what type of auction process might achieve the best outcome.

Another factor that will affect the efficacy of an auction process is the supply and demand balance in a region. A market in which concentration is minimal, but in which there is little excess supply might result in higher prices than a market in which concentration is greater but there is also

---

[13] An example is the Cournot oligopoly model, which can be cast as an auction in which firms bid quantities. In that type of auction, price is directly related to concentration. In other auctions, concentration may not affect price. An example is Bertrand oligopoly, which can be cast as an auction in which firms bid prices. See McAdams (2001) for a discussion

greater excess supply. Provisions taken to ensure resource adequacy vary significantly across regions, and so does the supply-demand balance.

3. IMPACT OF LOT SIZES, SWITCHING RULES AND SEQUENCING

Electricity procurement auctions are often for multiple uniforms shares of one or a few types of products needed to meet the load requirements. Among the more important details of the design of any such auction is the division of what is to be auctioned into lots and across auctions. And within a single auction event, the provisions for bidders to switch across products to arbitrage price differentials can also affect the outcome. This section explains *some* of the theory as to how these decisions can affect the outcome.

3.1 Sequential auctions

This subsection first analyzes a simple sequential auction. This simple situation is intended to illustrate how the decision to divide what is to be auctioned across lots across time. As is explained in the next section, it turns out that the experience in both the European and US spectrum auctions suggest that the theoretical concerns illustrated here may be of practical relevance.

The example supposes there are two firms, and two lots. Each firm has a cost of C to serve one lot. If the firm serves two lots the cost is D, where $C < D < 2C$. This means that there are economies of scale. Let $\Delta = D - C$. Now suppose that each lot is purchased one at a time in a standard (English) descending price auction. In particular, suppose price starts high, the auction manager gradually decreases it until no bidder is willing to go any lower. At that point the auction stops and the lot is awarded to the firm to accept a price announced by the auction manager.

Note, this type of descending price *procurement or reverse* auction is not the same as a descending price forward auction for selling one object. The optimal bidding strategy in a single lot descending price reverse auction is to stop bidding when the price reaches the bidder's costs. In an ascending price forward auction, the strategy is essentially equivalent, in that the optimal bidding strategy is to stop bidding when price reaches value.

The equilibrium of the sequence of (English) descending price procurement auctions is easily derived. One bidder will win the first auction for a price of Δ, and the *same* bidder will win the second for a price of C. If D is only a little larger than C, then Δ can be very low, i.e. very significant

scale economies, and the winning price of the second auction much larger than the first. For instance, if $C = 100$, $D = 110$, the price in the first auction will be 10 or a slight bit more. The first auction winner will incur a loss of 90 if that party were to only win that auction. However, the first auction winner can now afford to bid (a shade less than) 100 for the second lot. A first auction loser could not afford to bid less than 100, as that is that the cost of serving one lot. Therefore, the second auction price will be 100, and the first auction winner will also win the second auction. Summarizing Example 1, the first auction price will be 10, the second auction price will be 100, and one firm will win both auctions, collect revenues of 110 and incur costs of 110.

This example can be extended to longer sequences. Suppose, now that there is a fixed cost F, and a cost per lot of C. If there are N auctions, then in equilibrium the winning price in the first auction will be $C - (N\text{-}2)F$, and in subsequent auctions, the price will be $C + F$. In equilibrium the auction winner in the first auction must win all subsequent auctions. The losses from the first will effectively be offset by the profits of the subsequent auctions. This intuition extends to any finite sequence of auctions, or any sequence of auctions in which there will be a last auction with probability one (see Vickers (1986) and Riordan and Salant (1994)).

The fact that the prices are not uniform does not mean that the outcome is inefficient. However, it does mean that the outcome of any one auction has very little correlation with the marginal costs of the firm winning that auction. The next section provides some data from previous auctions that illustrate that sequencing of auctions may have affected prices in the U.S. PCS auctions.

3.2 Switching rules

Many electricity auctions are both multi-product and multi-lot. For example, the California auctions for financial transmission rights (FTRs) contain multiple numbers of lots (MW) on a variety of paths. The NJ BGS auctions include multiple tranches for each Electric Distribution Company, and the second BGS auction included contracts of 10 month and 34 month durations, as well as fixed price (FP) and hourly electric price (HEP) contracts. The Texas capacity entitlements auctions included entitlements for four types of products, in three, and later four, different zones and from three Power Generation Companies. Most SMR and clock auctions allow switching. However, some auctions, such as the Texas entitlements auction in the first year, and the California FTR auctions and the more recent NJ BGS auction do NOT allow switching.

In a multi-round auction, a switching rule allows bidders to switch from one substitute to another across rounds in the auction. The first Texas capacity entitlements auctions had no switching rules. Near the end of that auction, the price of the TXU baseload entitlements was higher than the price of the Reliant baseload entitlements in the same zone, and there was excess demand for the TXU entitlements, but not for the Reliant ones. However, bidders on the TXU entitlements could not switch to the Reliant entitlements. What this meant was that the price of the TXU entitlements kept increasing, while the Reliant price did not, widening the gap. If a low value bidder was fortunate to have bid for Reliant entitlements early it may have won despite the fact that others bidder for TXU had higher valuations but were stuck in the TXU auction. Switching rules allow bidders to switch across markets during the auction and allow bidding to arbitrage price differentials.

The following simple example illustrates the impact of switching in multi-round, multi-product auctions. There are two sealed-bid auctions, both being conducted at the same time. In this example, the two lowest cost firms have costs for supplying lot j of $c_1(j)$ and $c_2(j)$. The costs are those for winning a single lot. Each firm's cost for serving one lot is unaffected by whether it serves the other.

Other firms may bid, but have higher costs, say $d > \max\{c_1(j), c_2(j)\}$. In this example, bids are restricted to the range $[a,b]$ and $a > \max\{c_1(j), c_2(j)\}$. The winning bid receives the second lowest price. Ties are broken randomly. Each firm is assumed not to know the other's costs, and that the low cost firm for one lot can be the higher cost firm on the other lot. Specifically, $c_i(j) = c + \varepsilon_{ij}$. Assume that ε_{ij} is uniform on some interval $[-e,+e]$ for 'e' small.

This auction has no pure strategy equilibrium. To see this suppose, firm i bid $p_i(j)$ for lot j with probability 1. First, unless $e = 0$, or $c_i(j) = c + e$, each firm will want to offer a price $p_i(j) > c_i(j)$, but, then if the other firm knew, for certain, that its rival set prices at $[p_i(1), p_i(2)]$ it would want to undercut the price offered for each lot whenever its costs were lower than the offered price. However, this auction will have mixed strategy equilibrium.[14] The equilibrium will tend to favor the lower cost firm, but won't result in the lower cost firm always winning. Therefore, the outcome will not always be efficient. Part of the reason for this inefficiency is a coordination problem. The low cost firm for each lot will not know, in advance of bidding, that it is the low cost firm. In contrast, a descending price clock auction with

[14] Equilibrium can be computed by noting that $F_j(p)$, the distribution chosen by each firm must be such that $F_j(p)$, the distribution firm j chooses must make firm k indifferent as to its price.

switching will always result in the lower cost firm winning, but may not achieve lower costs. A clock auction solves the coordination problem that arises in sealed-bid, multi-product auctions by allowing bidders to revise bids and switch across lots between rounds.

The situation is more complicated when the firms have capacity constraints. In this case there are coordinated equilibria, i.e., equilibria where each firm bids for one lot but not the other, and they essentially split the market. This example also has multiple mixed strategy (random) equilibria. One possibility, which can have a 25% probability of occurring, is that one of the two lots will not draw any aggressive bids, i.e., offers less than d. The structure of the above example is essentially the same as that of the Maryland RFP process.

On the other hand, a clock auction with a switching rule would result in efficient allocations with near certainty. If there is a third firm whose costs are near c, then prices in a clock auction with a switching rule would almost surely be near costs. This example appears to have direct relevance to the initial Texas Capacity Entitlements auctions, the most recent New Jersey BGS auction as well as a number of spectrum auctions. None of these auctions allowed switching across all the available lots. The next Section presents some data from the Texas capacity auctions and the U.S. PCS auctions.

4. IMPACT OF SWITCHING RESTRICTIONS AND SEQUENCING DECISIONS

This section reviews experience in two sets of auctions, the Texas Power Generation Company (PGC) capacity auctions and the U.S. FCC PCS spectrum auctions. Both involved multiple auction events for essentially identical lots.

4.1 The Texas PGC Capacity Auctions

For the past two years, the Texas PGCs, AEP, Reliant and TXU, have been selling entitlements to 15% of their generating capacity in a series of auctions. The auctions are for 25 MW entitlements of various types of resources: peaking, gas cyclic, gas intermediate and base load. Purchase contracts, subject to review of the Texas Public Utility Commission, PUCT, define the rights of the owners, which include some rights and restrictions on dispatch and responsibility to cover fuel costs. Each PGC had separate entitlements, and there were different entitlements for each of the four zones,

North, South, West and Houston. Entitlements varied in duration, some monthly, some annual and some for two years.

All the contracts of a given duration were sold, initially, in separate parallel auctions. In a few cases, identical entitlements, offered by different PGCs were available at the same time. Bidders could bid on one or both, but during the course of the auction, they could not switch. Consider, for example, a bidder wanting to purchase up to fixed number of base load entitlements for the South zone. Reliant and TXU may both offer such entitlements. If at the start of the auction, the bidder chose to bid only on one, e.g., Reliant, it could never switch to TXU, or conversely.

To avoid getting stuck paying more for one PGC's entitlement when the other PGC's entitlement of the same type sells for less, a bidder might choose to bid for more than the desired number of entitlements at the start of the auction. For instance, the bidder may start by bidding for the total desired number of entitlements of each PGC (and therefore twice the desired demand in aggregate), and only dropping one as the prices rise. In the actual auctions, the Reliant prices went up faster, due to differences in activity and the rules for adjusting bid increments. Therefore, bidders dropped out of the Reliant auctions earlier. Activity lasted longer on the TXU entitlements. In the end, prices of the TXU entitlements ended up being significantly higher, in some cases over 50% higher, than the Reliant entitlements. This is illustrated by the results of the initial Texas Capacity auctions for one and two year entitlements, as shown in Table 2.

Table 2: Results for Comparable Entitlements in Fall 2001 Texas Capacity Auction

	Reliant South	TXU South	Δ	%Δ
1 year base load	$7.32	$10.59	$3.27	45%
2 year base load	$6.59	$10.33	$3.74	57%
1 year cyclic	$0.75	$1.70	$0.95	127%

Table 3: Results for Comparable Entitlements in Spring 2002 Texas Capacity Auction

	Reliant South	CPL South	Δ	%Δ
June base load - capacity only	$13.13	$6.85		
June base load - capacity + fuel	$22.57	$18.76	$3.81	41.3%
July base load – capacity only	$6.76	$10.60		
July base load – capacity + fuel	$26.86	$23.80	$3.06	45.3%
August base load – capacity only	$16.76	$9.85		
August base load – capacity + fuel	$26.86	$22.79	$4.07	55.6%

The other capacity auctions prior to the introduction of switching rules produced similar results.[15]

In most SMR and clock auctions, switching rules would allow bidders to switch eligibility across entitlements. Bidders could bid for their desired number of base load entitlements and would not have to choose, early on, which to bid for. These switching rules, which were subsequently adopted in Texas, would allow competition to arbitrage price differentials and would cause them to vanish.

The New Jersey BGS auctions did have switching provisions for the first auction. In that auction, there were four products, one for each of the utilities or Electric Distribution companies (EDCs). The second year auctions included twelve products, a 10 month and a 34 month FP product for each EDC (essentially for *fixed price* service load) and one HEP product for each EDC (essentially for *hourly electric prices* for larger commercial customers). There were significant differences between the requirements for FP and HEP contracts. The value of switching provisions may be limited. However, the two sets of auctions were conducted concurrently with largely the same bidders, and the combining all products in one auction should improve efficiency of the outcome, and also reduce costs for both administration and participation.

4.2 Other Sequential Multi-Lot Auctions

4.2.1 Spectrum Auctions

Spectrum auction experience provides an illustration how the sequencing of lots can affect prices. In 1994, the FCC began selling 120 MHz of spectrum in a series of auctions for personal communications services (PCS) licenses. For a variety of reasons, the FCC divided the 120 MHz into six bands, three of 30 MHz each and three of 10 MHz each. In addition, the FCC issued licenses for two of the 30 MHz bands divided into 51 geographic areas called Major Trading Areas, or MTAs, and the other bands for 493 geographic areas, call Basic Trading Areas or BTAs. The MTAs form a

---

[15] From December 19, 2001 Memo from Commission Brett Perlman, re. "Rulemaking Proceeding to Revise Substantive Rule §25.381 Capacity Auctions to Chairman Max Yzagguirre and Commissioner Becky Klein. The Texas utilities adopted switching rules late in the second half of 2002. The Spring 2002 auction did not have switching rules and exhibited similar price differentials. Bids were for the capacity charge component of the energy costs. Fuel costs were separate and based on actual plant dispatch. As there were fuel price differentials, the totals needed to account for those differentials.

partition of the U.S. and its territories, and the BTAs are subsets of the MTAs.

The first auction of the 30 MHz A and B block MTA licenses began in December 1994. The final scheduled auction of the 10 MHz D/E/F block licenses ended in January 1997. In addition, the FCC has conducted several re-auctions, in part due to bankruptcies. Table 4 shows the results of these six auctions.

Table 4: Bidding Competition in the PCS Auctions

Auction # and name	Eligibility ratio	Average price per MHz POP
4 (A block)	1.92	$0.52
5 (C block)	6.73	$1.33
10 (C re-auction)	7.11	$1.94
11 (D/E/F bloc)	1.68	$0.33
22 (C-F re-auction)	1.93	$0.16
35 (3rd re-auction)	7.71	$4.37

Before each auction, bidders were required to submit deposits in proportion to the *maximum* number of licenses, as measured by population coverage times the spectrum bandwidth, they might want. The eligibility ratio is the ratio of demand, as measured by aggregate deposits prior to the auction, to the amount of spectrum available. For example, a ratio of two means, at the start of the auction, two bidders were competing for each license.

As can be seen, the demand was not uniform across the auctions, and prices tended to higher when the eligibility ratio was higher. It appears that the division of the lots and the sequencing affected competition in each and auction prices, as the theoretical analysis of the last section suggested could occur (see McAfee and McMillan (1996), Cramton (1997), Ausubel, Cramton, McAfee and McMillan (1997), and Lichtenberg and Salant (1999)).

A similar pattern arose in the European 3G auctions. Six European Union (EU) countries conducted auctions for the same 3G spectrum. Each auction was a multi-unit auction, and the prices of all comparable lots within each individual auction were similar. However, the prices across auctions were far from similar. The first two auctions, in UK and Germany attracted the most interest, the most aggressive bidding and the highest prices. The later auctions in Switzerland, the Netherlands, and Austria were almost

uncontested and attracted much lower prices.[16] This is yet a further illustration of how decisions to conduct separate auctions for multiple objects can result in price differentials which would be difficult, if not impossible to explain by value differentials and are likely to be best explained the differences in the level of competition for each auction induced by the decision to conduct a sequence of auctions rather than a simultaneous auction.

Table 5: European 3G Auction Prices

Country	# of bidders	# of licenses	Auction revenues	Auction revenues / population
UK	13	5	$35B	$588
Germany	7	4 – 6	$46.7B	$569
Italy	6	5	$12.8	$221
Netherlands	6	5	$2.5B	$159
Austria	6	4-6	$596M	$74
Switzerland	4	4	$115M	$16

4.2.2 Other Electricity Auctions

This section very briefly discusses the applicability of some of the auction design considerations that have been described above to a few other ongoing auction processes in the energy sector, and specifically the California FTR auctions and the Maryland default service auctions.[17]

CAISO has been conducting auctions for FTRs for the past several years. An FTR is a right to a pre-defined share of the congestion revenues the CAISO will receive for transmission on a specific path. A company needing transmission rights can hedge these costs with an FTR, as the congestion payments made to the CAISO will be refunded to them as the owner of an FTR.

Each FTR is sold in a separate clock auction. All FTRs for a path sell for the same price. However, two adjacent or nearby paths can be close substitutes and can sell for substantially different prices, as the auctions are separate, although simultaneous. This was the situation in the initial Texas Capacity auctions. To the extent that there are substitutes FTRs that sell for

---

[16] Another explanation of the impact of the sequencing in the European 3G auctions was that the first auction or two were to determine which firms would have European 3G footprints and the subsequent auctions were to fill in those footprints. This explanation could explain the declining prices across the auctions.

[17] Information Maryland default service procurement is at http://www.psc.state.md.us/psc/ and information about California www.caiso.com.

significantly different prices, the CAISO auction can be improved by introducing switching rules of the type described above.

The Maryland utilities are conducting a sequence of procurements for energy to meet the needs of their default service customers. The default service load requirements are divided into customer types. Each customer segment for each utility is subsequently divided into a number of tranches (each an average of 50MW). These tranches are then being procured in a series of four sealed-bid auctions. This type of design has been analyzed in Milgrom (2004). It is to be expected that identical tranches auctioned at different times will be purchased at different prices. Price reversals, in which a lower cost tranche is purchased for more than a higher cost one, are possible if not likely even between auctions of different customer types or for different utilities being held at the same time.[18]

This section concludes with a discussion of one other commonly used multi-lot auction format in the energy sector – auctions in which bidders submit supply schedules. In supply function auctions, price and quantity are determined where the aggregate supply schedule of the bids submitted intersects the amount required. The most common version of a supply function auction is one in which bids are step functions. The CALPX and CAISO required bids to be supply functions. PJM and NYISO capacity markets also require bids to be step functions. These types of auctions have been extensively analyzed (Klemperer and Meyer (1989), Green and Newbery (1992) and von der Fehr and Harbord (1993)). These auctions need not, and often do not, have pure strategy equilibria, stable equilibria. Moreover, as Green and Newbery remarked, the outcomes can be far from competitive.

5. CONCLUSIONS

Multi-lot auctions are common in regulated industries. Electricity procurement to serve default service load requirements is typically best accomplished by dividing the load into one or a few types of lots. Utilities in most states have followed this practice. This chapter is intended to explain how auction design can affect the outcome in multi-lot auctions. In particular, switching rules are quite important.

There is now a significant theory of multi-unit and multi-lot auctions and significant experience with them. The theory shows how auction prices can depend on the division of lots, the specific rules used for the auctions.

---

[18] See http://www.pepcoholdings.com/rfp/ for a complete description of the auction rules.

Multi-lot procurement auctions which force bidders to make separate decisions on individual lots can result in inefficient mismatching of bidding suppliers with the demand. This mismatching means inefficient supply in that the low cost bidders will not always win.

Within a single multi-round, multi-product and multi-lot auction, the lack of switching rules limits the ability of bidders to arbitrage price differentials. This is one situation in which the lack of a switching rule can adversely affect the outcome. Identical lots need not go for the same price even in the same auction. The low cost supplier may not be chosen for any fraction of the load being procured in the auction. Moreover, as was explained above, the division of the lots to be auctioned into a sequence of auctions is likely to result in a similar mismatch of suppliers with demand and adversely affects the economic efficiency of the resulting allocation.

The paper illustrates the impact of switching rules and sequencing decisions using the experience from two sets of auctions. The Texas capacity auctions show how a design that did not include switching rules resulted in identical entitlements selling for significantly different prices. As most simultaneous auctions in the electricity sector still limit switching, it would appear that there is potential for efficiency gains. The US PCS and European 3G auctions for spectrum licenses show that in sequences of auctions, with very sophisticated and prepared bidders and a great deal at stake, auction price differentials may be totally unrelated to cost and value differentials and the outcomes can be inefficient.

The focus here has been on the potential for inefficient assignments in multi-lot and multi-unit auctions, and simultaneous auction designs with switching rules, such as the SMR and clock auctions, which are designed to limit the risk mismatches. In particular, such auctions are less likely to result in inefficient matching of buyers and sellers than are more conventional sealed-bids or sequential auctions. This is not the only type of inefficiency that the design of an auction can affect. It is however a problem specific to multi-unit and multi-lot auctions.

Auction design should account for other sources of inefficiency. The SMR auction was initially designed to reduce the impact of one, the winner's curse. SMR and clock auctions may be less well suited for addressing other issues, such as the potential for bidder withholding or bidder collusion (see Cramton (2000)) or where there are complementarities across lots (Milgrom (2004)). Package bidding can be more efficient, although potentially very complex, way of providing bidders the opportunity to put all or nothing bids on packages. Other mechanisms, such as the volume adjustments used in the New Jersey BGS auctions, can be applied to limit incentives for bidder withholding and collusion.

The best practical design for a multi-unit or multi-lot auction depends on the specific circumstances, including implementation cost relative to potential efficiency improvements. Sealed-bids are the simplest to implement, clock auctions are less so, and package bidding can be much more complex. An auction that takes several days to complete may work well for periodic auctions, such as procurement of one-year contracts to hedge the costs of default service load. However, it might not work nearly as well in day-ahead markets, in which trading needs to be completed within a few minutes, or for transmission rights, which involve significant complementarities.

REFERENCES

Ashenfelter, O. 1989. "How Auctions work for wine and art." *Journal of Economic Perspectives* 3(3): 23-36.

Ausubel, L., P. Cramton, R.P. McAfee and J. McMillan. 1997. "Synergies in Wireless Telephony: Evidence from the Broadband PCS Auctions." *Journal of Economics & Management Strategy* 6(3): 511-517.

Beggs and K. Graddy. 1997. "Declining Values and the Afternoon Effect: Evidence from Art Auctions." *Rand Journal* 28(3): 544-65 .

Borenstein, S., J.B., Bushnell, and F.A. Wolak. 1992. "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market." *American Economic Review* 92(5):1376-1405.

Cramton, P. 1997. "The FCC Spectrum Auctions: An Early Assessment." *Journal of Economics and Management Strategy* 6(3):431-495.

Cramton, P. and J. Schwartz. 2000. "Collusive Bidding: Lessons from the FCC Spectrum Auctions." *Journal of Regulatory Economics* 17: 229-252.

Easley, R. F. and R. Tenorio. 1999. "Bidding Strategies in Internet Yankee Auctions." In *Proceedings of Computing in Economics and Finance 1999,* Meetings of the Society for Computational Economics, June 24-26.

Fehr, N.-H. von der and D. Harbord, "Spot Market Competition in the UK Electricity Industry." 103: 531-46.

Green, R. and D. Newbery. 1992. "Competition in the British Electricity Spot Market." *Journal of Political Economy* 100: 929-53.

Joskow, P. and E. Kahn. 2002. "A Qualitative Analysis of Pricing Behavior in California's Wholesale Electricity Market During Summer 2000." *The Energy Journal* 23(4): 1-35.

Jurewicz, J. 2002. "California's Electricity Debacle: A Guided Tour." *Electricity Journal* 15(4): 10-29.

Klemperer, P. and M. Meyer. 1989. "Supply Function Equilibria in Oligopoly Under Uncertainty." *Econometrica* 57(6): 1243-72.

Kwerel, E., and G. Rosston. 2000. "An Insiders' View of FCC Spectrum Auctions." *Journal of Regulatory Economics* 17(3): 253-289.

Lichtenberg, F. and D. Salant. 1999. in *Nextwave Personal Communications vs. Federal Communications Commission.*

Loxley, C. and D. Salant. 2004. "Default Service Auctions." *Journal of Regulatory Economics* forthcoming.

McAdams, David. 2001. "Essays in Multi-Unit Bargaining." Ph.D. dissertation, Stanford University.

McAfee, R. Preston and John McMillan. 1996. "Analyzing the Airwaves Auction." *Journal of Economic Perspectives* 10(1): 159-176.

McMillan, John. 1994. "Selling Spectrum Rights." *Journal of Economic Perspectives* 8 (3): 145-162.

Michaels, R.J. and N.T. Nguyen. 2001. "Games or Opportunities: Bidding in the California Markets." *Electricity Journal* 14(1): 11-108.

Milgrom, P. 2004. *Putting Auction Theory to Work,* Cambridge, Cambridge University Press.

Milgrom, P, and R.J. Weber. 1982. "A Theory of Auctions and Competitive Bidding." *Econometrica* 50 (November): 1089-1122.

Riordan, M. and D. Salant, 1984. "Preemptive Adoptions of an Emerging Technology." *Journal of Industrial Economics.*

Vickers, John. 1986. "The Evolution of Market Structure When There is a Sequence of Innovations." *Journal of Industrial Economics* 35(1): 1-12.

Weber, R.J., 1983. "Multiple object auctions" In *Auctions, Bidding and Contracting. Uses and Theory,* edited by R. Engelbrecht-Wiggans, M. Shubik and R.Stark. New York, NY: University Press.

Wolak, F. 2003. "Measuring Unilateral Market Power in Wholesale Electricity Markets: The California Market." *American Economic Review* 93(2): 425-430.

Chapter 4

The Anatomy of Institutional and Organizational Failure*

Karl A. McDermott and Carl R. Peterson
NERA

1. INTRODUCTION

The debate over the appropriate institutional structure to govern public utility transactions has been driven by considerations of failure from the beginning. Whether it was the common sense understanding that a single bridge over a river rather than competition between bridges, made economic sense[1] or the more technical discussion regarding the nature of the conditions that would lead to market failures resulting in natural monopolies,[2] some form of failure has motivated regulation. Likewise, it has been the failure of incentives to control costs or to promote innovation, which has motivated claims of regulatory failure and calls for adopting market institutions.[3] In either case the search for a sustainable equilibrium institutional form to serve

* The authors would like to thank the two discussants, Richard Michelfelder and Richard Clarke, as well as the editors of this volume, for their insightful comments. We, of course, take sole responsibility for the content of this article.

1 Kutler (1971) explores the details of this long and hard fought battle to protect a monopoly franchise granted by the state.

2 Not all discussions of failure rest on such a formal analysis i.e., the violation of the conditions for a competitive market to produce the social welfare maximizing production, allocation and distribution of resources. Bator (1957, 1958) provides a complete description of these technical issues.

3 McDermott and Peterson (2001) provide a summary of the references relating to criticisms of traditional regulation.

as a governance mechanism for these *public utility* industries has lead to a constant stream of proposals and experiments under the rubric of reform. Currently there appears to be a perception that retail market experiments have either not had enough time to develop or have simply failed.

The purpose of this paper is to explore the theoretical and practical aspects of this search for the illusive sustainable equilibrium institutional structure to govern utility transactions. Governments have experimented with virtually every option available and space considerations alone force us to restrict the number of examples addressed below. In order to understand the anatomy of this failed search we will need to examine not only the economics underlying these problems but also the political issues that arise in the publics' perceptions, that so often play a critical role in evaluating an institutions success or failure.

Unfortunately, the fact that public utility services are considered essential raises the political stakes in finding a solution to **any** disequilibrium experienced in the supply of these services. As a result there is a level of impatience, bordering on dysfunctional, that prevents us from giving an institution the chance of addressing the problems either causing or resulting from the disequilibrium. Such is equally true for regulatory or market solutions. The unfortunate political response to a crisis in these essential services has been the search for the *silver bullet* solution. Moreover, by ignoring the complexity associated with the type of network industries that constitute the bulk of our public utilities providing these essential services, the silver bullets have fallen wide of the mark. The combination of politics and the search for a *quick fix* has resulted in a world where everywhere there is a failure lurking and everywhere a solution but nowhere the patience to allow the choices that we have made the opportunity to bear fruit.[4]

---

[4] Consider for a moment all of the policy options considered by regulators since the passage of the US Public Utility Regulatory Policy Act of 1978 (PURPA) alone. This list is not all inclusive but if we consider: inverted block rates, Time of Day rates, real time rates, competitive bidding, all source bidding, decoupling of sales from profits, re-coupling, statistical re-coupling, re-engineering, downsizing, right sizing, green pricing, demand-side management (DSM), conservation, renewable energy, integrated resource planning (IRP), least-cost utility planning (LCUP), performance regulation, yardstick regulation, spot markets, futures markets, fuel cells, merchant generators and other proposals. Ironically, if we could objectively identify all of the proposed solutions offered in the literature since 1978 and divided that number into the total number of months that have elapsed since 1978 our guess is that each policy was given about three months to solve the problems of the industry before we moved to the next set of buzz words. This is indicative of the dysfunctional level of impatience that permeates the policy making process, an artifact we suspect of the essential nature of the commodities and services provided by utilities.

These conditions determining the boundary between regulation and markets continues to be influenced by technology and input market conditions. This implies that our efforts in this paper are directed not at solutions but at developing an awareness of the problems and how regulators need to conceptualize the issues before making policy decisions. In the remaining sections of the paper we place this central policy problem within a context that illuminates the complexities of engaging in reform in reaction to perceived failures of markets or regulation.

The paper proceeds as follows: Section 2 provides a discussion of failure and reform of regulation. Section 3 discusses the context for the choice of institutions. Section 4 introduces the analysis of the comparison of institutional structures followed by Section 5 that provides certain examples. Some concluding remarks are provided in Section 6.

2. FAILURE AND REFORM—AGAIN

The idea of failure and its' corollary, reform, have been a standard part of regulatory literature. The literature is replete with the identification of failure and calls for reform. Gray (1940) prematurely predicted the passing of the public utility concept, which Kahn (1983) revived and applied to a new generation of regulatory problems. Averch and Johnson (1962) formalized the criticisms leveled at the poor incentive for cost-regulated firms, and later calls for reforms such as Posner (1969) and Breyer (1982) helped usher us down the path to our current market experiments. The ideas of failure and reform have been and continue to be regulators' constant companions[5]. Reform of the regulatory institutions traditionally took the form of a *fix* to the perceived problem with regulation or to limit the scope of regulation by shifting some or all functions to markets. Reforms addressing the *salvaging* or repair of regulation have examined points ranging from the use of incentives to improved pricing and planning of operations.[6] Gray (in Sichel (1976, p. 5)) remarked on the effort to salvage regulation noting:

> It (salvaging) implies that regulation is in dire peril, being threatened with destruction by some sinister force; at the same time, it suggests that

---

[5] The number of papers employing the concepts of failure and reform are too numerous to list here, a sample of this literature includes: Breyer (1979), Joskow (1989), MacAvoy (1996), Selwyn (1996), Hogan (2002) and Sidak (2002).

[6] Evetts (1922, p. 134) examines the incentive regulations adopted in England beginning around the 1860s. This research found that 258 of the 520 gas companies in England employed some form of sliding scale regulation for approximately 70 percent of the gas supplied at that time.

this unnamed force is not all-powerful; that men, if clever enough, can avert this disaster. Thus, by implication regulation is "salvable"; most important, it assumes that regulation is worth saving- a dubious assumption in the eyes of many. Lastly, the title suggests that some terrible calamity may befall us if we fail to rescue regulation.

Moreover, in *failing* to rescue regulation we may in fact suffer from the unintended consequences of well-intended actions. The reforms may be as dangerous, if not more so, than the institutions they replaced. This is as true of poorly designed market experiments as it is of poorly designed replacement regulations.[7] The paradox may be even more profound if the *sinister force* destroying regulation is competition itself; and the competition occurring is the result of regulatory pricing that has induced uneconomic entry inducing bypass or cream skimming. The fact that failed reforms can compound the failures that the next generation of reforms face makes for a lasting source of employment for reformers.

This, of course, begs the question of exactly what form of regulation we are saving, or, in the alternative, if it is the concept of regulation in general that is found objectionable. As we will discuss below the context for discussing failure and reform is often conditioned by the existence of our *traditional* rate-base, rate-of-return regulation and the legal and political structure associated with this approach to regulation. Clearly this context can shift as it would after a number of years employing performance regulation or benchmarking approaches. Why some forms of regulation work well in some periods and not others is a central concern of this paper and hopefully of regulators.

At the heart of the discussion of institutional failure is the need to address the economic realities of uncertainty and expectations. Since the rise of neo-classical economics, the assumption of the rational decision-maker has been at odds with the observed data. In recent years the *new institutional economics* has incorporated the concept of the less than perfectly rational decision-maker into the analysis with some surprising results.[8] This decision-maker cannot foresee all possible contingencies in future states of the world and therefore cannot write contracts that are complete. The possibility that contractual relationships (read market transactions) are costly to enforce and negotiate was completely absent from the traditional view of the neoclassical economist. However, the nature and extent of transaction costs has a great

---

[7] For example, in Illinois, early experiments with electric customer choice were not designed to answer the fundamental economic questions, but, rather, were designed to provide evidence of success or failure of retail choice.

[8] This term appears to have been coined by Williamson (1975, p. 1). Furubotn and Richter (2000) provide a comprehensive overview of the new institutional economics literature.

influence on the form and structure of governance relationships. Institutions that arise to reduce transactions costs, whether they are internal organizations or formal regulations, are in some sense at odds with the market process. Therefore, there will always be a tension between non-market institutions and market institutions.

This approach provides us with a conceptual framework for thinking about institutional equilibrium. We define an institution as set of formal or informal rules that provide the guidance for behavior and exchange. The institution also includes the (formal or informal) enforcement mechanisms. Institutional equilibrium exists where the given institution adapts to exogenous changes such that the essence of the institution lives on. Since we cannot foresee future contingencies we must create our institutions to allow for an evolution of structures as conditions warrant. Institutions that can sustain this evolution without fundamental change to the original structure can be said, in some way, to be stable. Furubotn and Richter (2000, p.24) suggest a two-fold test for institutional stability that will be useful for our purposes. An institution can be said to be in equilibrium if (1) new informal rules evolve to reach a stable endpoint without destroying the original formal framework; or (2) after a disturbance of an initial institutional equilibrium a new equilibrium will be reached.

One fundamental question facing the traditional public utility industries is whether or not there are conditions peculiar to these industries that lend themselves to institutional stability or instability. While we can identify periods of relative stability in regulatory institutions, since the inception of modern regulation,[9] there has always been an undercurrent of perpetual transition. This undercurrent has culminated in what Kearney and Merrill (1998) has labeled the *Great Transformation* over the last quarter century to a market oriented approach to controlling regulated industries. However, even the recent move to markets has been faced with perceived failures raising the issue once more of how to reform traditional regulatory methods in order to *stabilize* our infrastructure industries.

Interestingly, this *Great Transformation* reflects a return to some of our earliest debates over our choice of institutional design to govern regulated industries. As Berk (1994) has pointed out, in the early stages of railroad regulation we had a choice between the centralization (or *end-to-end monopoly* form of organization) or *regulated competition* where open access and the use of gateways and cooperation between subsystems could be used to provide public benefits through a combination of regulation and

---

[9] Modern regulation is here defined as the post *Hope* era. *Hope* resolved the great valuation controversy, which plagued the utility industries for the first half of the twentieth century. *See FPC v. Hope Natural Gas Company (Hope),* 320 U.S. 591 (1944).

competition. Our return to regulated-competition in the form of current open access and unbundling policies ironically represents the idea of regulation as a method of enabling competition not as the enemy of competition. The fact that in a well functioning market requires regulations (i.e., institutions as a complement to competitive structures) is a point often lost on those who have turned the battle over institutional choice into an ideological struggle rather than a search for "truth" and the creation of institutional frameworks that solve problems.[10] This problem of ideology cannot be understated. Bell (1962, p. 405) noted "[I]deology makes it unnecessary for people to confront individual issues on their individual merit. One simply turns to the ideological vending machine and out comes a prepared formulae."[11] Many of the mistakes in our most recent experiments with markets can be attributed to placing ideology ahead of rational analysis to solve the problem.[12]

3. THE CONTEXT FOR INSTITUTIONAL CHOICE

The issue of context takes on a number of dimensions. There is of course the special legal context that exists in the United States. In addition there are the dimensions of market structure, and the objectives and expectations of the participants. Institutions interpreted broadly include the norms that arise with the organization of society. In the case of traditionally regulated utility services we have developed certain expectations regarding the reliability and stability of the services. While the fulfillment of these expectations is complicated by the technical engineering and economic parameters governing the choices of the owners of capital serving these markets, the very fact that firms exist, as Coase (1937) showed, is indicative of a market failure of a different kind and the transaction cost issues discussed below directly influence the success or failure of various forms of organization.[13]

---

[10] One need only note that maybe one of the most competitive markets is also one of the most highly regulated—the US Stock market. The purpose of that regulation is to maintain the market's competitiveness.

[11] Alfred Kahn (1998, p.xxxvii.) stated the issue most clearly "... the optimal mix of institutional arrangements for any one of them (the utility industries) cannot be decided on the basis of ideology alone."

[12] McDermott and Peterson (2002a) provide a description and criticism of the political and ideological issues that have influenced recent market reforms in electricity.

[13] Williamson and Winter (1993) have edited a volume that provides an explanation of the importance of the transaction cost approach to understanding organizational choices.

3.1 Goals and Expectations

Success or failure is relative to the goals and objectives that an institution was designed to achieve. In many cases institutions evolved in the face of changing goals and objectives. In the case of public utilities, it has been the case that the institution had to balance a number of goals and objectives that were incommensurate (e.g., efficiency and equity). For example, Table 1 provides brief summary of the legislative goals for three different restructuring laws. The inherent conflicts in the enabling law make implementation and evaluation of regulation difficult at best. In light of these conflicts, the process of regulation has often been viewed, as Owen and Braeutigam (1979) argued, as playing the role of a stabilizing force. By adjudicating the disputes and reallocating benefits and harms regulation served as a buffer, dampening the effects of change, good or bad. This is something we expect from our institutions, especially institutions that govern the allocation of essential services. The objectives in the minds of society may not have the same weights as economists would give them, thus reliability and continuity may play a greater role than efficiency in the public's evaluation of an institutions performance. It should not come as a surprise that customers tend to reject markets when they use price to allocate resources, when traditionally quantities were used to adjust to price. Conditioned by years of quantity adjustment to meet their needs, some customers may bristle at the need for them to make the adjustment in response to prices. As we shall discuss in greater detail below success or failure often hangs on how an institution handles adjustments to disequilibrium conditions.[14]

[14] The issue and importance of adjustment and the role it plays in evaluating markets and other processes is illustrated by Hahn (1970).

Table 1: Sample of Legislative Goals

California-Electricity (1)	Illinois-Electricity (2)	Federal-Telecom (3)
...the creation of a proposed new market structure featuring two state chartered, nonprofit market institutions: a Power Exchange charged with providing an efficient, competitive auction to meet electricity loads of exchange customers,	A competitive wholesale and retail market must benefit all Illinois citizens.	...promote competition and reduce regulation in order to secure lower prices and higher quality services...
An immediate rate reduction of no less than 10 percent for residential and small commercial ratepayers (no less than 20 percent by April 1, 2002). The financing of the rate reduction through the issuance of "rate reduction bonds" that create no new financial obligations or liabilities for the State of California	Consumer protections must be in place to ensure that all customers continue to receive safe, reliable, affordable, and environmentally safe electric service.	...encourage the rapid deployment of new telecommunications technologies.
Competition in the electric generation market will encourage innovation, efficiency, and better service from all market participants, and will permit the reduction of costly regulatory oversight	All consumers must benefit in an equitable and timely fashion from the lower costs for electricity that result from retail and wholesale competition and receive sufficient information to make informed choices among suppliers and services.	[ILEC's must] offer for resale at wholesale rates any telecommunications service that the carrier provides at retail to subscribers who are not telecommunications carriers
It is the intent of the Legislature that electric industry restructuring should enhance the reliability of the interconnected regional transmission systems, and provide strong coordination and enforceable protocols for all users of the power grid...It is important that sufficient supplies of electric generation will be available to maintain the reliable service to the citizens and businesses of the state.	The citizens and businesses of the State of Illinois have been well-served by a comprehensive electrical utility system which has provided safe, reliable, and affordable service.	An incumbent local exchange carrier shall provide unbundled network elements in a manner that allows requesting carriers to combine such elements in order to provide such telecommunications service.
Accelerated, equitable, nonbypassable recovery of transition costs associated with uneconomic utility investments and contractual obligations.	The State has an interest in providing the existing utilities a reasonable opportunity to obtain a return on certain investments on which they depended in undertaking those commitments in the first instance while, at the same time, not permitting new entrants into the industry to take unreasonable advantage of the investments made by the formerly regulated industry.	[Each telecommunications carriers has the duty] to interconnect directly or indirectly with the facilities and equipment of other telecommunications carriers

(1) AB 1890
(2) Illinois Public Utilities Act, Section 16-101A
(3) Telecommunications Act of 1996

3.2 The Special Nature of Property

Ever since the challenge presented in the Charles River Bridge case decided by the US Supreme Court in February 1837, the question of the proper nature, role and scope for regulation has been bound to the peculiar role that private property has played in the formation of our national philosophy.[15] No discussion of failure or reform and no comparison to other

[15] *See Charles River Bridge v. Warren Bridge,* 36 U.S. 420. Kutler (1971) provides the details of this case. Briefly, the Charles River Bridge Company (CRBC) had been granted a charter in 1785 by the commonwealth of Massachusetts to operate a toll bridge over the Charles River. The bridge was to obtain tolls for a number of years, initially set at 40 years

countries experiences can be conducted without reference to the special role of property and our constitutional law[16]. The property placed in the service of the public acquires a new form. While property is normally associated with the right to exclude users, in the case of public utilities we have created a hybrid form of common property, the common carrier, which has an obligation to extend service to all[17].

The issue of common carriers is made more complex as a result of the *network* nature of the industries in question. Originally these networks developed as systems that integrated various supply functions within a single firm. That is, the public goods nature of the network was addressed through the specific organizational structure of a single firm producing intermediate goods for the provision of a bundled product while supervised by the regulator. This structure had the obvious advantage of internalizing all of the externalities associated with production of the intermediate goods utilizing a public good. These networks were not necessarily designed to support competition within the system. This aspect emerged only as a result of changes in technology that altered the relative costs of various components.[18] The task of determining which elements of the networks can be organized as competitive and which must remain regulated is made even more difficult since the forces that altered the relative costs of one organizational form over another can easily shift in the opposite direction[19].

---

and later extended to 70 years, with ownership and the right to set tolls (i.e., free) subsequently returned to the commonwealth. In 1828 the commonwealth authorized the Warren Bridge Company to operate a bridge approximately 260 feet from the Charles River Bridge in Charlestown. The Warren Bridge was to charge tolls until it was paid for, but no longer than six years. The CRBC filed suit claiming that its charter represented an economic contract and that contract was impaired by the authorization of the second bridge in violation of Article I, Section 10, (Contract Clause) of the US Constitution. The US Supreme Court upheld the commonwealth's authorization of the Warren Bridge, holding that since the original charter did not specifically grant a monopoly, the ambiguity in the contract would operate in favor of the public, thus allowing a competing bridge.

16 All too often the approach to market liberalization taken in countries with formerly nationalized industries is compared inappropriately to the United States. Market restructuring is a vastly different process when a takings clause can be invoked and involves far more negotiation over compensation for takings and the effect of this *stranded cost* recovery on the process of opening markets.

17 Epstein (1998) discusses the importance of the common carrier concept.

18 For the example of electricity, competition in commodity is made possible primarily by modern communication technologies that allow for rapid disclosure of information concerning the operations and coordination of the network and not primarily, as some might argue, as a result of generation technology improvements.

19 The question of which institutional structure is appropriate to adopt as a governance structure has been made more difficult as a result of the economic forces that blurring of

Perhaps the most important effect of the *network commons* is the structural issues that it raises in conjunction with the shifting boundary between firms and markets in the public utility industries. While the commons problem has always existed in these industries, the solution traditionally has been to internalize the externalities through system development. This solution has been challenged as open access and structural separation of the network components has become technically feasible. How to integrate the commons and its use into the regulation of monopolies or the coordination of the interaction of competitive firms has become an important challenge for reform. As we will see in the discussion of the traditional regulatory model significant resistance and inertia is embedded in this legacy system of regulation.

3.3 Characteristic of the Production and Exchange: The Evolution of Governance Structures

The study of public utility regulation changed in the mid 1970s. Williamson (1976) published *Markets and Hierarchies,* Goldberg (1976) produced "Regulation and Administered Contracts," and Trebing (1976) was asking the question whether a greater reliance on market structure could replace traditional regulation. In trying to come to terms with the problems facing network public utilities, a combination of ideas such as networks, commons and common carrier aspects of these industries as well as the nature of the incentives associated with the these organizations and in particular the transaction costs and organizational aspects of transactions were becoming the emphasis in understanding the organization of the industries. The boundary problems, such as those discussed by McKie (1970) and in particular those associated with changing technologies and the need to allocate risks and rewards are shown to be of particular importance in these markets. While a literature has grown up around these issues in the field of regulation, their impact has yet to be fully felt in the design of effective policies.

In many cases it has been our inability to properly interpret the implications of individual and firm choices within the context of the risk and incentive nexus associated with transactions conducted in high fixed cost network industries, that the resulting policy decisions were often counter productive to the public interest. Attempts to push markets or regulation too far and in other cases not far enough have resulted in a catalog of failures.

---

distinctions between where markets, firms or contracts and the effects they have on the selection of the most efficient organization of transactions.

The search for an equilibrium set of regulatory institutions robust enough to adapt its policies to changing conditions without wholesale restructuring is one of the goals of public policy.

What is perhaps the most interesting aspect in the evolution of market failure and economic theory is the implication of the work of Ronald Coase on the theory of the firm and transactions cost. This work has lead to a significant literature that examines the conditions related to the boundaries between firms, contracts and markets as governance mechanisms employed to organize resource allocation under changing industry conditions. Much of this literature rested on the fundamental notion that understanding the organization of, or boundary between, firms and markets is the recognition that contractual relationships between economic players are costly in terms of writing, executing and enforcing.[20] Faced with the realities of complexity and uncertainty, the human mind has comparatively little ability to fully calculate an objectively rational behavioral strategy.[21] Simon (1957, p. 198) suggests that each economic actors' behavior is bounded in its objective rationality. This helps create an environment where the opportunism impacts trading structures and the institutional organization of firms, and in turn, the market. Opportunism is created when there is an incentive for one or both parties to alter their behavior in order to extract additional gains from a contractual relationship.[22] Furthermore, in such an environment, contracting over both the long term and through spot market transactions can be hazardous. These realities of economic life have much to say concerning the ability to enter into contracts of any length.[23] A distinction between where the firm ends and where the market begins must therefore be made. This is particularly important when restructuring markets that have previously been either vertically integrated, fully or partially, or have been heavily regulated such as the natural gas and electric industries. These issues are not simply academic in nature, when industries are characterized by asset specificity,

---

[20] Williamson (1976, especially chapter 2) provides a general discussion of the issue. McDermott and Peterson (2001) provide or a review of contracting issues in the context of regulating natural monopolies.

[21] Here lies one reason why firms arise--firms make *sequential* decisions and operate in a Bayesian fashion.

[22] While opportunism is a subset of the standard assumption of self-interest, its application in the case of contracting has implications for the *organization* of markets and firms. Furthermore, it is generally assumed here that a zero-sum game exits. Opportunism arises when one party to the contract can in some way hold-up the other party thereby extracting (appropriable) quasi-rents from the other party. See note 26 for a definition of quasi-rents.

[23] A firm can be simply defined as a relationship that provides for the allocation of resources by a method other than the price system. Markets are the institutions that determine the prices used to allocate goods and services not allocated directly by the firm. *See* Coase (1937)

whether that is in the form of human assets, physical assets, dedicated facilities or site specificity; there is a fundamental problem with anonymous spot market transactions.[24] The problem becomes one of *ex post* small numbers[25] in which one or both of the parties have some incentive to attempt to extract a portion of the quasi-rents associated with the specific asset.[26] Furthermore, incomplete contracting via long-term arrangements can leave a party open to the risk that as circumstances change the contract will not address all possible future events.

