the CATO JOURNAL

AN INTERDISCIPLINARY JOURNAL OF PUBLIC POLICY ANALYSIS

ARTICLES

Geoffrey Black, D. Allen Dalton, Samia Islam, and Aaron Batteen	The Coasean Framework of the New York City Watershed Agreement
Jason E. Taylor and Jerry L. Taylor	Marginal Tax Rates and U.S. Growth: Flaws in the 2012 CRS Study
John Merrifield and Barry W. Poulson	State Fiscal Policies for Budget Stabilization and Economic Growth: A Dynamic Scoring Analysis
Andrew Foy, Christopher Sciamanna, Mark Kozak, and Edward J. Filippone	The Medical Care Cost Ratchet
Robert Krol	Forecast Bias of Government Agencies
Paul Ballonoff	A Fresh Look at Climate Change
Joshua R. Hendrickson	Contingent Liability, Capital Requirements, and Financial Reform
Thomas L. Hogan and William J. Luther	The Explicit Costs of Government Deposit Insurance
Paul H. Rubin	Pathological Altruism and Pathological Regulation

BOOK REVIEWS

Trevor Burrus • Mark A. Calabria • Patrick J. Michaels • Peter Van Doren

VOLUME 34 NUMBER 1 • PUBLISHED BY THE CATO INSTITUTE • WINTER 2014

the CATO JOURNAL

VOLUME 34 NUMBER 1 • PUBLISHED BY THE CATO INSTITUTE • WINTER 2014

Editor
JAMES A. DORN

Book Review Editor
TREVOR BURRUS

EDITORIAL BOARD

TERRY L. ANDERSON
Property and Environment Research Center

CHARLES W. BAIRD
California State University, East Bay

RANDY E. BARNETT
Georgetown University School of Law

JOHN H. COCHRANE
University of Chicago

KEVIN DOWD
University of Nottingham

RICHARD A. EPSTEIN
University of Chicago

JAGADEESH GOKHALE
Cato Institute

JAMES GWARTNEY
Florida State University

STEVE H. HANKE
Johns Hopkins University

RANDALL S. KROSZNER
University of Chicago

DEEPAK LAL
University of California, Los Angeles

LEONARD P. LIGGIO
Atlas Foundation

STEPHEN MACEDO
Princeton University

ANTONIO MARTINO
University of Rome, Luiss

KEVIN M. MURPHY
University of Chicago

JEFFREY A. MIRON
Harvard University

GERALD P. O'DRISCOLL JR.
Cato Institute

SAM PELTZMAN
University of Chicago

JUDY SHELTON
Fredericksburg, Virginia

VERNON L. SMITH
Chapman University

JOHN B. TAYLOR
Stanford University

ROLAND VAUBEL
Mannheim University

RICHARD E. WAGNER
George Mason University

LAWRENCE H. WHITE
George Mason University

LELAND B. YEAGER
Auburn University

THE CATO JOURNAL (ISSN 0273-3072) is published in the winter, spring/summer, and fall by the Cato Institute, 1000 Massachusetts Ave., N.W., Washington, D.C. 20001-5403. Annual subscriptions are $22 for individuals and $50 for institutions. Single copies are $8. Two-year and three-year subscriptions are $38 and $55, respectively, for individuals, and $85 and $125, respectively, for institutions. Foreign subscribers should add $5 per year for regular delivery and $10 per year for airmail delivery.

Correspondence regarding subscriptions, changes of address, acquiring back issues, advertising and marketing matters, and so forth, should be addressed to the Publications Department. All other correspondence, including requests to quote or reproduce material, should be addressed to the editor.

Unsolicited manuscripts cannot be returned and will be acknowledged only if accompanied by a stamped, self-addressed envelope. There is no submission fee. Articles should focus on specific public policy issues, be scholarly, but also be intelligible to the interested lay reader. Accepted manuscripts must conform to the CATO JOURNAL's style requirements. A style sheet is available on request.

Authors may submit their papers electronically, preferably in MS Word, to the editor (jdorn@cato.org), along with an abstract of no more than 250 words.

Unsolicited book reviews cannot be returned and will be acknowledged only if they are accepted for publication. They should conform to the above requirements and should provide an in-depth analysis of major studies of public policy. Reviews should range from 800 to 1,000 words.

The views expressed by the authors of the articles are their own and are not attributable to the editor, the editorial board, or the Cato Institute.

THE CATO JOURNAL is listed in *Cabell's Directory*, *Journal of Economic Literature*, *Current Contents/Social and Behavioral Sciences*, *America: History and Life*, *Historical Abstracts*, *International Political Science Abstracts*, and the ABI/INFORM database.

Printed in the United States of America.

Copyright © 2014 by the Cato Institute.
All rights reserved.

CATO JOURNAL • VOLUME 34 NUMBER 1 • WINTER 2014

ARTICLES

Geoffrey Black, D. Allen Dalton, Samia Islam, and Aaron Batteen
The Coasean Framework of the New York City Watershed Agreement 1

Jason E. Taylor and Jerry L. Taylor
Marginal Tax Rates and U.S. Growth: Flaws in the 2012 CRS Study 33

John Merrifield and Barry W. Poulson
State Fiscal Policies for Budget Stabilization and Economic Growth: A Dynamic Scoring Analysis 47

Andrew Foy, Christopher Sciamanna, Mark Kozak, and Edward J. Filippone
The Medical Care Cost Ratchet 83

Robert Krol
Forecast Bias of Government Agencies 99

Paul Ballonoff
A Fresh Look at Climate Change 113

Joshua R. Hendrickson
Contingent Liability, Capital Requirements, and Financial Reform 129

Thomas L. Hogan and William J. Luther
The Explicit Costs of Government Deposit Insurance 145

Paul H. Rubin
Pathological Altruism and Pathological Regulation 171

BOOK REVIEWS

Free Market Revolution: How Ayn Rand's Ideas Can End Big Government
Yaron Brook and Don Watkins

Reviewed by **Trevor Burrus** 185

Political Bubbles: Financial Crises and the Failure of American Democracy
Nolan McCarty, Keith Poole, and Howard Rosenthal

Reviewed by **Mark A. Calabria** 189

The Bet: Paul Ehrlich, Julian Simon, and Our Gamble over Earth's Future
Paul Sabin

Reviewed by **Patrick J. Michaels** 193

The Great Rent Wars: New York 1917–1929
Robert M. Fogelson

Reviewed by **Peter Van Doren** 196

The Coasean Framework of the New York City Watershed Agreement

Geoffrey Black, D. Allen Dalton, Samia Islam, and Aaron Batteen

Over 50 years ago, in "The Problem of Social Cost," Ronald Coase (1960) attempted to reorient the economics profession's treatment of externalities. He wanted to draw economists' attention away from the world of pure competition as a policy standard and investigate the consequences of transaction costs and property rights for the operation of markets. In 1991, he was awarded the Nobel prize in economics "for his discovery and clarification of the significance of transaction costs and property rights for the institutional structure and functioning of the economy" (Royal Swedish Academy of Sciences 1991). The Academy cited both his 1960 article and his 1937 article "The Nature of the Firm."

Still, critics question both the relevance and applicability of the Coasean framework for analyzing, explaining, and ameliorating harmful effects associated with economic activities, reflecting the degree to which the profession's treatment has remained unchanged. Nalebuff (1997: 35–37), for example, has argued that for environmental problems "as the scope of the externality affects more and more people, it becomes increasingly difficult to assign property rights." Moreover, "even when property rights have been assigned, exclusion is difficult

---

Cato Journal, Vol. 34, No. 1 (Winter 2014). Copyright © Cato Institute. All rights reserved.

Geoffrey Black is Associate Professor of Economics at Boise State University. D. Allen Dalton is Adjunct Faculty in Economics at Boise State University. Samia Islam is Associate Professor of Economics at Boise State University. Aaron Batteen is a graduate of the undergraduate Economics program at Boise State University. The authors thank Jim Dorn and referees for helpful comments.

if not impossible." In this article, we argue that the New York City Watershed Memorandum of Agreement (MOA) proves the usefulness of the Coasean framework—even when there are a large number of affected parties from nonpoint source pollution.

In 1997, nearly a decade after the Environmental Protection Agency ordered New York City to filter its water to remove contaminants originating in upper New York State watersheds, NYC entered into a MOA involving the State of New York, a number of local governments and environmental groups in the Catskills, and the EPA. In Coasean terms, the MOA is the consequence of the State of New York's assignment of property rights to the Catskill/Delaware Watershed communities to continue with regional development and current practices, although some of those activities degrade NYC's drinking water. This rights assignment positioned NYC as the responsible party for initiating negotiations and programs protecting the Watershed system. Once responsibility was established, NYC opted to buy lands contributing to water quality degradation (for example, farmlands on riparian corridors), instead of building a multibillion-dollar filtration system. Residents and landowners upstream were compensated for development restrictions incurred from the MOA. New York State's role as mediator eliminated bargaining barriers and effectively reduced the transaction costs of arranging negotiations, demonstrating potential economic benefits to all parties, and providing alternative options to unilateral regulatory decisionmaking.

Externalities and the Pigouvian Tradition

Externalities exist when the effects of a transaction between two parties spill over to nonparticipants. Effects can be either beneficial (positive externalities) or harmful (negative externalities) to nonparticipants. When externalities are unaccounted for in the decisionmaking of the participants, transactions will result in either "too much" or "too little" consumption or production. This very language, which economists use, derives from modern welfare economics founded on Pigou's extension of marginal analysis to the utilitarian market-failure analysis of J. S. Mill and Henry Sidgwick (Medema 2009).

In the history of welfare economics, the classical approach to solving negative externality problems, such as pollution, was through the use of regulation, or what Mill (1871) termed "authoritative" solutions, "in which certain types of conduct are prescribed or proscribed"

(Medema 2009: 37). Sometimes referred to as "command and control" policies, authoritative solutions prohibit firms from producing a negative externality, require a certain standard of emissions of a negative externality be reached, or require firms to employ specific externality-reducing technologies, all backed by legal penalties. The costliness and difficulty of obtaining information tailored to specific firm production methods, in conjunction with the nondiscriminatory or universality requirement of legal standards, generally yield uniform standards across all firms in an industry. Yet, this type of policy response often imposes solutions that are more costly than alternatives, inhibit innovations in pollution-reduction technologies, and deter economic growth (Marlow 1995, Davis 1992).

Modern welfare economics, based on the work of Pigou in *The Economics of Welfare* (1932), offers "corrective" taxation as another solution. Pigou raised the solution specifically in relation to spillovers, where he assumed there was a divergence between private and social net products. Pigou (1932: 174) also addressed spillovers due to "the separation between tenancy and ownership of certain durable instruments of production." This divergence, he believed, could be removed by renegotiating the landlord-tenant contract. However, Pigou (1932: 192) argued that spillovers could not "be mitigated by a modification of the contractual relation between any two contracting parties, because the divergence arises out of a service or disservice rendered to persons other than the contracting parties." He did not address the reasons why a contract could not be concluded by the contracting parties with those "persons other than the contracting parties."

The Pigouvian tax solution to a pollution externality places responsibility on the firm producing the pollution. The government then imposes a tax on emissions of a magnitude equal to the divergence between social and private marginal costs, so that the external costs are internalized into the decisions of the polluting firm. Pigouvian taxation differs from authoritative regulation because firms that pollute above a set amount are forced to pay an emissions tax.[1] Although Pigouvian taxation reduces pollution at a very low cost, there are problems that may arise. First, Pigouvian taxation will achieve pollution reduction only as long as the tax level is set equal to or above the marginal cost of abatement (MCA). In the event the tax rate is set

[1]An alternative to taxation is a subsidy of the appropriate magnitude. For brevity, this alternative is not presented.

below the MCA, pollution reduction will not take place. Second, because some firms may have "modern, well-maintained pollution control equipment," while other (typically, older) firms possess outdated pollution control equipment, it is inefficient "for all firms to carry identical burdens" (Marlow 1995: 96). Third, even if those harmed by the pollution are compensated from corrective taxes, further complications may arise. Compensation is bound to be difficult when not all affected parties suffer from the problem equally. Fourth, it is costly to discover the appropriate level of the corrective tax, and the information costs may exceed the benefits from implementing the Pigouvian tax. Fifth, "compensation policies may also create a perverse incentive for individuals or firms to move into areas for the sole purpose of receiving compensation payments" (Marlow 1995: 96).

In sum, Pigou's analysis in both *The Economics of Welfare* and his earlier *Wealth and Welfare* (1912) provided a list of examples where self-regarding behavior creates divergences of private and social net product (benefits, costs) providing "a strong sense that market failure is a pervasive problem" (Medema 2009: 64).[2] The Pigouvian framework, as further developed by Bergson (1938), Lange (1942), Meade (1952), Graaf (1957), and Bator (1958) in mathematical form, gave to modern welfare economics a seemingly powerful tool and standard for welfare improvement:

> The rhetorical, persuasive force of this analysis should not be underestimated. What this theory demonstrated, in a nutshell, was that perfect markets work perfectly, imperfect markets work imperfectly, and perfect government can cause imperfect markets to also function perfectly. This became the textbook model [Medema 2009: 76].

The Coasean Argument and Framework

For Coase, the Pigouvian framework fails on several grounds, but most fundamentally, he notes:

> Analysis in terms of divergencies between private and social products concentrates attention on particular deficiencies in

[2]To be fair to Pigou, however, Medema has recovered a "lost" or, at least, relatively unknown essay that Pigou wrote in 1935 entitled *State Action and Laisser-Faire*. Therein, Pigou's examples of market failure produced by the existence of external benefits or costs are balanced by a consideration of problems in carrying out governmental solutions.

> the system and tends to nourish the belief that any measure which will remove the deficiency is necessarily desirable. It diverts attention from those other changes in the system which are inevitably associated with the corrective measure, changes which may well produce more harm than the original deficiency [Coase 1960: 42–43].

What is necessary, according to Coase (1960: 34), is an approach comparing the total social product of alternative measures that gives attention to alternative specification of property rights and the transaction costs associated with different property rights regimes. He rejects the Pigouvian framing of the problem in terms of "restraining the producer" of the negative externality:

> We are dealing with a problem of a reciprocal nature. To avoid the harm to B would inflict harm on A. The real question that has to be decided is: should A be allowed to harm B or should B be allowed to harm A? The problem is to avoid the more serious harm [Coase 1960: 2].

Indeed, making "restraining the producer" the problem to be solved dictates the solution —whether making the producer liable for damages, imposing a tax, offering a subsidy, or restricting the producer's location. Such "solutions" take for granted that restricting the producer causes less harm than restricting the recipient. Coase (1960: 3) contends that the Pigouvian "courses of action are inappropriate, in that they lead to results which are not necessarily, or even usually, desirable." According to Coase (1960: 3), "If we assume that the harmful effect of the pollution is that it kills the fish, the question to be decided is: is the value of the fish lost greater or less than the value of the product which the contamination of the stream makes possible."

If the proper procedure is to compare the total social product of alternatives, then the assignment of property rights matters. However, Coase shows that the assignment of property rights or the assessment of damage liability only matters in a world of positive transaction costs (that is, the world we actually inhabit):

> The argument has proceeded up to this point on the assumption . . . that there were *no costs involved in carrying out market transactions. This is, of course, a very unrealistic assumption.* In order to carry out a market transaction it is

> necessary to discover who it is that one wishes to deal with, to inform people that one wishes to deal and on what terms, to conduct negotiations leading up to a bargain, to draw up the contract, to undertake the inspection needed to make sure that the terms of the contract are being observed, and so on. These operations are often extremely costly, sufficiently costly at any rate to prevent many transactions that would be carried out in a world in which the pricing system worked without cost [Coase 1960: 15; emphasis added].

And yet, the Pigouvian framework is firmly anchored within a world of zero transaction costs. In order to identify or measure the divergence between private and social costs (benefits, or net product), it is assumed that all necessary information is known, or at least knowable, by the outside observer, regulator, politician, or welfare economist at zero cost—a result of what Coase terms "blackboard economics" where "all the information needed is assumed to be available and the teacher plays all the parts" (Coase 1988: 19). Following Hayek (1945: 519), one might add that such information is best produced by market participants through individual interactions within the price-forming market process.

In a world of positive transaction costs, institutions matter, especially in the delimitation of property rights. Two questions arise regarding rights which Coase takes up in sections VI and VII, respectively, of *The Problem of Social Cost*: First, taking the initial specification of rights and the costs of market transactions as given, what possible changes in the economic environment can produce more efficient results? Second, what considerations will make the initial delimitation of rights more favorable to efficient results?

Aslanbeigui and Medema (1998) illustrate three possible solutions to the problem of externalities as outlined by Coase, *if participants find it too costly to reallocate resources through contracts/markets*. The first solution, in the case of producer-producer externalities, would be to have a single owner take into account "all costs (internal and external) and allocate resources in that light." By doing this, the "allocation of resources by administrative fiat would allow the firm to overcome the transaction costs concerns that attend the use of the market" (Aslanbeigui and Medema 1998: 604). One way of alleviating the problem of externalities would be for the two separate firms to merge so that the externality is internalized and at a cost much less than if regulatory policies were instated. However, Coase (1960: 16)

recognizes that "it does not, of course, follow that the administrative costs of organizing a transaction through a firm are inevitably less than the costs of the market transactions which are superseded." In addition, if consumers are involved, this remedy would not suffice. In cases

> which may affect a vast number of people engaged in a wide variety of activities, the administrative costs [of firm organization] might well be so high as to make any attempt to deal with the problem within the confines of a single firm impossible. An alternative solution is direct government regulation. Instead of instituting a legal system of rights, which can be modified by transactions on the market, the government may impose regulations which state what people must or must not do and which have to be obeyed [Coase 1960: 17].

However, Coase (1960: 17) notes that while government has powers that might get things done at a lower cost than firms, the administrative costs of government itself can be "extremely costly." Restrictions made by fallible humans subject to political pressure without competitive checks are unlikely to increase economic efficiency. He concludes that

> direct governmental regulation will not necessarily give better results than leaving the problem to be solved by the market or the firm. But equally there is no reason why, on occasion, such governmental administrative regulation should not lead to an improvement in economic efficiency. This would seem *particularly likely when . . . a large number of people are involved* and in which therefore the costs of handling the problem through the market or the firm may be high [Coase 1960: 17; emphasis added].

While authoritative regulation of externalities is not likely to give better outcomes than markets or firms, it is more likely to do so in situations involving large numbers and high transaction costs.

In light of the imperfections of firms and authoritative regulation, Coase (1960: 17) suggests that a third solution might sometimes be best: "to do nothing about the problem at all." This is the equivalent of the policy conclusion derived from what is sometimes referred to as the "Coase Theorem," except that here it is derived not from an artificial assumption about transaction costs, but a comparative

analysis of the costs of organizing activity through the voluntary arrangements of markets and/or firms versus authoritative regulation. A policy of "doing nothing" is tantamount to leaving market participants to decide whether or not the costs of rights rearrangements exceed the benefits from rearranging those rights through market exchange. Coase argues that neither the Pigouvian solution of taxation nor the classical solution of authoritative regulation is necessary if transaction costs are low, allowing for an agreement that benefits all involved parties. Transaction costs are key, because even in the face of externalities, an efficient outcome can occur through bargaining as long as costs do not outweigh gains.

This conclusion leads to a consideration of the original delimitation of rights. When private parties find it too costly to rearrange rights via market transactions, courts and legislatures directly influence economic outcomes by their decisions on who holds what rights. Those decisions influence both the allocation of resources and the distribution of income and wealth. Coase presents evidence that courts and legislatures in both Britain and the United States have often understood this aspect of their decisions and have taken the economic implications of their decisions into account. Moreover, Coase (1960: 19) notes that "even when it is possible to change the legal delimitation of rights through market transactions, it is obviously desirable to reduce the need for such transactions and thus reduce the employment of resources in carrying them out." Likewise, "while statutory enactments add to the list of nuisances, action is also taken to legalize what would otherwise be nuisances under the common law" (Coase 1960: 28). On this latter point, Coase is once again highly critical of the "blackboard economics" approach of modern welfare economics and economists who too quickly see an externality as a result of market failure "requiring corrective government action [that] is, in fact, often the result of government action" (Coase 1960: 28).

"The Problem of Social Cost" was Coase's attempt to get economists to pay renewed attention to the actual institutional framework within which economic decisions are made and economic activities are coordinated through markets and firms. He focused on how transaction costs altered the economic analysis suitable to mitigating harmful effects. His analysis stemmed from his earlier work on the Federal Communications Commission, which identified the factors that prevent efficient allocation of rights and resources generally

(Coase 1959: 27, n. 54). Coase clearly showed that the Pigouvian basis of modern welfare economics and its concentration on divergent private and social margins was inadequate for the task. Unless the role of transaction costs are made explicit, the policy responses derived from Pigouvian analysis will continue to be inappropriate and often undesirable.

Responses to Coase

The response of economists to "The Problem of Social Cost" has almost uniformly been centered on the proposition that quickly came to be known as the "Coase Theorem," rather than the argument Coase was making concerning the comparative evaluation of institutional frameworks. The Coase Theorem is typically presented as a proposition about a zero transaction cost world. In what is regarded as the first expression of the theorem, Stigler (1966: 113) stated: "Under perfect competition private and social costs will be equal." Medema (2009: 177) gave an alternative rendition: "If rights are fully specified and transaction costs are zero, parties to a dispute will bargain to an efficient and invariant outcome regardless of the initial specification of legal rights." There are two explicit assumptions and two results embodied in his specification of the theorem. The explicit assumptions are (1) full specification of rights and (2) zero transaction costs. The results are (1) Pareto efficiency and (2) invariance as to the initial rights distribution. Some discussions of the theorem are concerned with both results, others with just one (typically Pareto efficiency). Many discussions inadequately render the explicit assumptions and "can be dismissed under appropriate (some would argue 'correct') conceptualizations of the meaning of fully specified rights and zero transaction costs" (Medema 2009: 179).[3]

[3]In *The Applied Theory of Price*, McCloskey (1985: 335–39) correctly (in our view) identifies the "Coase Theorem" as follows: "In the presence of transaction costs the location of a pollution tax or of other liability for damages *does* matter for efficiency." McCloskey also importantly emphasizes that traditional Pigouvian taxation schemes, by using damages rather than the appropriate minimum opportunity cost of abatement, are inefficient. Finally, McCloskey observes that what the profession usually terms the "Coase Theorem" is really nothing more than "Adam Smith's Generalization" that if everything is owned and transactions are costless, then efficiency is obtained ("the Invisible Hand theorem"). The attempt to draw attention to Coase's central message and away from Pigouvian analysis, unfortunately, has been resisted by the profession as a whole.