These two problems with contracting in both the short-and long term provide certain restrictions on the ability of regulators and policymakers to force competitive market solutions on these industries. Vertical integration is one solution to the complex problem of contracting under uncertainty.[27] Therefore, the size of the firm, and then by implication the structure of the market, is generally dependant on the degree of these problems and not necessarily the technology that each firm exhibits. It very well may be that the standard technology exhibits traits that would lead one to believe that some form of competition can arise and be effective at regulating the market. However, the degree to which these markets face uncertainty and complexity will have more influence on the final efficient organization than the technologies themselves.[28]

The fact that firms exist at all is a testament to the fact that markets "failed" to satisfy the conditions of being the most efficient form of organization or in the alternative the traditional economics was simply ignoring important costs when labeling a particular organizational scheme "efficient." On a more fundamental level, the boundary issue addresses the age-old question of the appropriate role for government in the economy. While in the US, production of public utility services (e.g., gas distribution, electricity and telecom services) has traditionally been in hands of private companies, the role of government in restraining, via regulation, the property

---

[24] Another aspect of spot markets for commodities is that the market tends to clear at a single price. This price is set by the least-efficient producer that obtains a buyer.

[25] The problem of *small numbers* is so-called in reference to the oft-cited model of perfect competition in which it is generally assumed that *large numbers* of producers and consumers are interacting in the marketplace.

[26] In this context, quasi-rents are the return to temporally specialized productive services. The appropriable or extractable quasi-rents arise as the result of investment with little alternative value outside of the specific transaction. *See e.g.,* Furubotn and Richter (2000, p. 493).

[27] Joskow (1993) provides a good exposition of vertical relationships in markets characterized by asset specificity.

[28] Of course, the technologies can impact the degree to which assets are specific to the industry and even in some cases the degree to which complexity and uncertainty arise.

rights of those companies remains the open question. It is far easier, in principle, to apply the concepts of the boundary of government's role in network industries than it has been in practice." Move the footnote number to the end of the sentence.[29] Moreover, this boundary concept implies that changing conditions in underlying industry parameters can lead to shifts between the organizational forms that are most efficient given the prevailing conditions. This in turn may give rise to natural situations where competition and markets may be superior at any given time to regulation and visa versa. Alternatively, unnatural conditions may arise from the form of regulation itself altering the incentives for inefficient competitive entry or create artificial competition. The possibilities are in some sense endless and this fact has made the search for effective institutional structures a dysfunctional characteristic of recent regulatory history.

4. COMPARATIVE INSTITUTIONAL ANALYSIS

4.1 Institutions and Institutional Equilibrium

Institutions are critical to understanding a market, industry or economies performance. If there is one thing we can take away from the new institutional economics it is that cross comparisons and evaluation of performance cannot ignore the variation in institutional structure. We cannot, for example, simply suggest that because markets work reasonably well in the restructured electricity industry in England then it will work equally well in the United States. The property rights issues alone prevent such sweeping views from being correct. Furthermore, "What is to be Done?" at least in the electric industry, cannot be done in the United States due to our Federalist approach to dividing up the duties of regulation between state and Federal governments.[30] Imitating an institution, outside of the institutional context in which it was developed is a recipe for disaster. We have spoken about institutional equilibrium, Furubotn and Richter (2000) have defined this equilibrium as "an original set of formal rules remains in active use despite the fact that a supplementary set of informal rules and enforcement characteristics has grown up to complete the total structure." Furubotn and

---

[29] The principle traditionally applied to this question is to allow competition for those sectors of the market that no longer exhibit the "market failures" (read natural monopoly) and continue regulation in those sectors that continue to exhibit market failures.

[30] For example, Hunt (2002), astutely observers the necessary conditions for the final push to competition. Unfortunately, it appears that no one government body in the US currently has the authority to require implementation of all of these recommendations.

Richter (2002, p. 23-24) suggest that this institutional structure is stable "if it is achieved 'automatically' in the sense that the informal rules reach some stable endpoint of new (complete) institutional arrangement without destroying the original formal framework" alternatively, "if after a disturbance of an initial institutional equilibrium, a new institutional equilibrium will be reached."

The reason stability (or equilibrium) is a key concept is that it is the ability to adjust or as Hayek (1945, p. 524) has argued: "The economic problem of society is mainly one of rapid adaptation to the changes of particular circumstances of time and place." Schultz (1990) hit upon the true litmus test of the value of dealing with disequilibrium when noting that it is the mark of a good entrepreneur as well as institution that the ability to minimize resource waste and losses during periods of disequilibrium is critical. In some respect our dynamic system is always in a state of flux and one of the supposed virtues of the market was its dynamic stability in responding to this turbulence. As Fisher (1983, p. 3) has noted:

> "The view that equilibria are the points of interest must logically rest on two underlying properties about the dynamics of what happens out of equilibrium. First, the system must be stable; that is, it must converge to some equilibrium. Second, such convergence must take place relatively quickly".

Our experience in the regulatory field in the 1980s where reductions in demand resulted in price increases as opposed to price decreases is a telling case in point. If it is our ability to deal with and address the disequilibrium situations then administrative regulation failed miserably.[31] In some sense we should be developing a theory of institutional competence when assessing the effectiveness of any governance system. If the goal of justice, for example, is better served through taxation policy then those institutions should deal with the issue. If regulation in attempting to address this goal is forced to adopt a policy of socializing costs then the other goals of efficiency may be compromised as a result of this failure. Regulation and its enabling legislation need to develop and stick to assigning problems based on some form of institutional competency.[32] This issue is extremely important when we must operate in a pluralistic society with diverse sets of goals and objectives. If we are to meet these goals then we need to be as efficient as possible in employing policy tools matched to the problems at hand.

---

[31] This holds true for those states where a major disequilibrium existed. In the case of states that avoided these problems traditional regulation continued to work reasonably well.

[32] Dworkin (1977, p. ix) discusses the idea of institutional competency.

How do we reconcile the dynamics of innovation with the efficiency of routine? How do we meld the flexibility necessary for inventiveness with the orderly operation and systematic application of good managerial technique? In large network industries it is the balance between recognizing moments of transition where existing organizational paradigms and technology result in diminishing returns and new innovations present simultaneously threats and opportunities. The problem of avoiding obsolescence and accommodating change is particularly difficult in large network systems where fixed assets and expectations of owners, operators and consumers are for stability.

It is virtually impossible to build sufficient flexibility into such networks to accommodate the diverse needs of customers seeking to purchase customized services and at the same time provide the *uniform* services and routines valued by the majority. Routine is necessary to control cost. As the range of products expands so does the complexity of tracking costs. This is compounded in the network system where the vast majority of costs are common or fixed infrastructure costs.[33]

In analyzing the failure, not only of markets; but also of regulation, we must understand and appropriately characterize the traditional methods of regulation while understanding the evolution of these institutions over time. This section will provide some examples of institutional evolution and in turn provide some insight into institutional stability.

4.2 The Institution of *Traditional* Regulation

Historically, utilities were local providers of service and not national or even regional in scope. The legal rules and policy perspectives adopted were conditioned by these *facts.* Regulators operated in a world of autarky, adopting inward looking rules that examined the costs of the local company to serve its' *native* load. These costs were the costs of the firm and not reflective of a formal or informal market. The legal and regulatory guidance was predominated by the view that the utility-owned property involved with the supply of utility services was entitled to an opportunity to earn a reasonable return. In attempting to balance this return with the benefits of universalizing the gains created by the common carriers creation, regulation adopted the competitive equilibrium theory as its model. In this world total revenue had to always equal total cost which was modeled through the traditional regulatory equation: $R = E + k(V - D)$. Where R was the total revenue requirement, E operating expenses, k is the allowed rate of return, V

[33] Could it be that electric and other network industries will follow the path of railroads where the remaining rail traffic is bulk wholesale traffic and specialized carriers provide the custom services?

the total asset value of the firm and *D* the accumulated depreciation. The upside of this approach is its simple and transparent approach in reaching the appropriate balance between customers and owners. This static equilibrium framework was made operational through the use of a *test year* that compares cost to revenues on an equivalent annual basis (i.e., the matching of revenues with costs). These costs were normalized to avoid the special effects of special events. The test for what was included in *V* was the so-called *prudence test* and the level of *V* was decided by requiring the prudent investment to be *used and useful.* Retroactive ratemaking was prohibited as the focus was on setting prospective forward-looking rates. Individual cost factors could not be adjusted without reviewing the entire equation to avoid a mismatch of costs with revenues in any given test period. By adopting certain cost allocation methodologies, the cost of service were socialized to a certain extent through average prices for rate classes.[34]

This process worked reasonably well, in spite of the fact that the value of property was truncated. As long as the underlying parameters of the economy and the industry were stable the institution of regulation could allocate the small harms or losses without offending the customers or stockholders. In a stable economic environment historic test years remained reasonable forecasts of future costs of service and where technical change produced decreasing costs the regulatory lag provided rewards to be shared as opposed to harms. The system adjusted through supply anticipating demand; in effect it was an expected quantity adjustment process and not a market process that would allocate resources based on scarcity pricing.

As long as the economic environment remained relatively stable, institutional equilibria could be maintained since the error term on forecasted supply would not result in major rate impacts. Demand responsiveness, in the way we think of it today, was never an explicit part of the adjustment process.[35] In fact, since economies of scale and technical change enabled costs to fall, the only interest in the demand characteristics of customers was in discovering how elastic some customers' demands were in response to price cuts. The revenue and profit growth that occurred in the golden age of regulation represented the kind of disequilibrium that was easy to address[36].

---

[34] We recognize that this description involves some simplifications of the ratemaking process and the applicable tests. Several authors provide more detail on ratemaking including McDermott (1995), Kahn (1998) and Bonbright et. al. (1988). Lesser (2002) provides a discussion of the used and useful test. .

[35] Demand response in the traditional world was reliability driven and generally thought of as a voluntary program only to be used by a few customers.

[36] It is one of the paradoxes of regulation that when demand side management first became an issue, conservation and demand control alternatives were treated as supply options in

In some sense the pricing mechanism was anathema to regulators. Allowing prices to clear markets smacked of monopoly pricing and the profits that might arise under these conditions would obviously constitute monopoly profits. With this jaundiced view of profits, regulation rejected the option of harnessing the profit motive to work in the public interest. It was only when price increases stepped outside of the *normal* range that public pressure would result in experiments in incentive or performance regulation[37].

Even as inflation began to eat away at the stability of traditional regulation adjustments were made to patch the system and avoid a major institutional failure. The introduction of fuel adjustment clauses and various forms of "riders" that violated the single issue rate making prohibitions were adopted in order to provide flexibility in responding to cost inflation. The vertically integrated natural monopoly was not undermined by these changes. And even when the perception developed that regulation was failing we need to recognize that it did not fail in a systemic fashion. It was the mistakes of a few large utilities serving large urban populations, attempting to implement; in most cases significant nuclear construction programs, thatfailed.[38]For many states where utility management avoided these mistakes the shocks of the 1980s did not affect them in as pronounced a fashion as it did the nuclear companies.[39] The cancellation of plants and inertia embedded in the regulatory rules to complete plants, while significant, did not create a sufficient cost shock in itself to set in motion the restructuring process[40]. Ironically it was in part the unwillingness to cancel plants that eventually led to the excess capacity in the market for electricity. How this capacity would be treated by the US Federal Energy Regulatory

---

the planning process. With prices constant or somewhat declining it was the quantity or more accurately the potential to supply that bore the burden of market adjustment.

[37] Oddly enough all regulation is a form of incentive regulation when the process is viewed as fixing a price and employing regulatory lag between rate adjustments. Here again prices being fixed is used as a means of incenting the capture of cost efficiencies and not as a mechanism for addressing scarcity. Only with the newer forms of rate caps and baskets of services where Ramsey pricing and other approaches can be used are prices allowed to play their role in signaling relative scarcity and value associated with consumption.

[38] A similar fate awaited very small, relatively speaking, utilities that also jumped on the nuclear band wagon.

[39] In fact, many electric utilities remain vertically integrated under some form of traditional regulation today. The future of these institutional arrangements is less threatened by the failure of traditional regulation, than from the push of Federal and state restructuring efforts. The question as to whether or not the regulatory paradigm in these states will fail has yet to be answered.

[40] Pierce (1984) provides a detailed analysis of these issues. Tannenbaum and Hederson (1991) address FERC's market-based rates policy.

Commission (FERC) as opposed to state regulators created an incentive to shop jurisdictions. Utilities went to FERC to get permission to sell the excess on competitive terms. FERC in turn used this as leverage to obtain open access on transmission and promote more competition. In conjunction with the leverage FERC had in approving mergers open access in the transmission market could be achieved in the same way it was achieved in the gas industry[41].

4.3 The Search for an Equilibrium Institution

Having embarked on a path that attempted to circumvent state regulation of new plants and attempting to maximize the off system sales revenue to help pay for excess capacity the industry entered a slippery slope. Industrial customers, especially those facing the discipline of international markets, sought to maximize the competitive pressure in the utility industry by forcing the wheeling question. The US Federal Energy Policy Act of 1992 (EPAct) codified a new class of market participants called Exempt Wholesale Generators (EWGs) which facilitated competitive entry by new gas fired generation, spurred, in part, by the low cost of gas and relatively low capital costs and construction uncertainties, allowed competition to expand.[42] As native load grew and entry of independent merchant generators increased, prices fell and many generators faced bankruptcy. Having been the product of a disequilibrium situation that ultimately worked itself out, a new disequilibrium of the opposite nature arose. Were these failures of the market or regulation? Neither institution has handled disequilibrium in a fashion with which the public has been comfortable. The adjustment process is too damaging to our sense of fairness and disruptive given our expectations.[43]

Firms, markets and contracts have all had weaknesses that we find disturbing. Legislation has been adopted in some states to address the

---

[41] Fox (1988) has argued that the entire process of introducing competition into the natural gas industry was accomplished with out authorization from Congress. In many ways it can be argued that FERC pushed the limits of its authorization in the electric restructuring as well until EPAct was passed.

[42] Many gas-fired peaking units are essentially off-the-shelf technology which reduces the risk of construction cost overruns.

[43] Although one might suggest that while speculation in the generation market caused massive financial distress, prices, as one might expect, are generally falling and not rising as they did during the last major disequilibrium in the industry. Therefore to the extent that this latest round of mis-investment, however horrific for investors, has predictable results, might suggest that such a structure could be stable under our definition of institutional stability.

perceived risks and uncertainties by allowing pre approval of generation construction by utilities.[44] The boundary between firms and markets seems to be fluid once again as major cost parameters change and uncertainty takes a preeminent place in our planning process.

During this period of turmoil in the regulatory process state commissions began a long self critical period in which there was significant experimentation with alternative regulatory methods. In some cases these methods would have constituted a restructuring of regulation and in other cases they constituted a new set of policy patches to the traditional methods. This odyssey by state commissions represents an interesting study in institutional failure.[45]

While it is not as fashionable today, we will employ the more traditional industrial organization concept of the structure-conduct-performance paradigm in conjunction with a characterization of our organizational choice as between market and command and control policies to classify the forms of institutional experimentation that states conducted. Using these two dimensions we construct a typology of regulatory policy options that have been employed to one degree or another by regulators in the US over the last quarter century. This typology is shown in Table 2.

---

[44] For example, Iowa addressed this issue through a pre-approved ratemaking treatment *See Iowa Code* § 476.53 as applied by the Iowa Utilities Board in Docket No. RPU-01-9, *In RE: MidAmerican Energy Company.* Wisconsin provided the utility with the option of signing a contract that would be pre-approved by regulators. *See Wisconsin Statute* § 196.52(9) as applied in the Pubic Service Commission of Wisconsin's Order in Docket Nos. 05-CE-130 and 05-AE-118, *Application of Wisconsin Electric Power Company; Wisconsin Energy Corporation; and W.E. Power LLC; for a Certificate of Public Convenience and Necessity for the Construction of Three Large Electric Generation Facilities, the Elm Road Generating Station, and Associated High Voltage Transmission Interconnection Facilities to be Located in Milwaukee and Racine Counties* and *Application for Approval of Affiliated Interest Agreements Between Wisconsin Electric Power Company and Elm Road Generating Station*

[45] Jones (2001) provides an overview of the principles and rules that have come and gone in the recent past in public utility regulation. His conclusion is that a transformation of the rules and principles use in the institution of regulation is occurring toward a reliance on workably competitive markets.

Table 2: Characteristics of Regulatory Policy Options

	Structure	Conduct	Performance
Command and Control Regulation	Strong Pool	LCUP	Traditional Regulation
Incentives	Value-Based	Bidding	Performance-Based Regulation

In the electric industry after the shocks to the system in the late 1970s and early 1980s regulators experimented with least-cost (or integrated resource) planning or integrated resource. Developed at a time of excess capacity, cancelled plants, and rising costs regulators were suspicious of utility conduct and sought to implement processes that would supplement if not replace our reliance on utility information to make decisions. Unfortunately, for political and other reasons the approach focused on outcomes rather than process. In that sense it suffered the same fate as traditional regulation by being an equilibrium-oriented method rather than focusing on how to preserve our options to address change and disequilibrium conditions. For example, the very concept of optioning in planning for new capacity additions was often rejected because we still retained the total revenue requirement methods as a basis for evaluating choices. When considered in the context of minimizing the present value of revenue requirements (PVRR), any plan that required expenditure early on in the planning horizon in order to create optionality was rejected because it raised the PVRR. The value of flexibility was forsaken and exposure to uncertainty increased all to preserve the current level of rates. This focus on outcomes rather than on creating a flexible process helped to doom integrated resource planning since it could not address the key issue of how to deal with uncertainty and disequilibrium. Ironically it was these very issues that called for the change in utility regulation in the first place. The institution of LCUP, with its formal rules and less than optimal planning metrics, doomed this institution to instability, as we have defined it, from the beginning. Similar problems existed in addressing the need for policies for the cancellation of plants and the definition of and recovery of prudent costs. The inability of the regulator to commit future regulators to any policy destroyed the incentives to engage those activities that current information would deem prudent and eschew those same activities when new information deems them imprudent.[46] The issue of lost revenues associated

[46] The recognition of this lack of commitment on the part of regulators has driven new attempts to force commitment on regulators either via specific contracts or via pre-approved expenditures. However, even these institutional forms of regulation, while potentially providing a greater level of commitment, cannot fully commit regulators to a course of action.

with demand side management and conservation presented similar problems. Rather than create a profit motive by divorcing profit from sales we dabbled in experiments with incentives rather than address the issue head on.

Finding this approach unworkable, commissions opted to explore two alternatives. The first was to improve the incentives utilities faced via performance-based regulation. The second was to injecting more competitive forces into the generation markets via competitive bidding. Bad bidding design and rules regarding avoided cost calculations in many cases resulted in a "failure" of bidding. Locking in long term contracts in order to obtain financing created a contractual version of the traditional utility with a set of rigidities that exacerbated the disequilibrium problems rather than eliminate them. Must take provisions, fixed prices, and other characteristics of these contracts resulted in an adjustment mechanism that was no better than traditional regulation.

While some states battled with competitive bidding others examined the use of incentive mechanisms. Depending on the style of mechanism some of these experiments have been a success. Where incentives are broad based focusing on over all profits and in conjunction with sharing mechanisms both customers and stockholders have benefited.[47]

In the northeastern United States the PJM power pool stood as an anomaly by employing cost based dispatching in their pool. Some of the member states experimented with bidding and incentives as well as LCUP. While these experiments were being conducted PJM behaved as the classic Walrasian auctioneer employing cost based dispatching rules to achieve a least cost production result. The next step that was taken was the use of value-based bidding in place of cost based dispatch and the participation of third party generators with transactions governed by the Walrasian computer auctioneer. We have moved from command and control to markets with scarcity pricing play a greater role than it traditionally ever played in allocating resources in the utility sector.

The final move in states like Pennsylvania and California was the introduction of retail competition. Table 3 notes just a few of the characteristics of these two experiments.

---

[47] McDermott and Peterson (2002b) provide a summary of the rationale for earnings sharing in modern regulation of public utilities.

Table 3: Retail Electric Competition

Characteristic	California	Pennsylvania
Retail Price Freeze	Frozen at pre-restructuring levels until "stranded" cost was recovered or 2002.	Prices frozen
Portfolio Choices	Long-term contracts limited	Contractual relationships unlimited
Price to beat		Shopping credits inflated to "induce" competition.
Bankruptcy	IOUs	Marketers
Retail Competition	Direct access suspended in September 2001	Energy and capacity price increases in 2001 caused many retailers to exit market

The problem, however, was that many of the institutions created to govern the transition were not robust to exogenous or endogenous shocks. The fixing of retail prices while forcing the supplier to operate in an unregulated wholesale market was a prescription for disaster in California. Likewise in Pennsylvania the establishment of shopping credits, ostensibly to benefit customers, resulted in conditions that made marketers susceptible to losses from market price movements. Furthermore, the tacit acceptance of the failure of (some) retail markets is well illustrated by the movement toward a new version of the provider of last resort (POLR) function.[48] This new POLR service is not for a very small subset of customers as many had expected when restructuring was undertaken, but rather for the mass markets and even larger volume customers.[49] The post-transition pricing puzzle has been addressed in many different ways by various jurisdictions.[50] In some states the transition period has been extended such that difficult questions can be addressed later. These occurred in Ohio with Dayton Power and Light (DPL) and in Illinois with an extension of the rate freeze that was imposed at

[48] We use the term POLR to refer to the set of standard services (e.g., generation) provided to any customer that wishes to taken them. This is sometimes referred to as the default or standard offer service.

[49] POLR service was supposed to provide a back-stop service for customers whose supplier went under or for some other reason could not maintain supply until such time as the customer could choose another supplier. POLR was supposed to be for reliability purposes and not, as it has turned out, a basic service entitlement.

[50] The transition period was used by most restructuring programs to both allow incumbent utilities time to recover potentially lost revenue and prepare for competition and to allow the wholesale and retail markets to evolve the necessary institutions to support retail access.

the time of restructuring.[51] Currently Ohio is reviewing the proposal of First Energy that would also extend its rate freeze. In Ohio, the Commission specifically asked the utilities to propose a fixed rate option for customers. In the First Energy case, the Staff of the Pubic Utilities Commission of Ohio (PUCO) summed up the purpose for the extension of the transition period.[52] The Staff noted:

> The Ohio electric industry was restructured … because it was believed that, with respect to generation, customers would be better off under retail competition in a market system. However, the legislature recognized that it would take time to develop the appropriate markets… Most of the discussion regarding market development focused on retail markets. … it has become increasingly recognized that a well-functioning and competitive wholesale market is a necessary precondition for an efficient retail market. … [The actions necessary to create an efficient wholesale market] still remain in the future. Cahaan (2004, p. 3)

Other jurisdictions have used competitive bidding in one form or another to procure power and energy for the mass markets. New Jersey's recent auction for Basic Generation Service (BGS) provides an example of a formal competitive procurement process. Other jurisdictions such as Maryland, Connecticut, Massachusetts, and Maine have also used competitive bidding to procure supply for the POLR service. Arizona provides an example of the use of competitive bidding for incremental or *contestable* load,[53] but the utilities are required to maintain ownership of generation until competition

---

[51] In September of 2003, the PUCO approved an agreement that extended DPL's market development period (MDP) for two years and froze rates through the end of 2005. The agreement also contained a rate stabilization plan that froze delivery rates for three years beyond the end of the MDP and set a cap on the price of generation. In Illinois, the statuary transition period was ended from 2004 to 2006 by an amendment to the 1997 Electric Restructuring Act.

[52] Other states have had similar experiences. For example, the Arizona Corporation Commission (ACC) noted that when competition rules were first developed "the parties thought that retail competition was imminent and that the wholesale market would be competitive; that a significant number of retail competitors would be entering the market; and that customers would leave the incumbent utility and purchase power from the new competitors." However, the ACC found that "[C]ontrary to the parties' expectations and assumptions, the wholesale market has faltered, the new competitors have failed to materialize, and incumbent utilities have not lost customers in any meaningful number." *See* ACC *Opinion and Order* In Docket Nos. E-00000A-02-0051, E-01345A-01-0822, E-00000A-01-0630, E-01933A-02-0069and E-1933A-98-0471, at 28 ("Track A" proceeding).

[53] This is defined as the load that is unmet by utilities' own generation. *See* ACC *Decision No.* 65743 ("Track B" proceeding).

can take over in the wholesale market.[54] Other approaches are being proposed as well, these include re-integration, such as the recent Southern California Edison purchase power agreement with its own generation station.[55] Other examples include the Duquesne Light proposal to utilize its own (affiliated) generation for purposes of providing a POLR service.[56] Recent proposed legislation in Virginia would also provide a utility with the ability to build a coal-fired power plant if it is dedicated to providing POLR service.[57]

The seemingly inescapable conclusion is that retail markets remain immature for the mass markets and some larger customers as well. There is another looming question from the current approaches to post-transition procurement that relates to the fundamental difference between the central planning framework of the traditional utility environment and the de-centralized market environment. In a planning framework quantity is determined to meet the expected load at the minimum cost. In a decentralized framework an investor uses price signals to determine when and what level of investment is needed. Fraser (2001) explores the notion that, theoretically, both frameworks should produce similar results investment over time. However, currently none of the competitive approaches to procuring mass market POLR service have the kind of long-term arrangements that existed under the planning framework.[58] Long-term contracts appear to be necessary to finance the building of plants that have high fixed costs such as large base load coal and nuclear plants.[59] While it is not clear that the decentralized model cannot provide the incentive to invest

---

[54] The ACC delayed implementation of divesture after finding that the market had not sufficiently developed in its Track A proceeding.

[55] *See* FERC Order in *Southern California Edison Company, On Behalf of Mountainview Power Company, LLC.* Docket No. ER04-316-000.

[56] Duquesne Light Company, *Petition for Approval of Plan for Post-Transition Period POLR Service,* filed with the Pennsylvania Public Utility Commission.

[57] *See* Virginia Senate Bill (SB) 651. Specifically, SB 651 would allow any default supplier to build a coal-fired power plant in Virginia using local coal resources to have the costs of that power plant included in the default price of power and energy (including the cost of capital). The bill would require the Virginia Commission to "liberally construe" the language when reviewing proposals brought under the legislation.

[58] The long-term arrangement under traditional regulation is summed up in the term *regulatory compact.*

[59] Long-term contracts are necessary in the presence of uncertainty. Currently, a large degree of uncertainty is created by regulation itself, or more to the point, the inability of regulators (including legislatures) to commit to any particular long-term policy. It may be that medium term contracts and a well functioning spot market, as well as the other ancillary markets can take the place of a longer term contract. However, until investors are certain that policy changes will not dramatically alter the value of investments, long-term contracts will be necessary.

in the next round of baseload power plants, it is also not clear that it can, given the current environment.[60]

Ideology has unfortunately taken precedent during this last round of reforms. The necessary market reforms require not only detailed analysis and design, but also the political will to make the difficult decisions. Again, the Staff of the PUCO notes that "[I]t has become clear that the establishment of well-functioning and efficient markets for electricity is no simple matter." (Cahaan, p. 4) This can even be seen in the application of markets to emissions trading. In that case, even though we acknowledged that transport deposition models explained where emissions came from mattered, we ignored the science because it implied a constraint on trading. As a result what we achieved was a least cost production of emissions but not necessarily a reduction in the damages these emissions caused. As it turns out, since any plant could trade with any other if power plants in the New York Adirondacks were the least cost in reducing emissions they would do so and trade those permits to the Midwest plants. With these permits they could continue to emit and the acid rain damage could continue in up state New York. Once again the policy process divorced the goals and objectives from the methods employed and a failure once again is the result.

5. CONCLUSIONS

We have attempted to show that regulatory failures are products of institution designs where rigidities in the ability of the institutions to adapt to shocks lead to profound disappointment in the public's expectations. The politically sensitive character of these essential services makes them ripe for instant reform. Kahn (1988, p.xxxvii) has pointed out we really don't have a choice between perfect markets and perfect regulation as he admonished the

> "...central institutional issue of public utility regulation" remains the one that I identified at that time- finding the best possible mix of inevitably imperfect regulation and inevitably imperfect competition."

In selecting new institutions we need to learn from the lessons past and present that care in the design and implementation is critical. In restructuring efforts around the world today we still find rhetoric prevailing over analysis. There are still those who say all we need are property rights and everything

---

[60] Certain solutions to this issue have arisen including the above-mentioned Southern California Edison case. Other examples include the states of Iowa and Wisconsin that have provided contractual or regulatory certainty for building coal-fired power plants. See discussion in note 44.

else will fall into place. If there is one lesson to be learned from our regulatory history it is that there are no simple solutions, no silver bullets and no quick fixes.

REFERENCES

Bator, Francis M. 1957. "Simple Analytics of Welfare Maximization." *American Economic Review* 47(1): 22-59.

Bator, Francis M. 1958. "Anatomy of a Market Failure." *The Quarterly Journal of Economics* 72(3): 351-379.

Bell, Daniel. 1962. *The End of Ideology.* New York, NY: The Free Press.

Berk, Gerald. 1994. *Alternative Tracks: The Constitution of American Industrial Order, 1865-1917.* Baltimore MD: Johns Hopkins University Press.

Bonbright, James C. Albert, L. Danielsen, and David R. Kamerschen. 1988. *Principles of Public Utility Rates.* Arlington, VA.: Public Utility Reports, Inc.

Breyer, Stephen. 1979. "Analyzing Regulatory Failure: Mismatches, Less Restrictive Alternatives, and Reforms." *Harvard Law Review* 92, 549-609.

Breyer, Stephen. 1982. *Regulation and Its Reform.* Boston, MA: Harvard University Press.

Cahaan, Richard C. 2004. "Prepared Testimony." *In the Matter of the Application of Ohio Edison Company, The Cleveland Electric Illuminating Company and the Toledo Edison Company for Authority to Continue and Modify Certain Regulatory Accounting Practices and Procedures, for Tariff Approval And to Establish Rates and Other Charges Including Regulatory Transition Charges Following the Market Development Period.* Before the Public Utilities Commission of Ohio, Case No. 03-2144-EL-ATA.

Coase, Ronald. 1937. "The Nature of the Firm." *Economica* 4(16): 386-405.

Dworkin, Ronald. 1977. *Taking Rights Seriously.* Cambridge, MA: Harvard University Press.

Epstein, Richard. 1998. *Principles for a Free Society: Reconciling Individual Liberty with the Common Good,* Perseus Books, Reading MA.

Evetts, George. 1922. *The Administration and Finance of Gas Undertakings.* London, UK: Benn Brothers LTD.

Fisher, Franklin. 1983. *Disequilibrium Foundations of Equilibrium Economics* Cambridge, UK: Cambridge University Press.

Fox, William. 1988. "Transforming an Industry by Agency Rulemaking: Regulation of Natural Gas by The Federal Energy Regulatory Commission." *Land and Water Review* 23(1).

Fraser, Hamish. 2001. "The Importance of an Active Demand Side in the Electric Industry." *The Electricity Journal* 14(9): 52-73.

Furubotn Eirik G. and Rudolf Richter. 2000. *Institutions and Economic Theory: The Contributions of the New Institutional Economics.* Ann Arbor, MI: University of Michigan Press.

Goldberg, Victor. 1976. "Regulation and Administered Contracts." *Bell Journal of Economics* 7(2): 426-448.

Gray, Horace. 1940. "The Passing of the Public Utility Concept." *The Journal of Land and Public Utility Economics* 16(1): 9-11.

Hahn, Frank. 1970. "Some Adjustment Problems." *Econometrica* 38(1): 1-17.

Harvey Averch and Leland L. Johnson. 1962. "Behavior of the Firm Under Regulatory Constraint." *American Economic Review* 52(5): 1052-1069.

Hayek, Friedrich. 1945. "The Use of Knowledge in Society." *American Economic Review* 35(4): 519-530.

Hogan, William. 2002. "Electricity Market Restructuring: Reforms of Reforms" *Journal of Regulatory Economics* 21(1): 103-132.

Hunt, Sally. 2002. *Making Competition Work in Electricity.* New York, NY: Wiley.

Jones, Douglas. 2001. "Regulatory Concepts, Propositions, and Doctrines: Casualties, Survivors, Additions." *Energy L. Journal* 22(1): 41-63.

Joskow, Paul. 1989. "Regulatory Failure, Regulatory Reform and Structural Change in the Electric Power Industry." *Brookings Papers on Economic Activity: Microeconomics,* Special Issue, 125-199.

Joskow, Paul. 1993. "Asset Specificity and the Structure of Vertical Relationships: Empirical Evidence." in *The Nature of the Firm: Origins, Evolution, and Development,* O. edited by Williamson and S. Winter. Oxford, UK: Oxford University Press.

Kahn, Alfred. 1983. "The Passing of the Public Utility Concept: A Reprise" in Eli Noam (Ed) *Telecommunications Regulation Today and Tomorrow,* Law and Business Inc., NY.

Kahn, Alfred. 1998. *The Economics of Regulation: Principles and Institutions* Cambridge, MA: MIT Press.

Kearney, Joseph D. and Thomas W. Merrill. 1998. "The Great Transformation of Regulated Industries Law" *Columbia Law Review* 98(6): 1323-1409.

Kutler, Stanley. 1971. *Privilege and Creative Destruction: The Charles River Bridge Case.* Baltimore, MD: Johns Hopkins Press.

Lesser, Jonathan A. 2002. "The Used and Useful Test: Implications for a Restructured Electric Industry." *Energy Law Journal,* 23(2): 349-381.

MacAvoy, Paul. 1996. *The Failure of Antitrust and Regulation to Establish Competition in Long-Distance Telephone Services.* Cambridge, MA: MIT Press.

McDermott, Karl and Carl Peterson. 2001. "Designing the New Regulatory Compact: The Role of Market Processes in the Design of Dynamic Incentives." Working Paper, National Economic Research Associates, Chicago, IL. (Preliminary Draft presented at *Incentive Regulation: Making it Work,* Advanced Workshop in Regulation and Competition, Rutgers University, January 19, 2001.)

McDermott, Karl and Carl Peterson. 2002a. "Is There a Rational Path to Salvaging Competition?" *The Electricity Journal* 15(2): 15-30.

McDermott, Karl and Carl Peterson. 2002b. "The Essential Role of Earnings Sharing in the Design of Successful Performance-base Regulation Programs." In *Electricity Pricing in Transition,* edited by A. Faruqui and K. Eakin. Boston, MA: Kluwer Academic Publishers.

McDermott, Karl. 1995. "Illinois' Changing Regulatory Incentives." In *Reinventing Electric Utility Regulation,* edited by G. Enholm and J. R. Malko. Vienna VA: Public Utility Reports.

McKie, James. 1970. "Regulation and the Free Market: The Problem of Boundaries." *Bell Journal of Economics and Management Science* 1(1): 6-26.

Owen, Bruce M. and Ronald Braeutigam. 1979. *The Regulation Game: Strategic Use of the Administrative Process.* Cambridge: Ballinger Publishing Company.

Pierce, Richard. 1984. "The Regulatory Treatment of Mistakes in Retrospect: Cancelled Plants and Excess Capacity" *University of Pennsylvania Law Review* 132 (March): 497-560.

Posner, Richard A. 1969. "Natural Monopoly and Its Regulation." *Stanford Law Review* 21(3): 548-643.

Schultz, Theodore. 1990. *Restoring Economic Equilibrium.* Cambridge, MA: Basil Blackwell.

Selwyn, Lee. 1996. "Market Failure in 'Open' Telecommunications Networks: Defining the New 'Natural Monopoly'." In *Networks, Infrastructure and the New Task for Regulation,* edited by W. Sichel and D.Alexander. Ann Arbor, MI: Univ. of Michigan Press.

Sichel, Werner (ed). 1976. *Salvaging Public Utility Regulation.* Lexington MA: Lexington Books.

Sidak, Greg. 2002. "The Failure of Good Intentions: The Collapse of American Telecommunications after Six years of Deregulation." Beesley Lecture.

Simon, Herbert. 1957.*Models of Man.* New York, NY: John Wiley and Sons.

Tannenbaum, Bernard and J. Steven Henderson. 1991. "Market-based Pricing of Wholesale Electric Service." *The Electricity Journal* 4(10): 30-45.

Trebing, Harry. 1976. "Market Structure and Regulatory Reform in the Electric and Gas Utility Industries." In *Salvaging Public Utility Regulation,* edited by W. Sichel. Lexington MA: Lexington Books.

Williamson, Oliver E. and Sidney G. Winter (eds). 1993. *The Nature ofthe Firm: Origins, Evolution, and Development.* New York, NY: Oxford University Press

Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications.* New York, NY: The Free Press.

Chapter 5

Coopetition in the Telecommunications Industry*

Menahem Spiegel
Rutgers University

1. INTRODUCTION

It has long been understood that if the consumer is given the choice of buying their goods from a competitive industry or buying the same goods from an industry where the providers are cooperating they will prefer the competitive environment. As consumers they prefer to see more price and non-price competition among the providers of their consumption goods. They prefer the competitive behavior of the producers as it generates lower market prices. Similarly, under the other market structure where producers are cooperating in setting their outputs consumers would expected to end up paying higher (non-competitive) prices. In that respect, cooperation of this kind between independent providers has been perceived beneficial to the producers and detrimental to consumers. It is the main objective of this paper to draw the attention to the often observed market phenomenon of coopetition where the cooperation between the (otherwise) competitive providers is beneficial to both the consumers as well as the producers.

Coopetition is defined here as a market situation where non-cooperative providers are competing to increase the number of their customer according to the basic rules of market competition in order to maximize their profits

* I am grateful to the discussants of the paper Richard Clarke and Stephen Levinson and also to Richard Simnett, Jeremy Guenter and to the participants of the CRRI Research Seminar May 7, 2004 for their helpful comments.

while at the same time the providers enter into enforceable contracts of cooperation. In order for these contracts to be beneficial to consumers as well as to producers, the basic nature of these contracts has to be limited to activities in which they are helping each other to "produce" their intermediate and/or final outputs at a lower cost. Such a market coopetition is often observed, recently, in the telecommunication industry.

As a simplified example consider the following case of the wireless telecommunications industry. Assume that there are only two providers: Verizon and Sprint PCS. Thus, current and potential subscribers (consumer) would like to see these providers engaged in a fierce price competition to attract new subscribers to their respective network. These new customers might belong to one of the following groups:

1. New customers entering the market (first time users)
2. Customers from the other provider.

While the fierce price competition between the providers is going on, all subscribers (new and old) of these two providers of cellular network services are expecting the two networks to cooperate so that they will be able to call and receive calls from subscribers of the other network. Thus, in order to maintain this kind of service, the two providers of the telecommunication networks are expected to compete and cooperate at the same time. This kind of market situation is called 'Coopetition.'

Referring to the example of the two providers of network services, Sprint PCS and Verizon have made the *strategic choice* to cooperate in their competitive world by letting their subscribers communicate with subscribers of the other network. This kind of a strategic choice was not always the preferred strategy for phone companies. In the past, telephone companies were not always cooperating. In those days, consumers of the telephone services who were interested in obtaining a broad coverage and obtaining the ability to communicate with consumers across networks needed to subscribe, separately, to the different providers of the network services. As of 1996, cooperation (coopetition) between land-line phone companies is required by law. Since that time all the competing providers of network services are obliged to enable their subscribers to communicate with any other subscriber of any provider. At present, we do observe a similar situation where the providers of network services decided to make the strategic choice of keeping their network separated (not connected.) This strategic choice of the owners of the networks not to cooperate is mainly economic based decision. Therefore, in order any subscriber to extend the coverage of his communication capabilities that subscriber needs to subscribe to several networks simultaneously.

It is our basic assumption that the objective of a the provider of network services is to maximize his profits. Therefore, it goes without saying that in a

market with many providers, each one of the network providers would like to attract to his network many more new customers. In practice, these networks, no doubt, compete in a non-cooperative fashion to achieve this end of attracting customers. To achieve this end they compete by using the means of price as well as non-price competition. In this market, once a customer joins any network, say Sprint PCS, he can communicate with all the other customers of Sprint PCS. In addition, assuming the coopetiton between the providers, as a subscriber to Sprint PCS, that subscriber can also communicate with all the customers of Verizon. Clearly, to enable this inter-network communication requires an agreement of cooperation between these two competitors. Although we cannot rule it out, for the purpose of the current discussion and modeling of the market situation it is assumed here that this cooperation between the providers of the network services is not a cartel agreement aimed at raising the prices paid by consumers and restraining trade in this industry.

In general, we observe a large variety of activities of cooperation between competing firms operating in the same industry. Although there might be a long list of reasons to justify such cooperation, in this paper we concentrate on the "demand side" incentives for the cooperation between producers. In contrast to the typical cooperation of partners to a cartel agreement where the main objective of the group as whole is to control and limit the availability of output, in our framework here, the objective of the cooperation is to expand availability of the output produced. In particular, we assume here the existence of the long observed characteristic of the telecommunication services that the service provided is subject to significant network externalities. The concept of network externalities means that the quality of a network from the consumers' point of view is increasing with the coverage provided by that network. This quality variable was typically measured in terms of the number of consumers accessible via the network. This concept of network externalities was first introduced in the seminal article by Katz and Shapiro (1985). This concept was later applied to many other aspects of the network industries. Bental and Spiegel (1994a) analyzed the optimal network size under different market structures in particular, different modes of regulation and deregulation. Bental and Spiegel (1994b) note that this measure of quality of output is directly under the control of the producer (as the number of subscribers depend on the choice made by consumers not the producer). With a given population size the network quality is constrained by the total number of consumers. Therefore, the optimal network size and the total coverage strongly depend on market structure. Economidas (1994 and 1996) and Laffont and Tirol (1994 and 1999) applied it to competition in the telecommunication industry. In this paper, the coopetition between providers of telecommunication network

services enables to extend the positive externalities that are beneficial to consumers as well as to the network owners.

This type of coopetition is not limited to the cellular phone system only. Similar kinds of coopetition are often observed between other types of networks and non-network industries. A most striking example of coopetiton in non-network industry is the case of the financial sector located in lower Manhattan in the days immediately after September 11, 2001. As direct results of the catastrophic event, in addition to the tragic lose of lives many firms found themselves unable to continue their operations due to the lack of important inputs like space and computer facilities. Following the strict rules of competition one would expect that if a firm cannot perfume it should go out of the market. In the days after September 11, 2001, competitor cooperated in providing (sharing) the needed input to firms that otherwise would have needed to cease their operation. Another most relevant example from another network industry is the domestic and the international air travel industry. While the different airlines are competing in order to attract passengers at the same time they are cooperating to in scheduling, in code sharing, in transferring passengers and luggage. The electric utility industry is another example of ongoing coopetition between the different providers. In the postal services, the most recent and well-publicized agreement for coopetition between the two competitors in the mail delivery industry the USPS and FedEx is another example.

Another well-known example is in the field of the international re-mailing. Where the new entrant, a local contractor competes with the local national post in the sub-market for international deliveries. The new entrant, typically, collects the international (and sometimes local) mailing from 'big' enterprises and carries them all the way or part of the way to the receiving country. At the end point, the incumbent national post and the new entrant are using the same input, the end-country mail system for the distribution. This kind of competition by the re-mailing is used to bypass the 'expensive' domestic mail system.

For the telecommunication industry, while this type of co-opetition is the taken for granted by the customers of any given cellular networks and by subscribers of the other communication networks, as of yet the analytical modeling of this situation is not often presented in the literature. It is the main objective of this paper to develop the basic economic model that will consider the problem of network coopetition in a formal way.

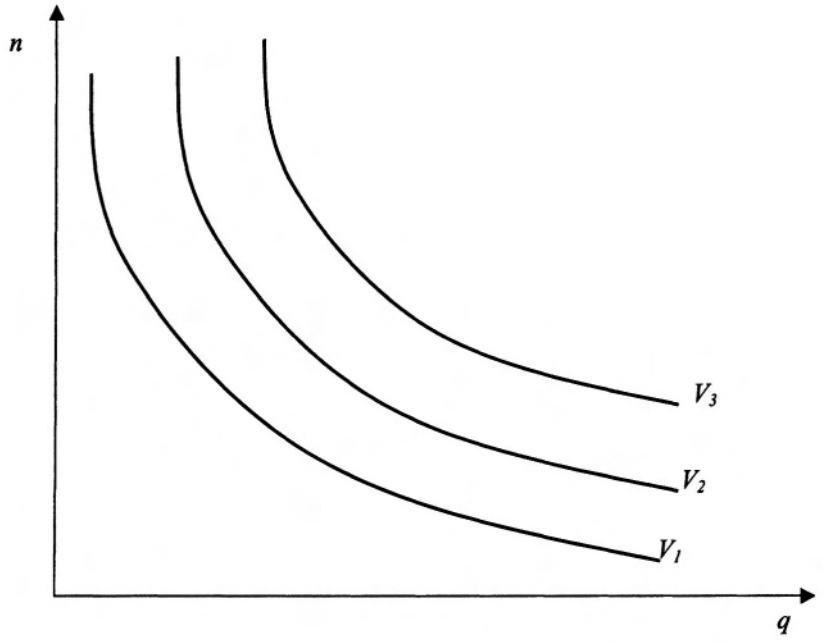
In the following section a simple model of network externalities is presented. Using this model, the benchmark for the analysis, the socially optimal structure is derived. Next the solution of a single (monopoly) provider of the network services is presented. In Section 5, the multi-network competition is introduced. At first the networks are disconnected

and then the competition between connected (cooperating) networks is presented.

2. THE MODEL

Consider an economy with a finite number N of identical consumers. Each of whom is endowed with a bundle of $\omega > 0$ units of an all-encompassing private consumption good (income). Each consumer maximizes her utility function $u(q,n,x)$. Where q represents units of the telecommunication services consumed, n represents number of consumers that can be accessible by the telecommunication network. And x represents units of the all-encompassing private consumption good (measure in terms of units of ω.) when p represents the unit price of the telecommunication services, the budget constraint of each consumer is $pq+x<\omega$. We also assume a well behaved utility function with positive and decreasing marginal utilities from consuming the telecommunication services and from consuming the all-encompassing private good. That is, $u_q(q,n,x), u_x(q,n,x)>0$ and $u_{qq}(q,n,x), u_{xx}(q,n,x,)<0$. By the introduction of n in the utility function, we explicitly assume the existence of network externalities in the consumption of the telecommunication good. For the positive network externalities it is assumed that the marginal utility of the network size is positive and decreasing. Thus, $u_n(q,n,x)>0$ and $u_{nn}(q,n,x)<0$ is reflecting the positive network externality. Having q, n and x as substitute goods is presented in Figure 1. Where, the indifference curves are negatively sloped.