The primary criticisms directed at (these versions of) the Coase Theorem point to its limited applicability and a restrictive set of assumptions without which the theorem breaks down. Canterbery and Marvasti (1992: 1180) argue that the theorem "borders on circularity" because externalities arise due to transactions costs, but externalities persist because of the transaction costs of removing them, leading to "Ptolemaic epicycles extending the 'Coase Theorem'." Cooter (1982) questions the existence of a Coase Theorem because Coase never stated the theorem and other economists have generalized the examples developed by Coase. Further, Cooter argues that in the absence of a third party's coercive force, parties will react to the harshest threats, so political power plays an enormous role in negotiation and may prevent efficient results. Samuelson (1963) contends that even in a zero transaction cost world, bilateral monopoly could prevent efficiency. Coase (1988: 161) responded that he believed "the proportion of cases in which no agreement is reached will be small," and has been faulted for not using real-world examples in support of this assertion (Black, Grant, and Islam 2008: 2). This is but the tip of the iceberg of a vast literature, much of which seems irrelevant to the world of positive transaction costs. This literature is dealt with at some length in Medema and Zerbe (2000).

David Friedman (1990) offers a defense of Coasean bargaining in a world of positive transaction costs, using the case of airplane noise plaguing a nearby residential area. Alternative solutions, Friedman suggests, are mandating airlines to make quieter planes, closing the airport when residents are sleeping, or soundproofing homes. "Charging the airlines for the cost of the noise they produce gives them an incentive to reduce noise, but that may be the wrong solution—it might be less costly to soundproof the houses or pay their occupants to move out." Unfortunately, when information is costly, it may be impossible to actually tell what will be the best solution for all of the parties involved. Since both parties essentially produce the problem, Friedman (1990) argues:

> Where we do not know who can solve the problem at lowest cost, the best solution may be to fall back on Coase's other idea: negotiations between the parties. . . . [T]he best solution to such problems may be for the legal system to clearly define who has the right to do what and then permit the affected individuals to bargain among themselves.

Regardless of whether the airline is held liable for noise damage or not, using Coasean bargaining may be the only way to discover the relevant alternatives toward finding a solution. Costly bargaining is preferable to uninformed guesses that mandate particular solutions, and may bring about a more efficient outcome once property rights are assigned. Using Coasean bargaining to resolve this issue will result in a more efficient allocation because whoever values the land the most will pay to use it.[4]

Criticisms of the viability of bargaining solutions within a Coasean framework, especially in addressing negative externality problems such as pollution, have been numerous. Some critics stress that it matters who owns the property rights in environmental pollution cases and that it may be difficult to assign such rights (Samuelson 1995, Abrego and Whalley 2004). Others point out that we do not live in a zero transaction cost universe (Varian 1995, Nalebuff 1997, Anderlini and Felli 2006). Moreover, Tybout (1972) argues that even in a zero transaction costs model, bribery to reduce pollution and compensation charges for it result in different total profits, so that long-run behaviors of parties may differ, negating any usefulness of the Coase Theorem. Frech (1973) refutes Tybout, arguing that zero transaction costs are based on a fundamental error in the analysis of costs for the polluting and polluted industries. He holds that a property right to the polluted area is a valuable asset with a definite market value. Because of this, when the rent of this asset is properly included in the costs of the industry holding the right, both marginal and total profits are shown to be identical regardless of who is assigned ownership of the asset.

Many detractors of the applicability of Coasean bargaining to environmental spillovers draw on Meade's (1952) case of honeybees. Meade argued that because bees cannot be convinced to respect property rights or keep contracts, contracts cannot be practically applied. The only solution would be to either subsidize farmers who grow nectar-rich crops or simply accept the inefficiency in the joint production of crops and honey. Two later articles, by Cheung (1973) and Johnson (1973), demonstrated that contracts

[4]Cooter and Ulen (1997: 89) argue that "the normative Coase theorem" consists of the claim that law should be structured to minimize the impediments to bargaining.

between beekeepers and farmers had been common practice since early in the 20th century, and Cheung also showed that the evident pricing pattern is actually efficient.

Hoffman and Spitzer (1985) published a paper empirically proving the relevance of Coasean bargaining in situations where there are two or three parties bargaining. Subsequently, the authors extended their work to report the results of controlled experiments designed to test the theorem for larger groups—an issue on which there is very little empirical work. Using 4, 10, and 20 parties with full and limited information, they found 93 percent of the bargaining outcomes to be efficient and that there was no deterioration of the results as the bargaining groups got larger. Although there was some deterioration in moving from full to limited information, almost 100 percent of the full-information bargains were efficient (Hoffman and Spitzer 1995). One implication of this result is that policymakers should recognize the power of voluntary negotiations to overcome negative externality problems, even in the absence of government intervention. But more relevant for our purposes, these results support the applicability of Coasean bargaining for disputes involving large numbers.

The world of zero transaction costs is, as Coase (1988: 15) remarks, "the world of modern economic analysis, and therefore economists feel quite comfortable handling the intellectual problems it poses." Once transaction costs are recognized to exist, however, the evaluation of alternative economic arrangements for possible improvement must include a comparison of those costs. This is the basis for Coase's argument for comparative institutional analysis and also the basis for his critique of Pigouvian taxation as a policy norm.

The Coasean framework has given rise to strong views both for and against its applicability. But the main weakness on the side of its supporters has been the scarcity of real-life scenarios of large number negative externalities where private Coasean bargaining resulted in externality mitigation. While a literature has developed in which "an assortment of cases . . . illustrates the voluntary internalization of externalities (both positive and negative) in history" (Higgs 2005: 407–08), no clear case of pollution mitigation involving large numbers of geographically dispersed affected parties has been forthcoming. This has prompted continuing debate on policy alternatives relying on Pigouvian taxation or forms of authoritative regulation to correct for inefficiencies caused by such large-number externalities. Interestingly, one recent public bargaining outcome confirms the

validity of Coasean bargaining in large-number situations that also involves nonpoint source pollution—factors that are traditionally considered to be inimical to its applicability. An examination of the recent New York City Watershed Agreement is a material example demonstrating the cogency of Coasean bargaining in providing a private solution to a negative externality problem, with only a limited role for government as mediator. In this solution, there was no need to implement any taxes, fees, or even "green payments" (Wu and Babcock 1996) to internalize the externality, as is the norm in addressing agricultural nonpoint source pollution.

History of the New York City Water Supply System

Unlike most other cities that rely on ground water, modern New York City is one of the few places in the world where the water comes from surface sources—namely, the Croton or East of Hudson (EOH) and the Catskill/Delaware or West of Hudson (WOH) watersheds in upstate New York. Together, these two watersheds comprise 19 reservoirs and 3 controlled lakes, storing 580 billion gallons of water (Daily and Ellison 2002). Early settlers of NYC obtained water for domestic purposes from private and public wells, and beginning in 1800, the Manhattan Company pumped water into a reservoir. In 1830, a tank for fire protection was added. As the population grew and well-water became polluted, NYC looked for alternatives (NYC Department of Environmental Protection [hereafter DEP] 2012).

After exploring alternatives for increasing supply, NYC built and placed into service what is now known as the Old Croton Aqueduct, which carried water impounded from the Croton River, in what is now Westchester County, into the Old Croton Reservoir. Distribution reservoirs were located in NYC, with locations varying and number increasing over time. The New Croton Aqueduct, built from 1885 to 1893, was placed in service in 1890. The various water systems of the NYC boroughs were consolidated into the present water supply system. The New York State legislature created the Board of Water Supply in 1905, and the Board moved to impound the waters of Esopus Creek, one of four Catskill watersheds, and develop what is known as the Catskill System, which includes the Ashokan Reservoir and Catskill Aqueduct. This project was completed in 1915 and turned over to the NYC Department of Water

Supply, Gas, and Electricity for operation and maintenance. The Catskill System was completed in 1928 with the building of the Schoharie Reservoir and Shandaken Tunnel (NYC DEP 2012).

In 1928, the Board of Water Supply also received approval to develop the upper portion of the Rondout watershed and tributaries of the Delaware River within New York State. Work was delayed by a lawsuit brought by the State of New Jersey to enjoin NYC and the State of New York from using waters of any Delaware River tributary. The U.S. Supreme Court upheld NYC's right to augment its water supply from the Delaware River, and construction of the Delaware System began in March 1937, to eventually be completed in stages (Delaware Aqueduct, Rondout Reservoir, Neversink Reservoir, Pepacton Reservoir, and Cannonsville Reservoir) by 1964 (NYC DEP 2012).

New York City controversially used eminent domain to build reservoirs in the early and mid-20th century. Between 1907 and 1915, several hamlets were taken using eminent domain and used to build reservoirs, displacing roughly 2,000 people. Another 350 residents were displaced a few years later, and finally 1,500 more after WWII (Catskill Watershed Corporation 2005).

The system impounds water in three upstate reservoir systems and three controlled lakes. The three water collection systems have various interconnections to increase flexibility by permitting exchange of water from one to another. Localized droughts and excess water supplies of the three watersheds are thereby mitigated (NYC DEP 2012). The NYC water supply system, with its sources (the "Watershed") that span 8 counties and 60 towns (CCCD 1997, Art. I, Sec. 4–5), is generally regarded as a monumental hydraulic and civil engineering achievement that provides approximately 1.5 billion gallons of water daily to the customer distribution system.

Background to the New York City Watershed Agreement

For many years, NYC's water has been among the highest rated in the nation for quality (USEPA 1996). The quality of the water coming from the surface sources of the WOH watersheds remained high, though concerns over development in the areas surrounding the watershed mounted. Development had particularly been a problem around the Croton system (which accounts for approximately 10 percent of the total supply), which was, by the early 1990s, already

filtered. In 1989, the EPA expressed concern about the developments around the NYC watersheds that were primarily driven by growth in tourism in the area. "In 1989, EPA's Surface Water Treatment Rule (SWTR), issued under the federal Safe Water Drinking Act, required filtration of all surface water supplies (rivers and lakes) to protect against microbial contamination of drinking water" (USEPA 1996). Specifically, the regulation required that all surface water systems serving populations greater than 10,000 filter their water for microbial contaminants and reduce turbidity (USEPA 1989). New York City officials were already concerned about water quality when the EPA released the SWTR. In January 1991, the New York State Department of Health, acting as the EPA's designated agency to carry out the Safe Drinking Water Act, ordered NYC to build a water filtration plant for the Catskill/Delaware system—the watershed that supplies 90 percent of NYC's water. A filtration system, however, would come at a high price. New York City "estimated that it would cost from $4 to $6 billion to filter the Catskill/Delaware supplies, which would more than double current water rates" (CCCD 1997: 3).

The high cost of filtration created economic incentives for NYC to seek alternatives. The regulation also provided an alternative to costly filtration. The "requirement can be waived if a water system's treatment processes and natural conditions provide safe water and if the watershed is actively protected to ensure that safety in the future" (USEPA 1996). In 1990, hoping to attain a Filtration Avoidance Determination (FAD), the Department of Environmental Protection for NYC drafted a Watershed Protection Plan that updated watershed rules, proposing regulatory changes to restrict construction of roads, parking lots, and storage facilities for hazardous substances and waste. NYC received a conditional FAD in January 1993. One condition of the FAD was that NYC issue final proposed Watershed regulations, begin to acquire land and conservation easements by June 30, 1994, and upgrade sewage treatment plants in the Watershed (CCCD 1997: 3). In addition, SWTR criteria required NYC "to show, through ownership or agreements with landowners, that it could control human activities in the watershed with the potential to harm the microbiological quality of the source water" (NYC DEP 2006: 3).

Acquiring the needed land or easements on land, and imposing new Watershed regulations would prove difficult, given that in 1991

only 27 percent of WOH land was publicly owned. Recalling NYC's use of eminent domain to build reservoirs in the early and mid-20th century, communities feared that the proposed plan would stifle economic growth and damage property values (NRC 2000: 26). These communities organized in 1991, forming the Coalition of Watershed Towns (CWT) comprised of more than 30 watershed towns, villages, and hamlets (CCCD 1997: 3), with the goals of ensuring proposed regulations would not prevent reasonable community development and limiting the regulations to the minimum needed to protect water quality (NRC 2000: 26). In late 1993, NYC filed an application for a water supply permit with plans to acquire watershed acreage and submitted to the EPA a long-term filtration avoidance program for the Catskill/Delaware System. Uncertainty over NYC's intent to use eminent domain and the perception by upstate communities that costs were being shifted onto them led the CWT and others to file lawsuits to prevent NYC's plans being implemented, leading to an impasse in efforts to reach a compromise on watershed management plans (NRC 2000: 27).

According to the 1990 plan, New York City assumed the right to regulate its watershed, pushing many of the costs onto the residents of watershed communities. The CWT, charged with protecting the interests of watershed communities, towns, villages, and hamlets, prevented this move through litigation. The State of New York had recognized the right of watershed communities and of private landowners in the watershed to be free of NYC regulation. Determined to pursue an FAD, New York City and the EPA asked the State of New York for help. In April 1995, Governor George Pataki organized a meeting among the involved parties (representatives from New York City, the EPA, the CWT, Putnam and Westchester counties, and others) to resolve the multitude of issues involving the Watershed (CCCD 1997: 4). New York City was desperate to strike a deal, while the watershed communities were not interested in making any deal that pushed costs onto them.

The Coasean Framework of the New York City Watershed Agreement

New York City and the State of New York wanted to assure the continued supply of high quality water for the eight million residents of NYC, the one million New York residents outside the City who

depend on the watersheds, and the millions of tourists who visit each year. However, New York City neither wanted to invest in the installation of the multibillion-dollar filtration system nor to incur future stream maintenance and upgrade costs. NYC suggested environmental regulations be prescribed for upstream, thereby reducing potential contamination. NYC also needed the watershed communities to cooperate with new rules and regulations in order to obtain the FAD which would lower NYC's costs of meeting the SWTR.

Because of the strict requirements of obtaining the FAD, any agreement reached between NYC and the watershed communities would have to be comprehensive. With the watershed communities having been recognized as holding the property right, NYC would have to compensate them for costs imposed by new rules and regulations. The Hudson RiverKeeper Fund, a local environmental group, argued that environmental protection could not happen without a strong economy. Farmers in the regional communities, represented by the CWT, were already feeling burdened by previous watershed restrictions and wary of new implementations such as nonpoint source pollution limits. Residents in the watershed communities also did not want to pay for upgrading their septic systems, a proposal from previous negotiations. The potential gains on both sides from bargaining were high, considering the immense cost of the proposed filtration plant.

Arranging negotiations between affected parties appeared cost prohibitive. The scope ranged from individual landowners to the State of New York, from residents of villages to the masses of New York City, and from environmental groups to the federal government. The magnitude of suggested pollution control efforts was on the order of billions of dollars. Nonpoint source pollution and water's physical nature amplified the circumstances. Agricultural runoff containing pesticides and fertilizers was increasing. Outdated septic systems were deteriorating, spilling contaminants into the flow. The regional residents portrayed a relaxed and uninformed stance about water pollution. Many parties were responsible for the degradation of water quality, but few were accountable. Generally, NYC found fault with the upstream landowners and government's lack of regulation. The Catskills region confronted a financial burden of maintaining livelihoods while meeting regulations that provided benefits only downstream.

This problem traditionally would have been relegated to Pigouvian taxation, subsidization, or authoritative regulation due to the scope,

magnitude, and dispersion of the pollutants. However, Pigouvian taxation would have been ineffective because the contamination is diffuse, making it difficult to pinpoint the responsible party. Subsidization would have resulted in inefficient and uncertain allocation of funds. Further authoritative regulation would have been fraught with market inefficiencies, monitoring difficulties, and enforcement costs.

Given that the State of New York recognized the right of watershed communities and of private landowners in the Watershed to be free of regulation from New York City, the foundation of a Coasean bargaining solution had been laid. Within the scope of a Coasean framework, property rights were decidedly held by the watershed communities. In this case, the multiplicity of externalities and the high potential gains for all parties indicates Coasean bargaining— contrary to the limitations described in literature—could provide for a more efficient solution, with government-assisted mediation, creating a partnership toward parallel, compatible goals.

Governor Pataki's bringing together of the interested parties eventually led to an announcement on November 2, 1995, that the parties "had reached an *Agreement in Principle* outlining the measures that would be taken to maintain and enhance the quality of the City's drinking water supply, while protecting the economic vitality and social character of watershed communities" (CCCD 1997: 4). The EPA played an important role during negotiations, remaining flexible with temporary FADs while the negotiations were under way, though still monitoring the drinking water in NYC, to verify it was within compliance with the Safe Drinking Water Act. The Agreement in Principle built the foundation on which the 1997 Memorandum of Agreement was erected. Table 1 provides a timeline of the important events in the MOA history.

The Memorandum of Agreement is a legally binding agreement enforceable in court. It was officially signed and accepted in January 1997 and gave New York City the comprehensive watershed management plan it needed to fulfill the requirements of the SWTR and obtain a five year FAD. The EPA issued a five-year Filtration Avoidance Determination on May 6, 1997, requiring "New York City to acquire environmentally-sensitive land in the Watershed, adopt strong watershed rules and regulations and institute and maintain a comprehensive watershed protection program" (USEPA 2010). During the negotiations of the MOA, the EPA set the minimum

TABLE 1
New York City Watershed MOA:
Timeline of Important Dates

Year	Events
1986	The Federal Safe Drinking Water Act Amendments are signed into law.
1989	The U.S. EPA promulgates the Surface Water Treatment Rule, requiring that all surface water systems serving over 10,000 people filter their water or obtain an exemption by ensuring the protection and quality of their water.
1990	New York City's Dept. of Environmental Protection releases a draft of its plan to protect the quality of the water (the Watershed Protection Plan).
1991	Coalition of Watershed Towns (CWT) comprised of 30 watershed towns is formed in response to NYC's proposed watershed management plan.
1992	New York State declares New York City may not use eminent domain or regulate watershed community activities.
1993	Jan.: EPA grants New York City a temporary Filtration Avoidance Determination. Dec.: EPA grants NYC an FAD effective until Dec. 1996.
1995	Governor Pataki brings together the CWT, NYC, and EPA to discuss the problem. Agreement in Principle is reached between New York City and other identified stakeholders.
1997	Jan: Memorandum of Agreement is signed and formally accepted. May: NYC is issued a five-year FAD, with ongoing monitoring and evaluation of the watershed.
2000	EPA conducts mid-term review of the 1997 FAD, provides feedback to the City, and makes further recommendations.
2001	NYCDEP submits to EPA its 2001 Long-Term Watershed Protection Program.
2002	Original five-year FAD is re-examined. WAP is expanded, and further regulations are recommended on storm water treatment and WWTPs, among other things. NYC receives another five-year FAD from the EPA.
2006	EPA passes new regulation requiring UV filtration of drinking water to prevent cryptosporidium outbreaks. Construction begins on the new UV filtration plant.

Continued

TABLE 1 *(cont.)*
New York City Watershed MOA:
Timeline of Important Dates

Year	Events
	NYC releases 2006 Long-Term Watershed Protection Program, which promotes better stream management.
2007	EPA reviews the FAD, and again makes recommendations. NYC receives a 10-year FAD, with regular reviews to be conducted and ongoing monitoring.
2011	NYC releases 2011 Long-Term Watershed Protection Program, which pushes for better riparian zone protection along streams and rivers.

standards and regulations necessary for obtaining an FAD, and facilitated provisions in the MOA that would allow NYC to attain it.

The specifics of the agreement demonstrate how the State of New York worked on each involved party's behalf. This assistance, particularly by Governor Pataki, facilitated meeting arrangements and negotiation, thereby reducing total costs below the perceived gains that would be garnered by an amiable resolution. The State allowed NYC to make direct payments to affected counties totaling $250 million, enhancing the regional economy while making NYC appear philanthropic. The MOA determined, also with the State's insistence, a requirement that NYC purchase vast amounts of land around the Watershed. This acquisition program deemed New York City responsible for its water quality and simultaneously averted its potential expenditure on a new water filtration facility. Finally, the State supported the watershed communities by creating programs such as the Catskill Fund for the Future, managed jointly by the Catskill Watershed Corporation (also created specifically for the Agreement) and the State Environmental Facilities Corporation.

The New York City Watershed Agreement's main components fall broadly under three themes: (1) the land acquisition program, (2) distinct watershed rules and regulations, and (3) protection and partnership programs. Additionally, the Watershed Agricultural Program provided incentives for farmers and ranchers to implement practices that reduce water pollution.

Land Acquisition Program

The purpose of the land acquisition program (Art. II of the MOA) is for NYC to acquire fee title or conservation easements on important land in the Watershed on a volunteer basis limited to 300,000 acres total or 30 percent of the Watershed. The MOA specifically prohibits NYC from obtaining land through eminent domain (Art. II, Sec. 59) as well as ensuring that all land is acquired from willing sellers (Sec. 61). In the MOA, New York City committed to set aside $250 million for the land acquisition program (Sec. 74 [a]), to pay fair market value for property acquired (Sec. 61), and to provide due notification to the town or village and appropriate county before commencing solicitation (Sec. 60). New York City is also required to pay local and county taxes on property acquired or a portion of taxes on land under conservation easement. The Agreement allows for development-limiting conservation easements to be purchased by NYC. Land near intakes into the water system or located in subbasins are highest priority while land located further from reservoirs and intakes are low priority.

Watershed Rules and Regulations

Article III of the MOA covers watershed rules and regulations on wastewater treatment plants to product use and storage. Article V creates several key partnership programs that administer NYC funds to the watershed communities and residents for the purpose of assisting them in observing the new regulations. These grants can be used to assist in upgrading and replacing equipment, or building new facilities necessitated by the new regulations. While the State of New York is responsible for enforcing the new restrictions, funds are available for upgrading septic systems, salt and other snowmelt material storage, as well as community education programs and forestry management programs. This portion of the Agreement served the original EPA stipulation for the City's drinking water standards. The regulations also require continued water quality monitoring and studies.

Watershed Protection and Partnership Programs

The protection and partnership programs (Art. V of the MOA) comprise an array of Agreement-initiated projects for land management, waste treatment, economic development, and education, each

cluster having access to financial aid. Many of the opportunities provide assistance creating best management practices for land and resource management. The MOA allocates $75 million for new sewage treatment infrastructure facilities for towns, villages, and hamlets. The Catskill Water Corporation (CWC) is allocated $3.5 million to fund its operating expenses, and administer or consult for New Sewage Infrastructure Funds, the Catskill Fund for the Future, the Stormwater Fund, the Septic System and Rehabilitation Program, Sand and Salt Storage Facilities, Stormwater retrofits, Stream Corridor Protection, the Alternative Design Septic Program, Public Education, and the Economic Development Study (CCCD 1997: 20–22). The aforementioned Catskill Fund for the Future is a loan and grant program for qualified development proposals. Fund-supported projects must enhance the economic base of the community in a manner not inimical to protection of regional water quality.

Finally, the MOA provides for "Good Neighbor Payments" by New York City that can be used for a variety of specified purposes within the municipalities to which the gift is paid, including "capital costs of designing, constructing and installing public works or public improvements, or purchasing public equipment, except for automobiles for six or less passengers, which will benefit the public at large" (Sec. 147[b][iii]).

Table 2 displays a list of the protection and partnership programs and the funds allocated to each respectively and indicates whether it is administered by or facilitated in consultation with the CWC.

Watershed Agricultural Program

Farm runoff, as nonpoint source pollution, is difficult to regulate (Hardy and Koontz 2008: 301–02). Instead of regulating agricultural practices in the MOA, the New York City Department of Environmental Protection and the EPA pushed for the establishment of the Watershed Agricultural Program (WAP). Article V of the MOA establishes and funds the WAP, the goal of which is "to refine and demonstrate an environmentally sound approach to farm management" (NRC 2000: 522). The function of the WAP is to provide economic incentives for farmers and ranchers of larger operations in the WOH watershed to develop Whole Farm Plans on a voluntary basis. "Whole Farm Plans are intended to reduce pollutant loadings by using innovative best management practices (BMPs)" (NRC 2000: 28).

TABLE 2
Overview of Major Partnership Programs

Protection and Partnership Programs	Allocated Funds
Sewage Treatment Infrastructure*	$75.0 M
Catskill Fund for the Future*	$59.7 M
Stormwater Fund*	$31.7 M
Septic Rehabilitation & Replacement*	$13.6 M
Sand and Salt Storage Facilities*	$10.25 M
Sewer Extensions	$10.0 M
Good Neighbor Payments	$9.765 M
Stormwater Retrofits*	$7.625 M
SPDES Upgrades	$5.0 M
Catskill Watershed Corporation	$3.5 M
Stream Corridor Protection*	$3.0 M
Tax Consulting Fund	$3.0 M
Alternate Design Septics*	$3.0 M
Public Education	$2.0 M
Forestry Management Program	$0.5 M
Economic Development Study*	$0.5 M
Total	$238.14 M

*Indicates a program in consultation with or administered by the Catskill Watershed Corporation.
Source: CCCD (1997: 19).

Implementation, Enforcement, and Early History of the MOA

Because the MOA is a legally binding agreement, it was necessary to implement some mechanism to ensure ongoing compliance and cooperation. The purpose behind the establishment of the CWC is for the nonprofit, locally-administered corporation to oversee, authorize funding for, or act as consultants for many of the programs provided for in the MOA (CCCD 1997: 26). Local enforcement of regulations and administration/implementation of programs is important to ensure compliance and cooperation.

In addition to the CWC, the Watershed Protection and Partnership Council (WPPC) is established in the MOA. This Council, which consists of representatives from city and state government agencies,

watershed counties, environmental groups, the CWC, the EPA, and the Watershed Agricultural Council, serves as a forum for the exchange of views, periodically reviews efforts undertaken by governments and private parties to protect the Watershed, and solicits input from agencies and private organizations and persons with an interest in the Watershed and New York City drinking water (NRC 2000: 123).

The oversight by the CWC and WPPC is supplemented by EPA Region 2 along with the NYC DEP and NYC Department of Health continuing to monitor water quality. Specifically, NYC implemented an ongoing water quality monitoring program, and the FAD is subject to continual review by the EPA, which ensures that sufficient progress is being made on MOA agreements and that EPA regulations are met. A May 2000 EPA review of the 1997 Filtration Avoidance Determination (FAD) concluded that, although NYC had made significant progress in many of its watershed protection programs, a number of corrective actions for specific FAD tasks as well as program enhancements needed to be implemented for the long-term viability of filtration avoidance (USEPA 2002). Due to the mid-term review and recommendations, NYC expedited its efforts on WWTPs in the Watershed region and acquired the needed land and easements to satisfy the EPA. On December 15, 2001, NYC DEP submitted to EPA its Long-Term Watershed Protection Program, emphasizing watershed protection as a long-term commitment (USEPA 2002). The EPA viewed the 2001 Long-Term Watershed Protection Program as a significant improvement upon the original MOA and applauded NYC's long-term commitment. The City released updated Long-Term Watershed Protection Programs based on mid-term reviews and recommendations in the year prior to the end of the FAD period in 2006 and 2011, respectively.

At the end of the five-year FAD in 2002, the provisions of the MOA were again reexamined. Among the notable changes in the 2002 MOA were (1) the expansion of the Watershed Agricultural Program to include EOH watershed farms as well as small farms that previously did not meet the program's gross farm income participation threshold;[5] (2) updates to the treatment of storm water; (3) increased outreach and education programs; and (4) a new focus

[5]As a result, by January 2006, 28 whole farm plans and 42 small farm whole farm plans had been approved by the Croton EOH watershed (USEPA 2011).

on the Kensico river basin, through which as much as 60 percent of the water in the Catskill/Delaware watersheds flows (USEPA 2002). Additional funds were committed to the Septic System Program to more actively address problems with failing septic systems, and active stream management was also encouraged in the 2002 FAD, creating a Stream Management Program to increase riparian zones and thereby decrease turbidity in streams feeding surface water sources. The EPA determined that NYC's 2001 Long-Term Watershed Protection Program "if complied with, will achieve the objectives of the Safe Drinking Water Act and the Surface Water Treatment Rule for unfiltered systems" (USEPA 2002). New York City received another five-year FAD from the EPA based on its own 2001 Long-Term Watershed Protection Program and its commitments to the recommendations in the 2002 FAD.

During the 2002–07 FAD timeframe, new EPA regulations required NYC to build a UV filtration facility. Construction began in 2006 and finished in 2010, allowing NYC to filter all of the water for cryptosporidium and other microbial pathogens (USEPA 2011). The revised FAD was granted to NYC based on the plans to construct the UV filtration plant and for the City to provide an additional $6 million to Wastewater Management Programs in the Watershed region (USEPA 2006a). A 2006 review of the progress on the various programs and implementation of MOA provisions notes, "Overall, the City has successfully satisfied the obligations specified in the 2002 FAD" (USEPA 2006b). Having made satisfactory progress during the 2002–07 FAD period, NYC was again issued an FAD, this time for 10 years (USEPA 2007).

Discussions surrounding the 2007 FAD received mixed input from various groups. Watershed communities expressed fear that further land acquisition by NYC would stifle local growth and economic activities while environmental groups urged an increase in land acquisition (USEPA 2007). The addition of the Stream Management Program was continued and bolstered in the 2007 FAD, along with a more comprehensive Riparian Buffer Protection Program designed to provide technical assistance to "streamside landowners who seek to implement stabilization and planting plans to enhance riparian buffers" (USEPA 2007).

In 2011, NYC released its renewed Long-Term Watershed Protection Program. The major update is a 15-year extension of the land acquisition program in conjunction with the New York State

Department of Environmental Conservation (NYSDEC) issuance of the City's Water Supply Permit. "The WSP resulted from extensive discussions among the City, NYSDEC, NYSDOH, the U.S. Environmental Protection Agency, Watershed counties, the Coalition of Watershed Towns, environmental groups, and other stakeholders" and extends financial and other commitments for MOA programs (NYC DEP 2011). Currently, NYC is in compliance with the MOA and the regulations promulgated by the EPA.

Conclusion

The New York City Watershed MOA is the first of its kind. While most discussions of the Coasean approach describe it as applicable only to externality cases in which the number of actors is identifiable and few in number, the MOA exemplifies the possibility of successful Coasean negotiations for large-number, large-scale, high-impact externality problems. The consequences of this MOA could be measured in billions of dollars, and the impact measured in over 1,900 acres and by well over 9 million people. The negotiations forged a new method for dealing with externalities showing how Coasean solutions could be facilitated and used in wide-reaching economic conflicts.

Every component of the Agreement reveals support for the regional economy and places the burden of water quality on NYC. In Coasean bargaining terms, the State assigned the watershed communities the property rights to continue with regional development and current practices, although degrading to NYC's drinking water. This assignment positioned NYC as the responsible party for initiating negotiation and programs protecting the Watershed system. Once responsibility was established, NYC opted to buy lands that were contributing to water quality degradation, such as farmlands on riparian corridors, instead of building a multibillion-dollar filtration system. The residents and landowners upstream are compensated for development restrictions incurred from the Agreement. Furthermore, the role of the State as a mediator eliminated bargaining barriers by effectively reducing the transaction costs of arranging negotiations, demonstrating potential economic benefits to all parties, and providing alternative options to unilateral decisionmaking.

The negotiations of the MOA set a new precedent in addressing externalities that involve a large number of stakeholders. The parties

to the MOA included stakeholders from various communities including the CWT, NYC, governmental organizations such as the EPA and New York State, and environmental groups including the Catskill Center for Conservation and Development. Many of the stakeholders were identified prior to the negotiation process because of the aforementioned conflict between NYC and the CWT, which is comprised of more than 30 municipalities and towns within the Watershed region. Environmental groups were brought into negotiations because of the environmentally sensitive nature of the land and system under consideration. The farming community of the Watershed area, although not directly involved in the negotiations, was represented by the CWT.[6]

The New York City Watershed MOA raises the possibility that governments can perform the role of simplifying negotiations in a way that will meet the overall social objectives and lead to Pareto improvement. Overall, when we observe externality problems, "We should ask not merely where the problem comes from, but what the transaction costs are that prevent it from being bargained out of existence" (Friedman 2000: 40). As demonstrated by the MOA, the government's role simplified negotiations for New York City by drastically reducing the costs of negotiation. Rather than the traditional regulatory approach to externality problems, a Coasean approach was facilitated by new forms of relationships between different branches of government and affected parties in the private sector.

Of interest here is the organic manner in which the reduction in externality losses was achieved. Although the participants were engaged in a process of the type envisioned by Coase, nowhere in the documents relating to the agreement and the follow-up is Ronald Coase or any version of the Coase Theorem mentioned. Rather, once property rights were established, participants did what people do whenever they have goals and must negotiate with others to meet them. The spontaneous order that arose is clearly Coasean, even though none of the participants were likely aware that they were engaged in activity suggested by a Nobel laureate.

The MOA and subsequent revisions serve as an example of how large-scale, large-number externalities can be successfully internalized through the kind of negotiation Coase advocated in his

[6]The originally proposed regulations created concern for watershed farmers due to the loss of as much as 25 percent of tillable land (NRC 2000: 28).

work. While the State of New York and the EPA facilitated the initial negotiations, subsequent discussions have taken place using the mechanisms established within the original MOA and from commentary solicited by the EPA. New programs and revisions are made on an ongoing basis with input from all concerned parties. Overall, the New York City Watershed Memorandum of Agreement is a huge success, and demonstrates the ability of the market to resolve externality problems with the help of minor state intervention.

References

Abrego, L., and Whalley, J. (2004) "Firm Property Rights, Bargaining, and Internalization." *Chinese Journal of Economic Theory* 1 (1): 1–13.

Anderlini, L., and Felli, L. (2006) "Transaction Costs and the Robustness of the Coase Theorem." *Economic Journal* 116 (508): 223–45.

Aslanbeigui, N., and Medema, S. G. (1998) "Beyond the Dark Clouds: Pigou and Coase on Social Cost." *History of Political Economy* 30 (4): 601–25.

Bator, F. M. (1958) "The Anatomy of Market Failure." *Quarterly Journal of Economics* 72 (3): 351–79.

Bergson, A. (1938) "A Reformulation of Certain Aspects of Welfare Economics." *Quarterly Journal of Economics* 52 (2): 310–34.

Black, G.; Grant, L.; and Islam, S. (2008) "The Coase Theorem and Large Number Externalities: The Case of the New York City Watershed Agreement." Mimeo.

Canterbery, E. R., and Marvasti, A. (1992) "The Coase Theorem as a Negative Externality." *Journal of Economic Issues* 26 (4): 1179–89.

Catskill Center for Conservation and Development (1997) "Summary Guide to the Terms of the Watershed Agreement." Available at www.cwconline.org/pubs/Summary%20Guide.pdf.

Catskill Watershed Corporation (2005) "Watershed History." Available at www.cwconline.org/about/ab_hist.html.

Cheung, S. N. S. (1973) "The Fable of the Bees: An Economic Investigation." *Journal of Law and Economics* 16 (1): 11–33.

Coase, R. H. (1937) "The Nature of the Firm." *Economica* (n.s) 4 (16): 386–405.

_____ (1959) "The Federal Communications Commission." *Journal of Law and Economics* 2: 1–40.

_____ (1960) "The Problem of Social Cost." *Journal of Law and Economics* 3: 1–44.

_____ (1988) *The Firm, the Market, and the Law.* Chicago: University of Chicago Press.

Cooter, R. (1982) "The Cost of Coase." *Journal of Legal Studies* 11 (1): 1–33.

Cooter, R., and Ulen, T. (1997) *Law and Economics*. 2nd ed. Reading, Mass.: Addison-Wesley.

Daily, G. C., and Ellison, K. (2002) *The New Economy of Nature: The Quest to Make Conservation Profitable*. Washington: Island Press, Shearwater Books.

Davis, B. (1992) "Bush Plans to Unveil a 90-Day Moratorium on New Regulations." *Wall Street Journal* (20 January): A1.

Frech, H. E. (1973) "Pricing of Pollution: The Coase Theorem in the Long Run." *Bell Journal of Economics and Management Science* 4 (1): 316–19.

Friedman, D. (1990) *Price Theory: An Intermediate Text*. Cincinnati, Ohio: South-Western Publishing Company. Available at www.daviddfriedman.com/Academic/Price_Theory/PThy_Chapter_18/PThy_Chap_18.html.

_____ (2000) *Law's Order: What Economics Has to Do with Law and Why It Matters*. Princeton, N.J.: Princeton University Press.

Graaf, J. de V. (1957) *Theoretical Welfare Economics*. Cambridge: Cambridge University Press.

Hardy, S., and Koontz, T. (2008) "Reducing Nonpoint Source Pollution through Collaboration: Policies and Programs across the U.S. States." *Environmental Management* 41 (3): 301–10.

Hayek, F.A. (1945) "The Use of Knowledge in Society." *American Economic Review* 35 (4): 519–30.

Higgs, R. (2005) "Coasian Contracts in the Coeur d'Alene Mining District." *Cato Journal* 25 (2): 407–21.

Hoffman, E., and Spitzer, M. L. (1985) "Entitlements, Rights, and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice." *Journal of Legal Studies* 14 (2): 259–97.

_____. (1995) "Experimental Tests of the Coase Theorem with Large Bargaining Groups." In S. G. Medema (ed.) *The Legacy of Ronald Coase in Economic Analysis*, Vol. 2, 129–51. Northampton, Mass.: Edward Elgar.

Johnson, D. B. (1973) "Meade, Bees, and Externalities." *Journal of Law and Economics* 16 (1): 35–52.

Lange, O. (1942) "The Foundations of Welfare Economics." *Econometrica* 10 (3/4): 215–28.

Marlow, M. (1995) *Public Finance: Theory and Practice*. Fort Worth: Harcourt Brace.

McCloskey, D. N. (1985) *The Applied Theory of Price.* 2nd ed. New York: Macmillan.

Meade, J. E. (1952) "External Economies and Diseconomies in a Competitive Situation." *Economic Journal* 62 (245): 54–69.

Medema, S. G. (2009) *The Hesitant Hand*. Princeton, N.J.: Princeton University Press.

Medema, S. G., and Zerbe, R. O. (2000) "The Coase Theorem." In B. Bouckaert and G. De Geest (eds.) *Encyclopedia of Law and Economics*, Vol. 1, 836–92. Northhampton, Mass.: Edward Elgar.

Mill, J. S. (1871). *Principles of Political Economy.* 7th ed. London: Longmans, Green.

Nalebuff, B. (1997) "On a Clear Day, You Can See the Coase Theorem." In P. Dasgupta and K.-G. Mäler (eds.) *The Environment and Emerging Development Issues*, Vol. 1, 35–47. Oxford: Oxford University Press.

National Research Council (2000) *Watershed Management for Potable Water Supply: Assessing the New York City Strategy*. Washington: National Academy. Available at www.nap.edu/catalog.php?record_id=9677.

New York City Department of Environmental Protection (2006) "2006 Long-Term Watershed Protection Program." Available at www.epa.gov/region2/water/nycshed/2007wp_program121406final.pdf.

_____ (2011) "2011 Long-Term Watershed Protection Program." Available at www.nyc.gov/html/dep/pdf/watershed_protection/2011_long_term_plan.pdf.

_____ (2012) "History of New York City's Water Supply System." Available at www.nyc.gov/html/dep/html/drinking_water/history.shtml.

New York City Watershed Memorandum of Agreement (1997) "MOA," signed January 21. Available at www.nysefc.org/Default.aspx?TabID=76&fid=389.

Pigou, A.C. (1912) *Wealth and Welfare*. London: Macmillan.

_____ (1932) *The Economics of Welfare.* 4th ed. London: Macmillan.

_____ (1935) "State Action and Laisser-Faire." In A. C. Pigou, *Economics in Practice: Six Lectures on Current Issues*, 107–28. London: Macmillan.

Royal Swedish Academy of Sciences (1991) Press Release: "The Sveriges Riksbank (Bank of Sweden) Prize in Economic Sciences in Memory of Alfred Nobel."

Samuelson, P. (1963) "Modern Economic Realities and Individualism." *Texas Quarterly* (Summer): 128–39. Reprinted in J. E. Stiglitz (ed.) (1966) *The Collected Scientific Papers of Paul A. Samuelson*, Vol. 2, 1407–18. Cambridge, Mass.: MIT Press.

_____ (1995) "Some Uneasiness with the Coase Theorem." *Japan and the World Economy* 7 (1): 1–7.

Stigler, G. J. (1966) *The Theory of Price.* 3rd ed. New York: Macmillan.

Tybout, R. A. (1972) "Pricing Pollution and Other Negative Externalities." *Bell Journal of Economics and Management Science* 3 (1): 252–66.

United States Environmental Protection Agency (1989) "40 CFR Parts 141 and 142 (Surface Water Treatment Rule)." *The Federal Register* 54 (124). Available at http://water.epa.gov/lawsregs/rulesregs/sdwa/swtr/upload/SWTR.pdf.

_____ (1996) "Watershed Progress: New York City Watershed Agreement." EPA840-F-96–005. Available at http://water.epa.gov/type/watersheds/nycityfi.cfm.

_____ (2002) "New York City Filtration Avoidance Determination." Available at www.epa.gov/region02/water/nycshed/2002fad.pdf.

_____ (2006a) "Final Modification of the New York City Filtration Avoidance Determination." Available at www.epa.gov/region2/water/nycshed/final_fad_modificationtransmittalletter.pdf.

_____ (2006b) "Report on the City of New York's Progress in Implementing the Watershed Protection Program, and Complying with the Filtration Avoidance Determination." Available at www.epa.gov/region02/water/nycshed/documents/epaeval_august2006.pdf.

_____ (2007) "New York City Filtration Avoidance Determination." Available at www.epa.gov/region2/water/nycshed/2007finalfad.pdf.

_____ (2010) "Filtration Avoidance." Available at www.epa.gov/region2/water/nycshed/filtad.htm.

_____ (2011) "Watershed Protection Programs." Available at www.epa.gov/region02/water/nycshed/protprs.htm.
Varian, H. R. (1995) "Coase, Competition and Compensation." *Japan and the World Economy* 7 (1): 13–27.
Wu, J., and Babcock, B. (1996) "Contract Design for the Purchase of Environmental Goods from Agriculture." *American Journal of Agricultural Economics* 78 (4): 935–45.

Marginal Tax Rates and U.S. Growth: Flaws in the 2012 CRS Study

Jason E. Taylor and Jerry L. Taylor

In September 2012, seven weeks before the presidential election—one in which top marginal tax rates were a major policy difference between the two major-party candidates—the Congressional Research Service (CRS) published a paper (Hungerford 2012) suggesting that there is no empirical evidence that top marginal tax rates impact U.S. economic growth. After all, top marginal tax rates were above 90 percent during the 1950s and early 1960s when the economy experienced rapid growth. Furthermore, marginal tax rate cuts in 2001 and 2003 were followed by the worst financial crisis since the Great Depression. The CRS study was widely reported in blogs, newspapers such as the *New York Times,* and *The Atlantic* magazine. It was portrayed as evidence refuting Republican candidate Mitt Romney's position that cutting the top marginal tax rate from 35 to 28 percent would spur economic growth and supporting Democratic President Barack Obama's position that top marginal tax rates could be raised to 39.6 percent with no cost to economic growth (Leonhart 2012, Thompson 2012).

Republicans claimed that the study was methodologically flawed and asked that the CRS report be pulled. On November 1, 2012, five days before the election, the report was pulled, and its content, as

---

Cato Journal, Vol. 34, No. 1 (Winter 2014). Copyright © Cato Institute. All rights reserved.