Figure 1:
Consumer's Valuation & Substitution Between network size *n* and *q*



The participation rule for each of the individual consumer is given by:

$$u(q,n,\omega-pq)>u(0,n,\omega). \quad (1)$$

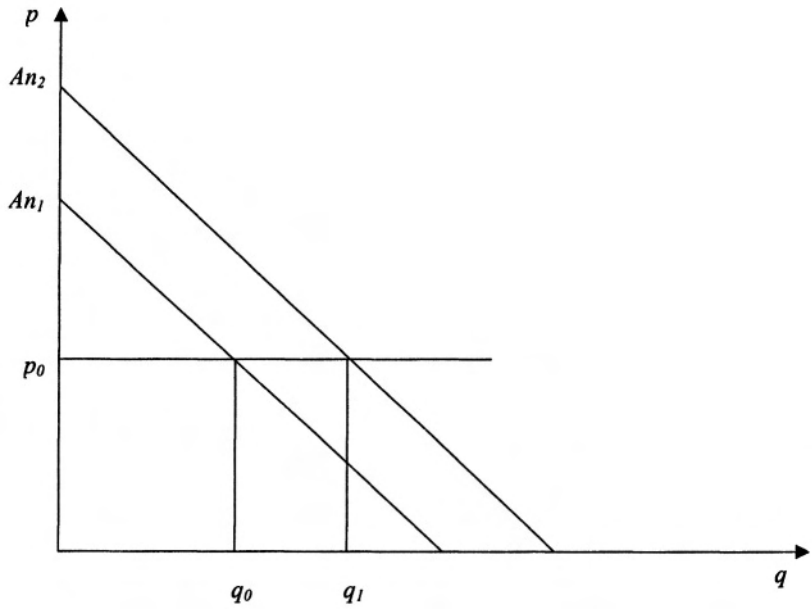
Where the LHS represent the utility derived when participating a network of size n charging the price p. The RHS represents the utility when consuming zero communication services. This participation decision depends on the market price *p*. If p is very high the LHS is larger than the RHS and consumers will be better off consuming zero telecommunication. Similarly, When RHS is larger than the LHS the consumer will subscribe for the telecommunication services. Thus, given that the RHS is monotonically decreasing in *p* and the LHS is a constant there is a p* such that for any $p<p^*$ the consumer will subscribe to the communication network and for $p>p^*$ the consumer will prefer not to consume telecommunication network services. It is clear that the same decision rule applies to all the identical consumers. Therefore, below we consider only cases where all consumers decide to participate in the network i.e; $p<p^*$.

For the sake of simplicity of the analysis we will limit the scope of our discussion and concentrate here on a partial equilibrium one market analysis. Therefore, we assume that the valuation function $V(n,q)$ (which, in this case also represents the total amount of consumer surplus (CS) derived) from consuming q units of the telecommunication services with other n members is given by:

$$V(n,q) = Anq - \frac{1}{2}bq^2 - pq \qquad (2)$$

where p is the price (in terms of the consumption good) paid for the telecommunication services. Since only q is under the control of the individual consumer he will maximize his consumer surplus with respect to q. From the first order condition of maximizing (2) we can derive the consumer and the market demand curve for the product communication as $p=An-bq$. This demand curve, which depends on the network size, represents the positive network externalities. An increase in the number of subscribers will increase the willingness of consumers to pay.

Figure 2: Consumer's Demand for Network Services (q) as n Increases



On the production side it is assumed that the cost of producing (transmitting) a telecommunication message when the sender and the receiver are subscribing to the same network is consists of a fixed set-up $S cost to establish the network and constant marginal cost c_d per unit of service (a call or per minute).

3. THE SOCIALLY OPTIMAL PROVISION (BENCHMARK)

Given the individual consumers private valuation of his consumption of the telecommunication services and the cost of providing these services, the social planner can construct and solve the maximization of the social valuation of the telecommunication services for the social optimal provision of these network services. This solution will include; k^s - the optimal number of networks, n^s - the optimal size of each network, p^s - the price per unit of service and q^s - the number of units to be provided. Formally this is done by:

$$\max_{q,n,k}: [(Anq - \frac{1}{2}bq^2 - c_d q)n]k - kS \qquad (3)$$

Where the *k* represents the number of equal size networks and thus, we have that $k=N/n$. (note that for a given N, k and n are negatively related.) The first order conditions of the maximization problem in (3) imply that the optimal consumption q^s will be achieved when $An-bq=c_d$. Thus, positive output will be produced when $An>c_d$. Clearly, due to the positive network externalities, the consumer's valuation described in (2) increases with the number of consumers connected to the same network (*n*). At the socially optimal solution n^s is achieved when all consumers *(N)* will be connected to a single telecommunication network. (Again, given that $AN>c_d$, and that the fixed setup cost *S* is small $k^s=1$.) At the socially optimal solution the price charged will be $p^s=c_d$. The socially optimal output will be $q^s=(1/b)(AN-c_d)$.

Below we consider the equilibrium provision of the telecommunication services under a variety of market conditions. The first market structure is the single producer of the network services.

4. THE MONOPOLY NETWORK

In the monopoly case we consider a single profit-maximizing provider the telecommunication services. Since all consumers are identical, at the price set by the monopoly they will all join the network provided that the price is $p<An$. Thus, the monopolist problem is to set the profit-maximizing price by solving:

$$\max_{q} : \Pi = TR - TC = (ANq - bq^2 - c_d q) - kS \ . \qquad (4)$$

Clearly, the profit function (4) is monotonic increasing in n. Therefore; the monopolist will operate a single network and connect all N consumers to that network.

The first order condition implies the regular monopoly profit maximizing condition of ($MR=MC$) and $AN-2bq=c_d$. Or $q_m=(1/2b)(AN-c_d)$. Like the socially optimal solution, the monopoly will provide its telecommunication services to all N consumers by connecting them to the single network. At the monopoly price (p_m) each consumer will choose to consume (q_m) that is only one half of the telecommunication services that he would consume under the socially optimal solution (q_s). The monopoly price is $p_m=(1/2)(AN+c_d)>c_d=p_s$.

5. NETWORK COMPETITION

Here we consider the case where two competing producers can provide the telecommunication services. Two cases will be discussed. First we consider the case of separated or closed networks where each network is limited to provide telecommunication services only to its subscribers. Next, we allow for interconnection between the two competing network.

5.1 The Closed (separated) Communication Networks

Consider a market consisting of the above-described N identical consumers where their valuation function for the telecommunication service is given in (2). On the production side assume that the telecommunication services can be provided by either one of the two independently owned networks. As these networks are separated (not connected) each one of them is capable of providing communication services only to its subscribing customers. That is, any customer

subscribed with network *A* can communicate with any other customer of that network but cannot communicate with customers subscribed with network *B*[1].

The consumer's decision can be represented as a 2-step decision. At first he has to decide whether or not to participate in the market according to condition (1). If he decides to participate he has to choose his preferred network. Therefore, the consumer's problem is to maximize his valuation by choosing to join and purchase his telecommunication services from the network *A* or *B* that will provide him with the highest consumer surplus. In terms of equation (2) the consumer's problem is:

$$\max_{A,B} : \{CS_A = An_A q_A - \frac{1}{2} b q_A^2 - p_A q_A,\ CS_B = An_B q_B - \frac{1}{2} b q_B^2 - p_B q_B\} \tag{5}$$

The result of (5) dictates the following consumer's choice: Select network *A* if $CS_A > CS_B$. Select network B if $CS_A < CS_B$. Flip a balanced coin to select the network provider if $CS_A = CS_B$. For simplicity let us assume that n_a and n_b represents market shares. Thus we have that $N = n_a + n_b = 1$. Using (5) we can derive the market share of each provider of the network services as follows:

$$n_A = \begin{cases} 1 & if \quad CS_A > CS_B \\ 0 & if \quad CS_A < CS_B \\ \frac{1}{2} & if \quad CS_A = CS_B \end{cases}$$

and (6)

$$n_B = \begin{cases} 1 & if \quad CS_B > CS_A \\ 0 & if \quad CS_B < CS_A \\ \frac{1}{2} & if \quad CS_B = CS_A \end{cases}$$

[1] Given the absence of the fixed subscription fee, we assume that no consumer will subscribe to the two networks simultaneously.

The profit functions of the two network providers are as follows:

$$\pi_A = \begin{cases} (p_A - c_d)q_A & if \quad CS_A > CS_B \\ 0 & if \quad CS_A < CS_B \\ (p_A - c_d)q_A \frac{1}{2} & if \quad CS_A = CS_B \end{cases}$$

and (7)

$$\pi_B = \begin{cases} (p_B - c_d)q_A & if \quad CS_B > CS_A \\ 0 & if \quad CS_B < CS_A \\ (p_B - c_d)q_A \frac{1}{2} & if \quad CS_B = CS_A \end{cases}$$

Under the current set-up of the problem, the price of the network services (p_i) is the key instrument determining the networks size. Furthermore, this variable is the only control variable available to the network provider. The Nash equilibrium of this non-cooperative price game will be at $p_A=p_B=c_d$.

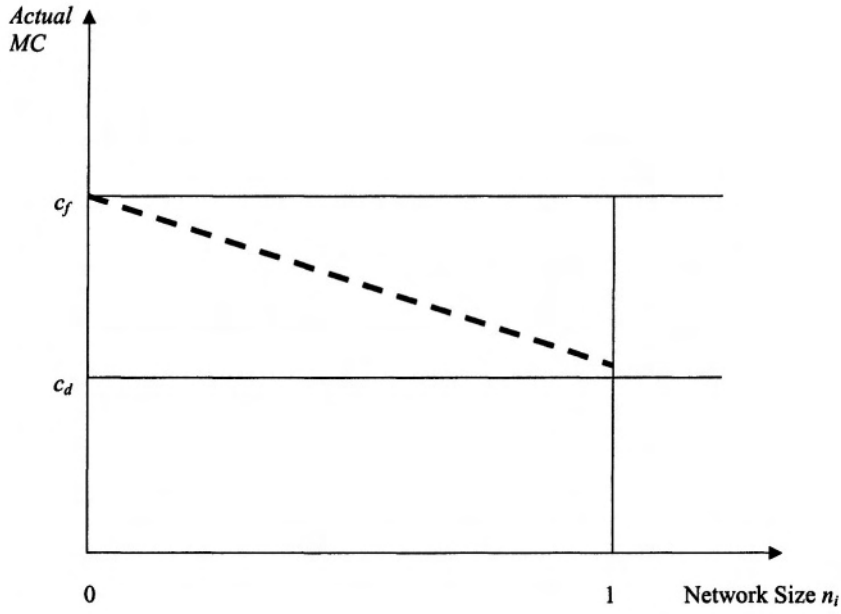
From the producers point of view this Cournot-Nash equilibrium resembles the outcome of a prisoners' dilemma situation. Since at this equilibrium the profits of the two providers are $\pi_A=\pi_B=0$. Clearly, there exists another (cooperative) symmetric equilibrium that is much preferred by the two providers. This is when: $p_A=p_B=P_m$. At this equilibrium the profits of the two providers $\pi_A=\pi_B>0$.

5.2 The Cooperating (connected) Networks

As before, assume now a two-network world. In this case we assume that the networks are connected. That is, any customers of network *A* can communicate with all other customers of the same network or she can communicate with all customers of network *B*. As before, assume that the total number of the identical consumers is *N*. Consumers objective function is described by their valuation function given by (2) and they need to choose whether to consume telecommunication services. If they choose to consume telecommunication services they have to select the network they wish to be connected to, network *A* or to network *B*.

On the cost side we assume, as before, that an intra-network communication between two customers of the same network (*A* or *B*) can be produced at the same constant marginal cost of c_d. Similarly, an inter-network communication between a customer of one network to a customer of the other network can be produced at a constant marginal cost of $c_f > c_d$. For simplicity, let the difference between the two constant marginal costs $(c_f - c_d)$ represents the marginal cost of inter-network access charges. It is assumed that the access charges paid between the networks is determined outside this market and it is fixed at the level of the marginal cost of access.[2] It is important to note that the access charges represent here the cost of coopetition between the networks. This is the cost of coopetition as it enables the network that originates the communication to produce its 'final' output.

Figure 3: The Actual Marginal Cost and the Network Size



The actual total cost of operation depends now on the number of intra-network communications and the number of inter-network communications generated by the consumers of the given network.

[2] For sake of convenience assume that the access charges is exogenously determined. See also Gans (2001) and Laffont and Tirole (1994) and (1999).

For simplicity, assume that each consumer connected to either networks (*A* or *B*) has an equal probability of communicating with any of other consumers. Thus, the actual marginal cost of network *i*, *i=A, B* is given by: $c_i=\alpha_i c_d + (1-\alpha_i)c_f$. Where α_i represents the share of intra-network communications generated by consumers of network *i*. This share is equal to the share of the total population of consumers connected to network *i*. Thus, $\alpha_A=n_A/N$ and $(1-\alpha_A)=n_B/N$ and $\alpha_B=n_B/N$ and $(1-\alpha_B)=n_A/N$. Clearly, since $c_f>c_d$ the actual marginal cost of a given network is decreasing with the size of the network.

As consumer, in their valuation, do not differentiate between intra-network communication and inter-network communication they will consider a single price per unit of communication *p*. As above, consumers' choice of their preferred telecommunication network will be dictated by their valuation function as follows:

Select network *A* if $CS_A>CS_B$.
Select network B if $CS_A<CS_B$.

Flip a balanced coin to select the network provider if $CS_A=CS_B$.

Given consumers' selection rule above, the market shares (or the allocation of consumers) of the individual networks will be as following:

$$n_A=\begin{cases}1 & if \quad CS_A > CS_B\\ 0 & if \quad CS_A < CS_B\\ \frac{1}{2} & if \quad CS_A = CS_B\end{cases}$$

and (8)

$$n_B=\begin{cases}1 & if \quad CS_B > CS_A\\ 0 & if \quad CS_B < CS_A\\ \frac{1}{2} & if \quad CS_B = CS_A\end{cases}$$

The valuation function in (2), $V(n,q)$ indicates that in this case of the connected networks the quality of both networks, measured by *n*, is equal. As all consumers, independent of their choice of which network to subscribe, can access all consumers the total population *N*

will represent the equal quality. Therefore, the profit functions of the two networks can be described by:

$$\pi_A = \begin{cases} (p_A - c_A)q_A & if \quad p_A < p_B \\ 0 & if \quad p_A > p_B \\ (p_A - c_A)q_A \frac{1}{2} & if \quad p_A = p_B \end{cases}$$

and (9)

$$\pi_B = \begin{cases} (p_B - c_B)q_B & if \quad p_B < p_A \\ 0 & if \quad p_B > p_A \\ (p_B - c_B)q_B \frac{1}{2} & if \quad p_B = p_A \end{cases}$$

Since the network quality N is identical to both networks, the only control variable left is the price p_i, $i=A, B$. Under such Bertrand price game, there are two possible Nash equilibria:

I. Equal treatment case.
II. Limit pricing case.

Under case I, the equal treatment, equilibrium will be reached only when both networks are of the same size and have the same number of subscribers $n_A=n_B=(1/2)N$. The price charged for telecommunication services will be: $p_A=p_B=0.5(c_d+c_f)$. And the profits are: $\pi_A=\pi_B=0$.

Under case II, if for some reason $n_A \neq n_B=(1/2)N$, and let $n_A>n_B$. Then, network A will obtain a cost advantage since $c_A<c_B$. Therefore, in the process of Bertrand competition, at every stage we will observe that $p_A<p_B$. At the Nash equilibrium we will have that $n_A=1=N$ and $n_B=0$. Under these circumstances, producer A will play the limit price strategy by choosing the price:

$$p_A=Min\{p_m, c_f\} \tag{10}$$

Since both the monopoly price (p_m) and the marginal cost of the inter-network communication (c_f) exceeds the marginal cost of intra-network communication, the owners of network A will realize positive profits. Some further analysis is needed in order to compare these two cases.

Consider first the equal treatment case. If the access charge is determined exogenously, the above solution ensures an efficient output level (at $mc=p$). But, if the access charge is determined endogenously, it might be set at a level higher than the marginal cost of the inter-network delivering of the communication. This might be because of the small number of players or because of some sort of a tacit collusion. The second case of the 'oligopoly' limit pricing might be socially less desired outcome.

6. CONCLUDING REMARKS

In a market for network communication when network externalities are present the socially preferred solution is the provision of the network product by a single firm. In this solution of the socially preferred single provider, the utilization of network externalities will be maximized. The main drawback of such a solution is that a single producer solution might result in the typical monopoly problem of (high) monopoly pricing and small output. On the other hand, a solution, which includes a large number of small providers, will under utilize the network externalities. One possible solution is to introduce the network competition with coopetition where the providers of the telecommunication network services are interconnected. This kind of coopetition is warranted by the profit incentive of each producer as well as by the consumers and therefore, the total social welfare generated is increased.

REFERENCES

Bental Benjamin and Menahem Spiegel (1990) "Externalities in Consumption of Telecommunication Services and the Optimal Network Size," In *Telecommunication Demand Modeling: An Integrated View,* (Contributions to Economic Analysis, 187) A. De Fontenay (ed.) North Holland, Amsterdam 1990.

Bental Benjamin and Menahem Spiegel (1994a) "Deregulation Competition, and the Network Size in the Market for Telecommunication Services," *Journal of Regulatory Economics,* 6, 283-296.

Bental Benjamin and Menahem Spiegel (1994b) "Market Structure and Market Coverage in the Presence of Network Externalities," *The Journal of Industrial Economics* v.6 pp 283-296.

Crew, Michael A, and Paul R. Kleindorfer (Eds). 2002. *Postal and Delivery Services: Delivering on the Competition,* Kluwer, Boston, MA.

Economides Nicholas and Lawrence J. White (1994) "Networks and compatibility: Implications for Antitrust, *European Economic Review,* vol. 38, pp. 651-662.

Economides Nicholas (1996) "The Economics of Networks" *International Journal of Industrial Organization,* vol. 14 #2

Gans Joshua S. (2001) "Regulating Private Infrastructure Investment: Optimal Pricing for Access to Essential Facilities." *Journal of Regulatory Economics,* vol. 20 #2 pp. 167-189.

Katz Michael and Carl Shapiro (1985) "Network Externalities, Competition and Compatibility" *American Economic Review* vol. 75 #3, pp. 424-440.

Laffont J. J. and J. Tirole (1994). "Access Pricing and Competition." *European Economic Review* 38 #4, pp.1673-1710.

Laffont J. J. and J. Tirole (1999). *Competition in Telecommunications.* Cambridge, MA MIT Press.

Chapter 6

Forward and Spot Prices in Electricity and Gas Markets

Does "Storability" Matter?[*]

J. Arnold Quinn,[1] James D. Reitzes,[2] and Adam C. Schumacher[2]
Federal Energy Regulatory Commission[1] *and The Brattle Group*[2]

1. INTRODUCTION

Forward sales of electricity and natural gas play an important role in developing prudent energy procurement programs and establishing robust wholesale energy markets. For both reasons, the relationship between forward (or futures) prices and spot prices is important to market efficiency.

To the extent that forward prices do not closely track expected spot prices, it may be costly to make advance energy purchases at fixed prices. In electricity markets, this is an increasingly important issue as more state public service commissions (PSCs) permit local distribution companies (LDCs) to meet their energy supply needs through advance-procurement auction mechanisms. In Maryland, for example, LDCs use an auction to procure energy supplies up to three years prior to actual delivery. The success of these auctions depends on the efficiency of the forward markets. If the firms procuring energy supplies through the auctions are highly risk

---

[*] The opinions and conclusions contained herein do not necessarily reflect those of the Federal Energy Regulatory Commission, its Staff, or any of its Commissioners. The authors would like to thank seminar participants at the Federal Energy Regulatory Commission, Research Seminar on Public Utilities, and the 23rd Annual Eastern Conference of the Center for Research in Regulated Industries. We thank Malcolm Ainspan, Greg Basheda, Pradip Chattopadhyay, Michael Crew, John Garvey, and Menahem Spiegel for helpful comments.

averse, then a substantial premium could be paid to secure a fixed price commitment. This excessive "hedging" cost is likely borne by consumers.

By contrast, in Nevada, the public service commission determined that forward power purchases prior to the summer of 2001 were transacted at excessively high prices, forcing the utilities to absorb the hedging cost. As such, this policy could make utilities wary about hedging against price volatility, fearing that public service commissions may subsequently disallow some of the associated cost if the spot price subsequently settles below the forward price. If utilities do not hedge their power purchases as a result, consumers will ultimately be exposed to potential price volatility through *ex post* changes in their retail rates.

For the reasons just mentioned, LDCs prefer efficient forward energy markets, and ideally a "tight" relationship between forward and spot prices. If forward energy markets are efficient, it is less costly for firms to develop a portfolio of physical and financial assets that conform to their risk preferences. Nonetheless, as we show in this paper, electricity markets are unlikely to exhibit a tight relationship between forward and spot prices. Moreover, as compared to other commodities, forward electricity markets may be "thin" and potentially inefficient.

Since electricity cannot be stored easily, forward electricity prices should be determined by supply and demand conditions expected to prevail at the "delivery" time for the power. Without product inventories to smooth out price fluctuations over time, the differential between current spot and forward prices in electricity markets may be quite variable when compared to storable commodities such as natural gas. Moreover, electricity markets face substantial temperature-related variations in demand, causing spot and forward prices to be highly variable by themselves. As a result, electricity markets inherently may be susceptible to periods of substantial disparity between forward prices and *expected* spot prices in the future. With a high concentration of physical power sellers and buyers in many regions, the forward markets may be subject to a lack of trading depth and potential volatility. This could create one of two potential situations: (1) *contango,* where forward prices are well above expected spot prices and hedging future price risk is relatively costly for power purchasers, or (2) *normal backwardation,* where forward prices are well below expected spot prices and hedging future price risk is relatively costly for power sellers.

In contrast to electricity, we would expect that natural gas would show less volatility in its spot-forward pricing relationship, since it is a storable commodity and market participants can accumulate inventories when they expect future prices to be high.[1] The ability to use inventory adjustments to

[1] The accumulation of gas in storage, particularly during summer and autumn, assists in meeting the winter peak demand. By contrast, electricity is prohibitively costly, if not

engage in inter-temporal arbitrage should bring forward prices, expected spot prices, and current spot prices into alignment in gas markets. Consequently, we would expect the concerns regarding possible contango or backwardation to be less in natural gas markets relative to electricity markets.

In this chapter, we test the above hypotheses by exploring the relationship between spot and futures prices for PJM electricity[2] and Henry Hub natural gas. Consistent with our predictions, we find generally that the differential between current spot and month-ahead forward prices is substantially more variable for electricity than for natural gas. While the PJM spot-forward price differential shows marked seasonal changes, these patterns have been absent from the Henry Hub natural gas market until somewhat recently, when winter demand apparently increased sufficiently to create constraints on storage capacity. In addition, electricity forward prices have substantially "overshot" and "undershot" actual spot prices, raising the prospect of possible contango or backwardation in the spot-future price relationship.

Our evidence also suggests that electricity forward prices are influenced very little by current spot prices, but do share a close relationship with spot prices during the delivery month (which we use as a proxy for *expected* future spot prices). Current market conditions have limited explanatory power with respect to the nature of future electricity prices. To the contrary, gas futures prices are apparently influenced to a much larger extent by current spot prices than by actual spot prices during the delivery month. This is consistent with the behavior of a storable commodity where the ability to engage in inter-temporal arbitrage implies that current equilibrium prices reflect both current and future supply and demand conditions.

On the whole, our evidence suggests that the forward-spot price relationship in electricity is volatile when compared with storable commodities such as natural gas. The individual spot and forward price series are volatile in themselves. Moreover, the lack of storability raises the inherent prospect of a thinly traded forward market for physical power, where substantial premiums may be paid for forward transactions under certain conditions. Although the relative immaturity of forward electricity markets may account for some of the currently observed price volatility and the possible appearance of price premiums, regulators should re-evaluate

---

technically impossible, to store. Electricity capacity is storable to a limited extent in that maintenance can be scheduled to maximize the available generation capacity during the summer peak demand. Hydroelectric power is also storable since reservoir levels can be raised to increase future generation capability.

[2] For the period under examination, the PJM electricity market covered Pennsylvania, New Jersey, Maryland, Delaware, and the District of Columbia.

their design of strategies for the advance procurement of electric power in light of potential market imperfections.

2. FORWARD CONTRACTS AND STATE ENERGY PROCUREMENT POLICIES

The ability to "lock in" the price of electricity or natural gas to be delivered and consumed in the future has the potential to drastically reduce the risk faced by local distribution companies (LDCs) and consumers. LDCs often sell to retail customers at fixed rates based on expected wholesale "power" costs. While some states have adjustment clauses that provide for recovery of certain deviations from expected wholesale costs, other states perform prudence reviews of the LDCs' procurement practices. In states that do not readily provide rate adjustments for recovering wholesale price changes, the LDC bears some financial risk as a result of price volatility. In states where LDCs obtain cost recovery more easily, this risk is ultimately borne by end users through *ex post* rate adjustments. Forward purchases provide a means for mitigating this risk.

While forward purchases can reduce the risk to which LDCs and end users are exposed, they do not necessarily lead to lower procurement costs. As discussed below, forward prices in efficient markets are essentially equal to the expected spot price for the delivery month, adjusted for inherent risk differences. On average, we would expect forward procurement to cost the same as spot procurement (after adjusting for risk). However, in thinly traded markets, this need not be the case. There may be additional "hedging" costs associated with "locking in" a fixed price.

Some states have decided that the benefits of reduced price risk outweigh the value of waiting for a potentially better deal closer to the energy delivery date. New Jersey, Maryland, and Maine all rely on auction mechanisms to secure fixed-price generation well in advance of its usage. All three states allow retail electricity competition, while providing "standard offer service" to all customers who do not choose an alternative provider. These states auction off the right to either provide retail service to, or wholesale supply for, standard offer customers. The winning bids effectively determine the retail price for this service.

Maine conducts an auction to provide retail standard-offer service for up to two years. Maryland allows bids covering up to three years to provide wholesale supplies for standard-offer service. New Jersey conducts an annual auction where wholesale supplies are obtained for standard-offer service for up to 34 months. While forward contracting for up to three years certainly reduces the price risk to end users, it is important to understand

what, if anything, it costs consumers to shed this risk. Ultimately, this is a question about the consistency of the forward and spot price relationship.

Other states, like Nevada and Pennsylvania, take a different view and consider whether forward contracting is more expensive than the expected cost of purchasing in the spot market. These regulatory policies raise interesting questions, such as whether it really is imprudent to pay even a small premium to shed price risk if end users are sufficiently risk averse. Moreover, the *expected* cost of forward purchases relative to future spot purchases may be quite different from the *actual* differential between forward and spot prices. Thus, the luxury of hindsight may bring a different perspective to the regulator's view of what constitutes prudent behavior.

The Nevada situation has drawn considerable attention as an extreme example of the potential for forward prices to deviate substantially from actual spot prices. In late 2000 and early 2001, Nevada Power Company (NPC) and Sierra Pacific Power Company (SPPC) purchased electricity through the broker market for delivery during the summer of 2001. When making these forward purchases, the western U.S. was still in the grips of the California energy crisis. Some observers were predicting that the western U.S. was in for another summer of extremely high electricity prices. In response to these concerns, Nevada Power bought some summer 2001 forward contracts for over $400 per MWh.

As it turned out, the Federal Energy Regulatory Commission took further action to constrain western electricity prices in June 2001. In addition, the price of natural gas (a key fuel for electricity generators) subsided. Consequently, the price for electricity during on-peak hours in the summer of 2001 was much lower than many had anticipated.

The Public Utility Commission of Nevada (PUCN) held a series of hearings to determine whether Nevada Power and Sierra Pacific Power should be permitted to recover their fuel and purchased power costs accumulated over the spring and summer of 2001. According to Nevada law, electric rates can be adjusted to "true up" any discrepancies between actual and forecast "fuel" procurement costs, subject to a prudence review. The PUCN found, among other things, that the Nevada Power Company's purchases of summer 2001 forward contracts at $420 per MWh in February 2001 were "imprudent" in part because the forward price was too high.[3] As a result of these and other findings, the PUCN disallowed recovery of $437 million—almost half—of NPC's $922 million of deferred fuel and purchased power costs.[4] The PUCN similarly concluded that SPPC made several imprudent forward power purchases, and as a result disallowed

---

[3] The PUCN also concluded that NPC bought more power than necessary to serve its expected summer load.

[4] Public Utility Commission of Nevada (2002a).

recovery of $53 million of SPPC's $205 million of deferred fuel and purchased power costs.[5]

Although the Nevada experience is unusual, it is important to acquire a better understanding of how forward and spot electricity prices can become so de-coupled. It is possible that a rare confluence of events drove forward electricity prices so high relative to the subsequently realized spot prices. Alternatively, as we show below, some structural elements of electricity markets may impede the convergence of forward and spot prices.

3. FORWARD CONTRACTS AND EFFICIENT ENERGY MARKETS

Healthy forward trading is instrumental to the development of robust wholesale energy markets. Indeed, the over-dependence on spot market transactions was cited by FERC as one of the primary drivers of the California energy crisis.[6] Market analysts and academic experts (see, for example, Borenstein, Bushnell, and Wolak, 2002) have argued that spot market demand for electricity is highly inelastic, facilitating the exercise of substantial market power. It is also argued that the presence of forward energy trading directly reduces market power. Since generators use forward trading to sell a portion of their supply capability at fixed prices in advance of the spot market, the incentive is lessened to withhold supply as a means of boosting spot market prices at the time of delivery (see Allaz and Vila, 1993, and Green, 1999). Given that forward prices and spot prices at delivery are related, the end result is the reduced exercise of market power in both forward and spot markets. Some analysts (see Joskow and Kahn, 2002) have suggested that certain participants in the California energy markets were less active in alleged market manipulation precisely because they had committed most of their capacity in the forward markets.

Forward contracts are essential to the execution of the risk management strategies used by active wholesale energy traders. Energy traders and merchant generators bear substantial risk in the course of doing business. If these risks can be managed, then the market will be more hospitable to trading and arbitrage activity, leading to increased liquidity. In turn, the increased liquidity would improve the quality of price signals in the market.

Indeed, if a fully functioning forward market were to develop, we would expect forward prices to serve as unbiased forecasts of spot prices.[7] Accurate future price forecasts offer numerous benefits, including improving

---

[5] Public Utility Commission of Nevada (2002b).

[6] Federal Energy Regulatory Commission (2000).

[7] Again, this is after an adjustment for inherent risk differences.

the valuation of investments in generation, transmission, and demand-reducing energy technologies. The bottom line is that a liquid forward market improves the transparency of the overall energy market and produces valuable price signals.

4. FORWARD-SPOT RELATIONSHIPS AND PRODUCT STORABILITY

In this section, we examine how forward markets and spot markets are related if they behave efficiently. The impact of product storability on this relationship is also analyzed.

4.1 Market Efficiency and the Forward-Spot Relationship

A forward market is said to be efficient if the forward price equals the risk-adjusted expected spot price. Algebraically, the relationship between the forward price and the expected spot price is as follows,

$$F_{t,T}/(1+r_f) = E[S_T]/(1+r_s), \quad (1)$$

where $F_{t,T}$ is the forward price at date t for delivery on date T, $E[S_T]$ is the expected spot price at date T, r_f is a (nearly) risk-free discount rate, and r_s is the appropriate discount rate for volatile spot prices.

The above result can be restated as:

$$F_{t,T} = E[S_T]((1+r_f)/(1+r_s)). \quad (2)$$

Note that this result implies that the forward price is frequently below the expected spot price because forward sales at a fixed price are typically less risky than future sales on the volatile spot market (*i.e.*, $r_f < r_s$). This result, of course, assumes that forward markets are perfectly efficient.

In reality, certain factors influence whether a forward market is efficient. Forward markets often need active participation by speculators and arbitrageurs, investors who have no "physical" need to hedge commodity price movements. Speculators may take a position in the forward market based on their belief that the forward price is either too high or too low relative to the expected spot price. Arbitrageurs are typically "risk-neutral" market participants, who respond to perceived pricing discrepancies by buying on the spot market and selling on the forward market, or vice versa.

If, for example, buyers of the underlying commodity are substantially more risk averse than sellers of that commodity, and hence willing to pay a "premium" to hedge their risk through a forward purchase, speculators and arbitrageurs would attempt to sell on the forward market (and buy on the spot market in the future) to take advantage of this situation. That would lower the forward price to the point where it equaled the expected spot price after accounting for inherent risk differences.

In the absence of active participation by speculators and arbitrageurs, the different risk tolerances of buyers and sellers determine whether the forward price contains an excessive premium or discount relative to the expected spot price.[8] The analysis of this situation was initially addressed by Keynes and Hicks.[9] Keynes argued that farmers were inherently more risk averse than consumers, as farmers typically sold one or two commodities to derive their income, while consumers could buy a combination of several commodities and satisfy their budget and diet requirements. As a result, he argued that farmers would accept an excessive discount from the expected spot price in order to sell their crop forward at a fixed price. Both Keynes and Hicks referred to this situation as *normal backwardation.* The opposite relationship, where futures prices are greater than expected spot prices, is called *contango.*[10]

The relative risk preferences of electricity buyers and sellers are not easily determined. It is possible that electricity buyers, facing highly volatile prices and defined customer service obligations, could be more risk averse than sellers. In this case, the forward electricity price would trade at a premium relative to the expected spot price *(i.e.,* contango). As a result, forward purchases would not represent the *least-cost* procurement option for electricity distribution companies.

This example illustrates the potentially perverse incentives created by state regulatory requirements that disallow costs in excess of the *"least cost"* option. Electricity futures prices may deviate most from expected spot prices during periods of high spot price volatility. Of course, these would be the times when forward purchases provide the most risk reduction. Thus, regulatory requirements requiring "least cost" energy procurement create an incentive for LDCs to wait for the spot market in times of volatile prices. In turn, this behavior may place end users in the situation of bearing substantial

---

8 For further discussion, see Brealey and Myers (2000).

9 See Keynes (1930) and Hicks (1946).

10 The use of the terms contango and normal backwardation is not entirely consistent within the economics and finance literature. In some cases, backwardation refers to situations where the forward price is less than the current spot price.

ex post price risk, as they ultimately pay energy procurement costs through their specified electricity rates.[11]

For example, it is possible that in early 2001, forward prices in the western U.S. for energy delivered during the summer included a substantial risk premium relative to the expected spot price. In hindsight, waiting to buy on the spot market represented the procurement option with the lowest expected cost. It also was the *ex post* optimal decision. However, in the wake of the California market meltdown, almost no utility in the western U.S. would have comfortably relied on the use of the spot market to meet its energy needs for that time period.

Of course, buyers of electricity are not necessarily more risk averse than sellers. In markets with substantial excess capacity, sellers arguably may be more risk averse relative than buyers, as some of the generating capacity will ultimately be idle in real-time. In this situation, the forward electricity price may be substantially less than the expected spot price, assuming there are not a sufficient number of speculators and arbitrageurs. In this case, the least cost procurement option would be to purchase electricity through forward contracts.

4.2 Storability and the Forward-Spot Relationship

Electricity potentially differs from many commodities in its forward-spot price dynamics because it is not a storable commodity. In an efficient market, the forward price of a storable commodity should have a close relationship with the *current* spot price. If a product can be stored and sold later, then the forward and spot prices are linked through *inter-temporal arbitrage.* That is, if a producer believes that prices will be substantially higher in the future, or observes that forward trading prices are relatively high compared to current spot prices, then that producer should store some of the commodity with the intention of selling at a later date. This will remove supply from the current spot market, increasing the current spot price and lowering forward prices (as well as expected future spot prices). Similarly, a buyer of the commodity may buy the product now and pay for storage if it observes that existing forward prices are high relative to current spot prices.

The "opportunity cost" of selling forward is the current opportunity cost (*i.e.,* the current spot price) plus the cost of storing the product for future sale. In an efficient market with inter-temporal arbitrage, the forward price should not be greater than the opportunity cost of selling forward. In other

[11] Also, as mentioned earlier, relying on the spot market (including the day-ahead market) for most energy needs may facilitate the exercise of market power, since substantial energy purchases occur under conditions where market demand is quite inelastic.

words, the forward price should not exceed the current spot price plus storage costs, adjusted for an appropriate discount factor since the proceeds of a forward sale are not received until the delivery date. When efficient inter-temporal arbitrage is constraining forward and spot prices, the following relationship must hold,

$$F_{t,T} = S_t(1+r_f)+k_{t,T}, \tag{3}$$

where S_t is the current spot price and $k_{t,T}$ is the cost of storage. Alternatively, this relationship can be expressed as:

$$S_t(1+r_f) - (F_{t,T} - k_{t,T}) = 0. \tag{4}$$

Frequently, in analyzing the efficiency of forward and spot markets, we examine the behavior of the *marginal convenience yield,* $CY_{t,T}$, which is defined as follows:[12]

$$CY_{t,T} = S_t(1+r_f) - (F_{t,T} - k_{t,T}). \tag{5}$$

In efficient markets for storable commodities, the expected marginal convenience yield should be zero.[13] Sometimes, since storage costs may be relatively small or difficult to measure, analysis is frequently performed on the *net convenience yield,* $S_t(1+r_f) - F_{t,T}$, which equals the marginal convenience yield net of storage costs.

The ability to engage in inter-temporal arbitrage suggests that storable commodities may be less prone to contango or normal backwardation. If expected future spot prices are high relative to forward prices, and forward prices are tied to current spot prices through inter-temporal arbitrage, then one would expect that producers would hold onto additional output for sale into the future spot market. Alternatively, consumers could buy on the current spot market and hold for future consumption or sale. This process

---

[12] Definitions of the *marginal convenience yield* differ. Some use the spot price alone, instead of the spot price multiplied by one plus the discount rate. In this case, "discount" costs are implicitly included as part of the cost of storage. Other definitions use the "present value" of the convenience yield, where the spot price is not multiplied by one plus the discount rate, but the forward price is instead divided by one plus an appropriate discount rate. See Brealey and Myers (2000).

[13] While many argue that market efficiency requires that the marginal convenience yield should hover around zero, some suggest that the convenience yield must be positive since the sales decision itself represents a "real option." Product prices are inherently volatile over time, and a producer has an option to delay and sell at a potentially higher price sometime in the future. To counterbalance that option value and induce current sales, some argue that current prices must produce a positive convenience yield. For further discussion, see Robert Pindyck (2001).

would drive down future spot prices in line with forward and current spot prices.

4.3 What Does This Mean for Electricity, A Non-storable Commodity?

Since electricity is largely non-storable, there is no ability to engage in inter-temporal arbitrage to ensure conformity of forward prices (and future spot prices) with current spot prices. Neither are there inventories that can be used to dampen the effects of short-term supply and demand shocks. Thus, the relationship between spot and forward prices should be substantially more variable for electricity than for a storable commodity such as natural gas. Also, the behavior of spot (and forward) prices by themselves may be more volatile, relative to a storable commodity such as natural gas.

Moreover, forward prices for a non-storable commodity such as electricity should depend on expectations regarding supply and demand conditions at the time of delivery. Current spot electricity prices should depend on current supply and demand conditions. Thus, current prices should provide only limited information toward determining forward prices, since current conditions may have little connection with expectations regarding future market conditions.

With natural gas, a commodity prone to inter-temporal arbitrage, current spot prices and forward prices should reflect current market conditions as well as expectations about future conditions. Thus, current prices contain much of the information relevant to determining forward prices.

This discussion raises these testable hypotheses:

1. Do electricity forward and spot prices exhibit a weaker relationship, compared to a storable commodity such as natural gas?
2. Are the marginal convenience yields for electricity highly variable, while the convenience yields for natural gas hover around zero?
3. To what extent do forward electricity prices depend on current electricity prices, as compared to other variables that are suggestive of future market conditions?
4. Do electricity markets exhibit periods where forward prices substantially deviate from spot prices, as indicative of either contango or normal backwardation?

In the next section, we attempt to empirically examine these hypotheses by analyzing forward and spot prices for electricity and natural gas.

5. EMPIRICAL ANALYSIS OF ELECTRICITY AND NATURAL GAS FORWARD-SPOT PRICE RELATIONSHIPS

To analyze differences in the relationship between forward and spot prices for electricity and natural gas, we focus on trading at two particular points: PJM's Western Hub for electricity and Henry Hub for natural gas. PJM, which coordinates the transmission system and facilitates electricity trading for the region consisting of Pennsylvania, New Jersey, Maryland, and Delaware, is one of the more developed regional electricity markets. Henry Hub is a specified delivery point for many spot and forward gas transactions.

With respect to forward (or futures) transactions, gas trading for delivery at Henry Hub is a "thicker" market than electricity trading for delivery to PJM's Western Hub. PJM forward trading is largely a physical market where industry participants take product delivery. However, some purely financial forward trading also arises as a result of speculative or arbitrage behavior.[14]

On the other hand, Henry Hub gas trading, which includes many participants seeking physical delivery, is characterized to a much larger degree by the involvement of financial speculators and arbitrageurs. The futures gas product is a largely financial vehicle, where the trading volume is a substantial multiple of the physical gas volume exchanged at the hub.[15]

5.1 Data Sources

Our proxy for "spot" prices is the prevailing day-ahead price for electricity or natural gas. In this paper, PJM "spot" prices (in $/MWh) are a volume-weighted average of the day-ahead prices for 16-hour blocks of

---

[14] PJM includes about 280 members and customers. These parties physically trade power through the organized day-ahead and day-of "exchange" markets, as well as through forward transactions that may involve a broker. The peak electricity demand in PJM is 87,000 MW which is generated from 800 different supply sources. However, purely financial trading of PJM electricity is of sufficiently low volume that NYMEX has "de-listed" PJM futures products during particular time periods.

[15] Henry Hub is arguably the most liquid U.S. trading hub for wholesale natural gas. More than 180 customers receive gas at this point where fourteen interstate pipelines converge on a large storage cavern. The maximum delivery capability at this hub is over 1,800,000 MMBtu per day. In August 2003, for example, physical natural gas delivery to Henry Hub averaged 450,000 MMBtu per day. By contrast the average open interest for monthly natural gas futures in the summer ranges from 30,000 to 40,000 contracts, where each contract represents 10,000 MMBtu of total monthly delivery. For further detail, see Energy Information Administration (2003).

electricity, as reported by *Power Markets Week* for PJM's Western Hub. Natural gas spot prices (in $/MMBtu) are reported by *Gas Daily* and reflect the day-ahead market price for wholesale delivery at Henry Hub. The daily midpoint price is used.[16]

With respect to constructing a forward (or futures) price series, the challenge lies in selecting products that are consistently traded and subject to common information limitations. For both electricity and natural gas, several durations of forward (or futures) contracts may be available at any one time, with transactions ranging from a one-month to a multiple-year advance purchase. However, some of these products are thinly traded.

In this paper, we use a "month-ahead" forward or futures price, which represents the price for next (entire) month delivery as reflected on the first trading day of the current month. PJM West's month-ahead prices were obtained from *Power Markets Week,* which reports "low deal," "high deal," and index prices for forward trades. If the index price was unavailable, we used the average of the low-deal and high-deal prices. Month-ahead gas futures prices for Henry Hub are published by NYMEX.

Electricity prices are examined over the period from August 1998 through November 2000. Before and after this period, PJM month-ahead forward contracts were subject to low trading volumes.[17] Prices for natural gas, a more robustly traded product, cover the period from January 1997 through December 2003.

5.2 Empirical Observations

Simple graphs of spot and month-ahead forward prices for electricity and natural gas provide some initial impressions of the differing inter-temporal price relationships for these commodities.

5.2.1 Electricity

Since electricity is not a readily storable commodity, and hence not subject to inter-temporal arbitrage opportunities, we would *not* expect to see a particularly strong relationship between current spot prices and month-ahead forward prices. Figure 1 appears to confirm this supposition, showing

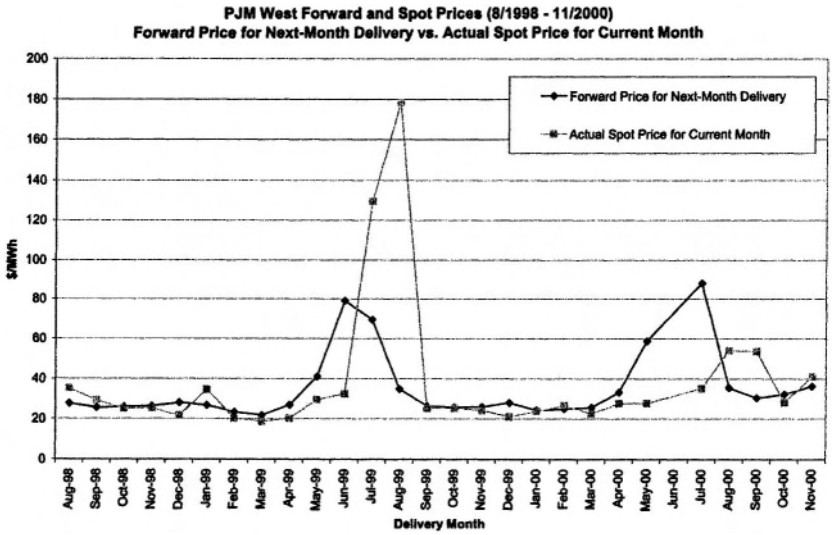
---

[16] This price is within a half-cent of the daily weighted average for all transactions.

[17] The PJM exchange market for "spot" (*i.e.,* day-ahead or day-of) electricity transactions did not begin until April 1998. Prior to that time, substantial power in PJM was self-provided or procured under long-term supply contracts. Subsequent to the price "spikes" in the central and eastern U.S. in the summer of 1999, and the meltdown of the California market during 2000, forward electricity markets in PJM and around the country became less liquid. By early 2001, NYMEX had delisted major electricity futures products.

first-of-the-month daily spot prices against forward prices for electricity delivered in the following month. For example, the initial pair of observations shows the daily spot price at PJM's Western hub on the first trading day of August 1998, compared with the forward price on that day for energy delivery to PJM Western Hub during September 1998.

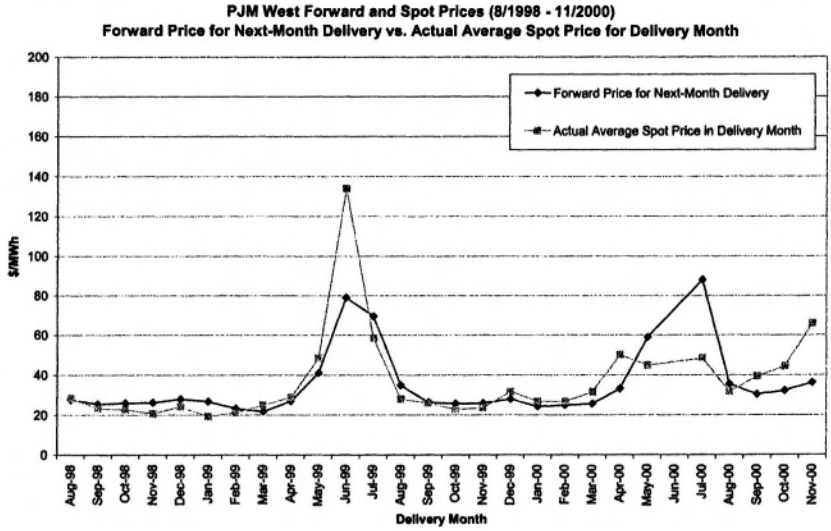
Figure 1



By contrast, the realized spot prices during the delivery month should be related to the forward price for that month, particularly if speculators are playing an active role in the forward market. This relationship is illustrated in Figure 2, which plots the forward price for PJM electricity (as reported on the first trading day of the prior month) against actual average daily spot prices during the delivery month. For example, the first pair of observations shows the September 1998 forward price (as reported on the first trading day of August 1998), compared with the actual average daily spot price during September 1998.

An examination of Figures 1 and 2 prompts three particular observations. First, as shown in Figure 1, forward electricity prices for next-month delivery appear to be little influenced by the prevailing spot electricity price at the time of the transaction. Though, as shown in Figure 2, a much closer relationship exists between actual average spot prices in the delivery month and the month-ahead forward price for that same month. This observation suggests that electricity forward prices are a function of market expectations of demand and cost conditions during the actual delivery month, and these expectations are not strongly influenced by current market behavior.

Figure 2



Second, Figure 2 indicates that forward prices for electricity in PJM were sometimes lower than actual spot prices in the summer of 1999. This relationship can be explained by substantial differences in expected and actual market conditions, with the result that expected market prices were much lower than the price "spikes" that arose during that summer. Alternatively, this relationship might represent evidence of normal backwardation, where sellers of electric power had to accept a substantial discount relative to expected future spot prices in order to hedge their pricing risk. This latter explanation is less likely, as we might expect sellers to extract a premium over expected spot prices when spot prices are anticipated to be high.

Third, by contrast, forward prices tended to be systematically higher than resulting spot prices in the summer of 2000. This is also consistent with substantial differences between expected and actual market conditions, where expectations of spot prices were much higher than the prices that actually developed. Alternatively, forward prices may have been substantially higher than actual spot prices as a result of contango, where buyers of electric power, subsequent to the price spikes in the summer of 1999, effectively paid a substantial premium in order to hedge their pricing risk.

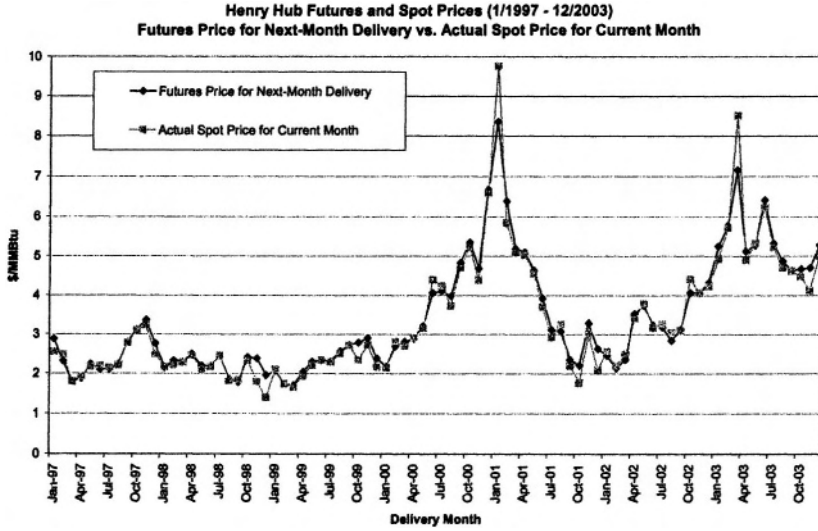
Based on these results, it would have been cost effective to buy electric power on the forward market prior and during the summer of 1999. At the same time, it would have been cost effective to eschew forward purchases

and buy power on the spot market during the summer of 2000. Thus, these figures suggest that an inflexible power procurement strategy, such as one where nearly all electric power is purchased well in advance of delivery, could end up being quite costly to consumers. Of course, this after-the-fact analysis ignores the value of reducing one's exposure to volatile spot prices.

5.2.2 Natural Gas

The relationship between current daily spot prices and month-ahead futures prices appears to be much closer for natural gas than for electricity, as illustrated in Figure 3. To the naked eye, the "tight" correspondence between current spot and month-ahead futures prices for Henry Hub gas reflects inter-temporal arbitrage activity involving a storable commodity.

Figure 3

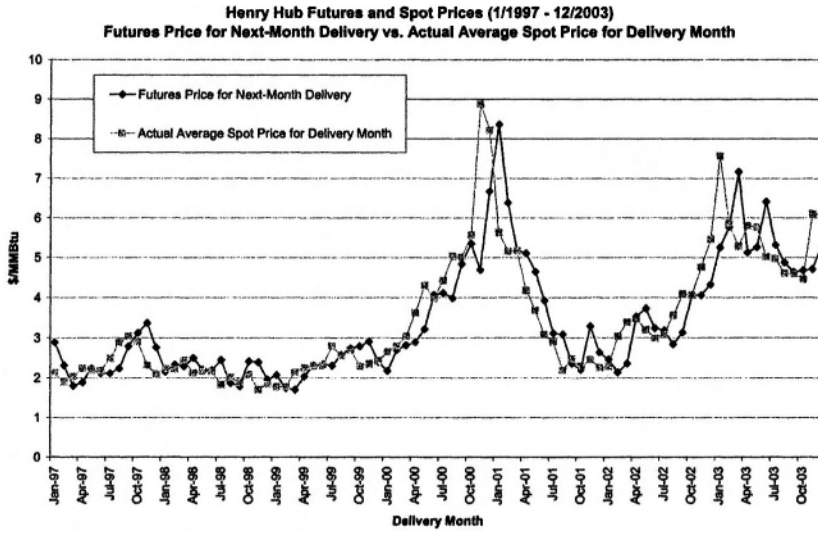


If inter-temporal arbitrage allows natural gas prices to remain near equilibrium, then movements in spot and futures prices would be related to "shocks" to the equilibrium state. These shocks may cause natural gas prices to behave like a "random walk," which implies that the best prediction of the future price is the current observed price level. At a glance, natural gas prices appear to conform to this behavior in Figure 3. By contrast, Figure 4 shows the month-ahead futures price and the actual average daily spot price during that delivery month, suggesting a much looser relationship.

Three contrasting observations may be made by comparing Figures 3 and 4. First, the futures price for next-month delivery of natural gas appears to

be driven more strongly by the prevailing spot price at the time of the transaction as opposed to spot prices during the delivery month.

Figure 4



Second, high winter gas prices in recent years appear to suggest that inter-temporal arbitrage has its limits, as price run-ups in the face of increased winter demand now seem unavoidable. One explanation for this pattern may be more seasonal variation in consumption by residential end users. Another explanation may be the increasing reliance of electricity generators on natural gas as a primary fuel source. This second effect implies that increases in electricity demand as a result of cold temperatures are a source of increased derived demand for natural gas. In a recent report, the Energy Information Administration (EIA) identified both factors as potential causes of recent winter price increases:

> Consumption of natural gas for electric power not only increased in 2002, but it also expanded in each of the last 5 years, 1998-2002. At present, electric power use of gas is the second-largest consuming sector, and, after moving ahead of residential consumption in 1998, now exceeds residential volumes by almost 800 Bcf. [...] The residential, commercial, and electric power consuming sectors exhibit seasonal variation in their consumption. Consequently, the share of the market driven by seasonal factors is growing."[18]

[18] Energy Information Administration (2004), p. 7.

In addition, the gas storage system may be used quite differently than it was prior to 2000 or 2001. For example, for much of 2002 and 2003, weekly storage levels reported by EIA have been at either 5-year highs or lows, suggesting that the storage system has been pushed toward its physical limits in recent years.

As a third observation, futures gas prices do not appear to fully reflect the spot price levels attained during the height of the winter season. Also, although seasonal patterns have emerged that depart from "random walk" price behavior, future prices still appear to react *ex post* to changes in current spot prices, rather than to anticipate seasonal increases in demand (see Figure 3).

Finally, to recap, Figures 1 through 4 offer evidence that conforms with two of our hypotheses. The relationship between forward and current spot prices for PJM electricity appears much weaker than the relationship between futures and current spot prices for Henry Hub gas. The PJM electricity market also exhibits sustained periods where forward prices differ substantially from the spot prices realized during the contract delivery month, which is broadly consistent with potential contango or normal backwardation. In gas markets, prices have increased during recent winters, and futures prices typically have risen by less than realized spot prices. This behavior suggests that the future-spot pricing relationship is changing in gas markets.

5.3 Net Convenience Yields

Net convenience yields[19] express the discounted difference between current spot and futures prices. As shown earlier, this relationship is expressed as

$$CY_{t,T} = S_t(1+r_f) - (F_{t,T}),$$

where $CY_{t,T}$ is the "net convenience yield" at time t with respect to a product delivered at time T, r_f is the imputed monthly interest rate,[20] S_t is the spot

[19] These convenience yields are net of storage costs. For electricity, storage is not a readily available option (except possibly for hydroelectric power and pumped storage). For natural gas, these costs are difficult to determine. Reasoning that a positive premium is frequently needed to induce current sales instead of waiting for future sales, Pindyck (2001) suggests using the (absolute value of the) largest negative spot-forward price differential as an estimate of gas storage costs. Our interest in the convenience yield concerns whether the spot-forward price differential (after discounting) in electricity and gas is systematically different from zero. That naturally focuses our attention on the behavior of *net* convenience yields.

[20] The interest rate is based on the rate for U.S. Treasury securities with a one-year maturity.

price at time t, and $F_{t,T}$ is the forward (futures) price at time t for delivery at time T. Net convenience yields for PJM electricity and Henry Hub natural gas are displayed in Figures 5 and 6, respectively.

Figure 5

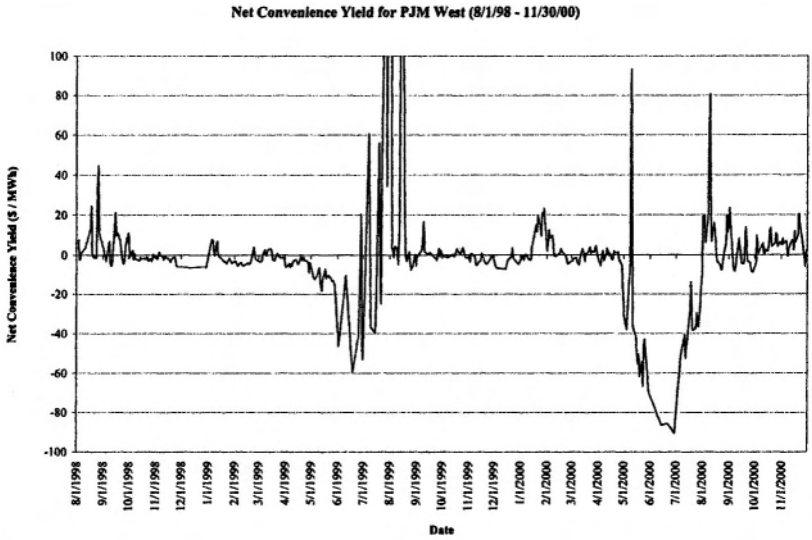
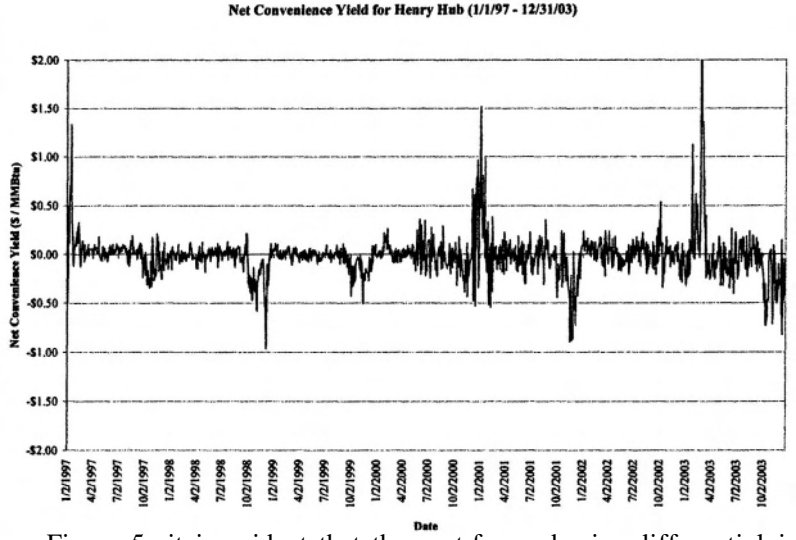


Figure 6



From Figure 5, it is evident that the spot-forward price differential in electricity markets is subject to substantial swings. Once again, this is consistent with a product that cannot be readily stored, so that at a given

point in time, futures prices are largely based on expected cost and demand conditions at time of delivery.

In PJM, negative net convenience yields appear in early summer as current spot prices are below month-ahead forward prices. This is likely explained by the seasonality of spot electricity prices. With the anticipation that temperatures will become hotter, and electricity demand will increase, future spot prices are likely to exceed current spot prices. As a result, forward prices rise above current spot prices, implying that a negative net convenience yield is observed (as shown in Figure 5). In late summer, a positive convenience yield arises due to the opposite effect. With temperatures becoming cooler, expectations are that electricity demand will fall. As a result, expected future spot prices, and hence forward prices, drop below current spot prices.

As illustrated in Figure 6, natural gas convenience yields fluctuated around zero prior to the end of 2000. This behavior is consistent with inter-temporal arbitrage that kept spot and forward prices together, likely near long-run equilibrium levels. Beginning in late 2000, increased winter demand often caused spot prices to increase while forward prices rose by a lesser amount. Consequently, convenience yields for Henry Hub gas have typically been positive during recent winters. There also has been a forward price increase in late autumn (prior to the winter increases in spot prices), producing negative convenience yields during that time.

Thus, the above results are consistent with our hypothesis that, due to a lack of storability, electricity convenience yields would be highly variable. Gas, as a storable commodity, would show convenience yields that hover around zero. While the behavior of gas prices in the past was consistent with this hypothesis, the more recent forward-spot pricing relationship in late autumn and winter does show substantial negative and then positive convenience yields.

5.4 Regression Results

To test our empirical observations more formally, we undertook regression analysis to examine the relationship between forward (or futures) prices and spot prices. Our analysis relied on the pricing data described in section 5.1, which is represented in Figures 1 through 6. Principally, we constructed three specifications where month-ahead forward prices $(F_{t,T})$ were regressed on different combinations of "independent" variables:

1. current daily spot prices (S_t) only;
2. current daily spot prices (S_t) and average daily spot prices during the delivery month (S_T);

3. current daily spot prices (S_t), average daily spot prices during the delivery month (S_T), and average daily spot prices during the delivery month in the previous year (S_{T*}).

Our regressions also included an intercept term (with coefficient *a)* and corrected for first-order serial correlation in the error process (with autocorrelation coefficient ρ).

Table 1

Summary Statistics of Key Variables

Product	*Variable*	*Units*	*Number of Observations*	*Mean Value*	*Standard Deviation*	*Minimum Value*	*Maximum Value*
PJM Electricity	$F_{t,T}$	$ / MWh	33	35.96	16.29	21.83	88.00
PJM Electricity	S_t	$ / MWh	33	39.27	32.74	18.68	178.44
PJM Electricity	S_T	$ / MWh	33	37.48	20.62	19.26	133.96
Henry Hub Natural Gas	$F_{t,T}$	$ / MMBtu	84	3.40	1.43	1.70	8.36
Henry Hub Natural Gas	S_t	$ / MMBtu	84	3.34	1.53	1.40	9.76
Henry Hub Natural Gas	S_T	$ / MMBtu	84	3.44	1.57	1.71	8.88

Table 2

Henry Hub Futures-Spot Price Relationship

(Values in parentheses indicate t-statistics)

Specification 1		Specification 2		Specification 3	
Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
a	0.3523	a	0.2617	a	0.1337
	(4.80)		(3.39)		(1.24)
S_t	0.9116	S_t	0.7721	S_t	0.7636
	(45.80)		(29.07)		(28.15)
		S_T	0.1621	S_T	0.1670
			(6.07)		(6.10)
				S_{T*}	0.0470
					(2.06)
ρ	0.062141	ρ	0.230017	ρ	0.200806
	(0.56)		(2.11)		(1.68)
DF	82	DF	82	DF	72
R-Square	0.9668	R-Square	0.9772	R-Square	0.9789

Table 1 provides summary statistics for our key variables. From this table, note that month-ahead forward electricity prices ($F_{t,T}$) have a much smaller standard deviation than daily spot electricity prices (S_t). For natural gas, the standard deviations of the forward and daily spot prices series are

similar. This also suggests that the forward-spot relationship is quite different for electricity than for natural gas.[21]

Table 3
Henry Hub Futures-Spot Price Relationship
(Values in parentheses indicate t-statistics)

Specification 1		Specification 2		Specification 3	
Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
a	0.3523	a	0.2617	a	0.1337
	(4.80)		(3.39)		(1.24)
S_t	0.9116	S_t	0.7721	S_t	0.7636
	(45.80)		(29.07)		(28.15)
		S_T	0.1621	S_T	0.1670
			(6.07)		(6.10)
				S_{T*}	0.0470
					(2.06)
ρ	0.062141	ρ	0.230017	ρ	0.200806
	(0.56)		(2.11)		(1.68)
DF	82	DF	82	DF	72
R-Square	0.9668	R-Square	0.9772	R-Square	0.9789

Tables 2 and 3 contain our regression results for PJM Western Hub electricity and Henry Hub natural gas, respectively, where t-statistics for our coefficients are shown in parentheses. With respect to the electricity regressions, current daily spot prices (S_t) are not statistically significant at the 5 percent level in any of the specifications.[22] Moreover, the coefficients for this variable range from 0.05 to 0.11, suggesting that changes in current spot prices have limited impact on forward electricity prices.

In contrast, the realized average spot price of electricity in the delivery month (S_T) is statistically significant, with a coefficient of about 0.5 (see

---

[21] The mean, minimum, and maximum prices reported in Table 1 suggest that the spot and forward price series are skewed, particularly for electricity. This is not unexpected given that the costs of electric generation vary greatly across plants. Electricity prices in low-demand periods may reflect the low incremental costs involved in operating "baseload" plants, while prices in high-demand periods reflect the substantially higher incremental costs in operating "peaking" plants, and possibly an additional scarcity (or market power) premium. Although the skewness of the prices suggests that the errors in our regressions may conform with a log-normal distribution, we chose to focus on price levels, rather than the log of price levels, in order to more precisely estimate the dollar-for-dollar relationship between forward and spot prices. Moreover, the distribution of our regression errors did not conform with a defined log-normal pattern.

[22] A t-statistic of 1.96 denotes significance at the 5-percent level (in a two-tailed test).

Specifications 2 and 3). Although actual spot prices during the delivery month cannot be observed at the time of the month-ahead forward sale, actual spot prices should be similar to expected spot prices on average. One might therefore view actual spot prices as a "crude" (*i.e.,* noisy) proxy for expected spot prices during the delivery month. If that is the case, futures prices adjust by approximately one-half of the increase in expected spot prices.

This suggests that future electricity prices may be substantially below expected spot electricity prices during periods when relatively high prices are expected. If true, this behavior would be consistent with the presence of normal backwardation, where sellers accept forward prices that are below risk-adjusted expected spot prices. However, before reaching this conclusion, different methods of estimating expected spot prices should be used other than using actual spot prices. We may merely be looking at a period where expected spot prices frequently were lower than actual spot prices.

Summarizing, without the ability to engage in inter-temporal arbitrage, current electricity prices principally should reflect current conditions. Instead, the market must rely on other information to form expectations regarding future conditions, and thus determine forward prices. Consistent with this notion and our prior hypothesis, our regression results do not indicate a strong association between forward prices and current spot prices for electricity. However, we do find a statistically significant relationship between forward prices and actual spot prices during the delivery month (where we use the actual future spot prices as a proxy for expected future spot prices).

To further illustrate this finding, it is likely that May spot electricity prices provide little information as to the weather conditions typically prevailing in June. To determine expected electricity spot prices in June, and hence month-ahead forward prices for June delivery, one must form expectations regarding June weather conditions (and other demand and cost factors). Past experience in June could prove helpful in this assessment, but our regression results find that the average daily spot price during the delivery month in the prior year (S_{T^*}) does not influence forward prices (see Specification 3).

As opposed to our electricity results, the natural gas results displayed in Table 3 confirm the importance of current spot prices in the formation of futures prices. Specification 1 illustrates that changes in current daily spot prices (S_t) have a nearly one-to-one relationship with changes in month-ahead futures prices. When actual daily spot prices during the delivery month (S_T) are added (see Specification 2), our findings suggest that month-ahead futures prices are more sensitive to changes in current spot prices

(with coefficient 0.77) relative to changes in actual spot prices during the delivery month (with coefficient 0.16).

As we previously hypothesized, this result is intuitively sensible since current and future prices should be aligned for a storable commodity such as natural gas. Given that current market participants can engage in inter-temporal arbitrage by choosing whether to sell now or later, current gas prices should reflect information on expected future supply and demand conditions.

Finally, Specification 3 suggests that futures prices for next-month delivery are potentially influenced by actual spot prices during that delivery month in the prior year (S_{T^*}). The coefficient for this variable, while statistically significant, is small in magnitude. One might expect this coefficient to grow in economic significance if the relatively new phenomenon of high winter gas prices persists.

6. CONCLUDING REMARKS

Several thoughts arise from our analysis of the forward-spot price relationship in electricity and gas markets. Our hypothesis that forward electricity prices are largely disconnected from current spot electricity prices appears supported by the data. This would be expected for a commodity such as electricity that is difficult to store, where current prices should depend on current cost and demand conditions while forward prices should depend on expected future cost and demand conditions.

At the same time, forward gas prices appear closely tied to current spot gas prices. This result is consistent with our hypothesis that these prices should be tightly connected, since natural gas is a storable commodity and market participants can engage in inter-temporal arbitrage by increasing or decreasing product inventories. Due to the ability to delay product sales, both current spot and forward prices should depend on similar information regarding current and future cost and demand conditions.

We hypothesized that electricity prices were potentially more volatile than gas prices, and the data generally supported this notion. The volatility was not only apparent through our examination of the forward-spot relationship based on "convenience yields," but it also was evident in the individual spot price series. While PJM electricity forward markets are relatively new and not nearly as liquid as Henry Hub gas futures markets, the limited storage opportunity for electricity may be a key contributor to this result. Since electricity is not easily stored, physical sellers into the forward market must hold claims on produced output in that future time period, as opposed to being able to source forward sales from stored inventory.

This may limit the set of potential participants in forward markets for physical sales, implying that the lack of product storability may bear directly on market liquidity. It is further possible that thinness in the physical forward market, which involves participants with information regarding physical market conditions, also would affect the depth of trading in any purely financial forward market.

Market thinness in forward electricity markets raises two possible adverse outcomes. First, it may lead to large deviations in expected future spot prices from realized spot prices. With market participants receiving different information signals, the market's assessment of future conditions may be more accurate when the number of market participants is larger. Secondly, market thinness may cause forward prices to deviate substantially from expected spot prices, as the different risk preferences of buyers and sellers result in a substantial forward price premium or discount. Two examples in PJM are consistent with these outcomes. In the summer of 1999, forward prices were substantially lower than observed prices in the spot market at the time of delivery. In the subsequent summer, forward prices were substantially higher than spot prices at the time of delivery.

Due to the possibility of contango, where forward electricity prices substantially exceed expected spot market prices, power procurement regulatory strategies need to be more carefully considered. The increasing use of auctions in some states to procure power well in advance of delivery could lead to increased power costs when forward markets are "thin." The limited number of participants in these auctions is consistent with this possibility. At the same time, the "least-cost" procurement policies favored by other state public service commissions can induce utilities to shun hedging strategies when electricity prices are most volatile. Also, as was suggested in certain analyses of the California electricity market, relying heavily on short-term power purchases may facilitate exercises of market power in concentrated markets.

Finally, trends in natural gas prices over time provide a contrasting example to electricity, but suggest increasing future volatility. The ability to store natural gas enables inter-temporal arbitrage, and prior to the winter of late 2000 and early 2001, gas spot and forward prices exhibited a relatively stable pattern, frequently lacking substantial seasonal effects. However, subsequent to that time, substantial price increases have occurred prior to and during the winter months. Storage constraints and increased demand for natural gas by wholesale electricity generators and weather-sensitive consumers may explain this developing seasonal pricing pattern. Thus, the forward-spot price relationship in gas markets is evolving in a manner suggesting that inter-temporal arbitrage is becoming more difficult.

REFERENCES

Allaz, Blaise, and Jean-Luc Vila. 1993. "Cournot Competition, Forward Markets and *Efficiency." Journal of Economic Theory* 59(1): 1-16.

Borenstein, Severin, James Bushnell, and Frank Wolak. 2002. "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market." *American Economic Review,* 92(5): 1376-1405.

Brealey, Richard and Stewart Myers. 2000. *Principles of Corporate Finance – Sixth Edition* Boston: Irwin McGraw-Hill.

Energy Information Administration. October 2003. *Natural Gas Market Centers and Hubs: A 2003 Update.* (available at http://www.eia.doe.gov).

Energy Information Administration. February 2004. *The Natural Gas Industry and Markets in 2002.*

Federal Energy Regulatory Commission. November 2000. *Staff Report to the Federal Energy Regulatory Commission on Western Markets and the Causes of the Summer 2000 Price Abnormalities.*

Green, Richard. 1999. "The Electricity Contract Market in England and Wales." *Journal of Industrial Economics* 47(1): 107-124.

Hicks, John R. 1946. *Value and Capital.* London: Oxford University Press.

Joskow, Paul and Edward Kahn. 2002. "A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market during Summer 2000." *Energy Journal* 23(4): 1-35.

Keynes, John M. 1930. *A Treatise on Money.* London: Macmillan.

Pindyck, Robert. 2001. "The Dynamics of Commodity Spot and Futures Markets: A Primer." *Energy Journal* 22(3): 1 -29.

Public Utility Commission of Nevada. 2002a. "Order." Docket No. 01-11029, March 29, 2002.

Public Utility Commission of Nevada. 2002b. "Order." Docket Nos. 01-11030, 01-11031, and 02-2002, May 28, 2002.

Chapter 7

Combinatorial Interlicense Competition

Spectrum Deregulation Without Confiscation or Giveaways[*]

Michael H. Rothkopf and Coleman Bazelon
Rutgers University and Analysis Group

1. INTRODUCTION

In the United States, the radio spectrum is, by law, the property of the public, but with varying degrees of private use rights. The radio spectrum is extremely valuable. For over three quarters of a century, the government has been making policy with the aim of having this valuable asset used in the public interest—a nebulous standard that has been subject to many different interpretations (U.S. Congress, CBO, 1997.) At present, some frequencies are reserved for government uses—defense, air traffic control, public safety, etc.—and some are licensed to companies for a variety of particular uses such as broadcasting and fixed and mobile communications. Almost all the valuable bands of spectrum—those that propagate well through walls, trees, and weather—have already been assigned for some use (Kobb, 2001.)

The current system of spectrum regulation is based largely on a command-and-control framework. The Federal Communications Commission (FCC) manages the allocation of private and state and local government uses of spectrum, while the National Telecommunications and Information Administration (NTIA) coordinates the federal uses of

[*] We are grateful for extremely helpful comments from Professors Sunju Park, Stephen A. Smith, and Timothy Brennan and from Michael Calabrese and members of his program at the New America Foundation including Troy Kravitz and J.H. Snider.

spectrum. For non-federal uses, traditionally the FCC allocates blocks of spectrum to types of uses, such as broadcasting or satellite, creates channel assignments and then assigns license rights to users. Licenses often specify where, when and how the licensee may use the radio spectrum. For instance, a typical television license will specify a transmitter tower location, height, power levels, channel assignment and broadcast technology.

Licenses initially were distributed on a first come basis. When more than one applicant wanted a particular license, the FCC was forced to choose among competing applicants. For most of its history it used comparative hearings, commonly referred to as beauty contests. This became an expensive and inefficient procedure and was replaced with lotteries in the 1980s. In 1994, the FCC began conducting auctions to assign licenses that had mutually exclusive applications. The FCC has pioneered innovative auction formats to assign rights to use radio spectrum. The assignments to date generally have been for bands of spectrum where either there were no significant incumbent licenses or there were clear rules for removing the incumbents.

1.1 Distributing Expanded License Rights

Currently, the FCC allocates spectrum on a licensed or unlicensed basis. Examples of licensed services are mobile telephone, broadcasting, and direct broadcast satellite. The licensee pays the government or promises to serve the public interest in return for use of the public airwaves. Examples of license-exempt services are cordless phones, garage door openers, Wi-Fi, and other consumer devices. On license-exempt bands, consumers share the spectrum without paying a fee to either the government or a licensee. This paper will not address the issue of when access to the spectrum should be under a licensed or unlicensed regime. Instead, we take the decision to expand the user rights in some currently licensed bands of spectrum as given and look to how those expanded, and hence more valuable, rights are distributed to private entities. An expanded right could free a broadcaster to cease broadcasting and offer a cellular phone service or allow a satellite operator to offer a terrestrial service. There is a general consensus at the FCC and among policy experts that the commercial use of spectrum should be largely deregulated, giving users far greater flexibility to determine the service provided on a band, or even to sell or sublease access to other firms through secondary market transactions.

Many interesting questions are raised in trying to define the scope and nature of the rights that should be attached to licensed radio spectrum. These range from fee simple property rights to time-limited, royalty-based rights leases. These are important questions, but this paper is agnostic with

respect to them. It is concerned with the method of distributing expanded rights, however they are defined.

There are at least two problems inherent in distributing expanded license rights in spectrum. First, there is a desire (or, at least, a political imperative) to respect the current use rights granted to current licensees, including the presumption a license will be renewed, even when those licensees received their licenses free. Indeed, the Communications Act of 1934 stipulates that licenses are temporary and confer no residual ownership interests. Second, both fairness and efficiency require that the government receive most of the value of the liberalization of the licenses. Since the right to use the spectrum for commercial purposes is worth hundreds of billions of dollars, the fairness aspect of a spectrum giveaway probably requires little comment beyond Senator McCain's observation that "They used to rob trains in the Old West. Now we rob spectrum" (Snider *et al.* 2003). However, the efficiency argument is subtler, and it is critical since the case for "privatizing" the spectrum is based upon efficiency.

The essence of the efficiency argument against a giveaway is that if the government fails to get full value for assets it gives away, the money it does not receive must be raised with taxes. There is a substantial economic literature documenting the marginal inefficiencies associated with raising money from income taxes.[1] A conservative estimate is that for every three dollars in federal revenue forgone (requiring, therefore, additional taxes to be raised) there is an additional dollar of lost productivity. Consequently, the added cost of the deadweight loss of raising government revenues—or worse, increasing the federal deficit—to compensate for lost spectrum revenue must be recognized as part of the price paid by the public when spectrum rights are given away.

Proposals exist to distribute spectrum relaxation rights. In the summer of 2002, the FCC established a Spectrum Policy Task Force (SPTF) with the mission to "provide specific recommendations to the Commission for ways in which to evolve the current 'command and control' approach to spectrum policy...." (U. S., FCC, 2002, p.1). In the end, the SPTF recommend that the Commission find a modest 100 MHz of spectrum below 5 GHz to transition from the current command and control regime to a market-managed regime based on flexible spectrum rights (*Op. cit.,* p. 51).

The SPTF does not recommend a specific process for distributing the expanded spectrum use rights, but two of the Task Force's members do. FCC senior economist Evan Kwerel and recently retired FCC senior

---

[1] While there may be unused opportunities to tax pollution or other externalities, these are likely to be relatively small, and the marginal source of tax revenue is the income tax. For details see Ballard *et al.,* 1985. Also see Fullerton, 1998 and Stuart, 1984.

engineer John Williams have proposed an auction to distribute rapidly significant amounts of spectrum relaxation rights, commonly referred to as the 'Big Bang' auction (U.S. FCC, Kwerel and Williams, 2002) Their proposal entices incumbents to put their existing spectrum license rights into the auction so that bidders will be able to bid on the full set of rights for a specific band of spectrum. Incumbent license holders are given three incentives to participate: first, they receive 100% of the auction receipts if their band is sold (or a prorated portion if the band is shared or combined with FCC reserve spectrum); second, if the band goes unsold, the licensee gets to keep the expanded rights for free; and third, the licensees get the right to match any competitive bid and thereby "buy-back," at zero additional cost, the expanded rights (thus discouraging others from competing for the rights). They propose auctioning bands totaling 438 MHz of spectrum under 3GHz. This ambitious auction proposal would likely distribute expanded use rights to incumbents for free or at far below their value. This is consistent with Kwerel and Williams' approach to spectrum management that focuses solely on the efficiency gains associated with distributing the expanded and valuable license rights to the largest amount of spectrum possible as soon as possible.

The likely low revenue outcome of the Big Bang proposal is driven by the presumed ability of incumbents to hold up the use of spectrum by new users. Hold up occurs when the incumbent can demand a disproportionate share of the benefits from the new, higher valued uses of a band of spectrum. By scaring away other bidders, the incumbent becomes the likely only bidder in many bands. It is a bit like trying to sell a valuable block of downtown real estate when someone has the right to have a lemonade stand on it. Who will offer to pay anything near its real value when the owner of the rights to the lemonade stand can block any potential use of the property? (This example is not contrived. The right to broadcast television on a UHF station in a major city where almost everyone who watches the station gets their signal over cable is probably worth a few percent of what the spectrum would be worth for mobile communications (Kwerel and Williams, 1992, and Snider, *et al.,* 2003.) Normally, if such downtown real estate were put up for competitive sale, the owner of the lemonade stand rights or someone in partnership with him would be the only serious bidder. With only one bidder, market forces could not be relied upon to set a price that comes anywhere close to the value of what is being sold. The purpose of this paper is to propose a way to overcome this difficulty.

1.2 Current Examples

To fill out the types of expansion rights we are proposing to be distributed, a few examples may be useful. The rights to be distributed by the proposed auction fall into two broad categories, both of which incumbent licensees can likely effectively block new licensees from using for new higher valued uses. The first is a filling out of the rights in currently licensed portions of spectrum. Examples of this type of expansion right would include:

Expanded rights for television licensees. These expanded rights would allow television broadcasters to cease television broadcasts and use their licensed spectrum for other uses.

Multichannel Multipoint Distribution Service (MMDS). MMDS licenses are for fixed wireless uses. Expansion rights would include allowing mobile uses in the band.

A second type of expansion right fills out licensing of currently unlicensed portions of bands allocated to private, licensed uses. Examples include:

Unused spectrum in the television bands. Broadcast television stations require that spectrum adjacent (both in spectrum and geography) to the licensee not be used. Therefore, channel 7 in Washington, DC prevents the use for broadcasting of channels 6 and 8 in DC and channel 7 in surrounding areas. The unlicensed portions of the television band could be licensed.

Unused spectrum in the fixed point-to-point microwave bands. Point-to-point microwave communications leaves much of the spectrum surrounding the links unused. This spectrum could be licensed. (Note that the PCS auctions were for the unused spectrum around existing links and the right to move the existing links to another band—containing features of both types of expansion rights.)

1.3 An Alternative: Interlicense Competition

We describe an auction procedure that can be used to sell relaxation rights that liberalize the use of spectrum while obtaining for the government the fair value of the licenses it is granting. The heart of the proposal is an adaptation of a procedure suggested by C. Bart McGuire and used in the early 1980s by the U.S Department of the Interior to auction coal rights to Federal coal tracts where the owners of adjacent coal deposits were the only logical bidders (U.S. Department of the Interior 1981). In the context of

coal, the approach was called "intertract competition." It made the bidders for different coal tracts compete with each other. This approach was authorized by Congress and evaluated favorably by the Commission on Fair Market Value Policy for Coal Leasing, chaired by David F Linowes (See Linowes, 1984) that was established by Congress to investigate a scandal that shut down the Department of the Interior's coal leasing in the early 1980's.

The proposal also draws on other ideas from the auction literature. One is to treat a constraint on the total amount to be sold as "soft." This idea dates back to discussions of "PURPA auctions" for electricity supply contracts (Rothkopf *et al.,* 1990). Since to be economic, new services may require combinations of current licenses, the proposal allows bids on combinations of licenses. (However, for ease of exposition, we explain first a noncombinatorial version that we first discussed in Rothkopf and Bazelon, 2003.)

Under this proposal, no licensee's rights will be damaged or limited in any way. However, under this proposal, no licensee or other party will get spectrum relaxation rights without competition. In particular, current licensees for a service that greatly under utilizes spectrum will have to compete with others to get their license restrictions eased even though they may be the only bidder for the particular rights that complement theirs. A critical point of this paper is that it is not necessary to give away spectrum rights in order to have the advantages of private ownership incentives.[2]

Section 2 presents the interlicense competition proposal, first in simplified form and then in a more complicated form that allows bidders to make offers for relaxation rights on combinations of licenses. Section 3 provides a discussion of the proposal, of implementation issues, and of its relationship to some specific concerns in spectrum management such as public interest obligations. An appendix gives a brief statement of the underlying mathematical problem.

2. INTERLICENSE COMPETITION

2.1 A Simplified proposal

Our interlicense competition proposal is first presented in simplified form and then in a more complicated form in which bidders can make offers on

---

[2] C.f., "Efforts to extract gains from licensees ... should not be permitted unduly to hinder or delay realization of the public benefits from promoting greater competitiveness through spectrum liberalization." Gregory L. Rossten and Thomas W. Hazlett, 2001, p.6.

relaxation rights on combinations of licenses, followed by discussion. This is a simple version of the interlicense competition proposal to expand spectrum license rights without either giving them away for much less than their value or forcing the holders of existing rights to release them (with or without compensation).

Under this simplified proposal, Congress will authorize the FCC to announce an annual or perhaps biannual series of auctions of "overlay" or "relaxation" spectrum rights. Each auction will relax the current regulatory constraints on a given amount of spectrum (measured in units of bandwidth times the population area covered, i.e., in MHz-Pops[3]) for essentially unrestricted use subject to responsibility for noninterference with licenses for other frequencies and other geographic areas as well as any existing license on the spectrum. However, the amount to be sold in a single sale will be a relatively small fraction, perhaps 10% to 20%, of the amount upon which bids will be accepted. While for national security, public safety, or other special purposes some spectrum may be excluded from bidding in these sales, "relaxation rights"[4] for most privately licensed spectrum will be eligible for sale and sold if the offer for it is high enough. Any currently licensed spectrum offered will be subject to the rights of the current spectrum license holder. Rights to currently unlicensed spectrum will also be included. For example, TV channel 2 in Washington would be included (but subject to the requirement that its use not interfere with channel 2 in Baltimore).

The current license holder may bid to relax the restriction on her license. Others may also bid for these relaxation rights, although other bidders may well be at a disadvantage relative the current rights holder. Similarly, the holder of the rights to TV channel 2 in Baltimore may have an advantage over other bidders for the currently unlicensed right to TV channel 2 in Washington. The auction will be a sealed-bid, market-clearing-price auction. In this simple version of the auction, there will be no combinatorial bids and spectrum with the highest bids per MHz-Pop will be sold up to the cut off limit on MHz-Pops for the sale. The important consequence of this is that a license holder wishing to relax the constraints on a license will have to

---

[3] The units here may be unfamiliar to some. Dollars per MHz per Pop is the same as dollars per MHz-Pop. Both refer to the per capita cost of 1 MHz of spectrum. However, MHz-Pops, which are appropriate here, refer to the amount of bandwidth (MHz) multiplied by the population in the geographic area of the license.

[4] We use the term "relaxation rights" to denote the right to ignore restrictions on a license other than interference with other licenses. Note, however, that we do not propose any diminution of the rights of the holder of the current license. Thus, if someone other than the current license holder were to win relaxation rights on a license, he would not be able to interfere with rights of the current license holder without bargaining for permission.

compete for the right to do so with holders of other licenses who also wish to relax the constraints on their licenses.

In this simple version of the auction, in order to select the winning bids the FCC will first rank order the bids with respect to the amount offered per MHz-Pop. Starting with the highest ranked bid, the FCC will award eligible bids that do not conflict with previously accepted bids until it reaches a bid that would put the total sold over the limit set in advance of the auction for MHz-Pops. This bid is the marginal bid and will set the price per MHz-Pop for all accepted bids (whether it itself is accepted or not). If accepting the marginal bid would make the total MHz-Pops sold exceed the announced target by less than a pre-announced tolerance percentage, the bid will be accepted. If accepting the bid would result in exceeding this tolerance limit, then the FCC will reject the bid. All bids offering a price per MHz-pop less than the marginal bid will be rejected. If the bid acceptance process ends without reaching the target number of MHz-Pops, then all bids that have not been rejected will be accepted and the price per MHz-Pop will be the minimum allowable bid.

Three theoretical results that can be readily demonstrated are worth noting. First, if the tolerance limit exceeds the size in MHz-Pops of the largest license, then the auction will always end with the acceptance of the marginal bid. Second, whether the auction ends with the acceptance of the marginal bid or its rejection, there are no price anomalies; all accepted bids offer higher unit prices than all rejected bids. Finally, whether the auction ends with the acceptance of the marginal bid or its rejection, the total value expressed in the accepted bids is the maximum possible for the number of MHz-Pops sold. Thus, this is essentially a market-clearing-price auction.

2.1.1 Interlicense Competition: More Design Details

In auction design, the devil is in the details. It is vital that a number of procedural details be set up correctly. Substantial deposits should be required of bidders, and there should be prompt payment by winners and prompt awards to them upon completion of the auction. If citizenship or other qualifications are required, bidders should be required to assert under oath at the time the deposit is made that they meet those requirements. All eligibility challenges except ones connected with criminal prosecutions for perjury should be limited to the period before the auction.[5]

Immediately after the auction, the FCC should return deposits on unsuccessful bids. Successful bidders will pay the remainder of the price of

---

[5] The purpose of this proposed procedure is to prevent competitors of the service to be offered by the new licenses from delaying their competition.

what they have won, and licenses will be awarded to them. If they fail to pay, they will be in default, should lose their deposits and get no rights, bankruptcy laws notwithstanding.

Before each periodic auction, the FCC will announce to potential bidders the geographic units (and their populations) that will be used and any frequencies that are not available. If some frequencies are available in some geographic regions but not others, this too will be announced. For simplicity, we will call the units the FCC announces "licenses." All frequencies not explicitly excluded will be available subject to the specific rights of existing licensees. The FCC will also announce the tentative target total number of MHz-Pops to be sold. Bidders should not be surprised by the announcement since a long-term plan for making frequencies available will have been adopted.

The FCC will also announce the deposit required from bidders per MHz-Pop of bidding eligibility. The deposit will be a substantial fraction of the anticipated price per MHz-Pop in the sale. It may also serve as the minimum bid per MHz-Pop, which should also be a substantial fraction of the anticipated price. In order to avoid a noncompetitive auction, after the deposits are received the FCC will, if necessary, announce a reduced total of MHz-Pops to be sold so that the amount to be sold is no more than some pre-announced fraction, say one-fourth, of the total eligibility.

Lower band and upper band relaxation rights should be sold in separate auctions, because not all MHz-Pops are the same. For example, lower frequencies that are suitable for mobile communications are more valuable than the upper frequencies (above 3 GHz) that do not readily propagate through walls, foliage and precipitation. It is important for the auction that bids be on the same basis so that they can be meaningfully compared. In addition, some further refinements in the $/MHz-Pops based on the frequency of the band in the bid may be considered useful.

Note that each of the periodic auctions can be treated as a one-time, sealed-bid auction. Hence, there is no need to restrict the bids to round numbers to prevent signaling. Since the bidders have the possibility and incentive to use lots of significant digits in their bids, ties should be exceedingly rare. If ties become common, collusion should be suspected. To discourage tacit collusion, bids at the exact same price should be treated as a single bid. If accepting this "bid" would result in too much spectrum being sold, all of "it" should be rejected. This also means that no rule is needed for resolving tie bids.

2.1.2 Example 1. Noncombinatorial Bids

Before going further, it may be useful to consider an example. Table 1 gives the highest nine of a large set of bids. For convenience, the bids have been numbered in decreasing order of bid amount

Table 1: A Set of Bids

Bid #	$/MHz-Pop	License #	MHz-Pops ($x10^6$)
1	6.0121	4321	60.2
2	5.8327	5432	43.5
3	5.7511	4321	60.2
4	5.6330	6543	12.7
5	5.5112	7654	44.0
6	5.5081	8765	32.6
7	5.0423	9876	25.8
8	4.8899	1234	10.4
9	4.8001	2345	10.9
Etc.			

Suppose that the government has announced that it will sell relaxation rights for **200 ($x10^6$) MHz-Pops** with a tolerance of 5%. In this case, it will accept bid 1 and bid 2. It will reject bid 3 because it has already sold the relaxation rights to license # 4321 to bid 1. It will then accept bids 4, 5, and 6. This brings the total MHz-Pops of accepted bids to **183.0 ($x10^6$).** Bid 7, if accepted, would bring the cumulative number of MHz-Pops of accepted bids to **208.8 ($x10^6$).** Since this is within 5% of the target of **200 ($x10^6$),** the bid will be accepted and its price will set the price of all accepted bids at $5.0423 per MHz-Pop. If the tolerance were only 2.5%, bid 7 would cover too many MHz-Pops to accept. It would be rejected, but it would still set the price for all accepted bids. For reasons discussed in the next paragraph, bids 8 and 9 would not be accepted even though accepting bid 8 would leave the total MHz-Pops sold below **200 ($x10^6$) MHz-Pops** and accepting bids 8 and 9 would leave the total at **203.3 ($x10^6$) MHz-Pops,** below 102.5% of that amount

In this example, accepting bid 8 or bids 8 and 9 after rejecting bid 7 would create two related anomalies. First, a bid has been rejected that would have offered a higher unit price than an accepted bid. Second, a bid offering a unit price below one in a rejected offer (viz. bid 7) would become the lowest accepted bid. This would create a dilemma. If the price remains the one set by bid 7, bidders 8 and 9 would have indicated an unwillingness to pay that much. If alternatively, these bids are accepted and allowed to set the price, they are lowering the price unfairly for all of the other successful

bidders. The effect of this second anomaly could be quite large if, for example, there was a very low bid, say \$0.01 per MHz-Pop, for relaxation rights on a license with just **1 ($x10^6$) MHz-Pops**. It could be accepted in addition to bids 8 and 9 and still leave the total of accepted bids below **205 ($x10^6$) MHz-Pops**. If it sets the price, it would essentially deprive the government of all revenue even though there is great demand for MHz-Pops.

2.2 Interlicense Competition with Combinatorial Bids

It is quite possible that the relaxation rights on an FCC license are worth more if the relaxation rights on other licenses are also obtained. This effect could be mild or it could be critical as when a proposed communication service would absolutely require the relaxation rights of more than one existing license. In addition, it is possible that relaxation rights on alternative sets of licenses would allow a proposed service. In such a situation, a bidder may well want to offer bids in the alternative on these sets of licenses. Finally, it is possible that bidders are capital limited and would like to limit their total expenditures in an auction. Thus, it is potentially quite useful to allow bidders to bid for combinations of relaxation rights rather than just for individual rights and to place constraints on their bids. However, allowing bids on combinations and such constraints makes selecting the winning bids more difficult, something we must deal with.

We cannot prove that we are dealing "optimally" with these complicated tradeoffs, but we can propose an auction form that, we argue, addresses such tradeoffs in reasonable manner. In the presence of the synergies that exist, it generally is better than allowing no bids on combinations. The basic idea is simple. Bidders may make combined bids for different licenses. For example, a bid may offer \$4.00 per MHz-Pop for license 1 and license 2. Because this is a combinatorial bid, it is not separate offers to buy either license 1 or license 2 at \$4.00 per MHz-Pop; it is just an offer for the combination.

The following example may be instructive. Two items of the same size, A and B are for sale. Bidder 1 values A at 3, B at 3, but the combination of A and B at 9 – 4.5 per unit for both items. In the absence of combinatorial bidding, her only safe course is to offer 3 for A and 3 for B. If she offers more, she risks winning one and not the other and suffering a loss. In the proposed auction, she can offer a bid of 3 for A, a bid of 3 for B and a bid of 9 for the combination of A and B. If the best competitive bids are 4 for A and 4 for B, her bid of 9 will win both. However, if the best competitive bids are 1 for A and 8 for B, only her bid of 3 for A will win. Note that if she had not bid 3 for A, her bid of 9 for A and B would have won. Thus, to

some extent, she is bidding against herself and will have incentives to act strategically.

Here is the specific auction we propose. It is a one-shot, sealed bid auction. In it, each bid is a price per MHz-Pop for the relaxation rights to a particular set of FCC licenses covering given geographical areas and ranges of frequencies. Each bid is made subject to several possible constraints by the bidder. The first such constraint is the eligibility constraint. Based upon her deposit, each bidder is limited to a given number of MHz-Pops. This does not constrain the number of MHz-Pops that she can bid upon, but it does constrain the number she can win. She may, if she chooses, make her deposit sufficiently large so that this constraint will not be binding.