Jason E. Taylor is Jerry and Felicia Campbell Professor of Economics at Central Michigan University, and Jerry L. Taylor is Professor of Economics and Finance at Kaplan University. The authors thank participants at the 2013 Economic and Business History Conference and anonymous referees for their comments, and also Thomas Hungerford for generously sharing his data.

well as the controversy surrounding it, were back in the headlines again. The *New York Times* quoted Sen. Charles Schumer (D–NY) saying, "This has hues of a banana republic. [Republicans] didn't like a report, and instead of rebutting it, they had them take it down" Weisman (2012). The study was reissued by the CRS in an "updated" form on December 12, 2013, with no major changes to the original.

Entin (2013) claims that the CRS study's model is flawed in that it does not control for several other factors that could have affected growth and thus "poisons its results by not holding other factors constant." Furthermore, Entin notes that it takes years for firms to fully adjust to changes in tax structure and that looking only at the effects of tax changes one year out "misses the point." In fact, the CRS study analyzes year-over-year changes rather than levels (because the data are not stationary), and hence it effectively asks whether GDP growth rates were different in years such as 1964, 1987, 1993, or 2003, when there was a change in the top marginal tax rate, relative to years in which there was no change in top marginal tax rates. But the key issue of interest is not whether a tax rate change has an effect on economic performance during that same year, but whether it changes the growth trajectory in subsequent years. Even very small changes in the rate of economic growth, if they are persistent, can have a very large impact on the size of the economy over time because of compounding.

Indeed, the vast literature examining tax rates and economic growth strongly suggests that marginal tax rates and GDP growth rates are negatively related. This result is well established both through the use of time series data for the United States and via large panels of international data. In this article, we employ the exact data and specifications from the CRS study but change the methodology to analyze how changes in top marginal tax rates affect growth over the following three to five years rather than just the year of the change. After this modification, the regressions suggest that tax cuts have brought faster economic growth in subsequent years in the postwar United States, consistent with the theoretical and empirical literature.

Literature Review: Marginal Tax Rates and Growth

A large literature exists in which the theoretical "optimal tax" is sought (Mirrlees 1971; Diamond and Mirrlees 1971; Saez 2001;

Mankiw, Weinzierl, and Yagan 2009). It is widely recognized in this literature that there is a tradeoff between income redistribution and efficiency. Proponents of progressive taxation (graduated taxes) argue that social welfare may rise when resources are more equitably distributed.[1] Furthermore, Conesa and Krueger (2006) argue that a progressive tax acts as a partial substitute for missing insurance markets. Still, taxes that vary with income distort behavior since they place a wedge between the market values of effort and reward. If taxes are highest on the successful drivers of growth, such as with a progressive tax, this will cause particularly large efficiency losses by distorting their labor supply and capital accumulation decisions. Along these lines, Cullen and Gordon (2002) suggest several avenues through which taxes affect entrepreneurial activity. Economists have long noted that a lump-sum tax, in which tax liabilities are independent of behavior, is the most efficient form of taxation since there is no distortive effect. Still, such a tax would be highly regressive, thus working strongly against the goal of fairness. Clearly the most efficient tax is unfair while taxes geared toward income redistribution are inefficient; high marginal taxes distort behavior and affect growth, even if they may be considered desirable from a perspective of income redistribution.

A large empirical literature has arisen to ascertain the importance of tax rates in determining growth in the real world—that is, how much of a tradeoff there is between income redistribution and efficiency. For example, Koester and Kormendi (1989) examined the relationship between effective tax rates and GDP of 63 countries during the 1970s. They found that although marginal tax rates do not affect GDP growth rates, a 1 percent tax cut would raise the level of per capita GDP by between 0.6 and 1.3 percent—creating a parallel shift in a nation's growth path. Following up on Koester and Kormendi, Engen and Skinner (1992) examined 107 countries between 1970 and 1985 and found a negative correlation between average tax rates and economic growth. Padovano and Galli (2001) expanded the time frame to 1950 to 1990, and examined a panel of 23 OECD countries. They found that effective marginal income tax

[1]This, of course, involves a value judgment and interpersonal utility comparisons. For the case against progressive taxation from a moral, rule of law, perspective, see Hayek (1960: chap. 20). Also see Blum and Kalven (1952).

rates are negatively correlated with economic growth. Lee and Gordon (2005) found that increases in corporate tax rates lead to slower economic growth. Some studies, such as Easterly and Rebelo (1993) do not find empirical evidence for any correlation between taxes and growth. Still, in a meta-analysis of 93 published studies on the effects of fiscal policies and long- run growth, Nijkamp and Poot (2004) conclude that there is broad empirical support for the hypotheses that higher taxes lead to slower growth. A more limited, but also more recent, review of tax studies by McBride (2012) found that 23 out of 26 studies have uncovered a negative relationship between taxes and growth while the other three found no significant relationship. With respect to tax rates and U.S. growth, Romer and Romer (2010), Barro and Redlick (2011), and Mertens and Ravn (2013) have further confirmed that changes in tax rates have a negative relationship with growth.

Studies of tax rates and growth have employed several different measures for taxes in their regressions. Theory suggests that *marginal* rates are particularly important since they distort relative prices and misallocate resources, resulting in welfare losses. An individual who is in the 50 percent marginal tax bracket gains only 50 cents on each extra dollar of income earned from work, saving, or investment, even if the average tax rate for that individual (total taxes paid divided by income) is only 20 percent. However, marginal effective tax rates are difficult to observe across the entire economy. Many studies, like Engen and Skinner (1992), have used average tax rates as a proxy by dividing tax revenues by GDP. However, Padovano and Galli (2001) estimated effective marginal tax rates by regressing total government revenues on gross domestic product, over 10-year intervals; the coefficient then yields the change in revenue for a one-dollar change in output. Another approach is to examine the top marginal tax rate, as Hungerford (2012) did in the CRS study. But that approach is not without its shortcomings, as it does not account for exemptions, deductions, evasion, and other strategies used by high-income earners in progressive tax regimes (Frenkel, Razin, and Sadka 1991).

Another aspect of the literature on the impact taxes have on growth examines differences in tax structures within the United States. Genetski and Chin (1978) found that growth in gross state product was negatively correlated with changes in state and local taxation. Dozens of studies have followed up or extended this seminal

work and the majority of them have concluded that state tax rates matter. Vedder (2001) provides a summary of this literature, while also concluding that states with lower tax burdens saw faster growth in the last half of the 20th century. Most recently, Laffer, Moore, and Williams (2012) have confirmed the consensus of this literature that low-tax states outperform high-tax states in terms of population growth, job growth, growth in gross state product, and growth in tax revenues.

Empirical Analysis: The CRS Study and Extensions

The motivation for this article is to explore the controversy behind the widely cited September 2012 CRS study by Hungerford suggesting that there is no evidence that changes in top marginal tax rates have impacted U.S. economic growth in the postwar era. Hungerford runs regressions in which the dependent variable is the growth rate of real per capita GDP and the independent variables include the change in the top marginal tax rate, the change in the top capital gains tax rate, the change in the percentage of the population who are college graduates, the change in the population growth rate, and the change in the real federal current expenditures ratio (real federal expenditures divided by potential real GDP).[2]

Hungerford's empirical analysis uses first-differenced data since the data in levels are not stationary and thus can lead to spurious results. However, he only asks whether the growth rate of real per capita GDP was different in years in which the top marginal tax rate changed. Table 1 reports the top marginal income tax rate from 1913, the year the income tax began, to 2013. Years in between the ones listed had the same rate as the prior year. The way Hungerford's regressions are specified, the tax rate variables take a zero value for all years when the top marginal tax rate did not change (and, hence, are not listed in the table). But this methodology is an oversimplification of the model: it suggests that changes in marginal tax rates only affect GDP growth in the year during which they were enacted. In fact, in many cases, tax rates were changed deep into the year in

[2]Hungerford also runs regressions with three other dependent variables: change in private savings ratio, change in private fixed investment ratio, and change in labor productivity growth rate. In no case does he find that the primary variable of interest—change in the top marginal tax rate—is statistically significant.

TABLE 1
Top Marginal Income Tax Rates and the Year They Went into Effect

Year	Top Marginal Rate (%)
1913	7
1916	15
1917	67
1918	77
1919	73
1922	58
1924	46
1925	25
1932	63
1936	79
1941	81
1942	88
1944	91
1964	77
1965	70
1982	50
1987	38.5
1988	28
1991	31
1993	39.6
2001	39.1
2002	38.6
2003	35
2013	39.6

Note: From 1968 to 1970, a Vietnam War surcharge was assessed on top rates as well. If these are considered, the top marginal rate was 75.25, 77, and 71.75 percent, respectively, during these three years. Some years during the late 1940s and 1950s were subject to maximum effective rate limitations equal to between 85.5 and 90 percent of "taxable income." In some cases this may have slightly altered the effective top marginal rate.
Sources: Data are from "Personal Exemptions and Individual Tax Rates, 1913–2002" (www.irs.gov/pub/irs-soi/02inpetr.pdf) and "Federal Individual Income Tax Rates History, Nominal Dollars, Income Years 1913–2013" (taxfoundation.org/sites/taxfoundation.org/files/docs/fed_individual_rate_history_nominal.pdf).

which they (often retroactively) took effect. One way to overcome this weakness is to examine whether or not GDP growth rates were different in the three, four, or five years after a change in top marginal tax rates occurred.

Table 2 reports the results of five regressions, which, following Hungerford, use 61 observations of annual data from 1950 to

TABLE 2
Dependent Variable: Growth Rate in Real GDP per Capita

	(1)	(2)	(3)	(4)	(5)
Constant	0.022059	0.01779	0.01038	0.01594	0.01554
	(4.34)***	(3.57)***	(1.77)*	(3.30)***	(3.54)***
1−Top Rate	−0.098				
	(−0.96)				
1−Cap Gains	−0.043				
	(−0.61)				
Percentage	−0.2699	−0.2351	0.2337	0.1997	0.1629
College Grad	(−0.32)	(−0.32)	(1.44)	(1.69)*	(1.47)
Population	−5.83	−5.7948	−0.0622	−0.0419	−0.0499
Growth	(−1.55)	(−1.59)	(−1.37)	(−1.25)	(−1.71)*
Fed Expenditures	−0.532	−0.5138	−0.1326	−0.0586	−0.0064
Ratio	(−0.95)	(−0.92)	(−1.17)	(−0.82)	−0.09)
Tax Cut Dummy		0.01014	0.01157	0.01203	0.01027
4 Years		(1.78)*	(2.02)**	(2.65)***	(2.25)**
Tax Increase Dum		0.00259	0.0027	0.0039	0.0046
4 Years		(0.43)	(0.50)	(0.95)	(1.26)
Growth Rate				−0.0772	−0.0779
Monetary Base				(−6.27)***	(−6.29)***
Change Stock				−0.0534	−0.0496
Mrk Return				(−5.03)***	(−4.37)***
Growth Labor					0.9408
Force/POP					(1.75)*
R-squared	0.0838	0.088	0.129	0.431	0.464
F-Statistic	1.01	1.06	1.62	5.72	5.65

Notes: T-statistics reported in parentheses. * Indicates statistical significance at the 10 percent confidence interval. ** Indicates statistical significance at the 5 percent confidence interval. *** Indicates statistical significance at the 1 percent confidence interval. Specifications (1) and (2) use the first difference of percentage of college graduates, population growth, and the federal expenditures ratio, while specifications (3), (4) and (5) use the log difference, or growth rate, of these variables.

2010. Also following Hungerford, all regressions use Newy-West corrected standard errors that allow for heteroskedastic and autocorrelated error terms. Specification (1) duplicates Hungerford's result. The primary variables of interest in Hungerford's regression are 1 minus the top marginal income tax rate and 1 minus the top capital gains tax rate; thus, they represent the leftover percentage of marginal income an earner in the top bracket would keep. The change in the percentage of the population who are college graduates, change in the population growth rate, and the change in the real federal current expenditures ratio are control variables. The r-squared is very low, as is the F-statistic on the regression. Additionally, as widely reported in the media, the coefficients on the tax variables are not statistically different from zero.

In response to the claim that the CRS study was flawed because it did not allow enough time for tax changes to have effects on behavior, specification (2) replaces the tax variables with two dummy variables: *Tax Cut Dummy 4 Years* takes on a value of 1 the year the top marginal tax rate is cut and the three years that follow, while *Tax Increase Dummy 4 Years* takes on a value of 1 the year of an increase in the top marginal rate and the three years that follow.[3] The coefficient on the *Tax Cut* dummy is positive and significant at the 10 percent level. The coefficient suggests that real per capita GDP grew about 1 percentage point faster in the four years following a tax cut (counting the year of the cut as the first year). The *Tax Increase* dummy is insignificant. Again the r-squared and F-statistic are low.

An alternative would be to look at *growth rates* in the control variables rather than just the year-over-year difference in them. Specification (3) is identical to specification (2) except that it examines the log difference of the control variables rather than just the difference. The *Tax Cut* dummy remains positive and statistically significant, now at the 5 percent level. The r-squared and F-statistics rise, but are still very low.

[3]The dummies turn on when the top marginal rate changes by more than 1 percentage point. The *Tax Cut* dummy takes on a value of 1 during 1964–67, 1982–85, 1987–90, and 2003–2006. The *Tax Increase* dummy takes on a value of 1 during 1951–54, 1968–71, and 1991–97.

Entin's (2013) major criticism of the CRS study was that it suffered from an omitted variables bias—namely, it did not control for enough other factors (such as monetary policy) that could affect real GDP growth, and thus isolate the effect of tax changes. Specifications (4) and (5) are an attempt to alleviate at least some of this concern. In specification (4), we add two new control variables—the growth rate in the monetary base and the growth rate in the S&P stock market—with the goal of explaining more of the variation in the growth of real per capita GDP. The r-squared jumps considerably as does the F-statistic, which is now statistically significant. This specification suggests that in the four years after a tax cut, the growth rate in real per capita GDP is 1.2 percentage points higher than in years in which no cut occurs. This result is significant at the 1 percent confidence interval. Finally, specification (5) adds the growth rate of the labor force to population ratio, to help control for demographic trends (women entering the labor force, changes in working age population structure) that could have affected the real per capita GDP growth rate. Again the *Tax Cut* dummy variable is positive and significant at the 5 percent confidence level.

To test the robustness of the finding that tax cuts brought faster growth in the postwar United States, we also tried dummy variables that controlled for 3 and 5 years around a tax change, rather than four, and the results were similar. In each case, the coefficient on the *Tax Cut* dummy was positive and significant at the 10 percent level or better, except in the case of using the differences (specification 2) for the 5-year dummy. Another issue is that Hungerford used tax data from the IRS that included some tax increases in 1951 and 1968, when the statutory top rates were not changed but surtaxes and surcharges were imposed. For example, 1968 to 1970 included Vietnam War surcharges that applied to the highest tax rate. We ran the regressions again with an *Alternative Tax Increase* dummy that only took on a value of 1 from 1991 to 1996, which were the years of and after the tax increases of 1991 and 1993. The major results are unchanged: the coefficient on the *Tax Cut* dummy remains positive and statistically significant at the same confidence interval, or better, in each specification. We also tried including the growth rate of real federal transfers as a percentage of potential GDP, and found that tax

cuts brought faster economic growth in the years following the tax change.[4]

Certainly, a dummy variable approach also has its shortcomings as it assumes that all tax cuts (large and small) are empirically identical. For another important robustness check, we replaced the 1s in the binary dummy variables with the change in the top marginal tax rate in the year of the change and the three following years. For example, for 1964 to 1967, rather than the *Tax Cut* variable taking on a value of 1, it took on a value of 14 for 1964 (reflecting the cut from 91 percent to 77 percent) and then a value of 21 for 1965, 1966, and 1967 (reflecting the cut from 91 to 70 percent once fully phased in). Consistent with the earlier results, the coefficients on the *Tax Cut* variable were positive and statistically significant at the 10 percent confidence level or better in specifications (2) through (5).

In one final robustness check, we combined the *Tax Increase* and *Tax Cut* variables into one *Tax Change* variable. This variable duplicated the earlier results, but took on negative values for tax cuts (e.g., −21 for 1965) and positive values for tax increases (e.g., 3 for 1991), again for the year of the cut and the following three years. This variable's coefficient was negative and statistically significant in specifications (2) through (4), generally confirming the notion that tax rates and growth are inversely related. In sum, when we allow for a time lag, the result that cuts in marginal tax rates brought faster growth in the postwar United States is quite robust, even using the exact data employed by the CRS study.

[4]While Hungerford's main regression dealt with the impact of tax rates on growth in real per capita GDP, he ran three other regressions whereby the dependent variables were change in the private saving as a percentage of potential GDP, change in fixed private investment as a percentage of potential GDP, and change in the labor productivity growth rate. For the investment and savings ratio regressions, change in AAA bond rates, and change in the S&P Stock market return were used as control variables. The investment regression also had lagged investment while the savings regression had change in disposable personal income. The productivity regression had only two controls: change in college graduates as a percentage of the population, and the change in the ratio of federal transfer payments as a percentage of potential GDP. We duplicated all these regressions replacing Hungerford's change in 1 minus the top marginal income and capital gains tax rates with our *Tax Cut* and *Tax Increase* dummies for four years. Consistent with Hungerford's findings, the tax dummies in these regressions were not statistically significant, meaning that from these specifications we cannot reject the null hypothesis that a change in marginal tax rates has no effect on these three variables, holding the specific controls constant.

Conclusion

In September 2012, a Congressional Research Service study claimed that there is no evidence that changes in top marginal tax rates have had any impact on economic growth in the United States since World War II. In the weeks leading up to the election, the CRS study was spun as evidence that President Obama's proposal to raise the top marginal tax rate to 39.6 percent could "spread the wealth around" without forgoing economic growth.[5]

Republican presidential candidate Mitt Romney's economic platform centered on cutting marginal tax rates to spur growth in order to help solve the nation's short- and long-run debt and demographic problems. The mainstream media, politicians, and political groups favoring higher taxes on the wealthy widely cited the CRS study as evidence against Romney's economic program and in favor of President Obama's plan to raise top marginal rates. Republicans claimed that the study was driven by ideology rather than economics and asked that it be pulled from the *Congressional Record*, which it was five days before the November 6 election. Critics accused Republicans of suppressing the study because they did not agree with its findings.

We find that the CRS study does have a serious methodological flaw—it examines differenced data so that the coefficients on the tax variables are zero except during the year in which the top marginal tax rate is changed. By employing this methodology, the CRS study does not allow tax changes to have lagged effects on growth. Economic theory, however, suggests that a change in marginal tax rates can impact the economy in the time frame beyond just the calendar year in which it goes into effect. We use the CRS study's data and find that if dummy variables are used for the three to five years around a tax change, rather than using the one-year growth rate in the top marginal rate, there is strong empirical evidence that real per capita GDP grows faster in the years after a tax cut. This finding is robust to several additional modifications in the empirical approach, including one that addresses another major criticism by adding more control variables that help explain GDP growth.

[5]While campaigning in September 2008, Obama told Joe Wurzelbacher, a small business owner who had become known as "Joe the Plumber," that "when you spread the wealth around, it's good for everybody" (Hardwood 2011).

Our results are consistent with what economists have long understood: that a tradeoff exists between income redistribution and economic growth.

References

Barro, R,. and Redlick C. J. (2011) "Macroeconomic Effects of Government Purchases and Taxes." *Quarterly Journal of Economics* 126: 51–102.

Black, W. K. (2013) "Robert Reich Is a Good Man but He Falls into the Austerity Trap." Alternet (www.alternet.org/economy/robert-reich-good-man-he-falls-austerity-trap).

Blum, W. J., and Kalven Jr., H. (1952) *The Uneasy Case for Progressive Taxation*. Chicago: University of Chicago Press.

Brownlee, E. (2004) *Federal Taxation in America: A Short History.* 2nd ed. Cambridge: Cambridge University Press.

Conesa, J. C., and Krueger, D. (2006) "On the Optimal Progressivity of the Income Tax Code." *Journal of Monetary Economics* 53: 1425–50.

Cullen, J. B., and Gordon R. H. (2002) "Taxes and Entrepreneurial Activity: Theory and Evidence for the U.S." NBER Working Paper No. 9015.

Diamond, P. A., and Mirrlees, J. A. (1971) "Optimal Taxation and Public Production I: Production Efficiency." *American Economic Review* 61: 8–27.

Easterly, W., and Rebelo, S. (1993) "Marginal Income Tax Rates and Economic Growth in Developing Countries." *Journal of Monetary Economics* 50: 409–17.

Engen, E. M., and Skinner, J. (1992) "Fiscal Policy and Economic Growth." NBER Working Paper No 4223.

Entin, S. J. (2013) "CRS Study on Tax Rates and Growth Still Flunks the Test." Washington: Tax Foundation. Available at http://taxfoundation.org/sites/taxfoundation.org/files/docs/ff350.pdf.

Frenkel, J. A.; Razin, A.; and Sadka, E. (1991) *International Taxation in an Integrated World.* Cambridge, Mass.: MIT Press.

Genetski, R. J., and Chin, Y. D. (1978) *The Impact of State and Local Taxes on Economic Growth.* Chicago: Harris Bank.

Hardwood, J. (2011) "'Spreading the Wealth' in Democrats' Favor." *New York Times* (27 November).

Hayek, F. A. (1960) *The Constitution of Liberty*. Chicago: University of Chicago Press.

Hungerford, T. (2012) "Taxes and the Economy: An Economic Analysis of the Top Tax Rates Since 1945." Washington: Congressional Research Service (14 September).

Koester, R. B., and Kormendi, R. C. (1989) "Taxation, Aggregate Activity and Economic Growth: Cross-Country Evidence on Some Supply-Side Hypotheses." *Economic Inquiry* 27: 367–86.

Laffer, A. B.; Moore, S.; and Williams. J. (2012) *Rich States, Poor States*. 5th ed. Washington: American Legislative Exchange Council.

Lee, Y., and Gordon, R. H. (2005) "Tax Structure and Economic Growth." *Journal of Public Economics* 89: 1027–1043.

Leonhart, D. (2012) "Do Tax Cuts Lead to Economic Growth?" *New York Times* (16 September).