The second kind of constraint is a budget-like constraint specified by the bidder to apply to the sum of her successful bids (not, however, the market-clearing price that the bidder would have to pay). This constraint prevents her from winning relaxation rights that will, in total, cost more than she is willing to spend. Its use by a bidder is voluntary. Bidders would probably prefer to have a budget constraint on actual expenditures, but this might cause computational difficulties.[6] Since budget constraints are often "soft" constraints, this kind of budget constraint may prove useful. It would allow bidders freedom to bid on many different alternatives. In its absence, a bidder might find its management imposing a more drastic "exposure" constraint, i.e., a constraint on the total of all of its bids, as is common in simultaneous sealed-bid auctions for offshore oil rights. (See Rothkopf 1977.)

The third kind of constraint is an "exclusive or" or alternative constraint. A bidder can always make two bids mutually exclusive by including the same license in both. Thus, if a bidder has a bid for relaxation rights for licenses A and B, and another bid for relaxation rights for licenses B and C, both bids cannot win since relaxation rights on license B can only be sold once. Thus, the bids will, in fact, be treated as bids for {A and B} or {B and C}. In addition, the FCC should allow each bidder a limited number of "pseudo-items" it can bid on.[7] (Such a "pseudo-item" is essentially a license of no value that only the bidder to whom it is assigned is allowed to bid upon.) By including such a pseudo–item in two different bids, a bidder can make the two bids mutually exclusive even though they don't actually overlap. Thus, for example, a bid for A, B and pseudo-item 1 is in conflict with a bid for C, D and pseudo-item 1. The use of such alternative

---

[6] Such a constraint, if binding, would make the linear programming problems we solve into linear complementarity problems.

[7] The idea for doing this is due to Fujishima *et al.* 1999.

constraints will allow bidders to attempt two or more independent ways to reach a given goal.

In addition to these bidder-specific constraints, the selection of winning bids is constrained by the (soft) limit on the number of MHz-Pops to be sold and by the requirement that the relaxation rights on each license either be sold once or not sold at all. We propose that subject to all of these constraints, the FCC select the bids that maximize the value expressed in the bids it accepts.[8] This mathematical problem is stated in the Appendix. It is an integer programming problem, which implies that it is in a class of problems that are potentially computationally difficult.[9] However, just as in the simplified problem discussed above, we avoided exact solution of the problem initially faced, we plan to avoid this problem here. As discussed below, we plan, instead, to solve a series of linear programming problems. Unlike integer programming, linear programming problems are not in the potentially computationally difficult class of problems.

One computational concern deserves special mention. With this auction form, bidders may have little to lose by submitting many slightly different conflicting bids. If the FCC anticipates the total number of bids posing a computational problem, it can require a nominal processing charge with each bid. This will not inhibit any serious bids, but could head off computational problems. In addition or instead, the FCC could impose a generous limit on the total number bids a bidder could submit. This would let a bidder express all of her important values.[10]

---

[8] An alternative, which we do not endorse, would be for the FCC to select the set of bids that would maximize the revenue it receives. Doing so could provide strong incentives for undesirable strategic bidding. In addition, it might lead to the FCC rejecting bids in order to increase the revenue from the sale by preventing the marginal price from falling. We believe that the public will be served best if the FCC makes and sticks to an overall judgment on the best pace at which to release spectrum from regulation taking into account both the efficiency gains from public revenue (which will remove the need for an equivalent amount of taxation) and the ability of industry to finance and make available services to the public and the public's readiness to make use of these new services.

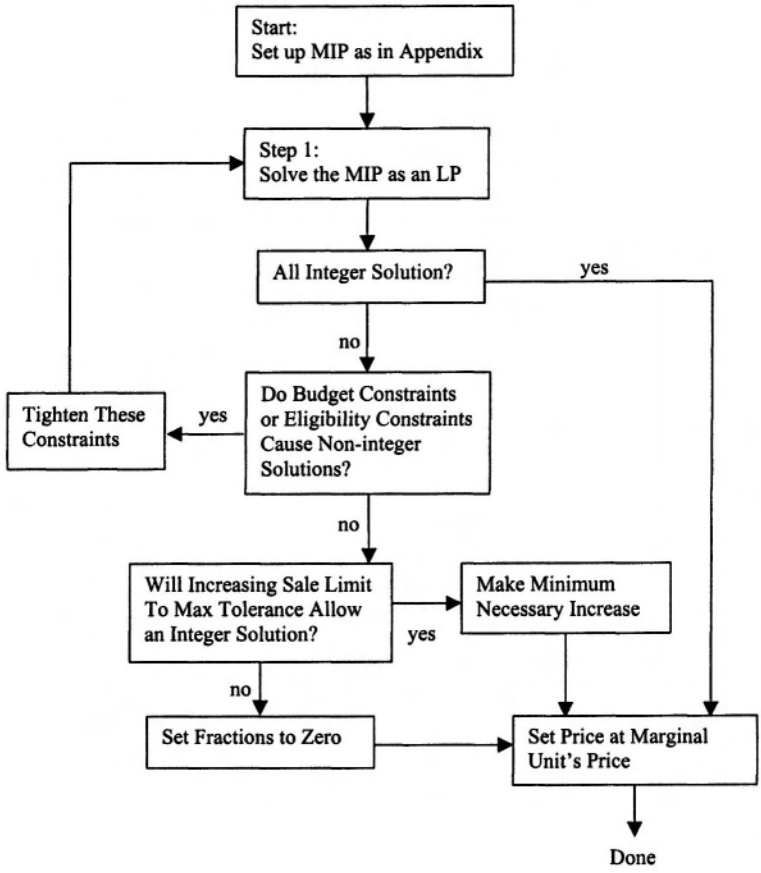
[9] Some mathematical problems have solution algorithms that even in the worst case grow in length as the size of the problem is increased no more than a given polynomial bound. Such problems are usually considered computationally easy. On the other hand, some problems have no known algorithm that is guaranteed for the worst case to grow no more than a polynomial bound as the size of the problem increases. Large instance of these problems are potentially unsolvable. Integer programming is in the latter class of problems. See Rothkopf, Pekec and Harstad 1998 for a discussion of this in the context of combinatorial auctions.

[10] See Park and Rothkopf 2004 for a discussion of the effect of limiting the number of combinations a bidder in an FCC auction can make and for a report of a related experiment in which economists who have advised bidders in FCC auctions bid against each other.

2.2.1 Selecting the Winning Bid

We will now describe the general computational procedure for selecting the winning bids and setting the market-clearing price for a given set of bids. Figure 1 illustrates this process. After we have done this, we will give a simple example.

Figure 1
Bid Evaluation Algorithm



The solution procedure begins by solving the integer programming problem given in the Appendix as a linear program. That is, the problem of maximizing value reflected in accepted bids is solved ignoring the constraints that force variables to take on integer values. We will call this Step 1. If this calculation happens to find a solution in which all of the variables are, in fact, integers, the solution also solves the integer

programming problem and is accepted as the solution to the bid acceptance problem. The lowest unit price (in $ per MHz-Pop) of any accepted bid is used to set the unit price of the relaxation rights.

If, as is quite likely at first, some integer constraint is violated, the procedure then goes on to modify the problem. This is Step 2. If the budget constraint of a bidder is binding and this results in the proposed sale to that bidder of a fraction of the relaxation right on a license, that budget constraint is tightened to eliminate the fractional purchases. If more than one bidder is in this situation, all of their budget constraints are tightened so that none of them are buying a fractional part of a relaxation right. Similarly, if the eligibility constraint of one or more bidders is binding and this results in the proposed sale to that bidder of a fraction of a relaxation right on a license, the eligibility constraints of those bidders are tightened to eliminate the fractional purchases. It is appropriate to make all of these changes simultaneously since reducing the fractional purchase of one bidder with a binding budget or eligibility constraint, will not eliminate a fractional purchase by any of the others. The calculation then returns to Step 1.

If no budget or eligibility constraint results in the purchase of a fractional relaxation right, then in Step 3 the calculation checks to see if relaxing the constraint on the total MHz-Pops to be sold, but not beyond the pre-announced tolerance limit, will result in the sale of all of those rights and do so without violating a budget-like constraint or an eligibility constraint. If so, the constraint on the total MHz-Pops to sell is relaxed in order to make that sale. The calculation is then complete, all of the bids in the optimal solution are accepted, and the price per MHz-Pop for all sold rights is set by the price of this marginal bid. If relaxing the constraint on the total number of MHz-Pops to the tolerance limit results in the violation of the budget-like or eligibility constraint of the bidder who made the offer on the marginal license(s), that offer is eliminated and we return to Step 1. If no budget or eligibility constraint is violated but the maximum relaxation still leaves the marginal offer or offers only partially filled, then all of the marginal offers are rejected, but their unit price is used to set the price for all accepted bids, and all of the other bids in the optimal solution without the relaxation are accepted. In effect, if there are multiple bids at exactly the same unit price on the margin, these bids are treated as a single bid. If this single bid fits within the tolerance limit, all of its components are accepted. If not, all are rejected. The reasons for rejecting the marginal bids and not going on to lower bids is the same as the reasons discussed above in the context of the simplified auction. The reason for rejecting all marginal bids if they collectively exceed the tolerance limit is that bidders are free to use many significant digits in their bids. Thus, bids by separate bidders offering the exact same price are suggestive of collusion, and rejecting such bids is a

good idea. Bidders can avoid equality in their own bids by adding or subtracting a different tiny amount to each bid, so there should be no problem in rejecting bids from one bidder with the same unit price. This also eliminates any arbitrariness in dealing with ties.

Readers may wish to note that since each time the procedure returns to Step 1 at least one bid is permanently discarded, the number of linear programming problems that must be solved is bounded above by the number of bids. Since the worst case bound for computational effort for solving linear programming problems is polynomial, so is the worst case bound on the total amount of computations involved here if the number of bids is limited. Worst-case bounds are usually conservative. In this case, they are likely to be extremely conservative. The reason is that what is being sold, in most cases, is relaxation rights on existing licenses. The value of these rights should be higher to the holder of the existing license or to someone with whom she strikes an exclusive deal. Hence, competitive conflicting combinations from *different* bidders should be rare.

2.2.2 Example 2. Combinatorial Bids

We now present a highly simplified illustrative example. It involves relaxation rights on the 18 existing licenses shown in Table 2.

Table 2: Licenses for the Example

License #	MHx-Pops ($x10^6$)
1	60.2
2	43.5
3	60.2
4	43.5
5	44.0
6	32.6
7	25.8
8	37.4
9	10.9
10	48.7
11	30.8
12	51.9
13	10.2
14	8.7
15	18.3
16	43.8
17	82.0
18	64.2

These licenses have a total of 715.7 million MHz-Pops. We will assume that the sale will try to sell 170 million, approximately one quarter of them, and have a 10 percent tolerance, thus allowing sale of up to 187 million MHz-pops. We assume that there is a minimum bid of $.01 per MHz-Pop, and we further assume that there are 10 potential bidders. The auction will be of the kind just described. Here is a description of the bidders' situations and of their choices. Note that the auction is essentially a market-clearing-price auction. This means that, except for bidders desiring a large number of items, there is little chance that a bidder's bid will affect the single price that prevails in the auction. (This will tend to be much a more realistic assumption in a large, practical situation than in small illustrative examples such as this.) Thus, in practice, bidders will generally not need to spend much effort on strategizing. Rather, they are likely to do relatively well if they bid their true values. In the example, all but bidder 6 (who wants to relax the restrictions on a lot of spectrum) do this.

Bidder 1 controls licenses 3 and 4 and would like to change the service offered on them. Doing so on 3 alone is feasible, but doing so on 4 alone is not. His values for relaxation rights are $100 million for 3 and $125 million for licenses 3 and 4. He will make two bids: $100 million for 3, and $125 million for 3 and 4.

Bidder 2 has conditional deals with the holders of licenses 6, 7, and 8. She needs to get rights to just one of these to provide a new service. Her value is $25 million for 6, $20 million for 7, and $30 million for 8. She will make bids in these amounts for the licenses and a "pseudo-item," P2, so as to be sure not to win more than one.

Bidder 3 controls licenses 1 and 2. She currently has no plans to change the service she is offering on them, but would like to lock in future flexibility. She is willing, independently, to pay $10 million for relaxation rights on 1 and $8 million for relaxation rights on 2 and will make separate bids of these amounts.

Bidder 4 is a "bottom-fishing" speculator. She controls no licenses. She decides to bid $1 million each on licenses 1 through 3, 5 through 8, and 10 through 12, but she wants to be sure not to spend more than $3 million dollars. Therefore, she will link her ten bids by a budget-like constraint of $3 million. To avoid having her bids for licenses 1 and 3 (which happen to cover the same number of MHz-Pops) tied, and thus treated as linked, she will add $0.01 to her bid for license 1.[11] In addition, she can only raise enough up-front money to cover deposits for 170 MHz-pops. Hence, she

[11] For simplicity of exposition, we are using round numbers for most bids. In practice, bidders would have an incentive to avoid round numbers to avoid unintended ties and to make their bids unpredictable to their competitors.

realizes that if she were high bidder on licenses 1, 3 and 12, her bid on license 3, the one with the lowest bid per MHz-pop, would be rejected. She also notes that her bids, while low, meet the minimum bid requirement.

Bidder 5 controls license 5, is willing to pay $8 million dollars for relaxation rights, and will bid this amount.

Bidder 6 controls licenses 10, 11, and 12. He wants to offer a service that will require two of the three licenses and would benefit from the third. His values are $100 million for 10 and 11, $130 million for 10 and 12, $110 million for 11 and 12, and $150 million for all three. He submits four conflicting bids. The three bids for pairs of licenses reflect his values. However, because the bid for all three licenses covers 131.4 MHz-pops, he thinks it has a significant chance of being the marginal bid. Hence, he shades his bid on this combination and offers only $140 million for it.

Bidder 7 want licenses 13, 14, and 15. She can pay $1.02 per MHz-pop for each of them and can pay a slight premium, $1.04 per MHz-pop if she gets all three. She bids this unit price plus differing small amounts on the three licenses and a unit price of $1.04 on the three of them.

Bidder 8 and ***bidder 9*** both have licensed spectrum that abuts currently unlicensed license 16 and both want license 16. Bidder 8 offers $70.3 million for it, and bidder 9 offers $81.2 million for it.

Bidder 10 wants to keep his options open on his operation on the complement of license 17. He doesn't particularly want to buy it, but it will bid $20 million to make sure that someone doesn't win it (cheaply) and then be able to block his future plans.

No one bids on licenses 9 and 18. Table 3 shows the bids.

The calculation to determine the winner proceeds directly to Step 3 since no eligibility or budget constraints are binding. In it, bid 28 for license 16 by bidder 9 is honored. This sells 43.8 MHz-Pops. Bid 1 for license 3 by bidder 1 is also honored. This brings the total MHz-Pops sold to 104.0. Bid 21 by bidder 6 for licenses 11 and 12, if honored would bring the total MHz-Pops sold to 186.7. Since this fits (barely) within the tolerance limit of 187 MHz-Pops, it too is honored. Since it brings the total number of MHz-Pops sold to more than 170, it is the marginal bid and sets the price at $1.33/MHz-Pop. Thus, bidder 9 pays $58.254 million for the 43.8 million MHz-Pops of license 16, and bidder 1 pays $80.066 million for the 60.2 million MHz-Pops of relaxation rights on license 1. Bidder 6 pays his bid of $100 million for licenses 11 and 12 and their 82.7 million MHz-Pops. Thus, the sale takes in $238.32 million for the 186.7 million MHz-Pops of licenses 3, 11, 12 and 16. All unsold relaxation rights will be offered again in next year's auction.

Table 3: Bids on Licenses in Table 2

Bid #	Bidder	Licenses	Amount ($10^6$$)	$/MHz-Pop	Budget Constraint
1	1	3	100	1.661	–
2	1	3,4	100	1.205	–
3	2	6,P2	25	0.767	–
4	2	7,P2	20	0.775	–
5	2	8,P2	30	0.802	–
6	3	1	10	0.166	–
7	3	2	8	0.184	–
8	4	1	1+	0.017	B4
9	4	2	1	0.023	B4
10	4	3	1	0.017	B4
11	4	5	1	0.023	B4
12	4	6	1	0.031	B4
13	4	7	1	0.039	B4
14	4	8	1	0.027	B4
15	4	10	1	0.021	B4
16	4	11	1	0.032	B4
17	4	12	1	0.019	B4
18	5	5	8	0.182	–
19	6	10,11	61.3	1.258	–
20	6	10,12	62.9	1.292	–
21	6	11,12	41	1.330	–
22	6	10,11,12	55.6	1.065	–
23	7	13	10.4	1.020	–
24	7	14	8.9	1.020+	–
25	7	15	18.7	1.020++	–
26	7	13,14,15	39.1	1.050	–
27	8	16	70.3	1.605	–
28	9	16	81.2	1.854	–
29	10	17	40	0.488	–

Note that because this is a simplified example involving only 18 licenses and 29 bids, some bidders had reason to think that their bids had a serious chance of being the marginal bid. However, in a large auction involving hundreds or thousands of licenses, shading a bid significantly from value in order to have a positive gain if it is the marginal bid will be an unattractive option for a bidder. The chance that a significant shading will add to profit will be dwarfed by the chance that it will lead to a profitable bid being rejected. Thus, in such large auctions, shading should be minimal and bids would thus approximate closely bidders' values. As noted above, a more significant incentive issue may involve bidding on combinations and on

parts of the combination. A bid on a part of a combination might combine with a competitor's bid for the other part to best a bidders bid on the entire combination. This could involve some significant strategizing by bidders, but it is not clear that any workable proposal could avoid this. In particular, Vickrey-Clarke-Groves procedures, which would do this in theory under some circumstances, are not practical. See Hobbs *et al.* 2000 and Sakurai *et al.* 1999.

3. POLICY DISCUSSION

Spectrum is a highly valuable public asset. There are strong arguments that U.S. spectrum is badly under used and over restricted and that a licensing system based upon expanded and flexible use rights would work better. While there is a legitimate need to protect temporarily non-licensees who have invested in equipment – such as owners of television sets – the overriding picture is one of misallocation and use of administrative procedures to block competition. The proposal in this paper would gradually make spectrum available on a property-rights-like basis. We believe its gradual nature is an advantage. It will take time for capital markets, and physical ones, to adapt, and non-licensee purchasers of equipment will have a chance for their past equipment purchases to be depreciated. A unique and important advantage of this approach is the use of competition rather than an administrative determination to decide which spectrum is freed up first. This will tend to assure that the spectrum first released from usage restrictions goes to meet the most pressing unmet needs.

One interesting perspective on spectrum rights comes from natural resource management. There is a long tradition in U.S. natural resource management of preventing speculative holding of publicly owned resources. This is often done through diligence requirements. Of course, one important difference between land or minerals and radio spectrum is that the lost value from unused spectrum is lost forever—it is a nondepletable natural resource. Nevertheless, there is precedent for the government being the custodian of a natural resource and holding on to ownership (in this case, the relaxation rights to spectrum) until the resource can be used productively.

In choosing an auction mechanism, the government faces two competing goals. On the one hand, the sooner a fuller set of spectrum rights are in private hands, the sooner they can be put to use (within the constraints on the ability of that spectrum to be used productively) with the concurrent increase in consumer welfare. On the other hand, the government wants to receive compensation for the public in return for distributing the valuable relaxation rights to the spectrum. Unfortunately, these two goals are somewhat in

conflict. That is, increasing the supply of relaxation rights decreases per unit prices the government will receive. Ideally, this trade-off is solved by transferring the relaxation rights to the private sector at a pace that equates the marginal cost to society in lost service from holding back a little more of the relaxation rights with the marginal cost to society of lost government revenues from slightly increasing the pace that the relaxation rights are distributed. That, rather than giving away the rights, is the efficient course.

The above trade-off illuminates an essential difference between the approach taken in this paper and the one proposed by Kwerel and Williams. Their approach does not consider the opportunity cost of government revenues. Supporters of their proposal may think this cost is not relevant for the analysis of the efficient use of spectrum license rights, or they may believe that the optimal trade-off between revenues and the speed of distribution of expanded license rights falls heavily on the side of the distribution of those rights. A third possibility that may apply to some supporters of the big bang approach is that they primarily care about reducing the size and scope of government by stripping it of resources. In any case, we disagree. As noted earlier, the marginal cost of a lost dollar of governmental revenue has been estimated conservatively at $0.33. This implies that the measured inefficiencies in the use of spectrum from slowing the pace of distribution of relaxation rights can get as high at 33 percent before they outweigh the revenue enhancing effects of that slower pace of spectrum rights distribution.

In the rich context of spectrum auctions, what is optimal is not known. However, while the auction we propose is not "optimal," it is reasonable. It should prove to be workable for fairly large auctions. It should allow bidders to represent important synergies. It should give good incentive signals to bidders whenever the chance that a given bid will be the marginal one is small. It should be relatively resistant to collusion. It should work particularly well in a situation where, for each license, one party already has the ability to prevent others from productively using the relaxation rights and thus is the only party bidding for those rights.

In general, the process should tend to pick out to sell first the most valuable rights. No administrative determination will be needed. Nonetheless, critical spectrum that should not be offered can be protected. The auction allows combinatorial bidding. Nonetheless, computational problems are avoided by placing mild constraints on the bidding. It can be argued that there may, in some cases, be problems because holders of different licenses need to cooperate in order to bid on useful combinations. To the extent that this is a problem, the existence of the auction should tend to ease it by imposing a credible deadline for agreement. Two further observations about this are apt. First, to the extent that these combinatorial

auctions are insufficient to solve the problem, the whole process of putting unrestricted licenses into private hands is called into question. The rational for this process rests on Coasian arguments that the market will make adjustments (Coase, 1959). To the extent that these Coasian arguments fail, progress may indeed require the FCC to clear unwilling incumbents. (Note that the big bang proposal of Kwerel and Williams aims to reduce this problem by giving the incumbents a strong inducement to put their spectrum at risk in the auction. That approach, however, does not eliminate the hold out problem and does not solve all coordination problems.) Second, if there are indeed problems involving specific spectrum rights that the private agreements envisioned by Coase will not solve, the existence of the auction should highlight this fact and focus regulatory attention on the management of this particular spectrum. The proposed process is independent of the application of the funds it generates. If desired, some of the funds can be used by the government to compensate for public interest obligations that are no longer present. Moreover, the interlicense process is also neutral with respect to the duration of rights auctioned. Congress could determine that the auctioned rights are permanent or could determine that a spectrum user fee (or lease fee) should attach after the initial license term.

In the past, the FCC has received less for some licenses than it might have because independent companies formed coalitions before entering the auction. This happened to an extreme extent in some European spectrum sales. Hence, it is tempting to suggest that legislation enabling the auctions should protect competitiveness by restricting joint bidding, not just by coalitions formed after bid deposits have been made, but also by joint ventures formed after the legislation is introduced. However, some new uses of spectrum may well require rights held by different parties. In such cases, coalition formation is natural and can be helpful. The solution is for the FCC to limit the amount of spectrum to be sold so that there is a high "eligibility ratio," — i.e., there are four or more serious bidders for each license that is to be sold. This should ensure that there is serious competition even in the face of coalitions and discourage coalitions that would pay off only by reducing competition in order to lower prices. High eligibility ratios will also reduce the incentives of bidders to strategize by bidding below their values.

There is significant political opposition to giveaways and many years of advocacy for liberalization of spectrum restrictions has had only modest results. Hence, we believe that those favoring such liberalization could gain political traction towards their goal by supporting this proposal and gaining the support of those opposed to giveaways.

In summary, with interlicense competition no licensee's current rights will be damaged or limited in any way, but no licensee or other party will get

spectrum rights without serious competition and some payment back to the public. In particular, those with rights for a use that greatly under uses spectrum will have to compete with others to get their license restrictions eased even though they may be the only bidder for the particular rights that complement theirs. It is both inefficient and unnecessary to give away spectrum rights in order to have the advantages of private ownership incentives and completely flexible license rights.

APPENDIX

This Appendix presents a mathematical formulation of the optimization problem discussed above. It assumes that all bids below the minimum allowable bid have already been deleted. It also assumes that only authorized bids on pseudo-items are included and that the size of each pseudo-item is defined as 0.

Let i index licenses, j index bids, k index bidders, and c index combinatorial bids. If bidder k makes bid j on license i, let b_{ijk} be the price per MHz-Pop offered by that bid. Let x_{ij} be 1 if bid j covers license i, and 0 otherwise; and let X_{jk} be 1 if bid j is by bidder k, and 0 otherwise. Let s_i be the size of license i measured in MHz-Pops. Let y_j be the fraction of bid j that wins; these are the decision variables over which we are optimizing. Let S be the number of MHz-Pops that are scheduled to be sold. Let E_k be the eligibility of bidder k. Let B_k be the budget-like limit of bidder k (infinite if bidder k specifies no budget-like limit).

The optimization problem is

$$\textbf{Maximize} \qquad Z = \Sigma_{ijk} b_{ijk} s_i x_{ij} y_j$$

$$\textbf{Subject to} \qquad \Sigma_{ij} s_i x_{ij} y_j \le S,$$

$$\Sigma_j x_{ij} y_j \le 1, \text{ for all } i,$$

$$\Sigma_{ij} s_i x_{ij} X_{jk} y_j \le E_k, \text{ for all } k,$$

$$\Sigma_{ij} s_i x_{ij} X_{jk} b_{ijk} y_j \le B_k, \text{ for all } k,$$

$$0 \le y_j \le 1, \text{ for all } j,$$

$$y_j \text{ integer, for all } j.$$

The objective function is to maximize the total value of the accepted bids. (Since bidders pay the market-clearing price, not their bids, the objective does not necessarily correspond to maximizing government revenue.) The first constraint assures that the relaxation rights to no more

than the allowed number of MHz-Pops is sold. The second set of constraints assures that the relaxation rights to no license is sold more than once. The third set of constraint assures that no bidder exceeds her eligibility. The fourth set assures that no bidder exceeds her budget. The fifth set of constraints assures that the fraction of each bid that is accepted lies in the range [0,1]. (Because of the second set of constraints, the upper bound in this set of constraints is redundant.) The final constraint changes what would otherwise be a linear programming problem into an integer programming problem by forcing the fraction of each bid accepted to be either 0 or 1.

REFERENCES

Ballard, Charles L., John B. Shoven and John Whalley. 1985. "General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States." *American Economic Review* 75: 128-138.

Coase, Ronald H. 1959. "The Federal Communications Commission." *Journal of Law and Economics* 2: 1-40.

Fullerton, Don. 1988. "If Labor is Inelastic, Are Taxes Still Distorting?" Working Paper. University of Virginia. Charlottesville, VA.

Fujishima, Y., K. Leyton-Brown, Y. Shoham. 1999. "Taming the Computational Complexity of Combinatorial Auctions: Optimal and Approximate Approaches." *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI).* Stockholm, Sweden: 548-553.

Hobbs, B.F., M.H. Rothkopf, L.C. Hyde, R.P. O'Neill. 2000. Evaluation of a Truthful Revelation Auction for Energy Markets with Nonconcave Benefits. *Journal of Regulatory Economics* 18(1): 5-32.

Kobb, Bennett Z. 2001. *Wireless Spectrum Finder: Telecommunication, Government and Scientific Radio Frequency Allocations in the U.S. 30 MHz to 3000 GHz.* New York: McGraw-Hill.

Kwerel, Evan R. and John R. Williams. 1992. "Changing Channels: Voluntary Reallocation of UHF Television Spectrum." U.S. Federal Communications Commission. Office of Plans and Policy. OPP Working Paper No. 27.

Kwerel, Evan R. and John R. Williams. 2002. "A Proposal for Rapid Transition to Market Allocation of Spectrum." U.S. Federal Communications Commission. Office of Plans and Policy. OPP Working Paper No. 38.

Linowes, David F., Chairman. 1984. *Report to Congress: Commission on Fair Market Value Policy for Federal Coal Leasing.* Washington, D.C.

Park, Sunju, and Michael H. Rothkopf. 2004. "Auctions with Endogenously Determined Biddable Combinations." *European Journal of Operations Research* forthcoming.

Rossten, Gregory L. and Thomas W. Hazlett. 2001. "Comments of 37 Concerned Economists." Comment on WT Docket No. 00-230. Federal Communications Commission.

Rothkopf, Michael H. 1977. "Bidding in Simultaneous Auctions with a Constraint on Exposure." *Operations Research* 25: 620-629.

Rothkopf, Michael H. and Coleman Bazelon. 2003. "Spectrum Regulation without Confiscation or Giveaways." Comment in the Matter of Issues Related to the

Commission's Spectrum Policies, ET Docket No. 02-135. Federal Communications Commission. January 9.

Rothkopf, Michael H., Edward P. Kahn, Thomas J. Teisberg, Joseph Eto and Jean-Michel Nataf. 1990. "Designing Purpa Power Purchase Auctions: Theory and Practice." In *Competition in Electricity: New Markets and New Structures,* edited by James Plummer and Susan Troppmann. Arlington, VA: Public Utilities Reports, Inc.

Rothkopf, Michael H., Aleksandar Pekec and Ronald M. Harstad. 1998. "Computationally Manageable Combinational Auctions." *Management Science* 44: 1131-1147.

Sakurai, Y., M.Yokoo, S. Matsubara 1999. "An efficient approximate algorithm for winner determination in combinatorial auctions." *Proceedings of the Second ACM Conference on Electronic Commerce (EC-00).* 30-37.

Snider, J.H., Michael H. Rothkopf, Bennett Kopp, Nigel Holmes, and Troy Kravitz. 2003. "The Citizen's Guide to the Airwaves." Washington, DC: The New America Foundation, July.

Stuart, Charles. 1984. "Welfare Costs per Additional Tax Dollar in the United States." *American Economic Review* 74: 352-362.

U.S. Congress. Congressional Budget Office. 1987. *Where Do We Go From Here? The FCC Auctions and the Future of Radio Spectrum Management.* 105th Congress. Washington, DC: The Congress of The United States.

U.S. Department of the Interior. Office of Policy Analysis. 1981. "Improvements to the Federal Coal Leasing Program Linked to the Use of Intertract Bidding." Report. April.

U.S. Federal Communications Commission. Spectrum Policy Task Force. 2002. "Spectrum Policy Task Force Report." ET Docket No. 02-135. November.

This page intentionally left blank

Chapter 8

Energy Trading Strategies in California
Market Manipulation?[*]

Michael DeCesaris, Gregory Leonard, J. Douglas Zona
Cornerstone Research

1. INTRODUCTION

There has been extensive press coverage of the trading strategies employed by Enron and others in the California energy markets during 2000 and 2001. These strategies first became widely known in May of 2002 with the public release of the so-called "Enron memos" as part of the FERC's fact-finding investigation of Western energy markets.

Most of the press on the strategies has been based solely on the limited descriptions contained in these memos, and there has been relatively little in-depth study from an economist's perspective. The strategies have proven hard to unravel and, to date, no party has successfully and accurately quantified their aggregate impact on Western energy markets.[1] One problem is that for most of the strategies, the potentially prohibited transactions

---

[*] The authors are, respectively, Associate, Manager, and Senior Advisor at Cornerstone Research, an economic and litigation consulting firm. The views expressed in this paper are solely those of the authors and do not represent those of Cornerstone Research or any client. An earlier version of this paper was presented at CRRI – Rutgers University's 22nd Annual Eastern Conference, May 21-23, 2003.

[1] The Cal ISO Department of Market Analysis has attempted to quantify these trading strategies in a couple of separate analyses but noted that "it is virtually if not absolutely impossible to disentangle the effects of the various strategies engaged in by disparate sellers in order to assign discrete market effects...", (CA ISO DMA 2003, p. 3).

cannot even be reliably identified and separated from legitimate transactions that have a similar signature in the data.

In large part "informed by the types of behavior that had been observed in the Western markets during 2000 and 2001" FERC issued an order conditioning market-based rate authority on compliance with six rules of market behavior (FERC Order 11/17/03). Indeed, this order is striking in how tailored it is to the specific trading strategies employed in CA. For purposes of this paper, we are mainly concerned with the second market behavior rule, which prohibits market manipulation generally (2), the submission of false information (2b)[2], and the creation and relief of artificial congestion (2c).

This paper aims to describe the Enron-style trading strategies, along with other variations brought to light in the investigation, and analyze them in the context of market manipulation and market efficiency. We find that there is an inherent tradeoff involved between the desire to ensure that markets are free of manipulation and the desire for efficient markets.

In terms of market manipulation, we attempt to apply FERC's definition to the specific strategies discussed. This is sometimes fairly straightforward to do, for example, when it comes to submission of false information or relief of artificial congestion. But we find rule 2, which prohibits "actions or transactions that are without a legitimate business purpose and that are intended to or foreseeably could manipulate market prices, market conditions, or market rules for electric energy or electricity products" (FERC Order 11/17/03, p. 13) to be vague and difficult to apply. It seems difficult to define what constitutes a legitimate business purpose, and almost any transaction could conceivably impact market conditions or market prices.

On the market efficiency side, we look for arbitrage-related behavior. For purposes of this paper, we define arbitrage as a trading strategy that takes advantage of relative mispricings in two or more markets (Hull 2002, p. 700). In the California energy markets, these price discrepancies occur between different time periods or locations. Some finance literature specifies that arbitrage must be risk-free in addition (Bodie, Kane, Marcus 2002, pp. 321, 978). However, all of the arbitrage strategies discussed here involve at least some degree of risk. Economists and finance experts generally agree that arbitrage is efficiency enhancing in the sense that it tends to smooth out price differences in related markets and reduce aggregate costs.

We find that many of the strategies studied might involve the submission of false information and could thus be prohibited under FERC's market

---

[2] Market behavior rule 3 also prohibits the submission of false or misleading information but is duplicative of rule 2b for our purposes.

behavior rules. About half of the strategies, including some that may rely on false information, appear to involve arbitrage. Markets which incorporate key aspects of the FERC's standard market design proposal either explicitly allow many of these efficiency-enhancing arbitrage strategies or are constructed so that traders would not profit from them. Finally, it appears that at least one of the strategies could run afoul of the general market manipulation prohibition under rule 2.

The remainder of the paper is organized as follows. Section 2 contains a description of how each strategy worked and examples of how it could make money. The discussion in Section 3 attempts to synthesize the information presented in the previous section and characterize the strategies. Section 4 concludes.

2. DESCRIPTION OF TRADING STRATEGIES

Most of the trading strategies discussed in this section are California-specific. That is, they are tied to and premised on the complex market design and rules put in place in California as a result of deregulation and the differences of the California market from other Western markets. As we will see, many strategies were designed either in response to differentials in price regulation in the West, or to take advantage of California's complicated and inefficient congestion management system. Therefore, some knowledge about the various Cal PX and Cal ISO markets is necessary in order to understand how the strategies functioned.

Before its closure, the majority of California power was bought and sold in the Cal PX day-ahead market.[3] The PX day-ahead market consisted of a single clearing price auction for each hour, conducted on the morning of the day before. The Cal ISO operates different markets in support of system reliability: the adjustment bid market is used to relieve transmission congestion; the ancillary services market is used to procure adequate reserves; and finally, the imbalance energy market keeps the system in balance in real-time with "incs" and "decs."[4] If the ISO faces a supply shortage, it may also purchase energy "out of market" at a negotiated price with a given supplier.

The strategies are grouped into three major categories for discussion: first are those dealing with trading in energy markets; the second category

---

[3] The Cal PX also ran an hour-ahead market, but this was primarily for deviations from day-ahead purchases and sales.

[4] In order to keep the system in balance, the ISO instructs resources to increment or decrement their generation or load in real time.

consists of congestion relief strategies; and the final category includes strategies related to ancillary services. We discuss eleven specific strategies as follows:

Table 1: Overview of California Trading Strategies

Strategy	Category	Discussed in Enron Memos	Discussed in ISO DMA Reports	Discussed in FERC Staff Final Report
Export of CA Power	Energy market trading	✓	✓	✓
Ricochet (Megawatt Laundering)	Energy market trading	✓	✓	✓
Underscheduling by Utilities	Energy market trading	✓	✓	✓
Fat Boy ("Inc-ing" Load)	Energy market trading	✓	✓	✓
Load Shift	Congestion relief	✓	✓	✓
Death Star (Circular Schedules)	Congestion relief	✓	✓	✓
Wheel Out	Congestion relief	✓	✓	✓
Non-Firm Export	Congestion relief	✓	✓	✓
Scheduling to Collect Congestion Charges	Congestion relief	✓	✓	
Get Shorty	Ancillary services	✓	✓	✓
Selling Non-Firm as Firm	Ancillary services	✓	✓	✓

The set of strategies included are those discussed in the Enron memos, the Cal ISO Department of Market Analysis (DMA) reports, or the FERC final staff report. While many of these strategies were first identified with Enron's use of them, other market participants have been accused of pursuing some of the same strategies.

2.1 Energy Market Trading Strategies

2.1.1 Export of California Power

While California typically relies on imports to meet its summer demand, the summer of 2000 saw an unexpectedly large amount of exports from the state relative to historical levels. One of the strategies listed in the Enron memos is called "Export of California Power". In short, power is bought at "capped" prices in the Cal PX day-ahead market and exported for sale at uncapped prices outside of the state.[5] For example, for August 1, 2000, on-peak Cal PX day-ahead prices averaged about $95/MWh (well below the $250 price cap in CA). Palo Verde (Arizona) on-peak price indices for the same day exceeded $500/MWh. Presumably this represented a huge

[5] While there were no price caps in place in the Cal PX market, prices there were effectively capped by price caps in the ISO imbalance energy market. This is because buyers in the PX market would simply refuse to purchase power priced above the cap that was effective in the ISO market.

arbitrage opportunity if a trader could export from California to sell at Palo Verde and make a profit of over $400/MWh. On the other hand, the fact that such large price differentials existed for days at times without being competed away leads one to believe that the export strategy was not pursued to the fullest degree.

This strategy represents pure price arbitrage based on location and is a consequence of uneven regulation in the West. As a result, load-serving entities in neighboring states could have bought power from California at capped prices to serve their native load. However, to the extent that power was exported by California generators (without first purchasing it in the PX), the strategy is not strictly arbitrage. Rather, it involves selling output for the highest available price. Another explanation given by generators for exports was their desire to make long-term or forward sales. Since the California IOUs were required to buy the majority of their power in the Cal PX day-ahead market, those generators desiring to sell a large portion of their power forward may have used exports for this purpose. Of course, the fact that prices outside of California were uncapped does not mean that they were always higher than prices within California.

2.1.2 Ricochet

The Ricochet strategy involves scheduling exports on a day-ahead or hour-ahead basis and re-importing the power for sale in real time or out of market in order to evade price caps or reporting requirements. It is essentially designed to arbitrage between the Cal PX markets and the Cal ISO markets and, of course, there is some risk involved. Since the strategy involves sending the power out of state in order to disguise the source, it has also been referred to as megawatt laundering.

The strategy has several variants depending on the source of the power, whether a second party provides parking services[6] out of state, and whether it is re-imported by the same party. The variant detailed in the Enron memo involves buying energy for export from the Cal PX day-ahead market, paying a second party for parking, and selling it back to the Cal ISO in real time. As a hypothetical example, Enron might purchase energy for export from the PX day-ahead market for $100/MWh, pay an Arizona utility $5/MWh for parking services, and sell the energy back into the real-time imbalance energy market for $250/MWh, clearing a sizable profit. Again, Enron would be taking the risk that real-time prices might actually be lower than day-ahead prices, in which case it would be losing money.

---

[6] An example of "parking services" is when entity A sends power to entity B day ahead, and then entity B sends the same amount back in real time.

The incentives to engage in Ricochet transactions varied over time with changing market rules. When hard price caps were in place, prior to December 8, 2000, the strategy could be used to attempt to sell energy out of market to the ISO at prices that exceeded the caps. It appears that out-of-market prices rarely exceeded the price cap until the second half of November 2000, so Ricochet was likely not a major strategy for in-state generators during the summer of 2000 (CA ISO DMA 2002, p. 29). With the advent of soft price caps, Ricochet could also be used to sell to the ISO above the price cap in the real-time market while benefiting from the more lenient reporting and cost justification requirements associated with imports and perhaps disguising the true cost basis of the power. Finally, starting in January 2001, the strategy could be used to sell imports directly to the state of California (instead of the ISO) to increase the chance of being paid promptly and reduce the chance of potential refunds being ordered.

2.1.3 Underscheduling by Public Utilities

It was widely known that the three major investor-owned utilities (and PG&E in particular) pursued a strategy of underscheduling their load in the Cal PX day-ahead market. In contrast to many of the profit-making strategies discussed elsewhere in this paper, this is a cost reduction strategy. The tactic consisted of shifting demand from the PX market to the ISO real-time market, where prices were capped.[7] This was accomplished by structuring PX portfolio bids to purchase energy only below the ISO price cap, knowing that the remainder could be bought at the real-time price cap in the worst-case scenario. This behavior is not surprising given the uneven regulation and sequential nature of PX and ISO markets.

For example, FERC staff reports that PG&E's forecast load for hour 13 on August 26, 2000 was approximately 9,000 MW. PG&E's bid curve in the PX day-ahead market was structured to buy only half of this expected load if prices hit $200/MWh. In reality, the market price was about $250/MWh in that hour, and PG&E purchased only about 3,800 MW in the PX, leaving about 5,200 MW of load to be supplied by the real-time market (FERC Staff 2003, p. VI-21).

2.1.4 Fat Boy (or Inc-ing Load)

Fat Boy was designed as a direct response to the underscheduling of load by the public utilities. The strategy involves overscheduling load with the

---

[7] ISO real-time prices were capped at $750/MWh from October 1, 1999. This cap was lowered to $500/MWh on July 1, 2000, and further to $250/MWh on August 7, 2000.

Cal ISO on a day-ahead or hour-ahead basis in order to be paid for excess generation in real-time. In other words, it is a way to pre-schedule real-time market sales. Fat Boy arises from the requirement that all participants submit a balanced schedule to the Cal ISO, *i.e.,* a schedule with supply equal to demand. It is important to note that this strategy applies only to importers, like Enron and British Columbia Power Exchange.[8]

If a participant has excess supply that it wishes to import into California, it has several options. First, it could bid its supply as an import into the Cal PX market. Since the public utilities underbid their demand into the Cal PX market in order to depress prices, this would not be an attractive option. Second, the imports could be bid into the Cal ISO real-time imbalance energy market. The disadvantage here is that there is no guarantee that all energy bid would be taken. It appears that the approach taken by Enron was to "dummy-up" load equal to any excess generation that it wished to sell, in effect guaranteeing a sale (as a price taker) at whatever the real-time price turned out to be. Enron traders could employ this strategy whenever the Cal ISO's load forecasts were smaller than their own and they believed that real-time prices would be favorable.

In the example cited in the memo (Enron Memo 12/6/2000, p. 2), Enron has 1,000 MW of power available. It schedules a 1,000 MW import to be used by Enron Energy Services (EES). In real time, Enron fulfills its schedule and transmits 1,000 MW. However, EES consumes only half of this power. The remaining 500 MW will be treated as positive uninstructed energy by the ISO and paid the uninstructed energy price. In effect, if Enron has purposely overscheduled EES's load, it has locked in a sale of 500 MW at the real-time market-clearing price.

There is ample evidence indicating that the Cal ISO was aware of the overscheduling of load and even acted to encourage it at times by creating fictitious load points.[9] This can be explained by the fact that Fat Boy helped offset the underscheduling by utilities and improved system reliability by reducing the supply shortfalls in real time. Indeed, prior to September 2000, there were no incentives in place to discourage the practice, as overgeneration was paid exactly the same price as instructed generation.[10]

---

[8] There is no need for in-state generators to employ the Fat Boy strategy and fabricate load, as they can simply overgenerate in real time and get paid for it. Imports, in contrast, would have to be either scheduled ahead of time or instructed in real time.

[9] FERC staff concludes in its final report that it appears the Cal ISO was aware of both the false underscheduling and overscheduling of load and ignored the information in constructing its load forecasts.

[10] Starting in September 2000, uninstructed generation was paid a different price (the real-time dec price) that was often lower than the price paid for instructed generation (the real-time inc price).

The strategy likely reduced real-time prices as it shifted the real-time supply curve outwards.

Markets such as PJM that do not have the artificial constraint of a balanced schedule requirement do not give participants the same incentives to engage in underscheduling of load or Fat Boy type bidding. Without a balanced schedule requirement, market participants can bid into the market in which they will receive the most favorable price. Removing this constraint increases the efficiency of the market and more tightly links the Day Ahead and real-time prices.

2.2 Congestion Relief Strategies

The Cal ISO uses a zone-based system and runs day-ahead and hour-ahead adjustment bid markets to control transmission congestion. If the desired energy flow between two congestion zones exceeds the available transmission capacity, adjustment bids are used to relieve the congestion and set a congestion price for the path (equal to the marginal value placed on the transmission). After payment for congestion relief to those participants whose schedules are adjusted in the congestion management process, any remaining payments from users of a congested interface are allocated to owners of Firm Transmission Rights (FTRs) on that line.

A number of strategies have been identified that manipulate the congestion management system to capture congestion payments without actually relieving any congestion. During the relevant time period, the ISO generally did not rescind congestion payments for energy flows that were not actually provided in real-time. These strategies are possible in part because the ISO's congestion management system is based on scheduled energy flows rather than actual, real-time flows. Further, due to the balanced schedule requirement, congestion is not optimized systemwide, but rather participant by participant. This creates the opportunity for a participant to create congestion and get paid for relieving it.

2.2.1 Load Shift

Load shift is a scheduling and bidding strategy designed to maximize the value of FTRs. It is best explained and understood in the context of Enron's use of the strategy. Enron purchased 62 percent of the FTRs on Path 26 in the north to south direction, a path necessary to move energy from Northern California and the Pacific Northwest to Southern California during summer peak demand periods. Enron submitted schedules that overstated its load in the southern region of the state and understated its load in the northern region by a corresponding amount. The result was increased congestion on

Path 26 as more energy tried to make its way south. In addition to collecting congestion relief payments (by reverting to its true load schedule), Enron's hope was to raise congestion prices for users of the path to maximize payments for its FTRs.

While it appears that Enron's FTRs for Path 26 were highly profitable (relative to their purchase price), the Cal ISO and FERC staffs conclude that the load shift strategy was not. The Cal ISO's Department of Market Analysis calculates that Enron earned only 2 percent of these revenues when it could have been pivotal in creating congestion, and only half of 1 percent when it could have both created congestion and relieved it (CA ISO DMA 2002, p. 14). FERC staff states that Enron was not generally successful in increasing congestion prices. Ironically, Enron appears to have profited from the underscheduling behavior of other load-serving participants in Northern California, primarily PG&E.

2.2.2 Death Star

Death Star is the general name used for a family of congestion strategies that appear to have been invented by Enron.[11] These strategies involve two or more simultaneous schedules (day-ahead or hour-ahead) that form a loop. In other words, the ultimate source and sink for the energy are the same such that no energy need flow. The key is that one of these schedules is on a congested line in the opposite direction of congestion so that a congestion relief payment is received. Further, the sum of any congestion payments made or transmission fees paid for other parts of the loop must be smaller than the congestion relief payment.

Of course, the schedules must be carefully engineered to avoid detection by the ISO. This can be accomplished in a number of ways. One possibility is to use municipal transmission lines, or other lines totally outside of the ISO control area for one of the legs of the circle. Another option is to involve a second or even third party as a "sleeve" to disguise Enron's involvement. The net impact of the transaction is that no energy flows and no congestion is relieved.

In the example cited in the memo (Enron Memo 12/6/2000, p. 4), Enron would import energy at Lake Mead for export to the California-Oregon border. It would earn a congestion relief payment because the energy was scheduled in the opposite direction of congestion. At the same time, Enron would buy transmission from the California-Oregon border back to Lake Mead across lines not controlled by the ISO. Therefore, the ISO is unaware

[11] Among the names cited for these strategies are Forney Perpetual Loop, Red Congo, NCPA Cong Catcher, Big Foot, and Black Widow.

of the circle: that power is being exported from Lake Mead and imported to Lake Mead at the same time.

It is interesting that there has been a lot of discussion within the ISO as to whether the Death Star strategy is beneficial or harmful on net. The main impact of this strategy is, of course, that congestion payments are received but no congestion is actually relieved. However, there is also some thought that this strategy might actually reduce congestion charges in the day-ahead and hour-ahead markets by allowing power to be diverted over unutilized paths. Nevertheless, the Death Star strategy makes it more difficult for the ISO to manage the grid in real time. Larger, more integrated markets or markets with more coordination in the congestion management process would be less prone the Death Star strategy.

2.2.3 Wheel-Out

Wheel-Out is a variant of the general Cut Schedule strategy that involves scheduling energy across a tie point that is known to be out of service in order to receive congestion relief payments. The ISO congestion management software accepts schedules across tie points with zero capacity. The key to the strategy is that when the schedules are submitted, the traders know that the line capacity is zero[12] and thus are certain that the schedules will be cut in real time. Due to this software flaw, the entity receives a congestion payment while never having to supply the energy.[13]

Hypothetically, Enron could use the Wheel-Out strategy in the following manner. Knowing that a tie point is out of service, Enron could schedule 500 MW of energy across this tie point in the opposite direction of congestion and then would receive the congestion payment equal to 500 times the congestion charge. However, because the tie-point is out of service the ISO then cuts all schedules over the tie-point. Enron gets to keep the congestion payment even though it never supplied any energy.

It must be emphasized that Wheel-Out is only objectionable when the party scheduling knows in advance that the schedule is infeasible. It appears that almost all potential Wheel-Out transactions occurred around a five-hour outage on May 27-28, 2000 and revenues from these transactions total about $3.4 million (CA ISO DMA 2002, p. 23).

---

[12] Information about line de-rates is available to all market participants through market notices and through the OASIS system.

[13] In 1999 the ISO wanted to change its software so that it rejected schedules on lines that were out of service. The PX rejected this proposal because it would have conflicted with the PX software. As of October 2002, the ISO was considering making such an amendment to the Tariff.

Wheel-Out highlights two weaknesses in the California market design: the market accepted schedules that were physically impossible and the market did not charge the participants who caused congestion the cost of relieving the congestion. In a market with the congestion management process fully integrated into the fundamental price determination process, participants would gain nothing from employing a Wheel-Out strategy. Thus, markets such as PJM and FERC's standard market design proposal eliminate the incentive for Wheel-Out.

2.2.4 Non-Firm Export

Non-Firm Export is another form of cut schedule strategy that involves scheduling a non-firm energy (energy not backed by reserves) export from California to earn congestion relief payments with no intention of actually exporting the energy. The supplier then cuts the schedule after receiving a congestion relief payment.

According to the Enron memo, this can be accomplished, for example, in the following manner (Enron Memo 12/6/2000, p. 4). Three hours before delivery, Enron schedules non-firm energy from SP15 to an area outside California. After two hours a congestion relief payment is received because the energy is scheduled in a direction opposite congestion. However, a trader then cuts the non-firm energy, and the congestion resumes.

It appears that the impact of non-firm export was quite small. According to the ISO, it was successfully employed (only by Enron) on three days in June and July of 2000 for total revenues of $54,000 before being prohibited (CA ISO DMA 2002, p. 7). Scheduling and cutting a non-firm export would not yield the same gain in a market with integrated price formation and congestion management processes. In addition, if schedules are financially binding, a trader cutting the non-firm energy would face a financial penalty calculated from the difference between the energy prices on either side of the congested transmission line. Such a penalty would force the trader to fully internalize the market cost of cutting the power and would wipe out any gains from the strategy.

2.2.5 Scheduling Energy to Collect Congestion Charges

In this strategy a participant schedules an imaginary counterflow across a congested line in the opposite direction of congestion in order to receive a congestion payment. In real time the ISO sees the non-delivery and charges the real-time inc price for the amount the entity is short, but the congestion payment is not rescinded. The price cap for the congestion market remained above that in the real-time market during part of the time period in question.

Therefore, it was possible that a participant could engage in profitable arbitrage whenever congestion prices exceeded real-time energy prices.

As a hypothetical example, Enron might schedule 300 MW of energy in the opposite direction of congestion although it only has 200 MW available. Enron might receive a congestion payment of $350 per MW on all 300 MW. However, in real time Enron only delivers 200 MW and is charged the inc price, say $250, for the 100 MW they did not deliver. Enron would make a profit here because the congestion payment is more than the charges incurred for delivering less energy than scheduled.

However, ISO analysis indicates that congestion prices have only exceeded the real-time price cap in about 50 hours since 1998 (CA ISO DMA 2002, p. 30). It is not clear how Enron or anyone else could have anticipated when congestion prices would be favorable and made this a highly profitable arbitrage strategy.

2.3 Ancillary Services Strategies

Ancillary services are reserve capacity required by the ISO for system reliability and include spinning reserves, non-spinning reserves, replacement reserves, regulation up, and regulation down. These services are held in standby and are called upon by the ISO in situations such as the loss of crucial generation or the loss of a transmission facility.

2.3.1 Get Shorty

Get Shorty involves selling ancillary services short in the day-ahead ancillary services market with the hope of buying them back at a lower price in the hour-ahead market. The ISO tariff recognizes the buyback of ancillary services as a legitimate form of arbitrage. What differentiates Get Shorty from legitimate buyback is that the selling entity never possessed the reserve energy and had no intention of ever supplying it. For this reason, this is referred to as paper trading of ancillary services.

It should be emphasized that this strategy applies only to importers. Generators within the ISO control area must identify the specific source of reserves, and the ISO is able to verify that this capacity is available. There is no way of knowing whether Enron or any other importer that bought back ancillary services really had them in the first place. Only a deliberate day-ahead sale of reserves that were never procured would fall under Get Shorty.

As a hypothetical example, suppose Enron sold 100 MW of ancillary services in the day-ahead market for $250/MWh. Then in the hour-ahead market it bought back 100 MW of ancillary services for $100/MWh. In this case it would make a profit of $150/MWh using the strategy.

Of course there is risk involved in this arbitrage that the hour-ahead prices will exceed the day-ahead prices. The ISO Department of Market Analysis has quantified total ancillary service buyback for the period January 1, 2000 through June 21, 2001. Total gains on potential Get Shorty transactions are $29.4 million, while total losses on these transactions are $1.5 million, for a net gain of about $29.8 million (CA ISO DMA 2003, p. 20).

2.3.2 Selling Non-Firm Energy as Firm Energy

This strategy involves deliberately selling or reselling non-firm energy to the PX or ISO while claiming it is firm energy, a practice prohibited by NERC. Since firm energy includes ancillary services and non-firm energy does not, a seller of non-firm energy would be charged for ancillary services. Therefore this is a cost reduction strategy by Enron or other importers to sell energy into California without paying for the associated reserves. There is some risk involved in that if the importer's supply were cut off it would be charged for the non-delivered energy at the real-time price. However, this is not likely to happen very often.

The Enron memo describes a complaint filed with the ISO by Arizona Public Service (APS) (Enron Memo 12/6/2000, p. 7). Evidently APS sold non-firm energy to Enron who then resold the same energy to the ISO as firm. APS then cut the energy. It is interesting that, in its report, the ISO states that it was not able to identify a single instance where this strategy had been used (CA ISO DMA 2002, p. 30).

3. CHARACTERIZATION OF TRADING STRATEGIES

The following table characterizes the strategies described above along several dimensions.

Table 2: Characterization of Energy Trading Strategies

	Prohibited by FERC Under			
Strategy	Market Manipulation (Rule 2)	False Information (Rules 2b, 3)	Artificial Congestion Relief (Rule 2c)	Arbitrage Related
Export of CA Power				✓
Ricochet				✓
Underscheduling by Utilities		?		
Fat Boy		?		
Load Shift	✓	✓	✓	
Death Star		✓		✓
Wheel Out		?	?	
Non-Firm Export		✓		
Scheduling to Collect Congestion Charges		✓		✓
Get Shorty		✓		✓
Selling Non-Firm as Firm		✓		

3.1 Market Manipulation

Load Shift is the only clear example of pure market manipulation we find. It was an attempt by Enron drive congestion market prices higher, involved the submission of false load schedules, and profited from the creation and relief of artificial congestion. Load Shift is probably the most offensive strategy as it violates rules 2, 2b, and 2c. Further, there is nothing in its description that suggests it has efficiency-enhancing arbitrage properties.

There are four additional strategies that others may deem questionable in terms of whether they fall under the prohibition of market manipulation under rule 2. The Export of California Power certainly had the potential to affect market prices and market conditions in California. However, as observed by FERC staff, "A merchant generator who exported power out of California in search of a better price or the opportunity to sell in forward (rather than spot) markets was behaving in a rational economic manner" (FERC Staff 2003, p. VI-16-17)." The purpose of the strategy was to sell power where prices were highest, not to manipulate market prices.

As a risky arbitrage strategy, Ricochet had a similar potential to affect market prices and conditions. However, the strategy was not necessarily detrimental to efficiency if the market price of the power exceeded the cap. Since there was no requirement to sell in California, the ability to circumvent price caps ensured that this power made its way back into the state when it was valued more highly there. Otherwise, the power would simply have been exported and never brought back.

Finally, Underscheduling by Utilities and Fat Boy were partially offsetting and are best described as scheduling strategies rather than trading strategies. Faced with the uneven price regulation of the PX and ISO markets, PG&E's minimization of its energy procurement costs appears to serve a legitimate business purpose and be socially desirable. Similarly, faced with PG&E's behavior and the resulting real-time shortages, it seems legitimate and socially desirable for importers to shift supply to the real-time market in response. However, since the ISO real-time market was designed to handle only about 5 percent of total load at most (and exceeded 25 percent of system load at times), these practices had consequences for system reliability, market outcomes, and strategic behavior of all participants. The net effects on market prices were likely large, with PX prices reduced and ISO real-time prices increased. However, neither strategy was designed to profit by manipulating prices. They could be viewed as supply or demand substitution.

3.2 Arbitrage

The Enron memo concludes that the Export of California Power does not appear to pose any problems except for public relations (Enron Memo 12/6/2000, p. 3). Indeed, there were no rules at the time to ensure that power sold in the PX day-ahead market ultimately served California load. The strategy was clearly designed to arbitrage prices within California and those in neighboring states. Given asymmetrical market regulations, it was perfectly natural profit-maximizing behavior.

Ricochet should be viewed as arbitrage between the day-ahead and real-time markets. Contrary to the way the strategy is portrayed, Enron and others who engaged in it were not minting money – this was risky arbitrage. If enough parties caught on and started doing the same, day-ahead prices would have risen and real-time prices would have fallen, making this a money-losing trade. Others have argued that power was purposefully moved into real-time because prices were somehow more easily manipulated there. It is likely a matter of degree. Smaller transactions were more likely to be arbitrage related, while very large movements of power were possibly attempts to raise real-time prices.

Although it might not appear so at first glance, Death Star was also an arbitrage strategy. Enron sold congestion relief to the ISO and bought transmission from other parties to complete the loop.[14] It involved arbitraging ISO transmission capacity with other transmission out of the

[14] Death Star may also have involved payment of some congestion charges to the ISO.

ISO's control. Death Star may have lowered day-ahead and hour-ahead congestion prices by allowing the ISO congestion management system to divert flows over under-utilized or un-utilized paths. However, this economic efficiency may have come at the expense of decreasing system reliability or causing other engineering problems.

Scheduling to Collect Congestion Charges was a pure arbitrage strategy between the congestion market and the real-time market. It appears to have been motivated by different price caps in the different markets.

Finally, Get Shorty was a risky arbitrage strategy between the day-ahead and hour-ahead ancillary services markets.

3.3 False Information and Artificial Congestion Relief

As noted earlier, many of the strategies appear to violate FERC's prohibition of transactions predicated on the submission of false information (rule 2b).

We believe it is questionable that the Underscheduling by Utilities or Fat Boy should fall under the province of the false information rule given the Cal ISO's knowledge of the practices, along with the absence of a clear requirement that a certain percentage of power be traded in the Cal PX market.

Load Shift, on the other hand, involved the knowing submission of false load schedules. Load Shift is the only strategy that both created and relieved (or attempted to relieve) artificial congestion, thus violating rule 2c as well.

While Death Star, Non-Firm Export, and Scheduling to Collect Congestion Charges did not create artificial congestion (and thus violate rule 2c), they did provide artificial congestion relief in the sense that no congestion was actually relieved in real-time. In that sense, they would all likely constitute the submission of false information. With Wheel Out, on the other hand, it is questionable whether offering adjustment bids on or scheduling energy across an out-of-service tie line crosses the threshold of false information and whether congestion must necessarily be created in order to profit from the strategy.

Get Shorty was premised on selling a service with no intent to deliver and, as such, would likely constitute a transaction predicated on the submission of false information. While selling short is a normal activity permitted in many other markets, it is not necessarily appropriate in the reserve energy market where system reliability is at stake. Selling Non-Firm Energy as Firm appears to be as close to pure fraud as any of these strategies get.

3.4 Collusive Acts

FERC market behavior rule 2d also contains a prohibition against engaging in "collusion with another party for the purpose of manipulating market prices, market conditions, or market rules..." (FERC Order 11/17/03, p. 27). None of the strategies described herein relied on collusive acts in order to function. However, there are specific variants of Death Star that would breach this rule if, for example, a second party knowingly agreed to provide a "sleeving" transaction to disguise one leg of the transmission loop.

4. CONCLUSIONS

We have examined eleven specific California trading strategies. The complexity of the strategies and their market impacts makes it difficult to classify them into neat bins. The approach taken is to characterize the strategies in terms of market manipulation and market efficiency.

Many of the strategies studied appear to potentially violate one or more of the market behavior rules adopted by FERC. We find FERC's general prohibition of market manipulation under rule 2 difficult to apply. However, in all cases except one (Load Shift) the potential infringement does not involve the market price manipulation that economists typically think of but rather the submission of false information. We find that many of the strategies involved arbitrage or rational product-substitution. Arbitrage should be encouraged, particularly in the context of the California energy markets.

It is possible that FERC overreacted to the situation in California and defined rules that are either too broad or vague, or too narrowly focused on the outcomes from the flawed California market design. Other electricity markets have been successfully set up with no reports of these types of trading abuses. There should be a clearer prohibition of, and penalties delineated for, submitting false information in future market design, but only to the extent these practices reduce economic efficiency and/or other social goals. It seems obvious that any loopholes that allow payment for contracts not performed should be closed in future market design.

REFERENCES

Bodie, Zvi, Alex Kane, and Alan Marcus. 2002. *Investments.* 5th Edition, McGraw-Hill.

CA ISO DMA. 2002. "Analysis of Trading and Scheduling Strategies Described in Enron Memos." Originally dated October 4, 2002, and released January 6, 2003.

CA ISO DMA. 2003. "Supplemental Analysis of Trading and Scheduling Strategies Described in Enron Memos." June 2003.

Enron Memo. 12/6/2000. Stoel Rives Memorandum from C. Yoder and S. Hall re: Traders' Strategies in the California Wholesale Power Markets/ISO Sanctions dated December 6, 2000. Released on FERC website in Docket No. PA02-2 on May 6, 2002.

Enron Memo. 12/8/2000. Stoel Rives Memorandum from C. Yoder and S. Hall re: Traders' Strategies in the California Wholesale Power Markets/ISO Sanctions dated December 8, 2000. Released on FERC website in Docket No. PA02-2 on May 6, 2002.

Enron Memo. Undated. Brobeck Memorandum from G. Fergus and J. Frizzel re: Status Report on Further Investigation and Analysis of EPMI Trading Strategies, undated. Released on FERC website in Docket No. PA02-2 on May 6, 2002.

FERC Order 11/17/03. Order Amending Market-Based Rate Tariffs and Authorizations, 105 FERC ¶ 61,218 (2003).

FERC Staff. 2003. *Final Report on Price Manipulation in Western Markets.* March 26, 2003.

Hull, John C. 2002. *Options, Futures and Other Derivatives.* 5th Edition, Prentice Hall.

Chapter 9

Economic Impacts of Electricity Outages in Los Angeles

*The Importance of Resilience and General Equilibrium Effects**

Adam Rose, Gbadebo Oladosu, and Derek Salvino
The Pennsylvania State University, Oak Ridge National Laboratory, and ICF Consulting, Inc.,

1. INTRODUCTION

In 2001, millions of customers of California electric utilities were subjected to a series of rolling "blackouts" generally attributed to poorly designed deregulation, weather conditions both within and outside the State, volatile natural gas prices, and lagging capacity expansion of both generation and transmission systems. Shortfalls of electricity supply have an economic cost, most often measured in terms of lost public utility revenues or of lost

---

* The authors are, respectively, Professor of Energy, Environmental, and Regional Economics, Department of Geography, The Pennsylvania State University; R&D Associate, Oak Ridge National Laboratory; and Associate with ICF Consulting, Inc. in Fairfax, VA. The research presented in this paper was supported in part by NSF-sponsored Multidisciplinary Center for Earthquake Engineering Research. Earlier versions of this paper were presented at the Annual Meeting of the International Association for Energy Economics, Aberdeen, Scotland, June, 2002, and the North American Meetings of the Regional Science Association International, San Juan, Puerto Rico, November, 2002. The authors wish to thank Joe Doucet and Gauri Guha for helpful comments on earlier drafts of this paper, and to Nusha Wyner and Mark Reeder for their discussant comments at the CRRI Research Seminar on Public Utilities, May 2004. We are also thankful to Stephanie Chang and to the Southern California Association of Governments for providing us with some of the necessary data. The authors, of course, are responsible for any errors and omissions.

sales by customers directly affected. However, the cost extends beyond these partial equilibrium aspects and impacts the entire economy through general equilibrium price and output effects imposed on direct and indirect customers and suppliers of firms that have their electricity service disrupted.

This paper presents the design, construction, and application of a computable general equilibrium (CGE) model to estimate total regional economic losses from electricity service disruptions. The model is used to analyze four rolling blackouts in Los Angeles County, California, in 2001. The methodology is, however, generally applicable to electricity and other utility service disruptions in the aftermath of other natural hazards and terrorist attacks (see, e.g., Rose, 2001; Rose and Liao, 2004).

We have incorporated several important refinements into the CGE model. First is the conceptual incorporation of adaptive behavior, or "resilience," by linking such actions as conservation, use of back-up generators, and substitution of other inputs to parameters of a constant elasticity of substitution production function. Second is the inclusion of these various adaptations into the empirical estimation of impacts. Third is the modeling of disequilibrium behavior by holding retail electricity prices constant and by including wage rigidities in the labor market. Fourth is the incorporation of a spatial dimension into the model by differentiating electricity supply availabilities across sectors (since firms in any given sector are not uniformly distributed across the region and hence not uniformly affected by localized blackouts).

Electricity outages are likely to be of even greater importance in the future. Their frequency can be expected to increase due to further impediments to capacity expansion, increased weather variability associated with climate change, and the advance of deregulation in other states. In addition, the economic system is becoming more vulnerable to electricity outages, in part due to greater interdependency that manifests itself in general equilibrium effects. The new "digital economy" means we are becoming increasingly dependent on computer networks, which require a continuous supply of electricity and which heighten sectoral linkages. Increased specialization makes it harder for firms to find substitutes for critical inputs in general. Business streamlining practices, such as "just-in-time inventories," make firms less able to cushion themselves against shocks from supply shortages. Finally, economic growth means more business enterprises will be at risk.

Electricity outage cost information is used for several important purposes including the determination of optimal capacity, reliability, pricing schemes, and load-shedding strategies (Tishler, 1993). Thus far, nearly every estimate of outages has omitted resilience and general equilibrium effects, and therefore could significantly misstate true outage costs.

2. PARTIAL AND GENERAL EQUILIBRIUM EFFECTS

Several approaches have been used to estimate the costs of electricity outages. Direct effects of outages manifest themselves in four major ways: lost sales, equipment damage or restart costs, spoilage of variable inputs, and idle labor costs. In addition, costs are incurred to reduce potential losses through the purchase of backup generators, permanent changes in production schedules, and utility capacity expansion to promote flexibility (Munasinghe and Gellerson, 1979). At the margin, the cost of outages should be equal to the cost of these adaptive responses. Hence, the most popular way of measuring electricity outage losses recently has been tabulating expenditures on back-up generation rather than measuring damages directly (Bental and Ravid, 1986; Beenstock et al., 1997). Still, measurement of just a single coping tactic, or type of damage, is likely to understate the direct dollar loss.[1]

Estimates of direct losses of electricity shortages range from $1.00 to $5.00/kwh in the U.S. (see, e.g., Caves et al., 1992). Many regions of the U.S. suffer outages of as many as ten to thirty hours per year due to ordinary circumstances of engineering failures and severe storms. The voltage disturbance blackouts of 1996, for example, are estimated to have cost California more than $1 billion (Douglas, 2000).[2]

---

[1] For a survey of approaches to measuring electricity outage costs see Munasinghe and Sanghvi (1988) and Crew et al. (1995). For a comprehensive estimation of four major types of outage costs, see Tishler (1993). The first major study to base direct sectoral outage costs on lost production was that of Munasinghe and Gellerson (1979). Telson (1975) was the first to consider broader impacts of outages. He used the ratio of gross product to industrial and commercial electricity consumption as an upper-bound and the ratio of the aggregate wage bill to industrial and commercial electricity consumption as a lower-bound. In effect, both of these measures include general equilibrium effects. However, they have three major shortcomings. First, partial and general equilibrium effects are not distinguished. Second, average, as opposed to marginal, ratios are used. Third, they omit various types of adaptation for electricity outages of brief duration. All three of these shortcomings are rectified in this paper.

[2] To date, there have been no rigorous economic analyses of the California rolling blackouts. Moreover, some casual observers have suggested their impacts might be rather minimal. Economists with the U.S. Treasury, the San Francisco Federal Reserve Bank, and a major investment firm interviewed by Berry (2001) generally downplayed the situation in the aftermath of outages in 2000. Part of the reason is that firms had preventative measures in place, such as backup generators and spatially dispersed facilities, in part due to the State's susceptibility to earthquakes. They noted the consumers were insulated against accompanying price increases by the continued regulation of retail electricity prices, though they emphasized the absorption of losses by utility stockholders and debtholders. Those interviewed concluded that the firms being affected were minor in the context of the overall California economy but warned that persistent outages could affect business and consumer attitudes and hence behavior. The

In this paper, we utilize economic output losses (also adjusted to net, or value-added, terms and consequently translated into welfare measures) as a common denominator for both partial and general equilibrium effects. This even enables us to include some capital and productivity costs into the measurement. Overall, general equilibrium effects consist of:

1. Output loss to downstream customers of a disrupted firm through its inability to provide crucial inputs. This sets off a chain reaction beyond the first line of customers of firms who have had their electric power curtailed.
2. Output loss to upstream suppliers of disrupted firms through the cancellation of orders for inputs. Again, this is transmitted through several rounds of suppliers.
3. Output loss to all firms from decreased consumer spending associated with a decreased wage bill in firms directly affected by the electricity outage, as well as all other firms suffering negative general equilibrium effects.
4. Output loss to all firms from decreased investment associated with decreased profits of firms suffering the electricity outage and other firms negatively impacted by general equilibrium effects.
5. Output loss to all firms from cost (and price increases) from damaged equipment and other dislocations (including uncertainty) that result in productivity decreases in firms directly impacted. Note that higher prices may not manifest themselves immediately during the outage period itself.

The direct and indirect costs of electricity outages thus do not just take place during the period in which power is curtailed. Backup generators are purchased in anticipation of outages, and the carrying cost of increased inventories of critical materials are incurred over a longer period as well. Equipment damage, spoilage, and idle labor costs may translate into an

---

outages have in fact persisted. Moreover, these initial assessments overlooked several important cost considerations explicitly modeled in this paper.

A survey by the National Federation of Independent Business (NFIB, 2001) found that over half of small businesses experiencing blackouts in California in January, 2001, had to curtail operations. Of these, 34.2% lost sales, averaging about 6.3% of their January sales total. Moreover, the Study indicated significant indirect effects. For example, 15.2% of businesses in California as a whole, and 10.5% in the Los Angeles area, noted that shipments or services to them were delayed because of a blackout affecting someone else. Also, 13.7% in California and 7.7% in LA lost sales "because customers directly impacted by a blackout either could not reach them or were otherwise preoccupied." California firms experiencing blackouts estimated that indirect effects cost them 16.9% of sales, more than double the direct effects. A significant number of firms suffered long-term effects, e.g., 13.6% curbed new hiring or delayed investments. Also, 12.8% of firms in California and 13.3% in the LA area responded that "the electricity problem has forced me to take concrete steps exploring the possibility of moving my business out of California."

immediate loss in profits, but they may not be passed through in the form of price increases until a later date. The same is true of electric utility cost and price increases that lag, even in a deregulated market. The three time periods, which we designate as *preparatory, crisis,* and *recovery,* will vary in length depending on the context. For estimation purposes, however, they may all be simulated simultaneously in cases where there are no significant dynamic (i.e., time-related) effects.

Note also that not all general equilibrium effects are negative. Some firms may benefit from the decreased prices associated with a shift in demand by other firms for various products. The analysis below indicates the existence of this possibility for several sectors, though the positive general equilibrium effects do not more than offset the negative ones.

For many years, input-output (I-O) models have been used to estimate the cost of utility service disruptions. These models are highly inflexible and likely to exaggerate both direct and indirect losses (see, e.g., Rose et al., 1997). For example, in their basic form, these models do not allow for adaptive responses such as conservation or input substitution. Moreover, they reduce general equilibrium effects to quantity interdependencies and are typically unidirectional (e.g., there are no offsetting effects through price reductions). It is not unusual for I-O models to yield multiplier effects that more than double the direct loss. General equilibrium models incorporate a broader range of interactions in the economy and more accurately measure regional economic impacts.[3] However, ordinary CGE models can be overly flexible (see, e.g., Rose and Guha, 2004), and, as we will demonstrate below, require serious refinement to avoid grossly *underestimating* losses due to electricity outages.

The marginal value of the loss of electricity is equivalent to the marginal value of electricity reliability (the obverse of the marginal adaptive response, or mitigation, noted earlier). This condition has been used to develop schemes to address the problem in the form of price discounts for curtailable service or premiums for priority service (see, e.g., Visscher, 1973; Chao and Wilson, 1987; Doucet et al., 1996). If structured properly, these pricing

---

[3] Computable General Equilibrium (CGE) analysis is the state of the art in regional economic modeling, especially for impact and policy analysis. It is defined as a multi-market simulation model based on the simultaneous optimizing behavior of individual consumers and firms, subject to economic account balances and resource constraints (see, e.g., Shoven and Whalley, 1992). The CGE formulation incorporates many of the best features of other popular model forms, but without many of their limitations (Rose, 1995). The basic CGE model has been shown to represent an excellent framework for analyzing natural hazard impacts and policy responses, including disruptions of utility lifeline services (Boisvert, 1992; Brookshire and McKee, 1992; Rose and Guha, 2004). Their applicability to other types of input supply disruption, such as the electricity blackouts was first suggested by Rose (2001).

provisions might also be used as a proxy for the partial equilibrium effects of electricity outages, in addition to serving as an efficient rationing mechanism.

A question arises, however, about the worthiness of interruptible service contracts in a general equilibrium context, both as an efficient rationing mechanism and as a means to estimate losses. For example, Firm A might pass up the interruptibility discount, thereby believing it has ensured reliable electricity service and continued operation. But Firm A may still be forced to shut down if Firm B takes the discount and fails to supply Firm A with the critical input. A poignant example of this took place in California in 2000. Kinder Morgan Energy Partners, which operated a 900,000 barrel per day pipeline, had signed up for a curtailment discount and repeatedly had to shut down. This almost caused San Francisco Airport to run out of jet fuel (Berry, 2001) and hence to shut down itself, despite ensuring itself a reliable supply of electricity (due to its criticality it had to pass up the interruptibility discount). Thus, interruptible service contracts as now structured do not reflect total interruption costs.

Thus, interruptible service contracts may be subject to a type of market failure in a general equilibrium context. This is closely related to public goods aspect of reliability identified by Spiegel et al. (2004) and the "contagion effect" identified by Kunreuther and Heal (2002) in their analysis of protection against terrorism in New York City neighborhoods. In the former, the socially optimal level of reliability is not provided because of the free rider problem. In the latter, the effort of one high-rise building owner may be all for naught if one of his neighbors fails to take adequate protective measures. We might also add market failure from the inability of individual firms to obtain and process information on direct and indirect supplier and customer reliability. In essence, these insights mean that considerations exist that would not allow the use of electricity reliability prices to represent an adequate measure of full outage costs. It also indicates the need for a system-wide approach, which is ripe for a modeling framework such as computable general equilibrium analysis.

3. RESILIENCY TO POWER OUTAGES

Will an X% loss of electricity result in an X% direct loss in economic activity for a given firm? The answer is definitely "no" given the various coping or adaptation tactics, which we will refer to as "resilience." Also, we use as our measure of direct resilience, the deviation from the linear proportional relation between the percentage utility disruption and the percentage reduction in customer output (see Rose, 2004). One of the most

obvious resilience options for input supply interruptions in general is reliance on inventories. This has long made electricity outages especially problematic, since this product cannot typically be stored. However, the increasing severity of the problem has inspired ingenuity, such as the use of non-interruptible power supplies (capacitors) in computers (Douglas, 2000). Other resilience measures include backup generation, conservation, input substitution, and rescheduling of lost production. In many business enterprises, these measures are adequate to cushion the firm against any loss of a rather short duration (definitely one hour or less), especially if advanced notice is given, as is often the case for rolling blackouts.

Will a Y% loss in direct output yield much larger general equilibrium losses? Again resilience adjustments suggest some muting of general equilibrium effects. These adjustments for lost output of goods and services other than electricity include inventories, conservation, input substitution, import substitution, and production rescheduling at the level of the individual firm and the rationing feature of pricing at the level of the market. Here we measure general equilibrium, or overall market resilience, as the deviation from the linear multiplier effect that would be generated from a simple input-output analysis of the outage (Rose, 2004).

Table 1 summarizes loss estimates from utility service disruptions. The number of studies is rather sparse, because we have limited inclusion to those studies that used output as the unit of measure and that have also included general equilibrium or region-wide effects. The first study noted in Table 1 is that of Tierney (1995), who received responses to a survey questionnaire from more than a thousand firms following the Northridge Earthquake. Note that maximum electricity service disruption following this event was 8.3% and that nearly all electricity service was restored within 24 hours. Tierney survey results indicated that direct output losses amounted to only 1.9% of a single day's output in Los Angeles County. A study by Rose and Lim (1996; 2002) used a simple simulation model of three resilience options to estimate adjusted direct losses at 0.42% and used an I-O model to estimate total region-wide losses of 0.55%. CGE analysis by Rose and Guha (2004) of the impacts of a hypothetical New Madrid Earthquake on the Memphis, Tennessee economy indicated that a 44.8% loss of utility services would result in only 2.3% loss of regional output; however, it should be noted that this model did not explicitly include resilience measures and was constrained from reducing major parameters, such as elasticities of substitution, to levels that truly reflected a very short-run crisis situation. A study by Rose and Liao (2004) for a hypothetical earthquake in Portland, Oregon, and for water, rather than electricity, utilities incorporated engineering simulation estimates of direct output losses into a CGE model. The first simulation, which represented a business-as-usual scenario,

indicated that a 50.5% loss of utility services would result in a 33.7% direct output loss, factoring in some resiliency measures. A second simulation, representing the case of $200 million capital expenditure initiative of replacing cast-iron pipes with modern materials, indicated that a 31% loss of utility services would result in a 21.3% loss of direct output in the region. In both simulations, only selected resilience factors were incorporated (e.g., production rescheduling was omitted), and this is one of the main reasons that adjusted direct output losses represent a higher proportion of loss of utility services than in the aforementioned studies (see column 9 of Table 1). Note that direct resilience declined following mitigation (direct output losses as a proportion of utility outage levels increased) because mitigation reduces initial loss of service, and hence ironically narrows the range of resilience options that can be brought into play.

Note also that all three studies that measured general equilibrium effects found them to be rather modest, ranging from 1.22 to 1.43. The I-O model of the Rose-Lim study did not allow for ordinary multiplier effects, because of assumed adequacy of inventories for goods other than electricity for the 36-hour outage period, and thus considered only "bottleneck effects" (see also Cochrane, 1997). Interestingly, the first simulation by Rose and Liao (2004) yielded general equilibrium effects on the order of 22% of direct effects, and the second simulation yielded general equilibrium effects 43% as great as direct effects. This means that pipe replacement actually not only lowers direct business resilience but also makes the regional economy as a whole less resilient, thus offsetting some of this strategy's benefits (see Rose and Liao for a further discussion).[4]

4. FORMALIZING RESPONSES TO INPUT SUPPLY DISRUPTIONS IN A CGE CONTEXT

The production side of the 33-sector CGE model used in this paper is composed of a multi-tiered, or nested, constant elasticity of substitution (CES) production function for each sector. The CES function has several advantages over more basic forms such as the Leontief (linear) or Cobb-Douglas (simple multiplicative) functions (see, e.g., Rutherford, 1997). It can incorporate a range of input substitution possibilities (not just the zero

[4] The only other formal study of general equilibrium or macro economic effects of electricity power outages was that of Bernstein and Hegazy (1988), who used a hybrid demand-driven/supply-driven input-output model. Their analysis yielded a total output loss/direct output loss ratio of 2.09. However, no resiliency considerations were incorporated, in part because developing countries have a relatively lower ability to adapt than do industrialized countries.

and unitary values of the aforementioned functions). The multiple tiers allow for the use of different substitution elasticities for different pairs of inputs (the elasticity is constant for a given tier, but elasticities can vary across tiers). The production function is normally applied to aggregate categories of major inputs of capital, labor, energy, and materials, with sub-aggregates possible for each (e.g., the energy aggregate is decomposed according to various fuel type combinations--electricity, oil, gas, and coal).

4.1 CES Production Function

Our constant elasticity of substitution (CES) production function has the following nested form for four aggregate inputs: capital, labor, energy, and materials.

$$Y_j = A_1 \cdot \left(\alpha_1 \cdot M^{-\rho_1} + \beta_1 \cdot KLE^{-\rho_1}\right)^{-1/\rho_1} \qquad 1^{st}\ \text{Tier}$$

$$KLE = A_2 \cdot \left(\alpha_2 \cdot L^{-\rho_2} + \beta_2 \cdot KE^{-\rho 2}\right)^{-1/\rho 2} \qquad 2^{nd}\ \text{Tier}$$

$$KE = A_3 \cdot \left(\alpha_3 \cdot K^{-\rho_3} + \beta_3 \cdot E^{-\rho 3}\right)^{-1/\rho_3} \qquad 3^{rd}\ \text{Tier}$$

$$E = A_4 \cdot \left(\alpha_4 \cdot (A_{4EL} \cdot EL)^{-\rho_4} + \beta_4 \cdot F^{-\rho_4}\right)^{-1/\rho_4} \qquad 4^{th}\ \text{Tier}$$

$$F = A_5 \cdot \left(\alpha_5 \cdot OG^{-\rho_5} + \beta_5 \cdot GU^{\rho_5} + \gamma_5 \cdot RP^{-\rho_5}\right)^{-1/\rho_5} \qquad 5^{th}\ \text{Tier}$$

where:

Y_j is output of sector *j*

A_i is the technology parameter of tier *i*, $A_i > 0$

$\alpha_i, \beta_i, \gamma_i$ is the factor distribution parameters of tier *i*, $0 \le \alpha_i, \beta_i, \gamma_i \le 1$

σ_i is the constant elasticity of substitution of tier *i*, $\sigma_i = 1/(1+\rho)$

K, L, E, M are capital, labor, energy, material aggregates

KLE is the capital, labor, and energy combination

KE is the capital and energy combination

EL is electricity

F is the aggregate of oil/gas, gas utilities, and refined petroleum

OG is the oil and natural gas aggregate

GU is gas utilities

RP is refined petroleum

The multi-tiered production function represents a type of hierarchical, or sequential, decision-making process. For a given level of output, the manager of a firm chooses the optimal combination of fuel inputs (non-electricity) in the bottom tier. He/she then chooses the optimal combination of electricity and the fuel aggregate in the 4th tier, etc. The model assumes homothetic weak separability, meaning that the substitution elasticities, and hence input choices, in one tier are invariant to those of another. Note that parameter values will vary across sectors.

4.2 Responses to Input Supply Disruptions

CGE models used for very short-run analysis, such as the case of electricity outages, are likely to yield estimates of direct business interruption for some if not all sectors of an economy that differ significantly from the direct loss estimates provided by empirical studies. This is because CGE model production function parameters are not typically based on solid data, or, even where they are, the data stem from ordinary operating experience rather than from emergency situations. Hence, it is necessary to explicitly incorporate the resilience responses below into the analysis. This is accomplished in our model by altering the parameters and the variables in the sectoral production functions of the CGE model.

Table 2 summarizes types of responses to input supply disruptions, linked to the production function tier and parameters to which each relates, as well as their likely changing intensity over time. The responses include:

1. Conservation of Electricity. This response can be implemented immediately and continued through the long run, i.e., be incorporated into the production process on a permanent basis. One of the silver

linings of recurring emergencies is that they force businesses to reconsider their use of resources. The parameter change for this response pertains to the technology trend variable for electricity in the fourth tier of the production function. More generally, in each tier of the production function except the fourth, the productivity terms are specified as general over all inputs in the given tier, i.e., factor neutral, for the purpose of generalization. In effect, adjustment of the productivity term for an individual factor, such as the A_{4EL} term in the fourth tier, biases the productivity improvement in the direction of that factor.

2. Conservation of Energy. This is a generalization of the adjustment for electricity and pertains to the third tier of the production function, where energy is substitutable with capital. Note that calculating energy conservation in the third tier involves distinguishing the technology parameter (A_3) by input category and significantly complicates the production function because the specification, if not reformulated, overlaps productivity and substitutability changes. A simpler way of approaching the problem is to change A_4, which covers energy conservation separate of other factors of production.
3. Conservation of Other Inputs. This is analogous to energy conservation and can be applied to any of the first three tiers. However, it can take on more permanence than energy conservation, and it is listed in column 4 of Table 1 as constant over the applicable periods rather than decreasing. An example of energy conservation would be to adjust factory temperature settings upward in the summer to save on air conditioning and downward in the winter to save on heating. Employees can sustain the harsher weather conditions for short durations, but it is unlikely that extremes beyond the 64-78 F° range can be made permanent. In contrast, a reduction in other inputs can be (e.g., reduction in number of trucks or maintenance personnel). One other adjustment option can be thought of as a sub-case—an increase in the use of inventories—if we focus on the very short run, or the *crisis* period itself (because the inventory purchase has been made in an earlier time period and will be replenished at a later date). If we consider all three periods together, the material is used and not saved, and the only production function modification is for the carrying cost of inventory holding.
4. Increased Substitutability of Other Fuels for Electricity. This response may require some grace period to implement, and is exemplified by hooking up a production process to a new source of power (with no major change in equipment). Substitution usually comes at a cost of implementation or more expensive operation, but options are likely to increase over time.

5. Increased Substitutability of Non-Energy Inputs for Energy. This is a generalization of response number 4. An example would be employees carrying boxes instead of using electrically powered conveyor belts.
6. Backup Generators. This adjustment is an increasingly popular way of coping with electricity outages, especially in California. If the generator is already part of the production function, it might be modeled as an electricity productivity improvement once it is activated. It is not exactly fuel factor neutral, because generators (other than small battery-operated devices) require some fuel input such as oil. The easiest way to model this during the *crisis* period is to offset the electricity input by an increase in one of the other fuels, a form of substitution. Methodologically, this could also be modeled by changes in the factor shares (α and β) in the 4th tier of the production function. As with the inventory item discussed above, there is some flexibility in how costs are considered temporally, in this case referring to the equipment and fuel oil purchases and their carrying cost. Thus, if the back-up generator is not already part of the firm's capital stock but is purchased in immediate anticipation of the outage, this option can be modeled by a change in σ_3. This convention can also be used to reflect a much earlier purchase of the generator if all 3 disaster time periods are telescoped into one.
7. Electricity Importance. This response refers to production activities that do not require electricity to operate (examples would be the growing of crops or production processes, such as some steel-making, that use other fuels). Thus, it refers to the inherent resiliency of a production process in the absence of any explicit adjustment and is simply reflected in the ordinary shape of the production function.[5] The presence of this factor is labeled as constant in Table 2; any increase in it over time would mean further technology adjustment and would come under the headings of responses 1 or 9.
8. Change in Technology. This refers to long-run (permanent) changes in the overall production process, such as replacing basic oxygen furnaces with electric-arc furnaces in steel making. It complicates the analysis because this response typically takes place long before or after the outage.
9. Time-of-Day Usage. This is a passive adjustment that pertains to cases where loss of electricity has no effect on output because the outage occurs during off hours. It is likely not relevant to planned outages, which typically take place during peak use periods, but is very important

[5] The term "electricity importance" stems from an engineering study (ATC, 1991) that measured it in the broadest terms as equal to a percentage change of sectoral output with respect to the percentage change in electricity availability. However, examples of this adjustment contained in the study are more narrowly construed in this paper.

in the case of unplanned outages (though, of course, these are relatively less frequent during off-peak hours).

10. Production Rescheduling. This is a major adjustment in the case of electricity outages and reflects the ability to make up lost production at a later date (or even at an earlier date if advanced warning is provided). It is most effective for manufacturing to the extent that production is not already operating at full (24/7) capacity. It is also relevant to most other enterprises, with the exception being on-the-spot services that cannot be readily purchased just across the regional boundary (if purchased from some other enterprise within the region not subject to the blackout, there is no cost except perhaps for increased gasoline consumption or inconvenience). As the outage length increases, the effectiveness of this option decreases (e.g., tourists canceling hotel reservations). Some outage costs might be incurred even for manufacturing firms, and these costs are likely to increase with the length of the outage if the production rescheduling requires overtime pay for workers or increased maintenance for equipment.

5. ROLLING BLACKOUTS IN CALIFORNIA

5.1 Background

In May of 2000, electric power reserves fell below 5% for the first time since California deregulated its electricity industry. Throughout the following twelve months, the stress on electric power generation persisted, with reserves falling below 5% on dozens of other occasions. On the six most critical days, rolling blackouts were implemented to ensure the integrity of the system. Rolling blackouts were implemented on January 17–18, 2001, in Northern California (PG&E's planning area) to avoid system-wide failure. Supply shortages also occurred on March 19–20, and May 7–8, affecting the entire deregulated portion of the State. These were the first times customers lost power due to supply shortages instead of physical damage to power system infrastructure in California in over 30 years.

While the emergencies ranged from one to eight hours on each of the days (all during regular business hours), affected customers only had electricity service interrupted for approximately one hour each because of the decision to rotate the outages. Customers were arranged into "circuits" representing all customer types (i.e., residential, commercial and industrial) except those providing public health, safety, or security services, which were exempt from the outages. Circuits were then aggregated into "Groups" containing approximately 50 megawatts of demand. When rolling blackouts

were in effect, each utility was required to reduce load by a specific number of megawatts. Utilities then interrupted service to the appropriate number of Groups in order to reduce demand by the specified amount. Once a Group had its service interrupted, it would not be selected again until all other Groups had experienced an outage.

During the four days of rotating outages that affected Southern California, approximately 640,000 Southern California (SCE) customers and 150,000 Sempra Energy (SDG&E) customers incurred blackouts, while during the six days of outages in PG&E's planning area, 1.1 million of its customers were affected.

5.2 Estimation of Electricity Outages by Sector

Estimating economic losses from electricity service interruptions requires data on the outage location and duration and on the normal level of economic activity that would otherwise occur. Data on the Stage 3 Blackouts in Southern California Edison's service territory are available in the form of maps for the specific areas affected (SCE, 2001). Data on the normal level of sectoral economic activity were obtained from the Southern California Association of Governments (SCAG, 2001) in the form of number of employees by 4-digit SIC and zip code. The complication in incorporating these data into the economic model is that no geographic or municipal boundaries (census tract, traffic analysis zone, zip code, etc.) are included in the SCE maps. Overcoming this difficulty required determining the zip codes that correspond to the outage areas described by each of the LA County maps.

The process used to identify the zip codes in the outage areas consisted of two steps (see also Salvino, 2004). First, intersections of major streets at the ends of the outage areas on the maps were identified using street names on the maps. Second, these intersections were entered into an electronic street map database that returned the city name and zip code for each intersection. As many as five intersections were checked for each map, depending on the size of the outage area, to ensure that all affected zip codes were properly identified. The process was repeated for each map, which produced 39 zip codes containing 799,000 jobs.

Next, the fraction of each affected zip code was determined by a three-step process. First, the total area of each affected zip code was acquired from Summary File 1 of the 2000 U.S. Census (2000). The areas for each zip code in this database are actually reported as Zip Code Tabulation Areas (ZCTAs), which are the sums of all census blocks associated with a postal zip code. ZCTA and postal zip code areas do not always match precisely, but are not substantially different, and so were deemed appropriate. Second,

GIS maps of each relevant zip code were obtained from Environmental Systems Research Institute, Inc. (ESRI, 2002). Each map contained the zip code boundary, interstate highways, local roads and highways, parks, rivers, and water bodies. Third, these detailed maps were used to assess the size of the areas indicated by the SCE maps. ArcExplorer 2.0, a GIS viewing program, was used to approximate the area of each outage as denoted by the SCE map. The dimensions of each outage were estimated using ArcExplorer's "measurement tool," which reported the full-scale distances, in miles, across affected streets and geographical features as indicated by the SCE maps. Once the appropriate dimensions were attained, the area of the outage was calculated using the formula for a geometric shape that best approximates the area of the outage (most often a rectangle). Finally, two zip codes originally thought to suffer outages in fact had 0% of their area affected, which reduced the total number of zip codes to 37 and the total number of jobs in affected zip codes to 731,000.

Having accurate measurements of both the outage areas and zip code areas allowed the calculation of a simple ratio of each zip code that experienced an outage. First, we assumed a uniform distribution of jobs throughout each zip code. However, when a large employer was known to have been affected,[6] the total number of jobs for that firm was added to the appropriate fraction of jobs in its SIC for the rest of the zip code to obtain the total number of jobs interrupted.[7] The data were then aggregated to the

---

[6] In addition, data on SIC, zip code, address, and the number of employees, for the case of large establishments (greater than 500 employees) was purchased from InfoUSA. A list of the number of large establishments by affected zip code and model sector was than compiled. Each address was then entered into an electronic database that generated a map indicating the location of the address. Each of the 112 large employer addresses was entered and compared to the maps of affected areas. Ultimately, 20 large employers were determined to be in the outage areas.