Mankiw, N. G.; Weinzierl, M.; and Yagan, D (2009) "Optimal Taxation in Theory and Practice." *Journal of Economic Perspectives* 23: 147–74.

McBride, W. (2012) "What Is the Evidence on Taxes and Growth?" Tax Foundation Special Report No. 207. Available at http://taxfoundation.org/sites/taxfoundation.org/files/docs/sr207.pdf.

Mertens, K., and Ravn, M. (2013) "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review* 103: 1212–47.

Mirrlees, J. A. (1971) "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic Studies* 38: 175–208.

Nijkamp, P., and Poot, J. (2004) "Meta-Analysis of the Effect of Fiscal Policies on Long-Run Growth." *European Journal of Political Economy* 20: 91–124.

Padovano, F., and Galli, E. (2001) "Tax Rates and Economic Growth in the OECD Countries (1950–1990)." *Economic Inquiry* 39: 44–57.

Romer, C., and Romer, D. (2010) "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review* 100: 763–801.

Saez, E. (2001) "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies* 68: 205–29.

Shlaes, A. (2010) "Obama Threatens to Follow in FDR's Economic Missteps." *Washington Post* (8 July).

Thompson, D. (2012) "Tax Cuts Don't Lead to Economic Growth, a New 65-Year Study Finds." *The Atlantic* (September).

Vedder, R. K. (2001) *Taxes and Economic Growth: Should States Adopt a Safehaven Strategy of Low Tax Burdens to Foster Growth, Capital Formation and Innovation?* Cedarburg, Wis.: Taxpayers Network.

Weisman, J. (2012) "Nonpartisan Tax Report Withdrawn after G.O.P. Protest." *New York Times* (2 November).

State Fiscal Policies for Budget Stabilization and Economic Growth: A Dynamic Scoring Analysis

John Merrifield and Barry W. Poulson

Economic downturns expose unsustainable fiscal practices. Widespread fiscal crises create opportunities to compare policy options that address especially adverse circumstances, especially pro-growth fiscal constraints that can stabilize state budgets over the business cycle. Our policy option assessments depart from the normal practice of assessing rules and policies independently. Our premise is that the fiscal policy mix determines its outcomes. We include dynamic scoring to provide a richer view of the policy interactions.

In this article, we assess reforms that address fiscal stress issues. We were driven, in part, by our conviction that stable spending growth over the business cycle curbs fiscal stress-induced pressures to raise taxes and weaken caps on spending growth. To generalize our findings as much as possible, we apply our dynamic scoring model to California, Montana, and Utah—states familiar to us that span the blue state–red state gamut, with Montana in the middle. Utah is "famously conservative" (Woo 2010), with one of the top business tax climates (Tax Foundation 2011). California's response to fiscal stress included large tax hikes, which helped create one of the worst business climates. With fiscal data and dynamic scoring, we simulate the economic growth and fiscal effects of income tax rate reductions and

---

Cato Journal, Vol. 34, No. 1 (Winter 2014). Copyright © Cato Institute. All rights reserved.

John Merrifield is Professor of Economics at the University of Texas-San Antonio (Downtown Campus). Barry Poulson is Professor of Economics Emeritus at the University of Colorado-Boulder.

fiscal rules designed to constrain the growth in state spending and stabilize the budget over the business cycle.

The Fiscal Rollercoaster

In the five years that preceded the still lingering 2007–09 Great Recession, spending growth topped personal income growth in 37 states, including those with fiscal caps more extensive than the balanced budget rule, absent only in Vermont (Poterba 1994, Merrifield 2000). The expansion in those 37 states was enough to achieve a 50-state average spending growth rate 5 percent faster than personal income growth. Large budget deficits and fiscal crises arose when the Great Recession sharply cut revenues (Chapman 2009, Eaton 2009, Kalita 2009, Vock et al. 2009). Legislators could not sustain the good times' rapid spending growth achieved, in part, by overriding their statutory tax and expenditure limits (Stansel and Mitchell 2008).

A key reason for rapid state spending growth has been widespread use of personal income growth to define fiscal discipline (Shadbegian 1996, Waisanen 2010). An income growth-based cap is a convenient, politically comfortable limit when economic growth is normal, but uncomfortable when personal income growth is modest, and a crisis when growth is negative, as it has been recently (Schunk and Woodward 2005, Wagner and Elder 2005). Fiscal instability and uncertainty seem to accelerate spending growth (Holcombe and Sobel 1997). Fiscal crises can be primary agents of tax hikes that typically survive into future high-growth periods; a process that ratchets spending upward over successive iterations of the business cycle. Fiscal stress also spills over into off-budget spending (Bennett and DiLorenzo 1982, Merrifield 1994), and into on-budget funding substitutes such as regulation and more responsibility for local governments.

Tax and Expenditure Limits

Early studies of tax and expenditure limits (TELs) found evidence that they had only a small effect on state budgets.[1] But more recent studies provide evidence that TELs can effectively constrain the

[1]See Abrams and Dougan (1986), Bails (1990), Poterba (1994), and Poulson and Kaplan (1994).

growth in state spending[2] TELs that link spending growth to personal income are often nonbinding, and for binding TELs, the instability of personal income growth erodes support for TELs by creating periods of costly fiscal instability and uncertainty.[3] Economic conditions and the phase of the business cycle are key determinants of the effectiveness of TELs. For example, they seemed to be more binding in low-income states. Florida introduced a tax and expenditure limit in the recession phase of the business cycle that was never binding. The cap rose more rapidly than actual growth in state revenue.

In 1971, Ronald Reagan, then governor of California, along with Milton Friedman, campaigned for the first state TEL. Though voters narrowly rejected the attempt to cap California's state spending at 7 percent of state income, the Reagan-Friedman effort set the stage for the TELs later introduced in California and 31 other states (New 2003, Poulson 2004). The first state to enact a TEL was New Jersey in 1976. The New Jersey TEL limited state expenditure growth to growth in state income. Like other statutory TELs, the New Jersey TEL did not notably constrain state spending growth, and expired after six years in 1982 (Bails and Tieslau 2000, Poulson 2004).

State constitution TELs have been the most effective spending rules (New 2003, Poulson 2004). California's Gann Amendment was a 1978 ballot partner of the more famous Prop 13 property tax limit. The Gann TEL yielded a large 1987 tax rebate, but a series of constitutional amendments gutted Gann, as proved by rapid spending growth thereafter (Vock et al. 2009, Poulson 2009a), leading eventually to large budget gaps, several budget gap crises, significant tax increases, and finally to some recent large spending reductions.

As the Gann Amendment started its slide into irrelevance, Colorado amended its constitution with a Taxpayer Bill of Rights (TABOR), which limits the growth of available revenue to population growth plus inflation. Surplus revenue generates tax rebates. Higher tax rates and new debt require voter approval. TABOR also prohibits imposition of unfunded mandates on local governments. In the

[2]See Elder (1992), Kousser et al. (2008), Merrifield (2000), Merrifield and Monson (2011), Mitchell (2011), Mitchell and Tusyznski (2011), New (2001, 2003), Poulson (2004), Stansel (1994), and Stansel and Mitchell (2008).

[3]See Crain (2003), Holcombe and Sobel (1997), Krol (2007), Mitchell (2010), Mullins and Wallin (2004), Schunk and Woodward (2005), Shadbagian (1996), Wagner and Elder (2005), and Waisenan (2010).

1990s, revenue growth topped the TABOR limit enough to yield $3.25 billion in tax rebates. When the 2001 national recession caused actual revenue to fall below the TABOR limit, the resulting new benchmark and its ratchet-down effect on future spending growth yielded a 2005 referendum that imposed a five-year time-out from the TABOR growth limit and adjusted the annual limit formula to avoid future ratchet-down effects.[4] Though the TABOR experience suggests that political support for spending caps erodes with fiscal stress, Colorado voters remained unwilling to create a budget-stabilizing, but tax rebate-reducing, rainy day fund (RDF), a key source of TABOR critique (Poulson 2009a).

Budget Stabilization and Emergency Funds

Some of the recently proposed TELs earmark surplus revenue for an RDF and an emergency fund (EF). Forty seven states have some kind of RDF or EF, but rules governing deposits and withdrawals vary widely.[5] Wagner and Elder (2005) found that states with strict rules for RDF deposits and withdrawals experience a 20 percent reduction in spending volatility, as measured by the cyclical variability of per capita spending over time. Stansel and Mitchell (2008) found that states with stricter RDF withdrawal rules experienced less fiscal stress during the 2001 recession.

Capital Investment Funds

Capital expenditures tend to be very volatile over the business cycle. In periods of recession and revenue shortfall, state capital spending is typically among the first items cut. A binding TEL lacking a well-designed RDF may actually raise capital expenditure volatility over the business cycle, which occurred in Colorado (Poulson 2004). A solution to the volatility problem is a well-designed business stabilization fund (BSF) and a capital investment fund

[4]With a TEL-permitted growth rate of 5 percent per year and spending at $1 billion in year t, spending can rise to $1.05 billion in year t+1, and $1.1025 billion in year t+2. But, if in t + 1, there is a severe recession, revenues plummet to $800 million. The $800 million is the basis for the next 5 percent spending increase, leaving spending at $840 million in t+2 rather than at $1.1025 billion if an adequate RDF balance averts the cuts of year t+1.

[5]See Holcombe and Sobel (1997), Reuben and Rosenberg (2009), Wagner and Elder (2005, 2007), and Wagner and Sobel (2006).

(CIF) designed to stabilize capital spending over the business cycle. A good case for countercyclical capital expenditure exists without reliance on the well-known Keynesian stimulus argument. Because of the cyclical nature of construction quality and price (Finkel 1997, Merrifield and Monson 2011), our simulations earmark some surplus revenue for a CIF to finance additional construction spending in slow growth periods.

TELs and Tax Policy

A binding TEL will yield a mixture of tax rebates and lower tax rates. Despite the tedious nature of tax rebates, controversy over the basis for estimating the appropriate rebate for each taxpayer, and evidence that permanent tax cuts have larger economic growth effects than one-time rebates (Padquit 2011, Poulson and Kaplan 2008, Taylor 2008), it will probably take some persistence in the payment of rebates to elicit the permanent cuts. Indeed, Colorado's TABOR yielded large tax rebates for several years in the late 1990s, before state legislators responded with several permanent tax cuts.

Dynamic Scoring Foundations

The evidence that Colorado's TABOR accelerated economic growth (Poulson 2009a, 2009b) is controversial.[6] The controversy is over whether a drop in the state's share of personal income accelerates economic growth.[7] For 1980–90, Peterson (1994) estimated a 22.1 percent private rate and a 7 percent public rate of return. The 15.1 percentage point gap is a proxy for the marginal cost of shifting resources from the private to the public sector. Some studies (most recently Bania and Stone 2008) suggest that shifting resources from the private to the public sector can increase economic growth. But their findings may not be that useful to our TEL simulation. Bania and Stone omitted the effect of higher taxes on growth, and their large, heterogeneous productive services and infrastructure spending category obscures the likely effects of how spending would

[6]See, for example, Amiel et al. (2012), Deller et al. (2012), Lav (2009), Lav and Williams (2010), Lyons and Johnson (2006), Mcguire and Rueben (2006), Stallman and Deller (2010), and Stallman (2011).

[7]See Bergh and Henrekson (2011), Dahlby (1998), Ladner and Schlomach (2007), McBride (2012), Peterson (1994), and Spencer and Yohe (1970).

vary with or without a binding TEL. Studies of K–12 spending changes indicate that the Bania and Stone finding of a small positive impact on economic growth from greater spending on productive services and infrastructure may be a net effect of conflicting factors. Every state's disappointing K–12 performance indicates that there is much room for improvement, which is theoretically possible with additional resources. Though states continuously identify promising K–12 projects, Tomjanovitch (2004) found a significant inverse relationship between education spending and economic growth. Other studies suggest at least a normal marginal opportunity cost of shifting expenditure from the private sector to fund K–12 expenditures (Hanushek 1997, 2006; Hoxby 2004). Thus, our study's use of dynamic scoring assigns a positive economic growth effect to reduced spending and taxation.

Dynamic Scoring of Fiscal Policies for Budget Stability and Economic Growth

In our TEL proposal, we simulate the outcomes of a TEL that caps general fund (GF) spending growth at population plus inflation.[8] Our simulation template includes four key features excluded from the Schunk and Woodward (2005) TEL simulation. First, consistent with key issues cited in some of the recent TEL debates, our simulation allocates funds to an emergency fund and an RDF. Second, it funds countercyclical spending increases, including acceleration of capital spending. Our study is the second to examine the impact of codifying the pursuit of construction bargains during slow growth periods.[9] Third, our study is the first to extensively simulate the growth effects of reducing state income tax rates in the context of fiscal policy reforms. Our simulation revises personal income and future tax revenues when dollars shift between the public and private sectors, tax rates drop, or idle fund balances accrue interest payments.

[8]We confined our TEL's scope to GF spending because measures of total expenditures vary considerably among the states and special funds typically spend earmarked revenue. We use GF spending mostly to maintain comparability in measuring the impact of TELs on expenditures in the different states.

[9]The first study was Merrifield and Monson (2011). See Finkel (1997) for a discussion of construction bargains during sluggish growth.

Without tax rate reductions, our population plus inflation TEL generates large surpluses and rebates in years with high economic growth rates. Large, frequent tax rebates signal an opportunity to cut income tax rates, which we do in our second set of simulations. This means that spending still rises in all years, typically at the population plus inflation cap. We assess the sensitivity of our simulation findings to the opportunity cost of shifting resources from private to public use, and to uncertainty about the effects of income tax cuts on economic growth.

A Dynamic Scoring Model with Tax Rebates

Our first dynamic scoring model (equations 1–5) only has tax rebates. The annual spending growth limit is the sum of inflation and population growth rates. In all of our simulations, surplus revenue above that limit is prioritized as follows: emergency fund, rainy day fund, capital investment fund, and tax rebates. So, surplus funds flow to the EF until the balance reaches its target level. To limit the scope of our effort, we do not explore the sensitivity of our findings to changes in the EF cap (2 percent of GF for Montana and Utah, and 5 percent for California), or the occurrence of a major emergency.

(1) If $SURP_t > EDEF_t$, then $EMERG_t = TARGEM$
If $SURP_t < EDEF_t$, then $EMERG_t = (1 + r_t)\,(EMERG_{t-1} - DISAST_t) + SURP_t$,

where $SURP_t$ = surplus (RREV − RSPEND) in year t;
$EMERG_t$ = emergency fund balance in year t;
$EDEF_t$ = emergency fund deficit (i.e., the gap between the balance and target);
$EDEF_t = TARGEM_t + DISAST_t - EMERG_{t-1}$;
$DISAST_t$ = disaster spending in time t;
and TARGEM = target balance of emergency fund.

Surplus revenue left over after mandated EF deposits flow to the RDF until it reaches the account balance target. Based on the Holcombe and Sobel (1997) recommendation that the RDF have enough money to cover three consecutive worst-case revenue declines, and because the Great Recession was more severe than the basis for RDF norms, our default RDF target is an unusually high 25 percent of the GF spending level. Later, we use a more conventional 10 percent limit in a sensitivity analysis test of the importance of that parameter.

We prioritize EF deposits ahead of RDF deposits, so if the EF is below its target level, the first line of equation 2 only assigns interest payments to the RDF, and subtracts debits. The remainder of equation 2 assesses the net change in the RDF based on the availability of surplus funds, debits for budget stabilization, and the RDF account level relative to the account cap.

$$
\begin{aligned}
(2)\ &\text{If } EMERG_t < TARGEM, \text{ then } RDF_t = RDF_{t-1}(1 + r_t) - RDFDEB_t \\
&\text{If } EMERG_t \geq TARGEM, \text{ then} \\
&\quad \text{If } RDF_{t-1} \geq TARGRDF_t, \text{ then } RDF_t = RDF_{t-1} - RDFDEB_t \\
&\quad \text{If } RDF_{t-1} < TARGRDF_t, \text{ then} \\
&\qquad \text{If } RDF_{t-1}(1 + r_t) + AVSURP_t - RDFDEB_t < TARGRDF_t, \\
&\qquad\quad \text{then } RDF_t = RDF_{t-1}(1 + r_t) + AVSURP_t - RDFDEB_t \\
&\qquad \text{If } RDF_{t-1} + AVSURP_t - RDFDEB_t > TARGRDF_t, \text{ then} \\
&\qquad\quad RDF_t = TARGRDF_t,
\end{aligned}
$$

where RDF_t = rainy day fund balance in time period t;
$RDFDEB_t$ = RDF debit in time t;
$AVSURP_t$ = available surplus after emergency fund deposits;
and $TARGRDF_t$ = target balance for RDF.

RDF debits occur when revenue growth cannot sustain spending growth at population growth plus the rate of inflation. When there is a decline in revenue, money moves from the RDF to the GF. In particular, the transfer will be equal to half the drop in revenue from the previous year so that spending can rise faster than population plus inflation to meet the increased demand for unemployment compensation, welfare, and health expenditures in severe economic downturns. Consistent with our assessment of the literature, a CIF finances extra construction when recent personal income growth is below average personal income growth over the prior 10 years, with an annual withdrawal cap of 67 percent of the CIF balance—a cap that approximately balances CIF withdrawals and deposits over the business cycle.

The money accumulated in the various funds does not remain idle, but rather is invested in the private sector. Thus, there is no opportunity cost incurred until the money is transferred to the GF, which shifts resources from the private to the public sector. The RDF and EF balances earn interest at the rates recorded by the State Treasurer's Office. The interest earned grows the account balance until the account target is reached, and then interest earnings accrue to the GF. The basis for revising GF spending in accordance with our

TEL (equation 3) looks more complicated than it is. The model has to test for several possibilities to make sure that spending growth is as close to the population plus inflation rate as revenue plus the RDF account balance will allow.

(3) If $RREV_t \geq RREV_{t-1}$, then
If $RSPEND_{t-1}(CAP_t) < RREV_t$, then $RSPEND_t = RSPEND_{t-1}(CAP_t)$
If $RSPEND_{t-1}(CAP_t) > RREV_t$, then
If $RDF_t > RSPEND_{t-1}(CAP_t) - RREV_t$, then
$RSPEND_t = RSPEND_{t-1}(CAP_t)$
If $RDF_t < RSPEND_{t-1}(CAP_t) - RREV_t$, then
$RSPEND_t = RREV_t + RDF_t$
If $RREV_t < RREV_{t-1}$, then
If $RSPEND_{t-1}(CAP_t) + b(RREV_{t-1} - RREV_t) < RREV_t$, then
$RSPEND_t = RSPEND_{t-1}(CAP_t) + b(RREV_{t-1} - RREV_t)$
If $RSPEND_{t-1}(CAP_t) + b(RREV_{t-1} - RREV_t) > RREV_t$, then
If $RDF_t > RSPEND_{t-1}(CAP_t) + b(RREV_{t-1} - RREV_t) - RREV_t$, then $RSPEND_t = RSPEND_{t-1}(CAP_t) + b(RREV_{t-1} - RREV_t)$
If $RDF_t < RSPEND_{t-1}(CAP_t) + b(RREV_{t-1} - RREV_t) - RREV_t$, then $RSPEND_t = RREV_t + RDF_t$,

where $RREV_t$ = revised GF revenue for fiscal year t;
CAP_t = cap on GF spending growth rate for fiscal year t;
$RSPEND_t$ = revised GF spending for fiscal year t;
and b = share of revenue decrease to convert into countercyclical spending increase, RDF account balance permitting.

Our simulations employ a conservative estimate of 6 percent for the opportunity cost rate (Dahlby 1998) for shifting resources from the private to the public sector. Consistent with Barro (1990), we assume that the opportunity cost rate applies to small changes typical of marginal transfers of resources from private to public use. When our TEL reduces resource transfers from the private to the public sector, personal income rises (equation 4):

(4) $RPI_t = RPI_{t-1} + (API_t - API_{t-1}) + (OCR \times (SURP_t - EMERG_t))$,

where RPI_t = revised personal income in fiscal year t;
API_t = actual personal income in fiscal year t;
$SURP_t$ = GF surplus (RSPEND – RREV) in t;
$EMERG_t$ = emergency fund spending in t;
and OCR = opportunity cost rate (0.06, 0.09).

Personal income growth generates additional tax revenue for the GF (equation 5):

$$(5)\quad RREV_t = AREV_t + \Sigma_T ([RPI_t - API_t] \times [MTR_{Tt}]) + SI_t,$$

where $RREV_t$ = revised revenue for fiscal year t;
$AREV_t$ = actual revenue for fiscal year t;
MTR_{Tt} = marginal tax rate for state tax T in fiscal year t from Reed et al. (2011);
and SI_t = RDF and EF surplus interest.

A Dynamic Scoring Model with Income Tax Rate Cuts

The second dynamic scoring model analyzed in this study incorporates tax cuts alongside rebates of still-remaining surplus funds. Tax cuts impact economic growth more than tax rebates because of different behavioral responses. Tax rebates are seen primarily as transitory private income rather than permanent income. Transitory income mostly pays down debt, with little impact on consumption or investment spending. When permanent tax cuts impact permanent income, people raise their consumption and planned investment spending to a greater extent, and increase productive activity. Permanent tax cuts in one state relative to those in another state will also create incentives for mobility of labor and capital into that state.

Poulson and Kaplan (2008) measured the relationship between marginal tax rates and state economic growth. The marginal tax rate (MTR) is the increment in taxes paid when personal income rises. MTRs vary with tax structure (Reed, Rogers, and Skidmore 2011). The nationwide MTR is the average of the marginal rates levied in each state. Poulson and Kaplan (2008) find that a drop in the MTR in state X relative to others is associated with higher economic growth in state X. Their regression analysis indicates that a 1 percentage point decline in a state's aggregate MTR relative to the nation's average MTR increases that state's growth rate between 0.251 and 0.374 percentage points.

For our dynamic scoring simulations, we modify equation 1 to include a growth adjustment (GA) factor that is greater than one for years in which a permanent income tax rate cut is made (equation 6). GA_t for our California, Montana, and Utah models came from multiplying the conservative 0.251 percentage point impact of a 1 percentage point MTR change (Poulson and Kaplan 2008) by each state's annual state MTR (Reed et al. 2011) percentage point change.

(6) $RPI_t = RPI_{t-1}(GA_t) + (API_t - API_{t-1}) + (OCR \times (SURP_t - EMERG_t))$,

where GA_t = growth adjustment factor for fiscal year t, with $GA_t = 1 + (RMTR \times MTRCH_t)$;
RMTR = marginal tax rate reduction effect from Poulson and Kaplan (2008); and
$MTRCH_t$ = change in the Reed et al. (2011) state-level MTR_t.

A revised RREV (equation 7) accounts for the static revenue losses (SRLS) caused by a 1 percentage point cut in the income tax MTR.

(7) $RREV_t = AREV_t(1 - SRLS_t) + \Sigma_T([RPI_t - API_t] \times MTR_{Tt}) + SI_t$,

where $SRLS_t = (MTRCH_t/MTRINC_t) \times INCTAXPCT_t$;
$MTRINC_t$ = income tax MTR;
INCTAXPCT = the income tax share of total state GF revenue;
MTR_{Tt} = marginal tax rate for state tax T, like income tax and sales tax, which varies with the Reed et al. (2011) estimates, and with our prior simulated MTR reductions.

Our revenue estimates reflect the static revenue loss from tax rate reduction, and the increased revenue that results when lower tax rates increase the rate of growth in personal income.

Empirical Results

The states chosen for our simulation analysis have quite different tax structures and tax policies. Table 1 provides a three-state overview, where ~ indicates that the number is a generalization of simulation outcomes arising from different parameters.

All three states see a nearly 2 percentage point drop in the GF's share of personal income. The potential for tax rate reduction and commensurate gains in personal income vary widely between Utah (with its low potential to reduce tax rates and increase economic growth while still avoiding Great Recession budget cuts) and California and Montana (with their high potential to lower tax rates and avoid GF spending cuts). The California and Montana MTR reductions of about 2.4 percentage points amount to an approximately 50 percent drop in income taxation.

Seemingly small boosts to the rate of economic growth (0.251 to 0.374 percentage points) per percentage point of MTR reduction compound into significant changes over the 1994–2012 period, and in FY 2012. The TEL and dynamic scoring of tax rate reductions

TABLE 1
Three State Overview

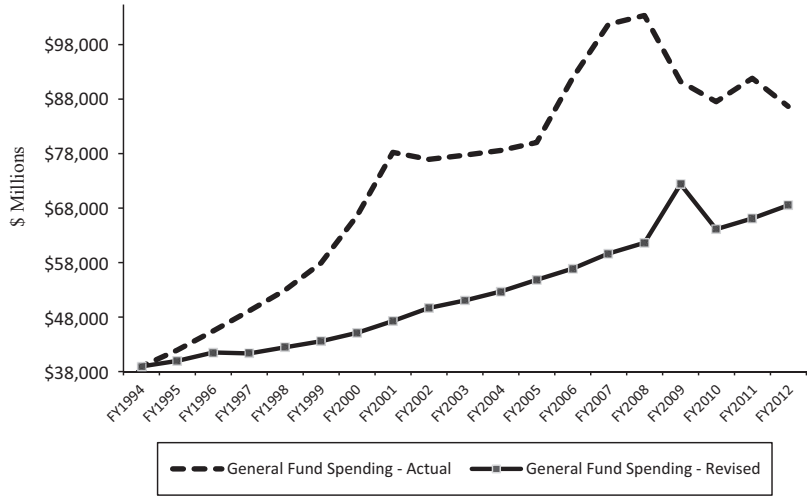
	California FY2012	CA 1994–2012 Cumul Δ	Montana FY2012	MT 1994–2012 Cumul Δ	Utah FY2012	UT 1994–2012 Cumul Δ
Marginal Tax Rate Drop (% points)	~0.4	~2.4	0	~2.3	0	0.4
Extra Personal Income (% points)	~8	~4	~6.5	~2.4	~0.33	~0.26
Drop in GF Share of PI (% points)		~1.67		~2		~1.67

yielded an approximate 8 percent rise in California's 2012 personal income, and 6.5 percent in Montana. The 1994–2012 cumulative gains were smaller because gains were smaller in the early years.

California

Our California simulation estimates what the state budget and personal income would look like had California added an RDF with an account balance cap at 10 or 25 percent and not abandoned the 1979 Gann spending limit in 1990. Our Gann-plus-RDF scenarios avoid tax hikes and budget crises while yielding large tax cuts and much larger state personal income. The 2009 spike in the Figure 1 revised/simulated spending reflects a temporary, counter-cyclical spending increase. So, the post-2009 drop does not violate our "no-cuts" constraint. Permanent spending increases every year.

FIGURE 1
California Actual and
Simulated Government Expenditures



Tables 2 and 3 describe simulated outcomes for FY 2012 and 1994–2012, respectively, with the growth adjustment rate (RMTR) at 0.251 percent.[10] In Tables 2 and 3 (4 and 5 for Montana and 6 and 7

[10]The higher estimates that result from the high-end RMTR of 0.374 percent, and the year-by-year California findings, are available at http://faculty.business.utsa.edu/jmerrifi/dsa.pdf.

TABLE 2
California, RTMR = 0.00251

FY2012 Counterfactual FY2012, Millions of 2012$	RDF Cap at 25% of GF 2.5 Inc Tax Rate Cuts	RDF Cap at 10% of GF 2.5 Inc Tax Rate Cuts	RDF Cap at 10% of GF 2.15 Inc Tax Rate Cut	RDF Cap at 25% of GF Rebates Only No Income Tax Cuts
Revised Personal Income, % of Actual	106.67%	106.69%	106.60%	101.25%
Additional FY 2012 Personal Income	$111,463	$111,675	$110,172	$20,796
Total Taxpayer Rebate	$0	$0	$0	$8,971
Revised Spending, % of Actual	76.43%	70.92%	77.02%	87.20%
Capital Investment Fund-End Balance	$554	$555	$555	$3,553
Budget Stabilization Funds,	$18	$9	$9	$18,826
% of Revised General Fund Spending	0.03%	0.02%	0.01%	24.99%
Emergency Funds	$3,410	$3,404	$3,410	$3,765
% of WC Emerg	99.16%	106.65%	98.40%	100.18%
Tax Cut-Induced Shortfall-Ratchet Down	8.51%	15.10%	7.80%	−4.39%

Personal Income Growth Rate Difference, Simulated Minus Actual, 1994–2012	0.364%	0.364%	0.360%	0.070%
Simulated Minus Actual, 1994–2008	0.336%	0.336%	0.336%	0.069%
Taxpayer Rebate per Capita 2012$ (not millions)	$0	$0	$0	$235
State Government Size				
Revised 2012 GF Spending, Share of PI	3.71%	3.44%	3.74%	4.46%
Actual in 1994	5.38%	5.38%	5.38%	5.38%
Actual in 2012	5.17%	5.17%	5.17%	5.17%
Actual in 2010	5.71%	5.71%	5.71%	5.71%
Actual in 2008	6.44%	6.44%	6.44%	6.44%

TABLE 3
California, RMTR = 0.00251

Cumulative Effects, 1994–2012 Millions of 2012$	RDF Cap at 25% of GF 2.5 Inc Tax Rate Cuts	RDF Cap at 10% of GF 2.5 Inc Tax Rate Cuts	RDF Cap at 10% of GF 2.15 Inc Tax Rate Cut	RDF Cap at 25% of GF Rebates Only No Income Tax Cuts
Cumulative Extra Personal Income, 1994–2012	$915,455	$915,867	$914,363	$170,424
% Increase over Actual 1994–2012 PI	3.27%	3.27%	3.27%	0.61%
Cumulative 1994–2012 GF Spending Decline	$470,174.72	$480,580.50	$475,311.91	$409,398.31
% of Total Actual, 1994–2012 Spending	28.00%	28.62%	28.30%	24.38%
Avg GF Spending Cap Shortfall	2.92%	5.23%	2.67%	0.00%
Avg GF Spending Cap Shortfall-SG	4.93%	5.34%	4.89%	0.00%
# of yrs Below GF Spending Cap	3	3	3	0

# of yrs Below GF Spending Cap-SG	16	16	16	0
Cumulative 1994–2012 Taxpayer Rebates	$59,037	$67,133	$67,133	$298,655
% of Total Actual, 1994–2012 Revenue	3.53%	4.01%	4.01%	17.84%
Cumulative 1994–2012 Tax Revenue Gain from Increased Growth	$48,918	$46,980	$47,304	$21,431
% of Total Revised, 1994–2012 Revenue	3.78%	3.64%	3.65%	1.26%

for Utah), the fourth (last) column describes the rebate-only outcomes with no income tax rate reductions. The heading of the first column in each table describes how many 1 percentage point cuts in the income tax MTR (Reed et al. 2011) are possible with the RDF cap set at 25 percent of projected GF spending without spending cuts in either the 2001–02 recession or the Great Recession, and growth in GF spending equal to population plus inflation in all other years. The second column shows how the outcomes of the simulation differ if the RDF cap is set at 10 percent. The third column indicates how many 1 percentage point cuts in the income tax MTR are possible with the RDF cap at 10 percent.

With the growth effects of the nearly 1 percentage point cut in the income tax MTR in 1996, a 0.6 percentage point cut in 1997, a 0.25 percentage point cut in 1998 and again in 1999, and another half percentage point drop in 2012 (2.5 total), with an RDF limit of 25 percent, without any spending cuts, California's simulated 2012 personal income is 6.67 percent ($111 billion) above the actual 2012 level—far above the no cut–rebate only 2012 gain of 1.25 percent, and something little changed by larger opportunity cost rates for shifting funds from the private to public sector. The 1994–2012 cumulative effect of the 2.5 percentage point MTR cut is $915 billion; 3.27 percent more than the actual personal income for the 18 years. With RMTR at the Poulson-Kaplan upper limit of 0.374 percent, the simulated 2012 personal income is $181 billion above actual personal income, with a 1994–2012 cumulative effect of $1.47 trillion. That growth yielded $72 billion in additional tax revenue over the 18 years. In 2012, in our most conservative scenario (RMTR = 0.00251, RDF at 25 percent), the California state government spends 1.67 percentage points less of personal income than in 1994.

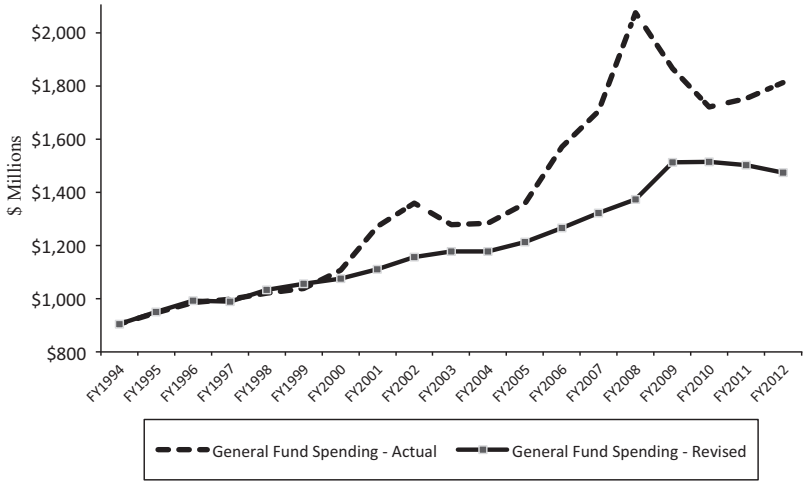
With RMTR = 0.00251, an RDF cap at 10 percent of GF spending yields some Great Recession spending cuts that the RDF cap of 25 percent would have avoided. To avoid those cuts, cumulative MTR reduction with an RDF cap at 10 percent must stop at 2.15 percentage points. Thus, in California, there would have been a trade-off between a 0.35 (= 2.5 − 2.15) point MTR reduction and the effects of much larger RDF account balances. Though actual cuts do not occur, spending rises by less than the cap in three fiscal years for a cumulative ratchet-down effect (less than the steady growth limit) of about 5 percent, varying slightly with the RDF cap and how much taxes are cut.

Montana

Montana is unusual in several ways. Over time, statewide elections yield a Republican-Democrat mix. Montana has no sales tax. It depends on volatile mineral severance and income tax revenue, typically spending up to those limits so that low-growth years, like the Great Recession, yield significant fiscal stress. Because the severance tax is state-income inelastic, the state income tax's MTR share is well above the income tax's share of state revenues. This fact underlies our simulation result that Montana could have managed a half percentage point cut in the income tax MTR in 2000, 2001, and 2003, plus a 0.25 percentage point reduction in 2002, 2004, and 2006 (2.25 points, total), with RMTR = 0.251 percent and the RDF account balance cap at 25 percent of projected GF spending. Those results would have been in addition to the actual 2005 income tax cut reflected in the data.

As in California, the population plus inflation spending cap yields large tax cuts and more income (Figure 2). The revised-simulated spending reflects temporary, Great Recession countercyclical spending increases, no violation of our "no-cuts" constraint.

FIGURE 2
Montana Actual and Simulated Government Expenditures



Tables 4 and 5, with RMTR = 0.251 percent, contain the 2012 simulated outcomes and cumulative effects for 1994–2012, respectively.[11] With the growth effects of the MTR reductions, with RMTR = 0.0025 and an RDF limit of 25 percent, Montana's simulated 2012 personal income is 4.96 percent ($1.8 billion) above the actual 2012 level, far above the no cut–rebate only 2012 gain of 0.47 percent. Again, the difference between cuts and rebates-only is little changed by larger opportunity cost rates for shifting funds from the private to public sector. The 1994–2012 cumulative effect of the 2.25 percentage point tax cuts and an RMTR of 0.00251 is $10.7 billion, 1.86 percent more than actual personal income. With RMTR at the Poulson-Kaplan upper limit of 0.374 percent, simulated 2012 personal income is $2.9 billion above actual personal income. The 1994–2012 cumulative effect is $16.6 billion. With the RMTR = 0.251 percent, the extra growth that would have resulted from the MTR reductions would have yielded $531 million in extra tax revenue over those 18 years. Because of a slight spending cut in 1997 (before the first simulated tax cut), and a Great Recession cap shortfall (not a spending decline), there is a 6.87 percent ratchet-down effect by 2012. In 2012, Montana's state government spent nearly 2 percentage points less of personal income than in 1994, nearly a full percentage point below the actual 2012 GF share of personal income. With RMTR = 0.00251 and the RDF cap at 10 percent, the cumulative MTR reduction has to be scaled back by 0.1 percentage points to avoid violating our no cuts constraint.

Montana's substantial severance tax revenues created a high-gain, low-pain combination for income tax cuts, which combined with spending restraint (at population plus inflation), allowed for a decline of nearly 50 percent in the income tax MTR, while maintaining budget stability and avoiding offsetting tax increases. The income tax cuts yielded significant additional economic growth.

Utah

For comparison to California and Montana, our definition of Utah's GF is the actual GF, funded mostly by the sales tax, plus the separate Education Fund, financed almost entirely by income taxes. Seemingly true to its conservative reputation, Utah had much less room for

[11]The higher estimates that result from the high-end RMTR of 0.374 percent, and the year-by-year Montana findings, are available at http://faculty.business.utsa.edu/jmerrifi/dsa.pdf.

TABLE 4
MONTANA, RMTR = 0.00251

FY2012 Counterfactual FY2012, Millions of 2012$	RDF Cap at 25% of GF 2.5 Inc Tax Rate Cuts	RDF Cap at 10% of GF 2.5 Inc Tax Rate Cuts	RDF Cap at 10% of GF 2.15 Inc Tax Rate Cut	RDF Cap at 25% of GF Rebates Only No Income Tax Cuts
Revised Personal Income, % of Actual	106.67%	106.69%	106.60%	101.25%
Additional FY2012 Personal Income	$111,463	$111,675	$110,172	$20,796
Total Taxpayer Rebate	$0	$0	$0	$8,971
Revised Spending, % of Actual	76.43%	70.92%	77.02%	87.20%
Capital Investment Fund-End Balance	$554	$555	$555	$3,553
Budget Stabilization Funds,	$18	$9	$9	$18,826
% of Revised General Fund Spending	0.03%	0.02%	0.01%	24.99%
Emergency Funds	$3,410	$3,404	$3,410	$3,765
% of WC Emerg	99.16%	106.65%	98.40%	100.18%
Tax Cut–Induced Shortfall-Ratchet Down	8.51%	15.10%	7.80%	−4.39%

continued

TABLE 4 (*cont.*)
MONTANA, RMTR = 0.00251

FY2012 Counterfactual FY2012, Millions of 2012$	RDF Cap at 25% of GF 2.5 Inc Tax Rate Cuts	RDF Cap at 10% of GF 2.5 Inc Tax Rate Cuts	RDF Cap at 10% of GF 2.15 Inc Tax Rate Cut	RDF Cap at 25% of GF Rebates Only No Income Tax Cuts
Personal Income Growth Rate Difference, Simulated Minus Actual, 1994–2012	0.364%	0.364%	0.360%	0.070%
Simulated Minus Actual, 1994–2008	0.336%	0.336%	0.336%	0.069%
Taxpayer Rebate per Capita 2012$ (not millions)	$0	$0	$0	$235
State Government Size				
Revised 2012 GF Spending, Share of PI	3.71%	3.44%	3.74%	4.46%
Actual in 1994	5.38%	5.38%	5.38%	5.38%
Actual in 2012	5.17%	5.17%	5.17%	5.17%
Actual in 2010	5.71%	5.71%	5.71%	5.71%
Actual in 2008	6.44%	6.44%	6.44%	6.44%

TABLE 5
MONTANA, RMTR = 0.00251

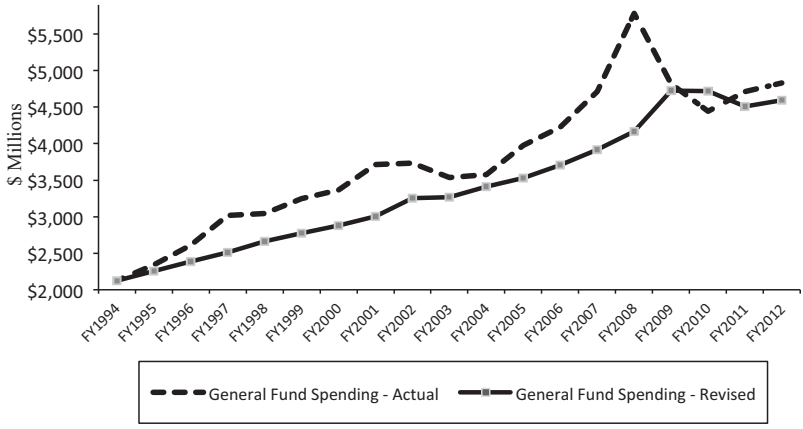
Cumulative Effects, 1994–2012 Millions of 2012$	RDF Cap at 25% of GF 2.25 Inc Tax Rate Cuts	RDF Cap at 10% of GF 2.25 Inc Tax Rate Cuts	RDF Cap at 10% of GF 2.15 Inc Tax Rate Cut	RDF Cap at 25% of GF Rebates Only No Income Tax Cuts
Cumulative Extra Personal Income, 1994–2012	$10,680	$10,697	$10,481	$984
% Increase over Actual 1994-2012 PI	1.86%	1.86%	1.82%	0.17%
Cumulative 1994–2012 GF Spending Decline	$3,484.66	$3,692.54	$3,602.22	$3,003.31
% of Total Actual, 1994–2012 Spending	11.12%	11.78%	11.49%	9.58%
Avg GF Spending Cap Shortfall	1.50%	2.13%	1.57%	4.26%
Avg GF Spending Cap Shortfall-SG	6.87%	7.65%	7.20%	4.67%
No. of yrs below GF Spending Cap	2	3	3	1
No. of yrs below GF Spending Cap-SG	12	12	12	16

continued

TABLE 5 (*cont.*)
Montana, RMTR = 0.00251

Cumulative Effects, 1994–2012 Millions of 2012$	RDF Cap at 25% of GF 2.25 Inc Tax Rate Cuts	RDF Cap at 10% of GF 2.25 Inc Tax Rate Cuts	RDF Cap at 10% of GF 2.15 Inc Tax Rate Cut	RDF Cap at 25% of GF Rebates Only No Income Tax Cuts
Cumulative 1994–2012 Taxpayer Rebates	$139	$324	$368	$2,436
% of Total Actual, 1994–2012 Revenue	0.44%	1.02%	1.16%	7.70%
Cumulative 1994–2012 Tax Revenue Gain from Increased Growth	$531	$542	$540	$113
% of Total Revised, 1994–2012 Revenue	1.89%	1.93%	1.92%	0.40%

FIGURE 3
Utah Actual and Simulated Government Expenditures



income tax reduction than California or Montana. Montana's maximum possible total income tax MTR reduction was 2.45 percentage points and California's was 2.5 percentage points. Utah could not maintain GF expenditure growth at population plus inflation because of a more than 0.4 percentage point drop in its income tax MTR in 1994–95, a revenue-neutral 2005 flattening of rates, and some minor cuts just before the Great Recession. Consequently, Utah suffered a relatively larger revenue drop in the Great Recession (Figure 3). Through 2008, Utah spent more of its personal income than California or Montana, and the Great Recession budget cuts put the personal income share of its GF at 5.01 percent, only slightly above where it would have been (4.75 percent) had Utah established an RDF with a 25 percent cap and steadily grown GF spending at population plus inflation—and well below the 6.53 percent share of personal income reached in 2008. With the growth that results from the 0.4 percentage points of reduction in Utah's income tax MTR, GF spending declines to 4.75 percent (varying slightly with RMTR and RDF cap) of personal income, more than half a percentage point higher than simulated 2012 levels for California and Montana. The results for Utah are presented in Tables 6 and 7 (with RMTR = 0.251 percent).