[7] In order to calculate jobs interrupted in SCE's service territory, each of the 459 zip codes containing economic activity in LA County, as reported by the Southern California Association of Governments, was assigned to either SCE's or LADWP's service territory, using one of five different procedures listed below in order of priority. First, a vector of zip codes with utility service territory assigned to them from a GIS analysis of 152 zip codes in LA County was used (Chang, 2002). Second, LADWP's website indicated that the utility serves only the portions of LA County that is considered "Los Angeles City," which allowed 38 additional zip codes to be classified as LADWP. Third, a list of zip codes from the "Los Angeles Almanac" website was used to identify all zip codes within the county limits. This indicated 142 zip codes within the SCE service area and provided a check for both Chang's estimate and the LADWP website data. Of the 190 zip codes classified using the first two methods described above, less than 20 conflicted with the data from the Los Angeles Almanac website. In the cases where conflicts did occur, priority was given first to Chang's data and second to LADWP's data. Fourth, remaining zip codes' community names were manually identified using an online database and then manually cross-checked on SCE's website for their blackout eligibility. Those

sector level by zip code, and finally into the total number of jobs interrupted by economic sector for the SCE service area of LA County.

Our preliminary estimate, prior to any resiliency adjustments, indicates that 87,730 jobs were interrupted, which amounts to 12.04% of total jobs in the affected zip codes (1.54% of total LA County employment and 2.82% of total employment in the SCE service area). Individual sectors were affected by differing amounts due to the heterogeneous composition of zip code business patterns (see column 1 of Table 3 in Section 7 and discussion below). Moreover, sectors such as Health Services were exempted from the planned blackouts, while others such as Electric Utilities were essentially impervious to them.

6. THE LA COUNTY CGE MODEL

6.1 The LA County Economy

Los Angeles County has one of the largest regional economies in the U.S. Total output in the County was about $570 billion in 1999, consisting of 55 percent value-added (roughly equivalent to Gross Regional Product), 40 percent intermediate inputs (including imports), and about 5 percent indirect taxes. Labor income makes up about 60 percent of all factor incomes. Exports from the County amounted to about $209 billion, a third of which is shipped to the rest of the U.S., but the majority of which is shipped overseas. Ninety percent of the imports into the LA County economy comes from the rest of the U.S., Household income amounted to about $282 billion, with the three highest income groups accounting for about 70 percent of the total (MIG, 2000). The economy is highly developed as exemplified by strong interdependencies between sectors, the prominence of manufacturing and service sectors, and a relatively high level of regional self sufficiency.

Energy is a major driver of the LA County economy both as an output and an input. For example, the economy produced $11.4 billion of Refined Petroleum Products in 1999. Private Electric Utilities (SCE) produced $2.3 billion, and State and Local Electric Utilities (primarily LADWP) produced $2.4 billion of power. The vast majority of energy intensive firms (typically heavy manufacturing) are located in the SCE service area. Residential customers comprised only 39.5% of SCE's 24.3 billion kwh sales in 1999,

---

communities eligible for rolling blackouts in the SCE database were assigned to SCE's service territory; 100 zip codes were classified using this method. Finally, the 12 remaining unclassified zip codes were entered into a mapping utility program to determine their geographic proximity to other already classified communities. All 12 of these zip codes were classified as SCE.

while they comprised 55.5% of LADWP's 12.9 billion kwh sales in that year.

6.2 Model Specification and Construction

We constructed a static, regional CGE model of the LA County economy consisting of 33 producing sectors (see the row labels of Table 3). The sectoral classification was designed to highlight the sensitivity of production processes to electricity availability. Institutions in the model are households, government, and external agents. There are nine household income groups and two categories each of government (State/Local and Federal) and external agents (rest of the U.S. and rest of the world).

1. Production

Production activities are specified as constant-returns-to-scale, nested constant elasticity of substitution (CES) functions as described in Section 4. The top level consists of a material inputs aggregate (in terms of fixed coefficients) and a capital-energy-labor inputs combination. On the second level, the capital-labor-energy combination is made up of labor and a capital-energy combination. On the third level, the capital-energy combination is made up of capital and energy aggregates. In order to capture the role of electricity more explicitly, we include an energy sub-nest consisting of fuels and electricity. Fuel use is a CES function of coal, oil (both crude and refined), and gas while electricity use is derived as a Leontief (fixed coefficient) aggregation of private electric utilities and state/local electric utilities. These electricity sectors correspond, respectively, to the Southern California Edison and LADWP electric utility service areas of Los Angeles County.

2. Supply and Trade of Goods and Services

We specify transactions between the LA County and the two external agents in the model using the Armington (1969) function for imports and the constant elasticity of transformation (CET) function for exports. The former is specified as a CES function to reflect imperfect substitution between domestic goods and competitive imports in demand. The latter is also a CES function that reflects the revenue-maximizing distribution of domestic output between exports and domestic markets, respectively. Regional export and import prices are based on exogenous external prices plus percentage tax/tariffs to reflect the open nature of the LA County economy.

3. Income Allocation, Final Demand, and Investment

Incomes from labor and capital employment in the economy are shared among institutions after the following deductions are made. Governments collect profit taxes on capital and employer-paid social security taxes on labor income, while industries deduct depreciation charges and retained

earnings before paying capital incomes. The remaining incomes are then distributed to households and external agents according to fixed shares. Institutions also receive inter-institutional transfers, such as subsidies, social security, and income taxes.

Households' consumption of goods and services are modeled using Cobb-Douglas expenditure functions, while government consumption is specified as a Leontief expenditure function. Thus, income elasticities are unity for both households and government, but price elasticities are one for households and zero for governments. Savings by households and governments are fixed proportions of disposable income, while external savings balance out this account. Households, government, and external entities also borrow from the capital account. Net savings by institutions, plus depreciation charges and retained earnings, are used to finance investment in capital goods. Investment in individual capital goods categories is a fixed proportion of total investment funding.

4. Equilibrium Conditions and Closure Rules

Equilibrium conditions balance supply and demand in goods and services markets. Capital endowments in the economy are fixed to reflect the short-run nature of our simulations. In the labor market, the Keynesian closure rule is used to allow for unemployment even in equilibrium.

5. Data and Solution Algorithm

Most of the data for the model are based on the detailed 1999 Social Accounting Matrix (SAM) for LA County, derived from the Impact Planning and Analysis (IMPLAN) database (MIG, 2000).[8]

Elasticities of substitution for regionally produced inputs and for imports were based on a synthesis of the literature (Oladosu, 2000; and Rose and Liao, 2004),[9] and other major parameters were specified during the model calibration process.[10] Because we are dealing with a very brief period of

---

[8] The IMPLAN system consists of an extensive data base of economic data, algorithms for generating regional input-output tables and social accounting matrices, and algorithms for performing impact analysis. The database follows the accounting conventions used by the U.S. Bureau of Economic Analysis. IMPLAN is the most widely used database for generating regional I-O models and SAMs in the U.S.

[9] Sources of these elasticities include: Prywes (1986), Dearoff and Stern (1986), Reinert and Roland-Holst (1992), Li (1994), and McKibbin and Wilcoxen (1998). Short-run substitution elasticities for the vast majority of sectors in the various production tiers are in the range of 0.65 and 0.9 before adjustment. Import elasticities between the focal region and the rest of the U.S. are generally 1.0 to 3.0, and those between the focal region and the rest of the world are generally 0.75 to 2.0.

[10] Parameters characterizing a CGE model are a combination first of those derived from external sources (either from the literature or by direct estimation), and second those specific to the economy under investigation. Once the first set of parameters have been specified, they are combined with the data for a particular year of the given economy to derive the second set of parameters. This is achieved by a model calibration process

analysis (less than one day), the short-run elasticities cited were further scaled down to 10 percent of their initial values.[11] We specify the model in the Mixed Complementarity (MCP) format using the MPSGE subsystem (Rutherford, 1998) of the General Algebraic System (GAMS) software (Brooke et al., 1995).

7. SIMULATING ROLLING BLACKOUTS

7.1 Estimating Partial Equilibrium Effects

In this section, we translate the sectoral electricity outage estimates presented in Section 5 into direct output losses. We noted earlier that outage costs consist of four categories: lost sales, equipment damage, spoilage of materials, and idle labor costs. The only study to date to have measured all four categories was that of Tishler (1993), who found that, on average for firms in Israel, the combination of material/labor costs accounted for the largest loss, followed very closely by lost sales, and lagged significantly by equipment damage. In this paper, we confine our attention to lost sales, but in our case this is likely to be the majority of the cost. The main reason is that California's rolling blackouts of 2001 were characterized by advance warning.[12] Although the warning notice time was sometimes as little as only ten minutes, the general electricity crisis environment and circulating policy statements gave all major electricity users a heightened awareness of the outage possibility and ample opportunity to initiate extensive contingency plans, especially by those firms with much at risk. This would imply firms

---

during which model functions are inverted to solve for the second set of parameters as a function of the externally derived parameters and variable values in the "base year" economy. The resulting parameters reflect conditions in the base year economy, ensuring that the model embodies the distinguishing features of the economy in question.

[11] The elasticity of substitution captures the ease of adjustment between input combinations in relation to changes in input price ratios. Such adjustment is easier for a longer time frame and visa versa. The elasticities in the original model were specified for an "short-run" timeframe (1-2 years) and were reduced to reflect a very short timeframe of our analysis. Unfortunately we are not aware of any studies that have estimated elasticities for a kind of "very short run" that we consider. Because the blackouts resulted in hard constraints on electricity availability and because electricity prices were held fixed in the simulations below (reflecting institutional limitations), the reduced elasticities prevent unrealistic substitutions of other inputs for electricity. Only empirical estimation of very short-run elasticities would enable us to assess the implications of our approach to the accuracy of the results.

[12] The NFIB (2001) Study found that, for California businesses as a whole, the average warning time was 2.5 hours, but that 83.8% of the firms had one hour or less. Businesses in the SCE service area had warning times considerably longer than the average.

would take precautions to avoid damage to equipment or materials either by temporarily shutting down production lines or by hooking up emergency generators.[13] Adjustments for idle labor time could also be initiated to reduce, though likely not eliminate, the associated cost by directing activities to training, maintenance, or "comp time" holidays.[14]

With respect to incorporating various resiliency options, at the time of this writing, we did not have adequate information on the sectoral distribution of backup electricity generating equipment, so we have not incorporated this option explicitly but rather have subsumed it in an assumed 10% increase in the input substitution elasticity between electricity and other fuels.[15] Our initial results can thus be interpreted as an upper-bound estimate of lost sales impacts in an instance where this resiliency option is underestimated.

Several of the other adjustments noted in Table 2 are incorporated into the analysis through the re-specification of production function parameters. These include: the productivity term (resilience adjustments 1-3) and the input and import elasticities of substitution (resilience adjustments 4-5). The productivity parameters are increased to reflect conservation possibilities.[16] Note that this conservation adjustment pertains not just to electricity but to all inputs, in part to reflect resiliency in the general equilibrium context. Substitution elasticities were lowered by 90% from their long-run values to

---

[13] Still, the NFIB Study found that firms had not necessarily been effective in making adjustments. In addition to the 34.2% who lost sales, 20.7% suffered material damage, and 37.5% were forced to absorb wage costs for work not done. However, dollar estimates of these impacts were not presented, and it is likely the latter two categories were of lower magnitude than lost sales. Moreover, these impacts are likely to have been lower for the SCE service territory because of the longer warning time.

[14] Note that the cost of estimating idle time is likely to involve double-counting and may not warrant separate estimation in any case. This is because wages/salaries are subsumed by foregone sales revenues.

[15] The NFIB Survey provides an indication of the prevalence of back-up electricity generating equipment. Of the small businesses surveyed, 10.6% in California and 8.3% in LA owned such equipment. Still, only 40.4% of the firms reported that these generators enabled them to operate normally during the blackout. Moreover, only 14.5% reported that a back-up generator was a practical option. There may, however, be a sample bias in the survey, as large firms are more likely to have their own source of electricity or contingency planning (including back-up generators).

[16] The California Governor's Office spearheaded a major conservation campaign in 2000-01 to prevent future blackouts. Electricity demand reduction targets were in fact exceeded, in part due to this initiative. By June 2001, California had achieved a 6.7% reduction in electricity and a 10% reduction during summer peak hours (CEC, 2001). Ironically, however, while such efforts reduced the frequency of blackouts, they may increase economic impacts of those that do occur. The remaining blackouts are more of a surprise, and increased ongoing conservation may reduce flexibility of conservation capabilities during a crisis.

reflect the very short-run period of rolling blackouts in the base case. For the simulations below, these lowered values are increased by 10% (thus offsetting somewhat the previous short-run adjustment) to account for ingenuity during the crisis.

More explicit modifications were made to reflect production rescheduling. In this case we used estimates from a study of economic impacts of the Northridge Earthquake by Rose and Lim (1996; 2002). No time-of-use adjustments were made because rolling blackouts take place during peak hours rather than off-hours. Recall also that "electricity importance" is already embodied in the shape of the production function. Also, no data were available for long-term changes in technology that intentionally or unintentionally cushioned against outage costs, so this response could not be simulated.

Finally, we note again the three periods relevant to electricity outage costs estimation: *preparatory, crisis,* and *recovery.* Our simulations combined all three time periods. With the exception of technological change, backup generators, and inventory purchases, all adjustments are likely to take place during the actual outage or in close proximity to it (e.g., most production rescheduling will take place soon after the outage). Moreover, we do not see any relevant time-related dynamics associated with the separate time periods, other than some minor time-value of money losses associated with idle back-up generators and inventories. The only major features that are compressed, and that we could otherwise have modeled separately, are the general equilibrium cost and price increases that are likely to otherwise manifest themselves in the *recovery* period.[17] However, we do not believe that this simplification has any significant effect on the accuracy of the results.

Note also that we were not able to formally model less concrete aspects of outage costs, such as producer and consumer attitudes and expectations that result in changed behavior and hence alterations in economic activity, including business/residential location or relocation decisions. To our knowledge, no study has empirically modeled these important features, and their analysis is beyond the scope of this paper.[18]

---

[17] Note that production rescheduling targets are likely to be affected by general equilibrium results. That is, a firm may have to reduce a day's production by 10% as a direct result of the blackout, but general equilibrium effects may affect the situation by lowering the demand for its product by 3%; hence the rescheduling target in the compressed simulation period would be only 7%. In actuality, the general equilibrium effects may not be known for some time, so the rescheduling could very well be at the full amount, with an appropriate production decrease taking place at a later date.

[18] Some studies have modeled expectations about outages but only on the part of utilities themselves and not their customers (see, e.g., Caves et al., 1992).

Table 3 presents our estimates of direct losses though a series of adjustment steps. Column 1 lists the baseline direct energy use by sector in the Southern California Edison Service Area. Column 2 lists the percentage reduced availability of electricity to each sector during the SCE (Private Electric Utility Sector) outage. Column 3 displays the average number of hours of operation of each sector in a given year, which we use to adjust our outage loss estimates to a one-hour basis. Column 4 shows this initial estimates of direct output losses in the absence of any resilience options, by simply reducing the electricity input in each of the sectoral production functions based on the proportion of the sector's output that would be lost in one hour if there was a fixed proportional relationship between electricity input and product output. Column 5 shows the influence of resilience options on direct output loss after making the parameter changes noted earlier in this section. For both columns 4 and 5, the production function of each sector is run separately in a constrained optimization (reflecting the lower electricity availability) to compute the partial equilibrium effects of the outage. Note that in several sectors the initial production function estimate of direct losses in percentage terms exceeds the electricity outage estimate in percentage terms. This stems from such factors as the decrease in the rate of return on capital (capital stock is fixed and its utilization rate decreases). The overall 2.81% reduction in electricity availability in the SCE Service Area for the four one-hour rolling blackouts results in only a $9.9 million initial partial equilibrium estimate of output losses (column 4). When we incorporate direct resiliency responses by recalibrating the sectoral production functions the total direct loss estimate drops to a county economy-wide total $1.2 million (column 5), or an 87.8% reduction in losses. Of course, in both columns 4 and 5, loss estimates differ by sector, reflecting sectoral differences in outage rates and resilience.

Column 6 shows the adjustment for production rescheduling, which again varies by sector. It is estimated by simply multiplying the result for each sector in column 5 by unity minus a production rescheduling factor expressed as a decimal fraction (see Rose and Lim, 2002). The rescheduling factors can be inferred by inspection from the results in column 5, and range from a low of 0.3 in the Entertainment sector to a high of 0.99 in Durable Manufacturing, Wholesale Trade, and several other sectors. No adjustment in production function parameters is necessary, and this response is implicitly modeled by simply increasing the available electricity supply for each sector to represent subsequent increased purchase of electricity.[19] Again, this does not violate the electricity supply constraint because we have

[19] This adjustment can be done at any stage of the direct loss estimation process as long as the production function exhibits constant returns to scale.

compressed the three business interruption time periods into one.[20] Note also that this response to outages means that, from our stylized partial equilibrium standpoint, electricity utilization does not decrease over the (compressed) three periods by the amount of the outage figures presented in column 2 because of the offset of production rescheduling when excess capacity exists. Thus, our final direct estimate of gross output losses in the SCE from the four service territory rolling blackouts in 2001 is only $266 thousand, or a 78% decrease in the direct losses due to other resilience adjustments.[21]

7.2 Estimating General Equilibrium Effects

In this section, we use the adjusted direct electricity outage impacts presented Table 3 as a driver for simulating general equilibrium effects. The sectoral electricity outages translate into constrained sectoral electricity availabilities in the calculation of a new equilibrium for the Los Angeles County economy. Decreases in electricity availability lead to reduced sectoral production, as in the previous subsection, and, in turn, stimulate price and quantity changes in all product and factor markets in the region.

Our general equilibrium simulations are based on a fixed price of electricity, reflecting the California regulatory structure of capped retail electricity prices (the blackouts themselves do not result in any subsequent regulated price increases, though the underlying conditions may more subtly influence future prices). The results are presented in Table 4 in terms of Los Angeles County as a whole and not just the SCE Service Area. Columns 1 and 2 are the analogue of the corresponding columns of Table 3 but for the larger region. Column 3 shows sectoral economic output (sales) in the County. Column 4 displays our initial direct, or partial equilibrium (PE),

---

[20] No modification is needed if retail electricity prices are the same in the outage and aftermath periods, which was the case in LA County (i.e., there was no peak-load nor time-of-day pricing and no rate increases). Otherwise, a price change must be entered into the sectoral production function simulations.

[21] Note that several factors were omitted from our analysis. First, we did not include the costs of outages to residential customers. Although, they are unlikely to be part of standard economic accounting, the inconveniences do detract from welfare, in addition to equipment damage that was also suffered. We also did not include the costs of restarting equipment or business operations in general. Also, we omitted any cost of temporary relocation, which might be major in a long outage but were unlikely to be significant here. Finally, we focused on the damage side and did not include the purchase or operating costs of back-up generators or other resilience options. In the other direction, we were not able to measure some of the long-term learning effects of adaptive responses applied to the 2001 outages and future events that would reduce costs to many customers. The net totality of these omissions suggest that our direct outage estimates understate the actual costs.

loss estimates in percentage terms for the County as a whole, not yet adjusted for production rescheduling; the 7.07% output reduction (on a one-hour basis) is comparable to the results of the NFIB (2001) survey. Column 5 shows the direct estimates after parameter recalibration to reflect the application of various resilience options in addition to production rescheduling; the 0.741% result is far below the survey results. Column 6 contains our estimates of indirect, or net general equilibrium (GE minus PE), effects; the overall 0.55% reduction in regional gross output represents a 74.3% increase in the recalibrated direct (PE) loss estimate, yielding an implicit multiplier of 1.743. This contrasts with the approximate multiplier of 2.7 of the purely linear LA County I-O model, and the implicit multiplier of 2.0 from the NFIB (2002) survey. The difference in values reflects our explicit incorporation of resilience responses, the non-linearity of the CGE model, and offsetting effects from price changes in product and factor markets.

All of the direct loss results in Table 4 are further significantly decreased once we adjust for production rescheduling. Essentially, as in Table 3 adjustments between columns 5 and 6, direct output losses in each sector and overall would be decreased by 78%. Because sectoral net GE effects are distributed differently than sectoral PE effects and because our CGE model is non-linear, we would not expect the same reduction in the absolute magnitude of the general equilibrium effects, thereby altering the size of the GE/PE ratio (implicit multiplier). The results (not shown) indicate this to be true. The bottom line total regional economic impact of the rolling blackouts in LA County are estimated to be less than $1 million. We can help put this relatively small loss in perspective by noting that less than 3 percent of the SCE electricity supply (or about 1.5 percent of Los Angeles County's total electricity supply) was affected, and then for less than 4 hours total. Also, the cumulative effect of various types of resilience (including market resilience) decreases the initial (linear) estimate of losses and is hence much stronger than the 74.3 percent increase due to general equilibrium effects.

Note also, that while utilities and businesses have devised numerous successful mitigation and adaptation tactics to reduce PE losses, a far different approach is required for GE losses. Here resiliency depends on inventory holdings and the ability to obtain imports quickly, for goods and services in general in both cases. Many of these considerations are much broader than just electricity outage concerns. However, other general equilibrium strategies can be problem-specific, such as enhancing information flows (e.g., through a clearinghouse that matches available suppliers and disrupted customers of a given product). Implicitly, the full market equilibrium adjustment in our model assumes such information flows are operative. We emphasize that this "re-contracting" and other types of

market equilibrium adjustment are not assumed to take place in one-hour. Rather, we have defined the period of analysis to include both the outage period and the recovery period, the latter possibly taking several weeks. Ordinary market adjustments would probably take longer, but several weeks is probably adequate for the very minor perturbations associated with the LA blackouts.

8. CONCLUSION

Electric power outages continue to disrupt many regional economies of the U.S. Although it is technically impossible, and likely economically impractical, to avoid all outages, utility managers and public officials can seek to identify an optimal level of reliability by comparing costs and benefits. The latter consists of both avoided direct (partial equilibrium) and indirect (net general equilibrium) economic losses. These two loss categories differ significantly in terms of their scope, measurability, and our ability to mitigate them.

This paper has attempted to improve loss estimation methodologies for electricity outages through simulation analysis. The approach to partial equilibrium (PE) loss estimation is to refine data on outage level, duration, and location as an input to non-linear, sectoral economic production functions. These production functions also embody various coping adjustments, or types of resilience (e.g., conservation, use of back-up generators, production rescheduling). Thus, rather than a linear proportional relationship between a loss of electricity and a loss of economic output (sales), our PE analysis reflects adaptive responses that cushions the blow significantly, though some gains are offset by lower profitability. Our application of the methodology to the 2001 rolling blackouts in Los Angeles County indicated that these resiliency tactics could reduce linear PE estimates of economic impacts by more than 90%.

Indirect, or net general equilibrium (GE), effects of a blackout are more complex, because they represent extensive chain reactions of price and quantity interactions for many rounds of customers and suppliers of disrupted firms. Here, additional resiliency options are available, including the use of inventories and substitution of imports for locally produced goods. Complications also ensue regarding the extent of price rigidities in goods markets (especially electricity) and factor markets (especially labor). Our computable general equilibrium analysis approach represents an improvement on linear models such as input-output analysis. The CGE model's inherent non-linear nature and ability to capture market interactions avoids the overestimation generally attributed to linear models. Our estimate

of net GE impacts, as approximately a 75% increase in PE losses, is lower than survey-based estimates and less than half as great as indirect loss estimates stemming from the use of an I-O model.

Overall, our results indicate total short-term economic losses from the LA blackouts to be less than $1 million! On the other hand, we have not measured long-term losses associated with discouraging business investment or the loss of consumer confidence, nor have we measured the cost of consumer inconvenience. Also, the minimal losses reflect "the best of a bad situation," including SCE efforts to minimize the geographic scope of outages and to provide advance warming. Only 2.81% of the SCE power system territory in LA was disrupted and for less than one hour for each of the four rolling blackouts. A disruption of the entire system would increase our $1 million estimate by a factor of 35. On top of that, if the disruption occurred for 24 hours, the loss would approach $500 million (even adjusting for time of day usage). Longer outages would have non-linear increasing impacts due to greater difficulty of maintaining emergency conservation and implementing production rescheduling. A full blackout of seven days stemming from, say, a major LA earthquake would result in several billion dollars of electricity outage-induced economic losses in just the SCE service territory alone. Thus, we ask readers to exercise care in trying to generalize our results to other contexts that are significantly different, such as the Northeast Blackout of the summer of 2003.

REFERENCES

Anderson, R., and L. Taylor. 1986. "The Social Cost of Unsupplied Electricity: A Critical Review." *Energy Economics* 8: 139-46.

Armington, P. 1969. "A Theory of Demand for Products Distinguished by Place and Production." *International Monetary Fund Staff Papers* 16: 159-76.

(ATC) Applied Technology Council. 1991. *Seismic Vulnerability and Impact of Disruption of Lifelines in the Conterminous United States,* Report ATC-25. Redwood City, CA: ATC.

Beenstock, M., E. Goldin and Y. Haitobsky. 1997. "The Cost of Power Outages in the Business and Public Sectors in Israel: Revealed Preference vs. Subject Evaluation." *Energy Journal* 18: 39-61.

Bental, B. and S. A. Ravid. 1986. "A Simple Method for Evaluating the Marginal Cost of Unsupplied Electricity." *Bell Journal of Economics* 13: 249-53.

Bernstein, M., and Y. Hegazy. 1988. "Estimating the Economic Cost of Electricity Outages: A Case Study of Egypt." *Energy Journal* 9 (Special Issue on Electricity Reliability): 1: ?.

Berry, J. M. 2001. "Impact of California's Crisis Muted for Now, U.S. Officials Say," *Washington Post,* January 20, page E1.

Boisvert, R. 1992. "Indirect Losses from a Catastrophic Earthquake and the Local, Regional, and National Interest." In *Indirect Economic Consequences of a Catastrophic Earthquake,* Washington, DC: FEMA.

Brennan, T. et al. 1996. *A Shock to the System: Restructuring America's Electricity Industry.* Washington, DC: Resources for the Future.

Brookshire, D. and M. McKee. 1992. "Other Indirect Costs and Losses from Earthquakes: Issues and Estimation." In *Indirect Economic Consequences of a Catastrophic Earthquake,* Washington, DC: FEMA.

(CEC) California Energy Commission. 2001. *The Summer 2001 Conservation Report.* Sacramento, CA.

Caves, D., J. Harriges, and R. Windle. 1992. "The Cost of Electric Power Interruptions in the Industrial Sector: Estimates Derived from Interruptible Service Programs." *Land Economics* 68: 49-61.

Chao, H. P., and R. Wilson. 1987. "Priority Service: Pricing, Investment and Market Organization."*American Economic Review* 77: 899-916.

Crew, M. A., C. S. Fernando, and P. R. Kleindorfer. 1995. "The Theory of Peak-Load Pricing: A Survey." *Journal of Regulatory Economics* 8: 215-48.

Crew, M. A., and P. R. Kleindorfer, 1986. *The Economics of Public Utility Regulation.* Cambridge, MA: MIT Press.

Deardoff, Alan and Robert Stern. 1986. *Michigan World Trade Model.* Cambridge, MA: MIT Press.

Doucet, J. and S. Oren. 1991. "Onsite Backup Generation and Interruption Insurance for Electricity Distribution." *Energy Journal* 12: 79-93.

Doucet, J., K. Min, and M. Roland, and T. Strauss. 1996. "Electricity Rationing Through A Two-Stage Mechanism." *Energy Economics* 18: 247-63.

Douglas, J. 2000. "Power for a Digital Society." *EPRI Journal* 25 (4): 18-25.

Environmental System Research Institute. 2002. "Zip Code Maps for Los Angeles County."

(GDT) Geographic Data Technology. 2002. GDT U.S. Street Data July 11, 2002 http://www.esri.com/data/download/gdt/index.html.

(IPCC) Intergovernmental Panel on Climate Change. 2001 *Climate Change 2001: The Scientific Basis.* New York: Cambridge University Press.

Kunreuther, H. and G. Heal. 2002. "Interdependent Security: The Case of Identical Agents." Columbia University.

Li, Ping-Cheng. 1994. *Environmental Policy, Coal, and the Macroeconomy: The Impact of Air Quality Regulations on the Pennsylvania Economy.* Ph.D. Dissertation, The Pennsylvania State University, University Park, PA.

McKibbin, Warwick and Peter Wilcoxen. 1998. "The Theoretical and Emprical Structure of the G-Cubed Model." Brookings Institution, Washington, DC.

(MIG) Minnesota IMPLAN Group. 2000. *Impact Analysis for Planning System (IMPLAN),* Stillwater, MN.

Munasinghe, M., and M. Gellerson. 1979, "Economic Criteria for Optimizing Power System Reliability Levels." *Bell Journal of Economics* 10: 353-365.

Munasinghe, M., and A. Sanghvi. 1988. "Reliability of Electricity Supply, Outage Costs and Value of Service: An Overview." *Energy Journal* 9 (Special Issue on Electricity Reliability): 1-18.

(NFIB) National Federation of Independent Business. 2001. "California Small Business Electric Supply Interruption Survey: Survey Results."

Prywes, Menahem. 1986. "A Nested CES Approach to Capital-Energy Substitution," *Energy Economics,* 8: 22-28.

Reinert, Kenneth and David Roland-Holst. 1992. "Armington Elasticities for United States Manufacturing." *Journal of Policy Modeling,* 14: 631-9.

Rose, A. 2001. "A Critical Issue in the Electricity Liability: Minimizing Regional Economic Losses in the Short Run." *IAEE Newsletter,* First Quarter.

Rose, A. 2004. "Defining and Measuring Economic Resilience to Earthquakes," *Journal of Disaster Preparedness and Management.* forthcoming.

Rose A., and J. Benavides. 1999. "Optimal Allocation of Electricity After Major Earthquakes: Market Mechanisms Versus Rationing," In *Advances in Mathematical Programming and Financial Planning* edited by K. Lawrence, Greenwich, CT: JAI Press.

Rose, A., and G. Guha. 2004. "Computable General Equilibrium Modeling of Electric Utility Lifeline Losses from Earthquakes," Forthcoming in *Modeling the Spatial Economic Impacts of Natural Hazards* edited by S. Chang and Y. Okuyama, Heidelberg: Springer.

Rose A., and S. Y. Liao. 2004. "Modeling Regional Economic Resiliency to Earthquakes: A Computable General Equilibrium Analysis of Water Service Disruptions." *Journal of Regional Science,* forthcoming.

Rose, A., and D. Lim. 2002. "Business Interruption Losses from Natural Hazards: Conceptual and Methodology Issues in the Case of the Northridge Earthquake," *Environmental Hazards* 4 (2): 1-14.

Rose, A., J. Benavides, S. Chang, P. Szczesniak, and D. Lim. 1997. "The Regional Economic Impact of an Earthquake: Direct and Indirect Effects of Electricity Lifeline Disruptions." *Journal of Regional Science* 37: 437-58.

Rutherford, T. 1995. "Computable General Equilibrium Modeling with MPSGE as a GAMS Subsystem: An Overview of the Modeling Framework and Syntax," www.gams.com/solbers/mpsge/syntax.htm.

Salvino, D. 2004. "The Economic Impact of Electricity Outages in California." Unpublished M.S. Thesis, Department of Energy, Environmental, and Mineral Economics, The Pennsylvania State University, University Park, PA.

(SCAG) Minjares, J. 2002. "RE: LA County Employment by Place of Work," electronic data files containing employment by SIC code and Zip Code compiled by the Southern California Association of Governments, June 15, 2002.

(SCE) Southern California Edison. 2002. *Mapsfor all Outage Group.* February, 2002, www.outagewatch.com/003_safety_first/003e2_ro_community,shtml.

Shoven, J. and J. Whalley. 1992. *Applying General Equilibrium,* New York: Cambridge University Press.

Telson, M. 1975. "The Economics of Alternative Levels of Reliability for Electric Power Generation Systems." *Bell Journal of Economics* 6: 679-94.

Tierney, K. 1997. "Impacts of Recent Disasters on Businesses: The 1993 Midwest Floods and the 1994 Northridge Earthquake," in B. Jones (ed.), *Economic Consequences of Earthquakes: Preparing for the Unexpected,* Buffalo, NY: National Center for Earthquake Engineering Research.

Tishler, A. 1993. "Optimal Production with Uncertain Interruptions in the Supply of Electricity: Estimation of Electricity Outage Costs," *European Economic Review* 37: 1259-74.

U.S. Bureau of the Census. 2002. "Census 2000 Summary File 1 (SF1) 100-Percent Data," April 1, 2000. hhtp://factfinder.census.gov/servlet/BasicFactsServlet.

Visscher, M. L. 1973. "Welfare-Maximizing Price and Output with Stochastic Demand: Comment." *American Economic Review* 63: 224-29.

Woo, C., and K. Train. 1988. The Cost of Electric Power Interruptions to Commercial Firms." *Energy Journal* (Special Issue) 9: 161-72.

Table 1. Summary of Loss Estimates from Utility Service Disruptions

Study	Location/ Event	Utility/ Duration	Method or Model	Loss of Utility Services (%)	Direct Output Loss Unadjusted (%)	Direct Output Loss Adjusted (%)	Total Output Loss from (Adjusted) Direct (%)	Direct Q Loss/ Loss of Utility Services (ratio)	Total Q Loss/ Direct Q Loss (ratio)
Tierney (1995)	Los Angeles/ Northridge	Electricity/ 36 hrs	Survey	8.3	—	1.9	1.9[b]	.23[b]	—
Rose-Lim (1996)	Los Angeles/ Northridge	Electricity/ 36 hrs	I-O	8.3	8.3	0.42[c]	0.55	.05	1.31
Rose-Guha (2004)	Memphis/ Hypothetical	Electricity/ First week	CGE	44.8	—	—	2.3[d]	.05[e]	—
Rose-Liao (2004)	Portland/ Hypothetical	Water/ First week	CGE	50.5	49.1	33.7[f,g]	41.0	.67	1.22
Rose-Liao (2004)	Portland/ Hypothetical	Water/ First week	CGE	31.0	30.7	21.3[f,h]	30.5	.69	1.43

a. Survey response incorporates direct resilience practices.
b. Includes only direct effects.
c. Resilience adjustments limited to time-of-day use, importance factor, and production rescheduling.
d. Model not able to incorporate very short-run elasticities; hence, simulates overly flexible result.
e. Numerator is total output loss, since direct and indirect output losses could not be distinguished in this model.
f. Resilience adjustments limited to conservation and inputs substitution for water.
g. Prior to any mitigation.
h. Following replacement of cast iron pipes by PVC pipes.

Table 2. Resilience Adjustments to Reduction in Electricity Availability to Business

Adjustment Type	Production Function Layer	Parameter	Applicable Time Frame
1. conservation of electricity	4th	$A_{4EL}\downarrow$	immediate to LR (decreasing)
2. conservation of energy	3rd (or 4th)	A_{3E} (or A_4) $\downarrow$	immediate to LR (decreasing)
3. conservation of various inputs	1st, 2nd, 3rd	$A_1, A_2, A_3 \downarrow$	immediate to LR (constant)
4. increased substitutability of other fuels for electricity	4th	$\sigma_4 \uparrow$	VSR to LR (increasing)
5. increased substitutability of non-energy for energy	3rd	$\sigma_3 \uparrow$	VSR to LR (increasing)
6. back-up generators	4th	$\sigma_{4EL}\uparrow$ (or $a_4 \downarrow$, $\beta_4 \uparrow$)	immediate to SR (constant)
7. electricity importance	4th	n.a.	immediate to LR (constant)
8. change in technology	all	potentially all	LR (constant)
9. time-of-day use	4th	n.a.	off-hours
10. production rescheduling	4th	n.a.	immediate to LR (constant)

Table 3. Direct Output Loss Estimates of Electricity Service Disruptions in the SCE Service Area, 2001

Sector		(1) Baseline Direct Electricity Use ($ million)	(2) Initial Direct Electricity Disruption (%)	(3) Total Work Time per Year (hrs)	(4) Initial Production Function Estimate of Direct Output Loss (thousand $)	(5) Recalibrated Production Function Estimate of Direct Output Loss (thousand $)	(6) Direct Output Loss Adjusted for Rescheduling (thousand $)
1.	Agriculture	2.84	3.66	8736	-6	-2	-1
2.	Mining	11.12	0.85	2080	-67	*	*
3.	Construction	16.38	4.21	2496	-1068	-2	*
4.	Food Processing	43.55	2.07	6240	-94	-1	*
5.	Petroleum Refining	29.35	1.93	8736	-74	*	*
6.	Other Non-Durable Mfg	167.95	4.67	6240	-426	-17	-1
7.	Primary Metals	36.93	2.34	6240	-84	*	*
8.	Semiconductors	4.06	55.37	6240	-163	-146	-1
9.	Other Durable Mfg	212.04	3.44	6240	-1169	-2	*
10.	Local Private Transportation	0.36	1.29	5824	0	0	0
11.	Other Transportation	27.92	4.89	5824	-579	-5	-4
12.	Communications	12.56	1.73	5824	-112	-55	33
13.	Private Electric Utilities	n/a	0.00	n/a	0	0	0
14.	Gas Utilities	2.12	0.46	8736	0	*	*
15.	Water Utilities	1.46	1.60	8736	-3	*	*
16.	Sanitary Services	0.43	1.73	8736	0	0	0
17.	Wholesale Trade	67.10	3.28	4368	-969	-87	-1
18.	Retail Trade	82.97	4.14	3744	-1128	-139	-28
19.	Real Estate	97.80	3.18	3744	-1161	-4	*
20.	Banking & Credit	14.85	1.50	2080	-338	-106	-11
21.	Security Brokers	2.18	1.10	2080	0	-39	-4
22.	Insurance	2.24	2.08	2080	-354	-70	-7
23.	Owner Occupied Dwellings	n/a	n/a	n/a	0	0	0
24.	Hotels & Restaurants	80.18	2.07	8736	-66	-2	-1
25.	Personal Services	15.17	4.10	2496	-163	-20	-8
26.	Business Services	66.16	2.25	2496	-1208	-137	-41
27.	Computer Services	2.65	5.02	3744	-192	-19	-6
28.	Entertainment	126.29	1.55	4368	-297	-95	-67
29.	Education	6.86	0.00	2080	0	*	*
30.	Health & Social Services	58.67	0.00	4368	0	-1	*
31.	S & L Electric Utilities	n/a	0.00	n/a	0	0	*
32.	Local Public Transportation	20.41	0.00	5824	0	-1	*
33.	Other Government	57.34	1.40	2080	-203	-261	-52
	Total	1269.94	2.81		-9924	-1210	-266

*Less than $500.
n/a not applicable.

Table 4. Total Economic Impacts Of Electricity Service Disruptions In Los Angeles County, 2001

	Electricity Input		Output	Output Change From Electricity Outage (%)			
	Baeseline	Direct	Baseline	Initial	Recalibrated	Indirect	Total[b]
Sector	(million $)	Disruption (%)	(million $)	Direct (PE)[a]	Direct (PE)[a]	(GE-PE)	(GE)
1.Agriculture	5.77	2.46	1398	-3.51	-1.07	-0.53	-1.6
2. Mining	22.62	0.61	2589	-5.37	-0.04	-1.58	-1.62
3. Construction	32.98	2.53	28770	-9.27	-0.02	2.11	2.09
4. Food Processing	88.47	1.28	14744	-3.97	-0.03	2.83	2.8
5. Petroleum Refining	59.50	1.38	11404	-5.68	-0.04	3.88	3.84
6. Other Non-Durable Mfg	341.07	2.39	33435	-7.96	-0.32	-0.55	-0.87
7. Primary Metals	75.06	1.78	3192	-16.38	-0.03	-35.78	-35.81
8. Semiconductors	8.26	33.26	1133	-89.92	-80.12	75.27	-4.85
9. Other Durable Mfg	430.66	2.48	63364	-11.51	-0.02	-11.28	-11.3
10. Local Private Transportation	0.71	0.72	1039	0.00	0.00	2.23	2.23
11. Other Transportation	55.51	2.90	21407	-15.77	-0.14	-1.47	-1.61
12. Communications	25.47	0.60	15674	-4.16	-2.04	1.96	-0.08
13. Private Electric Utilities	0.07	0.00	2349	0.00	0.00	-1.09	-1.09
14. Gas Utilities	4.30	0.08	4738	0.00	-0.02	11.98	11.96
15. Water Utilities	2.96	1.51	381	-6.30	-0.79	7.74	6.95
16. Sanitary Services	0.27	1.12	1149	0.00	0.00	4.96	4.96
17. Wholesale Trade	136.19	2.02	5676	-11.87	-1.07	1.52	0.45
18. Retail Trade	168.52	2.44	27761	-15.22	-1.87	2.95	1.08
19. Real Estate	197.58	1.74	31230	-13.91	-0.05	0.21	0.16
20. Banking & Credit	29.96	0.88	19759	-3.56	-1.11	1.63	0.52
21. Security Brokers	3.44	0.43	8153	0.00	-1.01	0.86	-0.15
22. Insurance	4.47	0.82	11733	-6.27	-1.24	1.16	-0.08
23. Owner Occupied Dwellings	n/a	n/a	31272	0.00	0.00	-0.21	-0.21
24. Hotels & Restaurants	162.55	1.20	14383	-3.98	-0.11	2.41	2.30
25. Personal Services	30.73	2.31	4300	-9.44	-1.14	2.71	1.57
26. Business Services	134.21	1.07	59026	-5.11	-0.58	0.60	0.02
27. Computer Services	5.38	2.88	6035	-11.93	-1.19	0.89	-0.30
28. Entertainment	256.38	0.73	39098	-3.32	-1.06	-5.87	-6.93
29. Education	13.92	0.00	5015	0.00	-0.02	2.94	2.92
30. Health & Social Services	118.73	0.89	30138	0.00	-0.01	2.86	2.85
31. S & L Electric Utilities	0.00	0.00	2425	0.00	0.00	-1.12	-1.12
32. Local Public Transportation	41.55	0.00	408	0.00	-1.72	16.85	15.13
33. Other Government	116.55	0.68	36916	-1.15	-1.47	2.42	0.95
Total	2573.84	1.61	570093	-7.07	-0.74	-0.55	-1.29

[a]From partial equilibrium analysis. [b]Following CGE simulation.

Chapter 10

Beyond Capture

A View of Recent U.S. Telecommunications Regulation[*]

Richard Simnett
Telcordia Technologies (TM)

1. INTRODUCTION

Regulators and their agencies have often been accused, over the years, of being 'captured' by the industries they regulate. The allegation is that the agency comes to understand the industry's problems so well that it helps the industry to deal with them, but loses sight of the of the ratepayers whose interests the commission exists to serve. Arguably, the FCC has moved 'beyond capture' with its recent actions in media, cable, telecoms and Internet.

The argument would run that in the series of industries that it regulates (or industries that it might regulate but chose not to, such as the internet and information services) the FCC has 'rolled over for' or been 'captured by' the major incumbents. This phase of FCC behavior contrasts with the policy prevalent earlier where incumbents, at least in the telephone business, were treated with suspicion and entrants, especially the internet-related businesses from data oriented local exchange carriers to internet service providers, were recipients of, if anything, positive discrimination by the FCC. Since the 1996 revision of the Telecommunications Act the commission has allowed consolidation of TV and radio stations into larger chains, and allowed multiple stations under common ownership in major markets. The FCC has

[*] This paper is a personal view, and does not reflect the views of my employers, nor of any clients. Errors are solely my responsibility.

also permitted integration of TV networks and the studios that supply programming, consolidation of cable system operators into much larger companies, and consolidation of the seven original Regional Bell Operating Companies (RBOCs) and GTE into four dominant companies.

All of these actions enabled the largest companies in their respective regulated markets to grow larger and develop stronger market positions by acquisition of adjacent (geographically or in related market spaces) incumbents rather than by competitive endeavors. The FCC has refused to regulate cable modems to ensure competitive access to the systems, is proposing to do the same for DSL or telephone company broadband services however provided, and deregulates telephone company services as 'competitive' once minimal competition can be demonstrated. The FCC has also failed to require performance on the various pro-competition commitments made in the merger proposals while FCC permission for mergers was sought. It allowed these companies into the long distance business, and is currently proposing to remove or reduce their obligations to make facilities available to their competitors through unbundled network elements specifically, (UNE-P or UNE-L).[1] This behavior contrasts with earlier FCC policies such as the doctrines of 'comparably efficient interconnection' and 'arms-length' relationships between monopoly and competitive businesses within a company.

Neither the FCC nor the respective state commissions has enforced performance of the investment commitments made by the Bell companies when they sought, and achieved, an end to rate of return regulation[2]. The effect has been to provide monopolists with the opportunity to earn economic profits if they can be achieved by cost reductions while subject to a price constraint.

In this paper I present a brief history of regulation and regulator-managed competition in the US, focused on the telephone and telecommunications business from the late 1950s until today. I try to place current controversies in enough context for the reader to decide whether 'capture' is an adequate explanation for the FCC's behavior.

---

[1] Where UNE-P refers to the entire platform including switching and UNE-L refers just to access to the local wires.

[2] For an unrestrained attack on these lines, directed particularly at Verizon, see "The Tell-The-Truth Broadband Challenge to Verizon vs. "The Verizon 100 Megabit Challenge"" at http://www.newnetworks.com/tellthetruthverizon.htm which has a list of references as well as a long series of quotes from various Verizon predecessor companies pledging fiber to the home rollouts in 'incentive regulation' proceedings in their home states.

2. REGULATORY BACKGROUND

The FCC reintroduced competition in the US telecommunications services market as a deliberate act of policy with its 1959 Above 890 Megacycles[3] decision. This authorized spectrum allocations for private microwave system use by corporations, providing an alternative to AT&T's private line services.

AT&T's defense of its long distance monopoly resulted in more than 25 years of regulatory controversy, and great advances in the theory and practice of regulatory economics and industrial organization theory. The forces released by the Above890 decision ultimately transformed the capital markets (the invention of junk bonds as well as other innovations enabling entrants in many industries to be financed). They also led to the divestiture of AT&T and the transformation of the industry worldwide from state or regulated private monopolies into regulator-managed competition, some with more active transition plans than others for 'unregulation' or normalization of the industry.

In economic terms we can abstract from the controversies of the period and the periodic antitrust actions against AT&T, and even from the policy controversies since the divestiture of the RBOCs in 1984, and find three continuing, unresolved, and fundamentally political, tensions:

1. A tension between social goals (mostly universal service, but also law enforcement access and emergency services, and capacity sufficient to meet emergency situations) and the desire for minimum efficient costs to be reflected in commercial service prices. (The efficiency-equity argument about proper industry governance captures the essence of this tension.)
2. A question about the existence and proper bounds of any natural monopoly in the industry, and a constant pressure to limit the scope of any natural monopoly to the greatest extent possible: the corollary to this is that the scope for competitive action should be as great as possible, consistent with economic efficiency.
3. A dynamic tension between static economic efficiency for current services and their prices (allocative efficiency), and the need to provide incentives for investment so that new services can be provided (dynamic efficiency). This is especially noticeable now in the areas of broadband, advanced wireless services, and the transition to digital television over the air and on cable. This dynamic issue could also be regarded as an extension of the issue in (1) above, because it can be stylized into a transfer issue: should today's ratepayers for yesterday's

[3] Formally, Allocation of Microwave Frequencies Above 890 Mc., 27 FCC 359 (1959)

services pay high prices so that some other set of users can get earlier access to expensive-to-provide new services? In other words, should the companies fund new investments internally, or should new investments be funded by newly raised capital as if they were separate businesses?

These tensions appear in many of the controversies if the actors in each are examined from the perspective of their expected rents from the alternative policies or, more broadly, what's at stake.

3. THE STARTING POINT: THE BELL SYSTEM MONOPOLY AND ITS BOUNDARIES

AT&T acted as if its monopoly over long distance service and the rented equipment connected to phone lines were essential parts of its bargain with regulators. The monopoly was seen as the compensation for its commitment to make universal access to telephone service available at affordable rates throughout the United States. In areas where AT&T owned the local service provider (a Bell Operating Company) the 'affordable' local service was subsidized by other services in several ways, 'hidden in plain view' and set out in regulatory proceedings. The first mechanism was the residual ratemaking process implicit in what was customarily called 'value of service' ratemaking.