With the very small MTR reduction, the rebate-only outcome is slightly better. Note that the rebate-only simulation yields the actual 2012 GF personal income share (5.01 percent). Thus, Utah's actual GF

TABLE 6
Utah, RMTR = 0.00251

FY2012 Counterfactual FY2012, Millions of 2012$	RDF Cap at 25% of GF 0.4 Inc Tax Rate Cuts	RDF Cap at 10% of GF 0.4 Inc Tax Rate Cuts	RDF Cap at 10% of GF 0.4 Inc Tax Rate Cut	RDF Cap at 25% of GF Rebates Only No Income Tax Cuts
Revised Personal Income, % of Actual	100.31%	100.32%	100.32%	100.56%
Additional FY2012 Personal Income	$294.61	$312.92	$312.92	$544.66
Total Taxpayer Rebate	$0	$0	$0	$0
Revised Spending, % of Actual	95.15%	94.51%	94.51%	100.58%
Capital Investment Fund-End Balance	$90.32	$94.36	$94.36	$100.42
Budget Stabilization Funds,	$0.00	$0.02	$0.02	$945.89
% of Revised General Fund Spending	0.00%	0.00%	0.00%	19.46%
Emergency Funds	$90.32	$94.36	$94.36	$100.42
% of WC Emerg	0.95	1.00	1.00	1.00
Tax Cut-Induced Shortfall-Ratchet Down	6.35%	6.98%	6.98%	0.00%

Personal Income Growth Rate Difference, Simulated Minus Actual, 1994–2012	0.02%	0.02%	0.02%	0.03%
Simulated Minus Actual, 1994–2008	0.03%	0.03%	0.03%	0.04%
Taxpayer Rebate per Capita 2012$ (not millions)	$0	$0	$0	$0
State Government Size				
Revised 2012 GF Spending, Share of PI	4.75%	4.72%	4.72%	5.01%
Actual in 1994	6.37%	6.37%	6.37%	6.37%
Actual in 2012	5.01%	5.01%	5.01%	5.01%
Actual in 2010	5.08%	5.08%	5.08%	5.08%
Actual in 2008	6.53%	6.53%	6.53%	6.53%

TABLE 7
Utah, RMTR = 0.00251

Cumulative Effects Millions of 2012$	RDF Cap at 25% of GF 0.4 Inc Tax Rate Cuts	RDF Cap at 10% of GF 0.4 Inc Tax Rate Cuts	RDF Cap at 10% of GF 0.4 Inc Tax Rate Cut	RDF Cap at 25% of GF Rebates Only No Income Tax Cuts
Cumulative Extra Personal Income, 1994–2012	$3,602.25	$3,647.60	$3,647.60	$5,339.78
% Increase over Actual 1994–2012 PI	0.25%	0.25%	0.25%	0.37%
Cumulative 1994–2012 GF Spending Decline	$9,001	$9,415	$9,415	$9,924
% of Total Actual, 1994–2012 Spending	10.39%	10.87%	10.87%	11.45%
Avg GF Spending Cap Shortfall	3.21%	3.54%	3.54%	−0.24%
Avg GF Spending Cap Shortfall-SG	5.74%	5.31%	5.31%	5.27%
# of yrs Below GF Spending Cap	2	2	2	2
# of yrs Below GF Spending Cap-SG	2	3	3	3

Cumulative 1994–2012 Taxpayer Rebates	$4,277	$4,494	$4,494	$7,688
% of Total Actual, 1994–2012 Revenue	4.92%	5.17%	5.17%	8.85%
Cumulative 1994–2012 Tax Revenue Gain from Increased Growth	$932.50	$869.27	$869.27	$1,142.21
% of Total Revised, 1994–2012 Revenue	1.12%	1.05%	1.05%	1.30%

spending ultimately increased by population plus inflation from 1994 to 2012, but without the economic growth effects of tax cuts and rebates found in our simulations, and via a stressful fiscal roller coaster that peaked in 2008 with GF spending at 6.53 percent of personal income.[12]

Conclusion

We used a dynamic scoring simulation model to explore state fiscal policies to stabilize budgets and promote economic growth. The simulations combine tax and expenditure limits and rules for the disposition of surplus revenue to an emergency fund, rainy day fund, capital fund, and taxpayer rebates. The simulations measure the potential for income tax cuts with those rules in place, and the impact of the cuts possible while maintaining budget stability.

All of the simulations indicate significant gains from budget stabilization with spending restraint, including improved emergency preparedness, tax relief, and accelerated economic growth. The fiscal rules simulated in this article also capture the unique tax structure and tax policies in each of the states, including changes in tax structure, a first for this type of study.

With these fiscal rules in place, California and Montana could have greatly reduced income tax rates. Those cuts could have raised the personal income annual growth rate by over one-third of a percentage point in California and in Montana. The 1994–2012 cumulative personal income gains could have been over 3 percent in California and over 2 percent in Montana, substantial rates consistent with McBride's (2012) survey findings. Utah needed greater savings to avoid budget cuts during the Great Recession, so the potential for tax cuts within the parameters of the model were more limited.

Our findings indicate that California would have seen the most dramatic state spending reductions—a cumulative reduction of about 28 percent over the period as a whole. The evidence of rapid growth in state spending after California abandoned the Gann Amendment helps explain why our fiscal rules have such a large impact in California. In Montana and Utah, the cumulative reduction in state spending over the period would have been more modest, about 11 percent and 10 percent, respectively.

[12]The higher estimates that result from the high-end RMTR of 0.374 percent, and the year-by-year Utah findings, are available at http://faculty.business.utsa.edu/jmerrifi/dsa.pdf.

A major issue in the debate over tax policy is the supply-side impact of income tax rate cuts. Using the Poulson and Kaplan (2008) estimates of the negative relationship between marginal tax rates and state economic growth, we show that income tax cuts could significantly increase economic growth and that the resulting tax revenue growth would have offset much of the static revenue loss. California could have seen a tax revenue dividend of about 4 percent, while Montana and Utah could have experienced dividends of about 2 percent and 1 percent, respectively.

In the years prior to the Great Recession, despite income tax cuts, RDF deposits, and capital fund deposits, California and Utah still had enough surplus revenue for large tax rebates. The tax cut and rebate-only simulation differences reveal the importance of income tax cuts.

For a policy choice between slight spending reductions in major economic downturns with RDF caps set at 10 percent of projected GF spending and RDF caps set at 25 percent to sustain GF growth through all major downturns, we lean toward the former. Large RDF balances raise the importance of state treasurer acumen as a private-sector investor, and large RDF balances provide tempting targets for rent-seekers whose actions could undermine the legal and political foundations of TEL-RDF regimes. Infrequent periodic pressures to trim spending are healthy provided they stay below the fiscal stress threshold that would prompt tax increases and TEL erosion.

References

Abrams, B., and Dougan, W. (1986) "The Effects of Constitutional Restraints on Government Spending." *Public Choice* 49 (2): 101–16.

Amiel, L.; Deller, S.; and Stallman, J. (2012) "Economic Growth and Tax Expenditure Limitations." *Review of Regional Studies* 42: 185–206.

Bails, D. (1990) "The Effectiveness of Tax and Expenditure Limits: A Reevaluation." *American Journal of Economics and Sociology* 49 (2): 223–38.

Bails, D., and Tieslau, M. (2000) "The Impact of Fiscal Constitutions on State and Local Expenditures." *Cato Journal* 20 (2): 255–77.

Bania, N., and Stone, J. E. (2008) "Ranking State Fiscal Structures: Using Theory and Evidence." *Journal of Policy Analysis and Management* 27 (4): 751–70.

Barro, R. (1990) "Government Spending in a Simple Model of Endogenous Growth." *Journal of Political Economy* 98 (1): 103–117.

Bennet, J. T., and DiLorenzo, T. J. (1982) "Off Budget Activities of Local Government." *Public Choice* 39 (3): 333–42.

Bergh, A. and Henrekson, M. (2011) "Government Size and Growth: A Survey and Interpretation of the Evidence," *Journal of Economic Surveys* 25: 872–97.

Chapman, S. (2009) "A Hole They Dug for Themselves." *Reason* (July 30). Available at www.reason.com/news/show/135123.html.

Crain, M. (2003) *Volatile States: Institutions, Policy, and the Performance of American State Economies*. Ann Arbor: University of Michigan Press.

Dahlby, B. (1998) "Progressive Taxation and the Social Marginal Cost of Public Funds." *Journal of Public Economics* 67 (1): 105–22.

Deller, S.; Stallman J.; and Amiel L. (2012) "The Impact of State and Local Tax and Expenditure Limits on State Economic Growth." *Growth and Change* 43: 56–84.

Eaton, L. (2009) "Pinched States Wrestle with More Cuts." *Wall Street Journal* (30 July): A3.

Elder, H. (1992) "Exploring the Tax Revolt: An Analysis of the Effectiveness of State Tax and Expenditure Limitation Laws." *Public Finance Quarterly* 20 (1): 47–63.

Finkel, G. (1997) *The Economics of the Construction Industry*. Armonk, N.Y.: M.E. Sharpe.

Hanushek, E. A. (1997) "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis* 19 (2): 141–64.

———— (2003 Failure of Input-Based Schooling Policies." *Economic Journal* 113 (485): F64–F98.

Hanushek, E. A., ed. (2006) *Courting Failure*. Stanford, Calif.: Hoover Institute Press.

Holcombe, R. G., and Lacombe, D. (2004) "The Effect of State Income Taxation on per Capita Income Growth." *Public Finance Review* 32 (3): 292–312.

Holcombe, R. G., and Sobel, R. (1997) *Growth and Variability in State Tax Revenue.* London: Greenwood Press.

Hoxby, C. (2004) "Productivity in Education: The Quintessential Upstream Industry." *Southern Economic Journal* 7 (2): 209–31.

Kalita, M. S. (2009) "Governors Try to Convince Voters that Budget Woes Are Theirs, Too." *Wall Street Journal* (19 October): A2.

Kousser, T.; McCubbins, M.; and Moule, E. (2008) "For Whom the TEL Tolls: Can State Tax and Expenditure Limits Effectively Reduce Spending?" *State Politics and Policy Quarterly* 8: 331–61.

Krol, R. (2007) "The Role of Fiscal and Political Institutions in Limiting the Size of State Government." *Cato Journal* 27 (3): 431–45.

Ladner, M., and Schlomach, B. (2007) *Government Growth or Poverty Reduction? Lessons from the States*. Texas Public Policy Foundation, RR 01–2007 Available at www.texaspolicy.com/pdf/2007-01-RR01-poverty-ladner-bs.pdf.

Lav, I. J. (2009) "Tabor Has Hampered Economic Growth and Reduced the Quality of Life in Colorado." Center on Budget and Policy Priorities (October).

Lav, I. J., and Williams, E. (2010) "Tax and Spending Limits." Center on Budget and Policy Priorities (March).

Lyons, K., and Johnson, N. (2006) "Education Investment, Not TABOR, Fueled Colorado's Economic Growth in the 1990s." Center on Budget and Policy Priorities (March).

McBride, W. (2012) *What Is the Evidence on Taxes and Growth?* Washington: Tax Foundation, Special Report No. 207. Available at http://taxfoundation.org/article/what-evidence-taxes-and-growth.

McGuire, T., and Reuben, K. S. (2006) "The Colorado Revenue Limit: The Economic Effects of Tabor." Economic Policy Institute (March).

Merrifield, J. D. (1994) "Factors That Influence the Level of Underground Government." *Public Finance Quarterly* 22 (4): 462–82.

_____ (2000) "State Government Expenditure Determinants and Tax Revenue Determinants Revisited." *Public Choice* 102 (1–2): 25–50.

Merrifield, J. D., and Monson, D. (2011) "Simulation of a Constitutional Spending Limit for a Conservative State with Dynamic Adjustment and Sensitivity Analysis." *Public Budgeting and Finance* 31 (3): 1–25.

Mitchell, M. (2010) "TEL It Like It Is: Do State Tax and Expenditure Limits Actually Limit Spending?" Mercatus Center, George Mason University. Working Paper No. 10–71.

Mitchell, M., and Tuszynski, N. (2011) "Institutions and State Spending: An Overview." Mercatus Center, George Mason University. Working Paper No. 11–39 (October).

Mullins, D., and Wallin, B. (2004) "Tax and Expenditure Limitations: Introduction and Overview." *Public Budgeting and Finance* 24 (4): 2–15.

New, M. (2001) "Limiting Government through Direct Democracy: The Case of Tax and Expenditure Limitations." Cato Institute Policy Analysis No. 420.

_____ (2003) "Proposition 13 and State Budget Limitations: Past Success and Future Options." Cato Institute Briefing Papers No. 83.

Padquit, K. (2011) *State Business Tax Climate Index*. 8th ed. Washington: Tax Foundation.

Peterson, W. (1994) "Over-Investment in Public Sector Capital." *Cato Journal* 14 (1): 1–6.

Poterba, J. M. (1994) "State Responses to Fiscal Crises: The Effects of Budgetary Institutions and Politics." *Journal of Political Economy*. 102 (4): 799–821.

Poulson, B. W. (2004) "Tax and Spending Limits: Theory, Analysis, and Policy." Independence Institute Issue Paper IP–2–2004.

_____ (2009a) "A Fiscal Roadmap for Colorado." Independence Institute IP–8–2009.

_____ (2009b) "What Is at Stake in the Current Battle over Colorado's Tax and Spending Limits." Independence Institute, *Issue Backgrounder* IB–2009–C.

Poulson, B. W., and Kaplan, J. G. (1994) "A Rent-Seeking Model of TELs." *Public Choice* 79 (1–2): 117–34

_____ (2008) "State Income Taxes and Economic Growth." *Cato Journal* 28 (1): 53–71.

Reed, R. W.; Rogers, C. L.; and Skidmore, M. (2011) "On Estimating Marginal Tax Rates for U.S. States." *National Tax Journal* 64 (1): 59–84.

Rueben, K., and Rosenberg, C. (2009) "State and Local Tax Policy: What Are Rainy Day Funds and How Do They Work?" Tax Policy Center *Tax Policy Briefing Book* (August).

Schunk, D., and Woodward, D. (2005) "Spending Stabilization Rules: A Solution to Recurring State Budget Crises." *Public Budgeting and Finance* 25 (4): 105–24.

Shadbegian, R. J. (1996) "Do Tax and Expenditure Limitations Affect the Size and Growth of State Government?" *Contemporary Economic Policy* 14 (1): 22–35.

Spencer, R. W., and Yohe, W. P. (1970) "The Crowding Out of Private Expenditures by Fiscal Policy Actions." Federal Reserve Bank of St. Louis *Review* (October): 12–24.

Stallman J. (2011) "State Tax and Expenditure Limitations and Economic Performance." *Public Budgeting and Finance* 31 (4): 109–35.

Stallman J., and Deller S. (2010) "Impact of State and Local Tax and Expenditure Limits on Economic Growth." *Applied Economic Letters* 17: 645–48

Stansel, D. (1994) "Taming Leviathan: Are Tax and Spending Limits the Answer?" Cato Institute Policy Analysis No. 213.

Stansel, D., and Mitchell, D. T. (2008) "State Fiscal Crises: Are Rapid Spending Increases to Blame?" *Cato Journal* 28 (3): 435–48.

Tax Foundation (2011) *Facts and Figures Handbook: How Does Your State Compare?* Available at www.taxfoundation.org.

Taylor, J. B. (2008) "Why Permanent Tax Cuts Are the Best Stimulus." *Wall Street Journal* (25 November).

Tomjanovich, M. (2004) "The Role of State Fiscal Policy in State Economic Growth." *Contemporary Economic Policy* 22 (3): 318–30.

Vock, D.; Prah, P.; Fehr, S.; Stephen, M.; Maynard, M.; Gramlich, J.; and Leonard, K. (2009) *Beyond California: States in Fiscal Peril*. Washington: Pew Center on the States.

Wagner, G. A., and Elder, E. M. (2005) "The Role of Budget Stabilization Funds in Smoothing Government Expenditures over the Business Cycle." *Public Finance Review* 33 (4): 439–65.

_____ (2007) "Revenue Cycles and the Distribution of Shortfalls in U.S. States: Implications for an Optimal Rainy Day Fund." *National Tax Journal* 60 (4): 727–42.

Wagner, G. A. and Sobel, R. S. (2006) "State Budget Stabilization Fund Adoption: Preparing for the Next Recession or Circumventing Fiscal Constraints?" *Public Choice* 126 (1–2): 177–99.

Waisanen, B. (2010) "State Tax and Expenditure Limits 2010." Washington: National Conference of State Legislatures.

Woo, S. (2010) "In Utah Race, Tea Party Has Already Won." *Wall Street Journal* (21 June).

The Medical Care Cost Ratchet

Andrew Foy, Christopher Sciamanna, Mark Kozak, and Edward J. Filippone

Since 1970, the annual growth in U.S. health care spending per capita has been more than double the real growth in GDP per capita: 4.3 percent versus 2 percent. Over that same time period countries belonging to the Organization for Economic Cooperation and Development (OECD) averaged an annual growth rate of 3.8 percent in health care spending per capita compared to only a 2.1 percent annual growth in GDP per capita. Eight of 20 countries had higher average annual growth rates in health care spending per capita than the United States (White 2007). In light of the pronounced institutional differences among these countries in medical financing arrangements, the similarity in the rate of health care spending growth is striking. Therefore, any explanation that seeks to account for the tremendous cost growth in health care over the last several decades must hold true across all OECD countries.

This article describes a construct for health care cost growth associated with social welfare loss that we refer to as the *medical care cost*

Cato Journal, Vol. 34, No. 1 (Winter 2014). Copyright © Cato Institute. All rights reserved.

Andrew Foy is a Fellow in Cardiovascular Medicine at Penn State Hershey Medical Center. Chris Sciamanna is Professor of Medicine and Public Health Sciences and Chief of General Internal Medicine at Penn State Hershey Medical Center. Mark Kozak is Associate Professor of Medicine and Radiology at Penn State Hershey Medical Center. Edward J. Filippone is Assistant Professor of Medicine at Jefferson University Hospital. The authors thank Robert Higgs for useful comments and suggestions. His original description of the "ratchet effect" in regard to government growth was integral to the development of the medical-care-cost-ratchet concept (Higgs 1987).

ratchet (MCCR). In this model, health care spending increases over time as new technologies are incorporated into the traditional standard of care that confer only modest clinical benefits. We explain how the current medical insurance model perpetuates the MCCR. We then explain how medical cost analyses are performed before presenting several clinical vignettes that validate our model. The article concludes by arguing that market-based approaches to health care reform would be effective at bending the cost curve over time by encouraging individuals to economize nonemergent health care decisions—doing so would upset the MCCR and reduce spending growth.

Technological Change and the MCCR

In the 1980s, the conventional explanation of health care cost growth emphasized the moral hazard from health insurance and particularly the tax treatment of health insurance (Newhouse 1992). According to this view, traditional health insurance reimburses as a function of expenditure or use. Because insurance drives the marginal price of medical care at the point of use to near zero, consumers—or physicians acting as their agents—demand care until the marginal product of additional care is nearly zero. Empirical evidence exists in support of the conventional view. Studies have found that a fully insured population spends about 40–50 percent more than a population with a large deductible and their status is not measurably improved by the additional services (Manning et al. 1987). This has been referred to by Enthoven (1980) as "flat-of-the-curve medicine," where spending on medical care increases even though additional gains from such spending are very low or nonexistent. This idea has been recently reaffirmed by a landmark analysis of the Oregon Experiment where Baicker et al. (2013) found that Medicaid coverage generated no significant improvements in health outcomes in the first two years, despite increased use of prescription drugs, office visits, preventive care services including mammograms, and annual spending per individual (by insurance plan) in excess of $1,100.

The conventional view was challenged by Newhouse (1992: 11) who argued that the bulk of health care cost growth "is attributable to technological change, or what might loosely be called the march of science and the increased capabilities of medicine." According to this view, increased medical spending is welfare enhancing. To support

this view, he offered that "patients are not going to the hospital more frequently . . . nor are patients staying longer. But the real cost of a day in the hospital rose by nearly a factor of 4 from 1965 to 1986. Thus, what is being done to and for people who are in the hospital is affecting hospital costs."

While Newhouse acknowledged that there was some validity to the conventional view when looked at over a single period, he argued that it was insufficient to explain health care cost growth over time: "To explain increasing expenditure, one needs to point to something that is changing, indeed to factors that have been changing for 50 years" (Newhouse 1992: 5). He reasoned that the factor-of-five increase in real expenditure per person over the period 1950 to 1980 was more than eight times as large as one could predict from the effect of increased insurance on demand in the context of the one-period model.

Newhouse (1992) dismissed the idea that increased insurance could lead to too much technological change that was not welfare enhancing. If technological change diminished welfare, he reasoned, then countries that make centralized decisions about how much to spend on medical care would not adopt certain changes. Hence, their health care cost growth would be less than the United States. But since cost growth is similar, technological change must enhance welfare.

In our opinion, Newhouse was incorrect to dismiss the argument that too much insurance could lead to too much technological change that does not enhance welfare. In this article, we demonstrate that technological change has indeed increased health care costs in many cases without significantly improving health outcomes. This has occurred because medical insurance in its current state discourages individuals from economizing health care decisions and incentivizes the adoption and overconsumption of services with progressively diminishing returns on investment. Through this process health care costs increase as each technological advance is added to the medical repertoire. We refer to this construct as the "medical care cost ratchet" because a ratchet is a mechanical device that allows continuous linear or rotary motion in only one direction while preventing motion in the opposite direction.

Despite differences in health care financing across OECD countries, medical insurance is basically the same; its goal is to make medical care free at the point of delivery. This is achieved in several

ways. The *Bismarck system*, which originated in Germany, is dominated by private insurers that are created by or connected with employers, financed by employees and employers, and heavily regulated by the government. In the *Beveridge system*, conceived in Great Britain, the state provides comprehensive health care insurance and services to all citizens with no intermediaries (Colombatto 2012). In many countries today, including the United States, health care delivery reflects a hybrid of these two systems. Over the years, the main principle underlying them has not been challenged—that health care is a social right rather than a service to be purchased. The consequence of this attitude is that people demand the standard of medical care be made available despite the cost. Unfortunately, the general public has no conception of how valuable (or beneficial to them) the standard of care actually is and how it is determined. In many cases, providing the standard of care to a patient at the expense of a third party or the public could reasonably be considered a social welfare loss.