'Value of service' ratemaking translates into economic terms as setting the rates for the set of unsubsidized non-basic services at profit maximizing levels, so as to maximize their 'contribution' to the subsidy pool. The process ran as follows: set all rates for discretionary residential and non-residential services[4] at close to monopoly levels. Compare the revenues generated by all of these services taken together to the total revenue requirement on the permitted rate base of regulated services. The rates for basic residential services can then be set to cover the 'residual' revenue requirement. The New York Public Service Commission, chaired by Alfred Kahn, explicitly went through this process in directing the New York Telephone Company to increase its proposed rates for Princess phone rentals, because the Commission-ordered studies conducted found that the proposed prices were still in the inelastic range.

---

[4] These services included all long distance services, business services, business telephone features such as key sets and switchboards, vertical features on voice services, non-basic telephone set rentals such as the Princess phones or Mickey Mouse phones, directory services such as Yellow Pages and unlisted numbers, and touch tone services when introduced.

The organization of these subsidy flows was complicated by the existence of independent (non-Bell) telephone companies and of both state and interstate jurisdictions. The methods used in traditional practice to share the revenues (and subsidies) among all these parties consisted of various rules for allocating the regulator-approved rate base of companies to the interstate jurisdiction, to be covered by interstate rates, and within states between local and long distance services. In effect, a high cost company in a rural area could transfer most of its investments into the long distance cost pool, and thereby earn the same rate of return on these assets as AT&T did on its long distance assets.

3.1 The Above 890 Megacycles Decision[5]

The FCC decided to allow private companies to set up their own microwave long distance networks. The actual proceeding concerned the award of spectrum licenses for the bands above 890Megacycles per second. Third party microwave manufacturers, who were among the petitioners, could not sell to AT&T, which bought only from Western Electric, and had no potential US commercial market unless private networks were authorized.

The Above 890 decision directly threatened the cross subsidy system. AT&T's rates for long distance service, including private lines, were set at levels well in excess of any measure of their direct costs, and certainly much higher than later theorists would describe as their 'stand-alone' costs or incremental costs.

The Above 890 decision raised the demand elasticity for private line services into the elastic range for large companies between the city pairs where they demanded most circuits. AT&T's response to the Above 890 decision was rational given the value of service ratemaking framework. In December 1960 AT&T introduced a bulk-rate discount tariff for private line services, called Telpak, which was intended to provide AT&T services to its users with the cost structure of a private microwave network. Discounts ranged from about 50% at the 12-circuit level to about 85% at the 240-circuit level. Later AT&T introduced route-based pricing differentials on the busiest routes (the Hi-Lo tariffs). These offers broke with the traditional practice of uniform rates regardless of volume or the locations of the end-points of calls or private lines, and led to a regulatory controversy on price discrimination and how to manage it that lasted for more than twenty years, and that in some form still continues today.

The issue for the regulators became the interpretation of the rate standard of the 1934 Telecommunications Act: 'just, reasonable and without undue

---

[5] Formally, Allocation of Microwave Frequencies Above 890 Mc., 27 FCC 359 (1959)

discrimination.' AT&T, in modern terms, was pricing selected services for selected customers at market-determined rates. Arguably, in fact, AT&T was then using what has since become the FCC's TELRIC pricing standard (use forward-looking long run average incremental cost of a total service) precisely to prevent an entrant from building an additional network to provide the service, since the entrant could not profit from it at these rates.

The FCC conducted an historic average (later called fully distributed) cost study and determined that the Telpak rates in the higher volume ranges were too low, producing less than the authorized 7.4% rate of return. The FCC ordered the low volume services withdrawn and raised rates for the 60 and 240 circuit options in 1964; AT&T appealed and, in a new twist, so did Telpak users, but in 1966 the Court of Appeals upheld the FCC order. The effective date of the order was deferred until August 1967. New tariffs were filed in 1968, became subject to legal challenges, and eventually became effective in February 1970. The FCC adopted fully distributed historic costing as its standard in 1976.

The decisions on private network authorizations evolved over time, with the FCC authorizing new carriers such as Datran to build intercity private networks for third-party use. This allowed smaller private companies to benefit from the reduced cost of microwave technologies, and increased the pressure on AT&T. Most of these carriers went bankrupt, unable to gather enough customers in the face of AT&T's various reduced rates and the allegedly less-than-enthusiastic provision by Bell companies of any termination services needed by their customers. The FCC ordered local carriers to interconnect, providing the circuits connecting these new carriers and their customers' premises, in 1970.

3.2 WATS and Execunet

The switched long distance service was also priced well above any measure of direct cost, but after the introduction of direct distance dialing in the 1950s, and with advances in both switching and toll transmission technologies, costs were falling rapidly and were expected to continue to do so. In the late 1950s Wide Area Telecommunications Service (WATS) was introduced as an unlimited use flat rate service for business customers. The subscriber rented a special WATS access line or group of lines, and all calls over that line or lines to the subscribed calling area (a specified set of geographic areas) were free. The customer thus had the problem of managing his calling by geography and offered load to get the best value from the service, but had the opportunity to reduce calling costs considerably if he did so well.

In the early 1960s Inwats service was introduced, with the same pricing concept, but offering a business the option of receiving calls, free to the calling party, from the specified calling area, over the rented set of Inwats access lines. Inwats later became 800 service.

MCI (which was originally an acronym for Microwave Communications Inc) received FCC permission[6] in 1969 to build a shared private line network between St Louis and Chicago. This was a high-density route, where even at discounted tariffs MCI could expect to earn high margins, competing with AT&T's post-Telpak nationally averaged rates, which included its contributions to the costs of local connecting networks. In 1975 MCI began Execunet: a shared private line service where the local end connecting private lines were Foreign Exchange lines rented from the local company. (A Foreign Exchange line is, in effect, a business telephone number in an exchange, whose other end was the MCI network.) MCI customers could access the MCI network by dialing a local number, and then dial any number in the US. The call would be completed by MCI as a local call from its distant end-point or if necessary as a long distance call. This meant that MCI was providing toll service functionally equivalent to WATS, if customers had full time private lines connected to MCI, and nearly equivalent to basic MTS for customers who dialed an FX number. The FCC had made no formal finding that competition for switched long distance services was in the public interest. In June 1975 the FCC ruled that MCI was not authorized to provide switched long distance service, and rejected the Execunet service. MCI appealed, and the courts reversed the FCC decision in 1978.

This service raised the stakes considerably for the subsidy structures in the industry, because AT&T's long distance services paid almost 30 cents per minute of use towards the costs of the originating and terminating local carriers, and using business lines MCI calls paid only local usage rates, if any, that were typically less than 5 cents per minute.

In 1979 the FCC ordered unlimited resale and sharing of AT&T's tariffed services, including WATS. This had the effect of limiting AT&T's ability to price discriminate on the basis of customer size, for MCI and Sprint could become AT&T's largest customers for any such tariff. AT&T would clearly lose if it set discounts so great that prices could drop below its incremental costs, so this decision was one of the first where the FCC consciously set out market rules to govern efficient pricing behavior. This ruling also allowed MCI, Sprint, and others to provide national services, including areas where their own networks did not exist, so it increased the elasticity of substitution between AT&T and its competitors in retail long distance services significantly.

---

[6] Microwave Communications Inc. 18 FCC 2d (1969)

The FCC proceedings trying to deal with the subsidy problems led to a series of actions defining carrier access services, introducing the Exchange Network Facilities for Interstate Access (ENFIA) tariffs proposed in 1978. These eventually offered an effective discount for MCI and other entrant carriers of around 75 to 80% compared to AT&T long distance, and to the program for equal access services. These arrangements were unaffected by the 1984 divestiture of AT&T.

The effect of high explicit per minute access charges was to encourage long distance carriers to migrate customers on to directly connected private lines wherever possible. The first competitive local exchange carriers, such as Teleport and Metropolitan Fiber Systems, were set up to build these 'bypass' networks serving large concentrations of business customers. Regulatory filings and FCC reports to Congress on Bypass Adoption Rates showed that the largest business customers, accounting for a disproportionate share of revues and minutes, were the earliest adopters of these services.

The bypass threat led the FCC to reduce per minute access charges by the introduction of a planned migration to flat rate Subscriber Line Charges, but this planned migration was not completed because of substantial political pressure from Congress.

3.3 Terminal Equipment Cases

These cases addressed the boundaries of the monopoly from a different angle, limiting the boundary in terms of the network edge rather than the ability to build parallel networks.

The first was the 1956 Hush-A-Phone case, where an FCC decision in AT&T's favor was overturned by a Federal Appeals Court. However, the FCC and the courts actually set out a number of then-innovative adjudicatory standards, akin to the various compensation principles in welfare economics: if the device was privately beneficial, without any public detriment (e.g. on the quality of service received by the person at the distant end, or by harming the network more generally) it would be 'unjust and unreasonable' to forbid its attachment to the network. After the remand the FCC ordered AT&T tariffs revised to permit 'harmless interconnection devices' to be attached.

The next significant case was Carterfone, which permitted the attachment of two-way private radio networks to the telephone system. In 1968 the FCC authorized connection of any non-Bell equipment to telephone lines so long as the telephone company was allowed to install a protective device at the edge of the network. The broader language of the decision found that the

subscriber had a *'right reasonably to use his telephone in ways that were privately beneficial without being publicly detrimental*[7]*'*

This decision created a large potential market for privately supplied equipment, and was followed in 1975 by an FCC registration program for independent equipment to be attached without any protective device, to a standard network interface (the familiar RJ11 and RJ45 jacks).

3.4 The Computer Inquiries

The FCC drew a boundary between telecommunications services on the one hand and computing and information services on the other in a series of proceedings called, for short, the Computer Inquiries. In 1971 the FCC drew a distinction between data communications, in which the content of a transmission was unchanged, and data processing, in which changes to the transmitted information might be made. Data transport across the network was permitted, but the data could not be changed or processed. Telephone companies could not perform data processing in connection with any regulated service.

In 1976 the FCC began a Second Computer Inquiry, recognizing the increased interdependency of computing and communications. The 1981 ruling permitted AT&T to provide data processing (enhanced services where the content of the communication was acted upon in some way, even if only by temporary storage) but only on an unregulated separate subsidiary basis, with strict separation from any regulated entity or service. Other providers were permitted to transport data on an unregulated basis.

The separate subsidiary requirement provided a lucrative opportunity for specialists in cost allocations and other regulatory arcana to argue that cross-subsidies flowed from the regulated business to the unregulated ones. A new regulatory standard was developed in these proceedings and their various appeals to define the necessary standard by which third parties could fairly compete with the Bells' separate subsidiaries: they must receive *'Comparably Efficient Interconnection'*. This held that any dominant carrier providing an enhanced service must offer the same basic services interface to third parties as it uses in its own enhanced services, and under the same terms and conditions. This included not just prices but network change notifications and so on.[8]

---

[7] Carterfone, 13 FCC 2d 439 (1968)

[8] It is interesting to note the parallels here, and the divergences, between the rulings made and enforced against telephone companies and the comparatively toothless antitrust ruling made against Microsoft, which has a greater position of dominance in its markets than AT&T had in the 1970s and the Bell Companies and AT&T have enjoyed since then. Another parallel situation with different treatment is the position of cable companies and

In practice it proved very time consuming and expensive for any separate subsidiary to get going, because its rivals argued that it was specially privileged, or alternatively that it was cross-subsidized, every step of the way to market. Any potential cost synergies between the telephone company and these subsidiaries were vitiated, replaced by expensive to maintain and audit 'Chinese Walls' that certainly delayed time to market and increased costs.

This policy and its periodic revisions did result in dynamic efficiency losses. AT&T had developed a voice mail service in the late 1970s, and proposed to introduce it in 1981, but scrapped it when the Second Computer Inquiry defined voice storage as an enhanced service. Many 'information services' on the same general lines as the French Minitel were planned too, but never introduced in the US because the difficulties of operating on a third-party CEI (Comparably Efficient Interconnection) basis were too great. In retrospect the FCC attributes the development of the Internet and online services in the USA to these decisions, but this may be too self-congratulatory. A Third Computer Inquiry was undertaken in 1989.

Another 'lost technology' was ISDN, a digital technology that made second lines available on a single wire pair with both lower capital and operating costs than adding additional plant, but at the expense of new consumer premise equipment. The separation of the equipment from the service made it essentially impossible to deploy this service on a large scale in the US. In Germany if customers wanted second lines they were provided by ISDN, which amongst other things made the network ready for a subsequent easy upgrade to DSL.

Another example of what might have been, but which was forbidden by FCC decisions, may help. Nynex proposed to build a bandwidth-on-demand fiber access network for major business customers in New York City. This service required that terminal equipment on the customer premises be secure from interference by third parties, and addressable from the Nynex network so that the bandwidth provided could be changed instantly and remotely. The FCC found that this terminal was customer premise equipment and that it could not be provided as part of, or required by, a tariffed service.

The FCC has repeatedly made, but never enforced, similar rulings against cable TV systems. Repeated rulings find that set-top boxes should be available at retail, and that normal television sets and other consumer video equipment should be 'cable-ready' in the fullest sense, by means of published cable industry interfaces. However, the set-top box market remains largely proprietary, and is controlled by cable operators, who typically rent the boxes to consumers as the Bell System provided telephones.

---

third party ISPs wanting access to cable modems or unaffiliated cable programmers wanting access to cable delivery systems for their programming.

4. THE DIVESTITURE OF AT&T

The Department of Justice antitrust action against AT&T was concluded by a consent decree in 1982, which was modified by the Court and implemented on 1st January 1984.

The DOJ had sought the divestiture of Western Electric, AT&T's manufacturing arm, but would allow the continued integration of AT&T and the Bell Operating Companies providing both local and long distance services. AT&T did not believe that continued integration of monopoly local and already competitive long distance telecommunications services markets was practicable: it was bound to lead to continuing multi-year legal and regulatory battles with no clear resolution at all likely.

The divestiture that resulted was originally intended to provide a 'clean break' between those services markets where monopoly power might be sustainable, and those where it would not. Given the FCC and Court decisions on long distance entry the theory was that the local exchange was and would remain a practical monopoly, and that the companies owning these exchanges and providing those services should be separated from AT&T. There would thus be a true separation between regulated monopoly Bell Operating Companies and what AT&T mistakenly believed would be an unregulated competitive long distance company. There would no longer be any incentive for BOCs to deny or delay interconnection with any long distance carrier or set of carriers, nor to discriminate in pricing.

The tension between allocative efficiency and cross subsidy for social purposes was felt even before the decree was entered. Judge Greene expressed concern for the financial viability of the BOCs, once separated from AT&T (which I find peculiar since his findings that motivated AT&T to negotiate an end to the case included one that there could be no valid regulatory defense for AT&T's past actions defending its monopoly). To address these concerns several profitable but potentially competitive businesses were assigned to the BOCs that were originally intended to go to AT&T, including the Yellow Pages and cellular services.

4.1 The Long Distance Market

The divestiture and FCC proceedings brought equal access tariffs and equal access services for all long distance carriers. However, it did not bring lessened regulation to AT&T for 11 years, until AT&T's market share had fallen below 60%, and AT&T could demonstrate that its national long distance rivals had sufficient capacity to carry all traffic if AT&T attempted to raise prices by withholding capacity from the market. The standard the FCC then applied was a short- and long-run price and supply elasticity

standard. As we shall see, this is quite different from whatever standard the FCC is implicitly applying today in its local competition rulings.

The divestiture did not change the legal framework controlling the industry. It remained the 1934 Communications Act, which stated its goal in the preamble as ". . .to make available, so far as possible, to all the people of the United States a rapid, efficient, nationwide, and worldwide wire and radio communications service with adequate facilities at reasonable charges, . . ."

The term 'efficient' in the above clause had been the basis of FCC pro-competitive policies for years, where efficient was read as economically efficient, and this desired state could be ensured only by competition or competitive forces. However, after the divestiture the FCC was concerned not only for the viability of the divested BOCs but also for the entrants competing with AT&T in the long distance market. As the transition to equal access occurred MCI and Sprint had to pay rising prices for their access services, and AT&T repeatedly filed volume discount plans for the residential and small business markets, the core markets for MCI and Sprint. These were repeatedly rejected although AT&T showed that they not only covered their own costs but could increase the company's overall profits: since AT&T was still rate of return regulated this meant that even smaller users would indirectly benefit from the discount plans under a residual ratemaking philosophy.

By the early 1990s MCI and Sprint had apparently survived the transition to equal access, and AT&T had lost so much market power (and was at a 55% market share) that it was finally declared non-dominant in 1995.

4.2 Regulation of the RBOCs

After the divestiture much of the regulatory action moved to the regulation of the ILECs (Incumbent Local Exchange Carriers). Several critical steps occurred which changed the regime markedly from that which had controlled the Bell System.

Equal access reduced one area of persistent controversy; formalized access charges and universal service contributions (in effect a gross receipts tax) removed another. The greatest change however was the adoption of so-called incentive regulation.

The basic driver for incentive regulation was a quest for economic efficiency, in the static X-efficiency sense. Incentives were sought to ensure that ILECs produced on or near the efficiency frontier. The regulators had long suspected, and entrants had long argued, that the Bell System was inefficient in this sense, and the proposed CPI-X price regulation scheme seemed to offer an attractive option. Both states and federal authorities

adopted this price cap regulation for the RBOCs, and authorized rate of return determinations became unusual. It is now the dominant form of regulation. In some states only basic residential local service remains regulated: e.g. Utah, where if vertical features are purchased the entire package becomes an unregulated service, and Ohio where only the first line is regulated. Wyoming legislation mandated that rates had to be rebalanced and cost justified, so business and residential rates converged.

This change in regulation was a bargain: the companies promised greater efficiency and innovation (new networks, new services, broadband for all[9], . . .) because of the greater returns they could earn on their investments. Verizon's achieved returns (including directory revenues) under incentive regulation in Pennsylvania have been estimated at 24.26% in 2001; 26.19% in 2000; 29.40% in 1999; and 25.33% in 1998, compared to a probable authorized rate of return, in a low interest rate environment, of some 7-8%.

As we shall see, the 1996 Telecommunications Act changed the rules governing the RBOCs very considerably. However, some active controversies have not really changed much at all. In particular, ten years after the first round of state proceedings promising broadband deployments in exchange for incentive regulation, the RBOCs are still promising broadband investments if they receive sufficient incentives, but now at the Federal level.

5. THE 1996 TELECOMMUNICATIONS ACT

5.1 Principal Clauses in the Act

The 1996 Act reflected a political compromise based on a number of problems with the post-divestiture arrangements. Local telephone competition had not emerged (and under the theory of the divestiture it should not have been expected to do so, since it was a natural monopoly) except for a small number of very high revenue areas and customers (central business districts, suburban and highway business centers, and very large enterprises where direct fiber–optic links could be justified by a CLEC or

[9] See, for example, Bell Atlantic announcements referenced at Pennsylvania PUC Docket No. P-930715F0002, a review of Verizon's biennial report on its network modernization plan. Verizon sought to amend its 1994 plan promising symmetric 45Mbps fiber service to customers throughout Pennsylvania by redefining the commitment to 1.544Mbps asymmetric DSL service. Verizon Pennsylvania, Inc. Petition and Plan for Alternative Form of Regulation under Chapter 30; 2000 Biennial Update to Network Modernization Plan, Docket No. P-00930715, Order at 22 (May 15, 2002), and the PUC order of 18 July 1995 accepting the company's commitment to building a statewide 45Mbps symmetric network.

IXC)[10]. However, these areas were the source of a disproportionately high share of ILEC revenues, and the ILECs sought an offsetting revenue opportunity to enter the long distance market.

The Act contains a few key clauses governing telecommunications regulation for the wireline carriers. It preempted state initiatives by setting out an FCC-centered framework for regulation, and while its stated purpose was to change the fundamental governance structure of the industry to one based on effective market-based competition it actually directed the FCC to conduct 80 (!) new regulatory proceedings to begin the transition. The possibility of different experimental market designs and rules based on state actions (New York, Illinois and California had begun to open their ILEC markets to competition, each in different rule-making contexts) was lost.

The deregulatory clauses are Sections 10 and 11 that set out the 'sunset' procedures for industry regulation. Section 10 allows the FCC to 'forebear' from any regulation if it decides that that regulation is no longer necessary 'in the public interest' and that no state may regulate what the FCC decides to forebear from regulating. Petitions to forebear may be brought by any interested party and are by default granted unless denied by the FCC within a year. Section 11 provides that the FCC shall review all regulations every two years and must decide which are still required to protect the public interest.

The basic pricing clauses (Sections 201-2) of the Act are the same as in 1934: 'just and reasonable' rates without 'undue' discrimination are required. This does not provide a very specific guide for FCC action, but pricing rules are specifically mentioned in three other clauses, and in giving meaning to these differences the courts have thrown FCC policy into disarray.

Section 251 c3 says that ILECs must "provide, to any requesting *telecommunications carrier* for the provision of a *telecommunications service,* nondiscriminatory access to network elements on an unbundled basis at any technically feasible point" with pricing under the rules of Section 252. It also orders ILECs to "provide such unbundled network elements in a manner that allows requesting carriers to *combine* such elements in order to provide such telecommunications service." [Italics added]. Section 251 c also orders that resale of ILEC services and collocation of other carriers' equipment with ILECs' be permitted. In another clause that has caused great legal controversy, Section 251 d2B orders the FCC to consider, in ordering unbundling, whether "the failure to provide access to such network elements would *impair* the ability of the

[10] Even today, the potential market for rival access networks is small. Cogent Communications, a fiber-optic based national carrier, estimated the potential market for its services at 60,000 buildings in the entire USA.

telecommunications carrier seeking access to provide the services that it seeks to offer" [Italics added]

Section 252 sets out the procedures for negotiation between carriers. Voluntary agreement is the first choice, but if agreement cannot be reached State commissions are to mediate. If agreement is not reached within 135-160 days, either party can request compulsory arbitration by a State commission. This whole process can last no longer than nine months. The state commission must ensure that the final agreement meets any FCC standards under Section 251, and for requests under 251c2 and 251c3 prices shall be based [Section 252 d1ai] "on the *cost (determined without reference to a rate-of-return or other rate-based proceeding)* of providing the interconnection or network element (whichever is applicable), and (ii) nondiscriminatory, and [252 d1B] may include a reasonable profit." [Italics added].

Section 252 d2 orders that interconnection rates must allow for each carrier to discover the other carriers' costs for traffic transport and termination, "on the basis of a reasonable approximation of the additional costs of terminating such calls."

Section 252 d3 orders that the wholesale services mandated by section 251(c)(4) shall be priced "on the basis of *retail rates charged to subscribers* for the telecommunications service requested, *excluding the portion thereof attributable to any marketing, billing, collection, and other costs that will be avoided by the local exchange carrier.*" [Italics added.]

Two more sections have contributed to the implementation failures of the FCC, after court appeals. The basic problem is the difference in language between these clauses and sections 251 and 252.

The first, Section 271, applies only to the Bell companies (and so not, for example, to major ILECs like Sprint or GTE) and provides that the Bells may enter the long distance business when they have satisfied a 14-point competitive checklist. This list provides for the unbundling of network elements, and provision of wholesale services in accordance with the requirements of sections 251(c)(2)(3) and 252(d)(1), but does not repeat the pricing rules or the requirement that ordering carriers be able to *combine* the network elements they purchase on an unbundled basis.

The second is usually referred to as Section 706, from its numbering in the Senate Bill S652 which amended the 1934 Act. This provides for Advanced Telecommunications Incentives, and orders Commissions to "encourage the deployment on a reasonable and timely basis of advanced telecommunications capability to all Americans … by utilizing, in a manner consistent with the public interest, convenience, and necessity, *price cap regulation, regulatory forbearance, measures that promote competition in the local telecommunications market, or other regulating methods that*

remove barriers to infrastructure investment." [Italics added.] Advanced telecommunications capability is defined as high-speed, switched, broadband telecommunications capability that enables users to originate and receive high-quality voice, data, graphics, and video telecommunications using any technology.

5.2 FCC implementation and current controversies

The FCC has run into numerous legal entanglements in implementing this Act. The key terms italicized above have proven to be particularly problematical. The Courts have overturned the FCC on numerous grounds, mostly having to do with the FCC giving insufficient weight to one or another of these terms or phrases, since the rules for statutory construction require that every term have meaning.

5.2.1 Cost Estimation

It is clear that the FCC is intended to deregulate the industry by promoting competition and implementing regulatory forbearance where competition permits. Rate of return regulation is clearly discouraged and price cap regulation is encouraged. Carrier negotiations leading to agreements are preferred to regulatory prescriptions, but state commissions can act as arbitrators if necessary, basing their decisions on 'costs *(determined without reference to a rate-of-return or other rate-based proceeding)'*.

The costs so determined have been a continuing source of controversy since 1996, because if traditional regulatory methods cannot be used then cost models must be substituted, and all litigants come prepared with their own experts, models, and relevant data. I shall not spend any further time on these cost model controversies.

5.2.2 Forbearance: Market Tests

The FCC has forborne from regulating markets where it has found competition exists, but the standards used for this determination have been subject to extended litigation and controversy.

A continuing controversy relates to markets for private line services between business users and inter-exchange carrier (IXC) networks (the same tail-end circuits that MCI had a hard time getting from the Bell Operating Companies before divestiture). The CLECs provide these services in certain areas, but their networks are by no means coextensive with the ILECs, even in business districts. Currently litigation is continuing over whether the FCC

has properly decided the issue: AT&T, MCI and others argue that private line rates have doubled in areas where the FCC has deregulated the rates, and that the facts show that there is no alternative to the ILEC monopoly. The argument, in economic terms, is one of supply elasticity for the particular address where the customer in question is located. Networks can only offer service to addresses they reach so the availability of several competing suppliers in, say, Wall Street, New York City, does not show that there is competitive pressure on Verizon private line rates in Queens, or even elsewhere in Manhattan. It does not even serve to show that the Verizon rates are limited by the threat of entry (i.e. that the market is contestable) elsewhere in the city.

As a practical matter, the details of the market analyses that would be required for the FCC to meet the proposed AT&T/MCI tests for forbearance are probably beyond the FCC's capacity to administer, and the transactions costs of forbearance in these circumstances would make it a difficult process to complete, for each market (however defined) in the allotted time under the Act.

The principal capital costs of building a local facilities-based network are in the permit and civil engineering (holes, trenches, ducts . .) processes, not the network components themselves. All these costs are sunk, once the network is built, and the network has no alternate use except for serving the particular addresses it passes. This is the same asset-specificity problem as a railroad or pipeline spur to a particular oil or gas well, or a major freight shipper: if the user chooses an alternative transportation medium the asset has no value.

In these circumstances incumbents have many opportunities to deter entrants, and the history of cable overbuilding shows this. Florida Power and Light funded a cable over-builder in Florida, which was bankrupted when the incumbent cable operator offered free service for a year to the households the over-builder could serve. The incumbent's salesmen followed the entrant's construction crews. The effect was that the entrant spent capital to almost no avail, and the incumbent actually raised prices in areas not directly threatened by competition.

In the continuing private line controversy there is little chance for an alternative network provider to emerge that will not be vulnerable to this kind of incumbent strategy. The only plausible contender is some new terrestrial radio technology with an appropriate spectrum allocation, but many companies in that space were bankrupted several years ago. AT&T, MCI and Sprint have all experimented with these technologies but have yet to find a winning technology/market combination, except fiber to very dense areas. The courts have found that the FCC must consider the particular circumstances of local markets to justify regulation, and even a finding of

'impairment' for CLECs in considering mandatory unbundled element availability under Section 251.

My view of this situation is that entry in limited markets has led the FCC to forbear from regulation, because high regulatory processing costs have made it impossible, administratively, to do otherwise. If the incumbents have raised prices in response, the FCC can point to that as an incentive for facilities-based competition, because higher prevailing prices offer higher potential profits for entrants. However, this incentive is weak if the incumbent may engage in price discrimination (as the incumbent cable companies did when faced with entry). In earlier times the FCC preempted long term contracts between the Bell System companies and their customers to make entry easier for PBX manufacturers and other carriers, so that the power of incumbency was weakened. It might be appropriate for the FCC to impose market conduct rules for incumbents to achieve the same market-opening results today. A requirement that ILECs maintain their announced prices as uniform minimum selling prices (umbrellas) for some period of years could make the linkage between pricing and entry incentives work, especially if joined with uniform pricing rules within limited geographic areas. The Act does not explicitly provide this power, and the FCC has not sought to test the limits of its freedom by claiming it.

5.2.3 Unbundling and UNE-P

There are two controversies here, related but not identical. The first is the existence of the UNE-P offering at all, in particular the availability of local switching as part of a combined unbundled local platform for CLEC services. The second is the alleged injustices of the rates for UNE-P as implemented by state regulators.

The FCC's latest unbundling order was reversed by the DC Circuit Court of Appeals in March 2004. There have been some years of controversy over the TELRIC pricing standard for UNE-P, and about the UNE-P requirements themselves. The Court order may end the controversy, but only time will tell. It is currently under appeal to the Supreme Court, but not supported by the Solicitor General. This controversy has some of the elements mentioned above: the practicalities of the regulatory process mean that the FCC cannot itself cope with the volume of proceedings required to examine each local geographic market in detail. The Courts found that the appropriate proceedings would have to examine the state of local competitive network supply to determine the continued existence of that *impairment* of competitors which is necessary for unbundling to be provided under the section 251. The FCC's previous rulings on unbundling failed because they did not define or examine impairment closely enough for the courts; this

time the FCC attempted to have State commissions make the determination, and the courts found this to be impermissible. The Court found that the FCC did not sufficiently define the applications and limitations of the impairment standard, and that the FCC could not delegate its authority to state commissions to examine impairment. The local switching and shared transport elements of the voice UNE-P combined platform used by CLECs need no longer be provided at TELRIC rates. They can only be obtained by commercial negotiation. This reduces the availability of Section 251 elements to CLECs on regulated terms.

The basic statement of the problem, in economic terms, is this: under section 251 entrant telecommunications carriers can obtain ILEC networks elements, separately or combined, if they would be *impaired* without them. The courts have sought a meaning for impairment other than simple commercial advantage, because providing incumbents' assets to entrants at less than an efficient entrant's cost would clearly not lead to efficient competitively governed markets, and this is the intent of the legislation. The FCC has yet to produce an administratively workable definition and now appears to have given up the attempt to do so.

The availability of UNE-P as required by Section 251 may now be time limited, but unbundling is still required of the Bell companies under section 271. However, section 271 does not state that unbundled elements must be provided under the same cost standard as Section 251, nor that they must be provided in a manner permitting CLECs to combine them. The court's view of this is that if Congress intended the two clauses to have similar standards it would have said so in the legislation, so unbundled elements will be available under section 271, but not automatically combinable into a package, and not at regulated rates.

The rate level controversy is interesting because of the amounts of money at stake, but its principles are simple enough. Given the variations in the retail regulated rates in telecoms, the UNE-P platform can offer entrants potentially profitable business for all types of customer only if UNE-P rates are so low that they fall below the (subsidized) basic residential rate. At that price level competitors could siphon off most business users with almost no risk.

If UNE-P rates are higher than this floor, then business and larger residential users may possibly be profitable for entrants, but lighter users will not. If the UNE-P rate structure differs from the retail rate structure the set of entrant opportunities it creates will be more complex.

In general the regulated UNE-P rates do not appear to reward indiscriminate entry, but require some degree of business skill and judgment

to support successful entry[11]. However, the UNE-P structure challenges the continuation of heavily cross-subsidized residential flat rate service because many of the internal cross-subsidy flows necessary to its support will be competed away. High business local rates become unsustainable if cost-based local switching and transport are available for entrants to use.

The economics of this issue are also fairly straightforward. Switches are now relatively cheap, so most entrants could afford them. However, collocation facilities are very costly and impose a minimum efficient scale of entrant in terms of lines per wire center that must be won for a viable business. UNE-P eliminated the need for collocation facilities, and this is probably its most important characteristic, not the local switching. Entrants without UNE-P also need connecting lines from each wire center to their own switches. Only if entrants attain substantial market share will these lines be as efficiently loaded as the comparable ILEC trunks, and otherwise there is an economy of scale disadvantage to entrants. Other countries have recognized this issue, and the differences between unbundling policies in different countries can be instructive. Unbundling has not been a very useful policy in most parts of Europe, where policies requiring the incumbent to offer wholesale services have been more effective in supporting the existence of retail competitors.

In the UK true facility-based competition from cable companies and urban CLECs has been fairly effective in restraining BT, but deregulation was only introduced for particular services and markets after BT demonstrably lost market power: a process more similar to the FCC's proceedings in the 1980s than to the post-1996 practices. In Japan investment incentives for broadband deployment appear to have been increased by the unbundling regime in place, which allowed a CLEC (Yahoo) to order fiber to the home and provide 100Mbps service[12].

---

[11] The most convenient reference site for this information is maintained by the West Virginia commission's office of the consumer advocate. Annual repeated surveys and compilations are available, the most recent (January 2004) can be found at http://www.cad.state.wv.us/JanIntro2004.htm.

[12] See Ovum's report *Business Models for Exploiting the Local Loop,* July 2002 ". . . in July 2000 the Ministry of Posts and Telecommunications (MPT) ordered NTT to open all local exchanges for co-location, to allow unbundlers access to central offices, and to lift the limitations on rack space.

Since then, NTT has been under continuing pressure from both the MPT and the government to ease access and cut prices for co-location. In December 2000, the charge for unbundled line sharing was reduced from ¥800 ($6.50) to ¥187 ($1.50) per month - the lowest in the world. Further measures reduced co-location costs, allowed for self-installation of equipment by unbundlers, shortened provisioning periods and prevented NTT from accessing competitive information. NTT was also obliged to unbundle backhaul to its local exchanges over its fiber network and to provide the necessary information to support competitors in getting access. NTT is obliged to provide facilities to competitors

5.2.4 Advanced Services and Unbundling

Sections 251 and 271 of the 1996 Act state that only telecommunications carriers can request unbundled elements, and only for the provision of telecommunications services. The FCC has found that broadband Internet access is not a telecommunications service but an information service. Consequently no provider of this service is entitled to request unbundled access to network elements to provide broadband services because they are not telecommunications carriers nor are they providing a telecommunications service.

The Section 706 language cited above is also being used to support this, because the FCC has (rightly) determined that cable modem services are competitive with DSL and that competition will be advanced if the ILECs do not have to share their new networks with other providers (since cable companies do not).

This policy debate (regulatory incentives to promote construction of broadband networks) is strongly reminiscent of the ILEC arguments for incentive regulation some ten years ago, and like those debated then it does not appear to have any enforcement teeth should ILEC investment fail to materialize as pledged.

It also leaves the broadband or advanced service market in the hands of at best a duopoly, and this structure does not generally lead to functionally competitive markets (for a telecoms example one need only look at the history of mobile services pricing and margins since the number of licensees was expanded).

The FCC need not allow this situation to persist. It could adopt incentive regulation with teeth on the same model used in many countries' spectrum licenses, which impose build-out timetables and penalties for failure to meet them. Local cable franchises in the U.S. commonly used similar requirements to ensure coverage of the whole franchise area. Perhaps an appropriate modification to the incentive regulation regime would be to use a rate of return on rate base proceeding to determine the monopoly cost of capital, and then authorize higher returns to provide the capital needed for the investment program. If investments are not made then rate reductions and return reductions would be imposed to 'claw back' the excess corporate funds by increasing the X, or productivity offset, factor in the price index change <= CPI-X formula.

---

under the same terms and conditions as it provides to its own divisions." Also see "The Unbundling of Network Elements: Japan's Experience" by IKEDA Nobuo1, at http://www.rieti.go.jp/jp/publications/dp/03e023.pdf.

5.2.5 Internet and Internet Service Provider (ISP) regulation

The FCC's Computer Inquiries framework separating terminal equipment and unregulated enhanced services from tariffed services prepared the ground for the emergence of online services using dial-up modem connections. Modems could be connected to the telephone network without any new service being required, at both ISP and consumer premises. Online services grew with most of their end-users paying flat rate local service tariffs (the subsidized residential rates intended to achieve universal access to telephony) to call the ISP modem bank reached through a local business number.

This created a number of problems for ILECs as online usage grew, because their local exchange switches were designed for traffic with normal voice call durations and distribution characteristics. Residential call durations averaged 8 minutes or less, and were to fairly dispersed sets of local numbers. Online access calls could, and did, last for hours on end, and were focused on blocks of numbers (the modem banks) and together these factors dramatically changed the economics of local switched networks. The designed line to trunk ratios no longer provided the desired quality of service for voice users, and the switches could be congested in their outgoing call direction modules because of the high concentration on particular blocks of numbers.

Typically, after a time, users themselves would experience voice service degradation because of the online use, and order a second line for their modem so that ordinary voice service was still effectively available to other members of the household. This led to a 1990s surge in ILEC investments to add loops in residential areas.

The first regulatory problems that showed a strongly protective FCC view of ISPs arose when ISPs began to integrate with CLECs. A CLEC providing the lines used for an ISP modem bank had very unbalanced traffic since almost all his calls were inbound from ISP customers. The FCC decided that all this traffic was interstate (because it was to access the internet or online information services over which the FCC asserted jurisdiction, preempting the states) and also that it was local. Calls were treated as inter-carrier local calls, completed on the CLEC network at the ISP modem bank, so Section 251 reciprocal compensation rules applied and the ILEC had to pay terminating access charges to the CLEC. ILECs thus found themselves in a situation where free local calls generated termination payments, and estimated the transfer from themselves to CLECs because of this at up to $1B per year.

This was a remarkable FCC decision for two reasons. An ILEC is supposed to treat all calls equally, and the only precedent for calls to

business lines being treated as interstate was in the ENFIA (Exchange Network Facilities for Interstate Access) decisions which resulted in MCI and Other Common Carriers paying access charges for the first time. These calls, like those to access an ISP, required customers to dial an access number, then authenticate themselves with a login code or access number, before going on to actually use the connection they established, but the FCC treated the online access calls quite differently.

This decision had two consequences before being reversed by the FCC. For ISPs it made the transition to DSL unattractive. DSL generated no call termination charges to the terminating local carrier, so the effective cost to ISPs was much higher than for dial-up. This was a partial cause of the unanticipated (by investors) failure of the data oriented local exchange carriers to gather customers in large volumes. DLECs deployed DSL using collocation and unbundled ILEC loops (before the FCC determined that this was not legally required). They mostly failed financially because DSL adoption was slow, and they could not reach the customer density per collocation site necessary to cover their costs. (This was part of the 'bubble economy' period and took some billions of investor dollars into oblivion. Major players included Rhythms, Covad, Northpoint, and NAS. Only Covad emerged from bankruptcy.)

Once reversed the new policy deprived CLECs of a large but declining revenue source, and combined with the new policy on unbundled plant access it deprives DLECs of any ability to sustain their businesses.

6. THE CURRENT STATE OF THE INDUSTRY, INTERMODAL COMPETITION, AND THE KNOWN ECONOMICS OF LOCAL COMPETITION

The 1996 Act is based on the idea that competition can and should govern the conduct of local network and service providers. Competition from alternative facilities-based networks is the ideal. Courts have construed the Act, in their various examinations of FCC Orders, to say that the presence of facilities-based competition in an area demonstrates that competition is possible. From this courts have relied on contestability theory to conclude that rate regulation can longer be justified unless based upon explicit findings that take local circumstances into account.

Most of the one hundred largest cities in the US have at least one facilities-based CLEC in operation, using its own access network. However, the geographic and customer reach limits of these networks mean that the great bulk of customers are still only served by ILEC facilities. However,

the private line market has been largely deregulated (CLECs after all target this market) and the IXCs in particular allege that the rates they now pay for private lines, as retail or wholesale services, are no longer just and reasonable but are monopolistically high. The contrast between the economics of constructing local access networks and the distribution of customers can be illustrated by the business case drawn up by Cogent Communications, a fiber-based network operator in the continental US[13]. Their business is based on the conclusion that the potential market for their service is limited to 60,000 buildings in the whole country, a very small proportion of the number of telecommunications-using buildings, so they will, even if completely successful, place almost no price discipline on incumbents' behavior in serving the remaining 100 million or so premises.

RBOCs have received permission to enter the long distance business in all significant markets, and compete directly with IXCs for both private line and switched services. We can contrast the FCC's actions in the 1970s and 1980s with today's FCC. Under 1980s policies ILEC rates for the local components of end to end services would be examined closely and compared to several different measures of cost; ILEC services beyond any core set of local monopoly services would have to be provided by a separated subsidiary, and calls for accounting transparency would be common. Imputation tests for ILEC services (demonstrating that they cover the sum of costs charged to rivals for similar component services) would be widely used. Unbundling rates and terms would be governed by the principles of equal ILEC treatment for rivals and internal users. No ILEC service would be deregulated, or even governed by relaxed price-cap regulation, until irreversible entry had been demonstrated (i.e. substitute infrastructure under different ownership and control actually exists to supply substitute services in case the ILEC should try to exercise market power.) The 1996 Act has changed this picture remarkably.

RBOCs have taken a substantial share in the long distance markets they have entered, and have lost significant share to rival local or local + long distance suppliers in many markets. However, most local mass-market competition relies on the availability of the RBOC network as a UNE-P offering and thus may be an artificial creation, depending on continuing arbitrage opportunities sustained by regulators. Cable operators providing service over their own networks have achieved 30% shares of the voice market and more than 60% shares of the high speed access market in areas where they have been most aggressive.

The economics of local telecommunications network overbuilding are fairly clear, and very unattractive in most areas. The capital cost of a fiber to the premises network has been estimated at $800 per home passed for an

[13] See http://fcke.fastcompany.com/fullfcke.html?cid=1851.

average suburban area, with another $800 or so per home connected. Most of the home-passed cost is for the civil engineering work (i.e. construction of ducts, trenches, or poles where permitted) and would be the same for a copper network. The additional capital cost per home connected is for the installation of a drop connection and fiber terminal equipment: the trenching and/or duct cost would also be the same for a copper network.

Costs per customer decline substantially with more customers in a given area, and this economy of scale suggests that there is a sustainable natural monopoly in any area where the revenue density per infrastructure (i.e. duct, trench, or aerial plant) mile is lower than in central business districts. Competition in the services market is unlikely unless competitive providers have access to an incumbent's local network. Intermodal competition is both possible and happening, with cable companies attracting ILEC customers.

Two forms of intermodal competition are occurring: from cable TV system operators and from mobile service providers.

Most current cable networks are capable of supporting high-speed data services, and have the market lead in these services over DSL. However, cable operators have no FCC-imposed obligation to accommodate multiple service providers on this platform (though Federal courts may still have the last word on this topic). Cable operators are now building on their lead in broadband access services and offering voice services in addition, in some cases using the cable modem and providing voice over the internet protocol (VOIP) service, and in other using traditional PSTN-style telephony over the coax plant.

The FCC had decided that two competitors were not enough to impose market discipline on mobile service prices, and took pains in spectrum allocation and auctions to ensure that there would be enough licensees for a cartel to be very difficult indeed to sustain. Foreign experience seems to show that this was the right approach. In countries where four or more mobile licensees compete prices drop towards costs quite quickly, but where there are three or fewer competitors prices stay higher for longer. Residential users in the US are increasingly moving their usage from wireline to mobile networks, and in some cases their access lines too. Currently less than 5% of households use only mobile service but this proportion could be expected to rise, although the consolidation of the wireless business in the hands of ILEC affiliates may slow the process. (The Cingular-AT&T Wireless merger puts the two largest mobile service operators into ILEC-affiliated hands.)

Other radio infrastructures remain as possible entry vehicles. However, the attempts to commercialize MMDS failed some years ago, but there may be another attempt using better technologies (in the IEEE standards 802.16 and 802.20 families), and the FCC is proposing to allocate frequencies in the

3GHz range for unlicensed broadband radio services. It is too early to tell if these potential competitors will actually emerge and provide any market discipline.

An indirect form of intermodal competition comes from voice over Internet protocol services. VOIP services range from PC to PC services with no intermediary service provider (such as Skype) through PC to PC services with service providers (such as voice-enabled instant messaging or the Pulver service the FCC recently considered) to services where telephone numbers are assigned and an instrument is provided to connect to the broadband service. Cable operators, Vonage, and now AT&T offer these services.

AT&T's deployment of its CallVantage mass-market voice over IP service uses its subscribers' cable modem or DSL service for access, without any disclosed compensation to the broadband service provider. It is the recreation of the initial MCI Execunet service where a paid local service is used as the tail at either end, although the FCC determined that AT&T will have to pay terminating access where it connects to the ILEC networks to complete calls.

An open issue at the FCC is whether, and if so how, VOIP providers should have to pay the underlying broadband access providers, and whether any broadband access providers who disable non-affiliate VOIP service providers' services should be subject to regulatory action.

7. CONCLUSIONS

The FCC does not appear to have been captured by any particular group of its client industries. However, policies now in effect are extremely unlikely to result in a telecommunications industry that is even workably competitive. They will not produce an industry whose prices and services approach either static or dynamic efficiency goals. The FCC is far from being the ideal welfare-maximizing omniscient regulator.

A revised Telecommunications Act should clarify that the FCC and state commissions should have the power to contract with regulated firms so that incentives to build out can be both negotiated and enforced. For example, instead of relying on pledges to build fiber networks if sufficient incentives are offered (and deregulating these networks and relaxing the requirements for unbundled access to provide that incentive) the FCC could negotiate a timetable for build out and deployment, and punish failures by using price cap measures such as increasing the productivity offset factor. The FCC could reward success by offering such things as time-bounded monopolies in using the new networks, analogous to the patent process for rewarding

innovation. The new network could be rewarded by a monopoly on its use until the equivalent of a compulsory patent license or patent expiration was imposed by regulation. (This should apply to cable operators too.)

The basic problem with the Telecommunications Act is that it loads the dice in favor of premature deregulation by making the procedural burden of sustaining regulation almost impossible to bear. The Act treats cable, wireless, Internet, and PSTN service providers quite differently, and so distorts the outcomes of their competition with each other in the current layered model of service provision. Welfare maximization, or national economic efficiency, requires that the Act be changed in certain respects that would eliminate many current avenues for the exercise of monopoly power, within all the industries governed by the current law: telecommunications, mass media, cable and satellite TV, and the internet.

Similar distortions appear to be emerging in the video/TV business. Cable companies currently bundle their programming services into a very limited set of tiers, so that customers cannot select individual channels a la carte. In my view it should be required throughout the programming distribution chain. End users, cable system operators and their rivals in satellite distribution should all have the right to purchase only those unbundled programming services they wish, on transparent terms and conditions. Recent disputes between Disney and Viacom as program suppliers, and cable and satellite system operators as their customers, have shown that the program suppliers currently leverage their control over some 'essential' programming (the ABC or CBS broadcast networks) to obtain carriage for other networks they want to distribute.

A modest change to the Act could restrain this exercise of monopoly power by forbidding any spectrum grantee (that is, a licensee who received free licenses instead of purchasing his spectrum at auction) to charge any distributor for their broadcast programming, in any format. Another change would be to require transparency in all the program supply contracts: they should be public, and subject to regulatory or court intervention to eliminate 'undue' discrimination among different parties similarly placed.

In conclusion, the 1996 Act appears to have intended to build a deregulated industry by means of the be the most regulator work-intensive process imaginable, but it lacks the apparatus necessary to ensure irreversible competitive processes exist before withdrawing regulation. The FCC may not have been captured by its 'clients', but it has been forced into unsuccessful policies by its controlling legislation.