Before building our case we would like to clarify several points. We do not believe that our concept of the medical-care-cost-ratchet and Newhouse's concept of technological change are mutually exclusive. Instead, they should be viewed as a continuum. For example, it is common for a new drug, diagnostic test, or procedure that significantly benefits a small subgroup of patients to be used in larger subgroups of patients who benefit from it much less or not at all. We will demonstrate an example of this later in the article. Also, we do not believe that insurance per se is the problem but rather, insurance that covers all medical expenses is. If the public desires comprehensive coverage, then its consequences should at least be clearly understood. However, public perception that health care costs are too high and insurance premiums are rising too fast suggests that many are seeking alternatives. While it is reasonable and socially desirable for individuals to insure against medical emergencies, the vast majority of health care decisions are not made in an emergent setting—they are not a matter of life or death. In an elective or nonemergent setting, personal economization should be encouraged. As F. A. Hayek (1960: 422) noted,

> There is no objective standard for judging how much care and effort are required in a particular case; also, as medicine advances, it becomes more and more clear that there is no

> limit to the amount that might profitably be spent in order to do all that is objectively possible. . . . As in all other decisions in which we have to deal not with certainties but with probabilities and chances, we constantly take risks and decide on the basis of economic considerations whether a particular precaution is worthwhile, i.e., by balancing the risk against other needs.

The current medical insurance model discourages individuals to economize nonemergent health care decisions and instead encourages them to do or accept whatever the physician advises. And, in most cases, whatever the physician recommends is the standard of medical care that insurers are obliged to cover. What is the standard of medical care? It is defined as the best available combination of benefit and risk; cost is a secondary consideration that is not formally recognized in most OECD countries. Where it is formally recognized, such as the United Kingdom, the acceptable cost is quite high. The standard of care for most major medical conditions is ubiquitous across OECD countries. From the physician's standpoint, anything that provides the prospect of benefit—regardless of how small—must be offered so long as the perceived benefit is larger than the perceived risk. For example, suppose a new drug offers the prospect of reducing the risk of a heart attack in one year by 1 percent. That is a relatively small risk reduction. However, suppose the known risk of a major side effect or complication of the drug is 0.5 percent. In this situation the potential benefit outweighs the potential risk by 0.5 percent per year, so the new drug should be offered. Whether the drug costs $10, $100, $1,000, or $10,000 dollars per year does not factor into the physician's decision.

To demonstrate the MCCR, we will present several prominent medical interventions. Each vignette will present an intervention that has been added to the standard of care, describe how it has affected clinical outcome measures, and discuss the cost of its inclusion. However, we first must explain how medical cost analyses are performed and address their limitations.

Evaluating the Cost Effectiveness of Medicine

A range of approaches exist to perform economic analyses in medicine. The two most widely used are cost-effectiveness and cost-utility analyses. Cost-effectiveness estimates are expressed in terms of "years

of life saved" (YLS) and cost-utility evaluations as "quality-adjusted life years" (QALY) gained (Meltzer 2001). These approaches aim to assess the cost, both direct and indirect, of any therapeutic intervention with respect to its predictable benefits—with the effectiveness or utility being measured as the mean YLS or QALY gained as a result of the intervention (Boriani et al. 2009).

First and most important, cost effectiveness and cost utility do not mean cost saving. Cost saving means that the costs of an intervention are *less to the payer* than the costs would be if the intervention was not performed. For example, if a payer did not cover drug X, which reduces the annual risk of a heart attack in patients with condition Y, the payer would end up paying more down the line to cover the costs of treating heart attacks that could have been prevented if drug X were used by all patients with condition Y. Little is done in modern medicine that is actually cost saving. It is also important to note that cost effectiveness and cost utility do not correspond to the price of the service in question. It is often presumed that if a service is cheap, then it is cost effective and vice versa. However, a very cheap medicine can have a very poor cost ratio if many people must be treated to stop one event from occurring and if survival or quality of life is not significantly altered by its use.

Ultimately, cost analyses are modeling exercises rather than scientific experiments. This makes them dependent on the input variables such as the cost of treatment or the cost of a hospitalization for an adverse event. They are especially sensitive to the validity of data obtained in clinical trials and are therefore likely to overestimate the cost effectiveness or cost utility of any drug or intervention. Positive outcome bias, also known as "publication bias," is the well-established tendency of investigators, reviewers, and editors to submit or accept manuscripts for publication that have positive findings and to ignore or reject negative studies (Dickersin 1990, Hasenboehler et al. 2007). Cost-effectiveness and cost-utility analyses rely on the validity of unbiased, balanced, and objective data from published studies, independent of the reported outcome. This is corrupted by the positive outcome bias of individual studies, rendering clinical recommendations and cost analyses flawed toward a positive effect of specific treatment strategies. In the era of evidence-based medicine, this prevalent, often unrecognized, positive outcome bias poses a severe challenge to cost analyses by promoting unjustified therapeutic concepts.

Other limitations of cost analyses include the narrowness of inclusion criteria used in clinical trials, which often exclude elderly patients and patients with more serious co-morbidities. This can bias the cost analysis in favor of an intervention when in real life it would be much less favorable due to its use in populations who were not represented in the clinical trial(s). Cost analyses may also be plagued by overly optimistic assumptions. For example, a recent cost analysis published in a major cardiovascular journal observed that the cost effectiveness of a new endovascular technique for lowering blood pressure was very favorable; however, the authors assumed that its proven benefit in lowering blood pressure directly translated to reductions in cardiac events, though the procedure had never been proven to reduce events (Geisler et al. 2012).

Despite these significant limitations, the numbers of cost-effectiveness and cost-utility analyses have increased steadily over the last several decades. These studies have covered a range of interventions, with drugs, surgical procedures, and various diagnostic procedures as well as a range of conditions, with particular emphasis on the cardiovascular system. A review of the literature up to 1997 found that the cost utility of medical interventions varied considerably with a median QALY gained ratio of $2,000 for vaccines, $6,000 for medical care delivery, $10,000 for surgical interventions, $11,000 for pharmaceuticals, $12,000 for screening, $20,000 for health education/counseling, $20,000 for diagnostics, and $40,000 for devices. Cost-utility analyses funded by industry had more favorable results than those that were non–industry sponsored (Neumann 2000).

Due to the limitations and complexity of formal medical cost analyses, some experts suggest thinking about cost effectiveness in another, simpler term—NNT, which stands for *numbers needed to treat* or the number of patients needed to apply a particular intervention to realize a benefit of that intervention. Simply stated, the higher the NNT is for a particular intervention, the lower the cost effectiveness is for that intervention.

Illustrative Cases

In this section, we refer to cost analyses as well as NNT derived from clinical trial data to demonstrate the MCCR.

Screening for Breast Cancer with Mammography

Screening mammography in women without any signs or symptoms of breast cancer has been studied in large randomized trials of nearly a half-million women. The theoretic basis for the intervention is sound. It is presumed that therapeutic intervention at a point when cancer is visible on mammogram but not yet palpable or visibly noticeable on the breast will result in earlier, ultimately life-saving therapy.

Gotzsche and Nielsen (2013) found that routine screening is likely to reduce the absolute rate of breast cancer mortality by only 0.05 percent over a 10-year period. However, screening led to an absolute rate of overdiagnosis and overtreatment by 0.5 percent over 10 years. This means that for every 2,000 women screened throughout 10 years, only 1 will avoid death from breast cancer, and 10 healthy women will be misdiagnosed and treated unnecessarily. Furthermore, more than 200 women will experience important psychological distress for months because of false positive findings.

Based on this data, a formal cost-effectiveness analysis of screening mammography is not required to appreciate the massive costs required to prevent a single death from breast cancer. Moreover, despite the miniscule reduction in breast cancer mortality, *screening mammography does not reduce overall mortality* (Baum 2013).

Coronary Stents

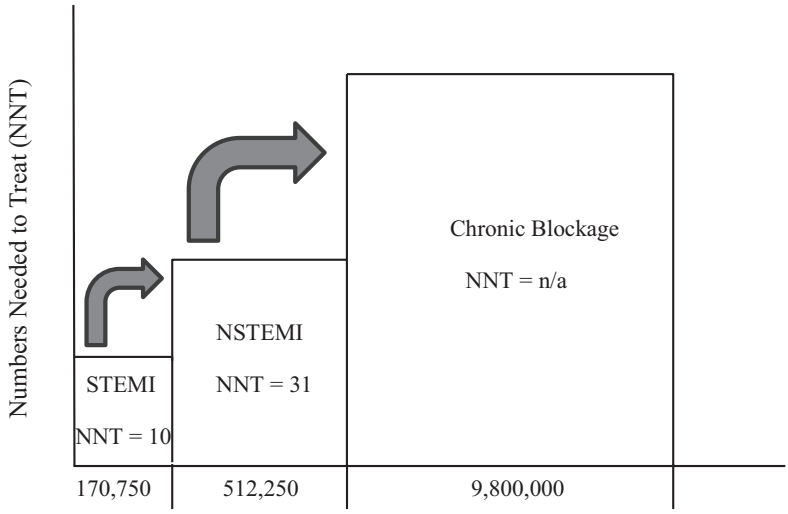
Coronary stents represent an excellent example of an intervention that is beneficial in subgroups of patients in emergent clinical settings but is much less so for the majority of patients who receive them electively. There are three broad categories of patients who receive a coronary stent. The first category is made up of patients who have a major heart attack that involves the full thickness of the heart muscle; this is termed an ST segment elevation myocardial infarction or STEMI. These patients are at significant risk of dying and being debilitated. In this setting, putting in a stent reduces the risk of death by 5–10 percent compared to the best available medical therapy alone (FTT Collaborative Group 1994, Huynh et al. 2009). Therefore, only 10 to 20 patients with an STEMI need to be treated to prevent one death and far fewer need to be treated to

prevent rehospitalization or significant disability. The overall number of STEMI patients is relatively small compared to the latter two categories.

The next category is made up of patients who have a heart attack that does not involve the full thickness of the heart muscle; this is termed non-ST segment elevation myocardial infarction or NSTEMI. In the setting of an NSTEMI, 31 patients need to be treated to prevent one death (Fox et al. 2010). STEMI and NSTEMI are both acute situations where patients often show up to the hospital sick and unstable. However, the third category is composed of patients who are not having an acute heart attack but who have a stable blockage in one or more of their heart arteries, also known as "stable angina." Patients with stable angina represent a significant percentage of those receiving stents in the United States and across OECD developed countries. Multiple large-scale clinical trials have demonstrated that in patients with stable angina, placing a stent does not reduce the risk of death or a heart attack. Therefore, an infinite number of patients with stable angina would need to be treated with a stent to save one life or to prevent one heart attack (Boden et al. 2006, De Bruyne et al. 2012). Weintraub et al. (2008) estimated that the cost per patient for a significant improvement in chest pain frequency, not prevention of death or a heart attack, was $154,580.

Figure 1 demonstrates how overall spending increases as coronary stenting is applied to subgroups of patients who derive progressively diminishing returns from it. The area of each box represents a hypothetical utility function that takes into account the number of eligible patients and the effectiveness of the intervention in that subgroup. A smaller area represents an intervention with higher utility because the cost would be low relative to the expected benefits. This is an excellent example of how the medical-care-cost-ratchet and Newhouse's idea of technological change can be viewed as a continuum. In certain subgroups of patients (STEMI and to a lesser extent NSTEMI), coronary stenting increases costs but significantly improves cardiovascular outcomes, consistent with Newhouse's technological change argument. However, in another large subgroup of patients with chronic blockages, coronary stenting has little effect if any on improving cardiovascular outcomes, consistent with our argument for the MCCR.

FIGURE 1
Medical Care Cost Ratchet (MCCR):
Coronary Stenting



Response to Newhouse's Criticisms

Newhouse (1992) specifically criticized the idea that too much insurance could facilitate technological change that was not welfare-enhancing. He argued that if consumers thought the cost of medicine did not justify its benefits, companies would provide policies offering coverage for outdated services at cheaper prices. But since insurance companies have not offered such policies, changes associated with increased costs must be welfare-enhancing. However, Newhouse himself recognized the flaw with that argument—companies would open themselves up to malpractice complaints for not providing the standard of care.

Newhouse also claimed that countries that make centralized decisions about how to allocate health care resources would not adopt changes to the standard of care that did not enhance welfare. But since they did adopt the same changes as the United States in many cases, and their cost growth was very similar, the changes must be welfare-enhancing. However, in cost-effectiveness analyses, services with

incremental costs less than $50,000 per QALY are considered low-cost, $50,000 to $100,000 per QALY are considered intermediate-cost, and greater than $100,000 per QALY are considered high-cost (Fang, Minichiello, and Auerbach 2005). In most OECD countries, there are no formal rules to define the upper limits for what is acceptable. One recent study has estimated the upper bounds for a cost-effectiveness decision rule in the United States to be $297,000 per QALY (Braithwaite et al. 2008). In the United Kingdom, the National Institute for Clinical Effectiveness (NICE) has a stated range for suitable cost effectiveness between £20,000 and £30,000 per QALY. However, research on NICE decisions reveals the threshold to be considerably higher (Devlin and Parkin 2004). This criterion assures that services with progressively diminishing returns on investment will be adopted indefinitely. Therefore, the fact that countries with more socialized health care systems continue to experience cost growth similar to the United States does not invalidate the MCCR concept—but rather, affirms it.

Policy Implications

Today there seem to be two prevailing schools of thought on how to reduce health care costs and bend the cost curve. One school, represented by Emanuel et al. (2012) and the Center for American Progress, seeks a systematic approach to containing health care costs. This approach includes global targeting of payment rates, replacing the fee-for-service model of payment, and simplifying administrative systems for payers and providers. At its core, this school believes that it can reduce costs by reducing waste and inefficiencies. This school takes for granted that technological change is welfare enhancing and that the standard of care should be provided for free at the point of service. By seeking approaches that would make health delivery in the United States more like other countries with centrally planned delivery systems, this school fundamentally fails to address the underlying problem of health care cost growth. Their solutions would work to lower costs in the context of a one-period model, for example, by reducing regional disparity in health care service intensity that does not significantly improve patient-level outcomes. However, they would not address cost growth over time. Furthermore, because their approach strongly relies on government to regulate health care delivery, it is subject to the heavy influence of industry lobbyists.

Our theory of the MCCR recognizes that health care costs around the world have not been increasing at a sensational rate because of inefficiencies within the system. Rather, the rapid cost increase is the result of incorporating new technologies into the standard of care and disseminating them to the public. The MCCR provides a useful construct for understanding how health care spending increases with the incorporation of new technologies as the clinical benefits to patients and the public progressively decline. However, medical insurance that encourages individuals to economize on nonemergent health care decisions would lower costs over time without diminishing social welfare. Reform efforts should focus on rejuvenating market forces that have been systematically suppressed. The market-based school offers several approaches.

Consumer-directed health care (CDHC) as defined by Goodman (2006) is a potential solution to control health care cost growth. By carving out areas in which it is appropriate and desirable for individuals to self-insure or pay out of pocket, CDHC encourages the economization of nonemergent health care choices. It recognizes that scarce resources must be allocated among unlimited wants. As the MCCR demonstrates, the costs of providing the standard of care in many cases, far exceeds it social benefits. Empirical evidence exists that patients with CDHC plans would reduce health care spending *without jeopardizing their health*.

The RAND Health Insurance Experiment (HIE) is the gold standard when assessing the impact of level of insurance coverage on health. The study randomly assigned individuals to different levels of health insurance generosity and compared utilization and health outcomes across experimental groups. It found no significant effect of insurance generosity on various measures of health status for the average adult—despite increased use of medical services in patients with higher levels of coverage (Levy and Meltzer 2008). Buntin et al. (2006) reviewed studies assessing the effects of CDHC on cost and quality that have followed the HIE. Despite limited data, the authors conclude that "the early evidence, consistent with the HIE, suggests that higher deductibles reduce total health care use and spending" (Buntin et al. 2006: w523). In some studies, this reduction in spending was attributed to deferral of appropriate medical care like screening mammograms. Some opponents of CDHC have cited this as a weakness of such plans. However, based on the data presented

in this article, deferral of screening mammography could as easily be considered prudent as lamentable.

A properly constructed CDHC plan would, for example, cover the cost of stenting in a patient with an STEMI, but it would not cover the cost of elective stent placement for a patient with stable angina. The latter patient would have to pay for the stent with money from his own pocket or could draw from a health savings account. A CDHC plan could be structured in such a way that it tiered co-payments for elective services so that those services with lower NNT would have lower deductibles than those with higher NNT. If this were the case, the co-pay required of an elective coronary stent would be very high. Ulrich, Brock, and Ziskind (2003) reported that from 1987 to 2001 the rate of stenting increased 128 percent. Given the high cost of revascularization, the growth in procedure volume, especially for elective cases, was only possible because the cost to the patient was near zero.

Moving toward CDHC requires a reversal of the policies and programs that have suppressed market forces. Antos, Pauly, and Wilensky (2012) offer proposals for doing this in both the public and private insurance markets. A premium-support model would shift Medicare from a defined-benefit to a defined-contribution plan by providing a fixed subsidy for each beneficiary's purchase of insurance. "Seniors would receive a uniform subsidy to purchase insurance from competing health plans (including traditional Medicare), with each offering at least a core set of benefits" (Antos, Pauly, and Wilensky 2012: 955). This is a reasonable approach, but it cannot work unless traditional Medicare is reformed in some way to encourage seniors to economize on health care. If traditional Medicare remains an option and is set as the low bid, there would be no incentive for seniors to pay for an alternative plan that approaches the CDHC model.

Antos, Pauly, and Wilensky (2012: 954) also propose that "the principle of defined contribution be applied to the currently unlimited tax subsidy for employer-sponsored insurance." Under the current tax code, employers offer insurance to employees with pre-tax dollars that encourages the purchase of health insurance policies with minimal cost-sharing, which helps fuel cost growth. It is also structured in such a way that is particularly unfair to low- and medium-income employees because "shielding premium payments from income taxes

is worth more to employees in higher income-tax brackets" (Antos, Pauly, and Wilensky 2012: 957). The existing tax exclusion could be turned into a predetermined tax credit that is made available to anyone purchasing insurance, whether through an employer or on the individual market. A fixed subsidy would eliminate the bias of the tax exclusion toward more coverage and higher spending.

Other reform proposals exist, but they are beyond the scope of this discussion. Our goal in this article was to describe a model for health care cost growth that has not received adequate attention. We believe that an understanding of the medical care cost ratchet (MCCR) is important to guide reforms that will lower health care costs—that is, bend the health care cost curve—over time without diminishing social welfare.

References

Antos, J. R.; Pauly, M. V.; and Wilensky, G. R. (2012) "Bending the Cost Curve through Market-Based Incentives." *New England Journal of Medicine* 367 (10): 954–58.

Baicker, K., et al. (2013) "The Oregon Experiment: Effects of Medicaid on Clinical Outcomes." *New England Journal of Medicine* 368 (18): 1713–22.

Baum, M. (2013) "Harms from Breast Cancer Screening Outweigh Benefits If Death Caused by Treatment Is Included." *British Journal of Medicine* 346: f385.

Boden, W. E., et al. (2007) "Optimal Medical Therapy with or without PCI for Stable Coronary Disease." *New England Journal of Medicine* 356 (15): 1503–16.

Boriani, G., et al. (2009) "Expenditure and Value for Money: The Challenge of Implantable Cardioverter Defibrillators." *QJM* 102 (5): 349–56.

Braithwaite, R. S., et al. (2008) "What Does the Value of Modern Medicine Say about the $50,000 per Quality-Adjusted Life-Year Decision Rule?" *Medical Care* 46 (4): 349–56.

Buntin, M. B., et al. (2006) "Consumer-Directed Health Care: Early Evidence about Effects on Cost and Quality." *Health Affairs* 25 (6): w516–30.

Colombato, E. (2012) "Is There a Health-Care Problem in Western Societies?" *Independent Review* 16 (3): 381–98.

De Bruyne, B., et al. (2012) "Fractional Flow Reserve-Guided PCI versus Medical Therapy in Stable Coronary Disease." *New England Journal of Medicine* 367 (11): 991–1001.

Devlin, N., and Parkin, D. (2004) "Does NICE Have a Cost-Effectiveness Threshold and What Other Factors Influence Its Decisions? A Binary Choice Analysis." *Health Economics* 13 (5): 437–52.

Dickersin, K. (1990) "The Existence of Publication Bias and Risk Factors for Its Occurrence." *Journal of the American Medical Association* 263 (10): 1385–89.

Emanuel, E., et al. (2012) "A Systemic Approach to Containing Health Care Spending." *New England Journal of Medicine* 367 (10): 949–53.

Enthoven, A. (1980) *Consumer Choice Health Plan: The Only Practical Solution to the Soaring Cost of Medical Care*. Reading, Pa.: Addison-Wesley.

Fang, M. A.; Minichiello, T.; and Auerbach, A. D. (2005) "Cost Considerations Surrounding Current and Future Anticoagulant Therapies." *Cleveland Clinic Journal of Medicine* 72 (S1): 43–49.

Fibrinolytic Therapy Trialists' (FTT) Collaborative Group (1994) "Indications for Fibrinolytic Therapy in Suspected Acute Myocardial Infarction: Collaborative Overview of Early Mortality and Major Morbidity Results from All Randomized Trials of More than 1,000 Patients." *Lancet* 343 (8893): 311–22.

Fox, K. A., et al. (2010) "Long-Term Outcome of a Routine versus Selective Invasive Strategy in Patients with Non-ST-Segment Elevation Acute Coronary Syndrome a Meta-Analysis of Individual Patient Data." *Journal of the American College of Cardiology* 55 (22): 2435–45.

Geisler, B. P., et al. (2012) "Cost-Effectiveness and Clinical Effectiveness of Catheter-Based Renal Denervation for Resistant Hypertension." *Journal of the American College of Cardiology* 60 (14): 1271–77.

Goodman, J. C. (2006) "What Is Consumer-Directed Health Care? Comparing Patient Power with Other Decision Mechanisms." *Health Affairs* 25 (6): w540–43.

Gotzsche P. C., and Neilsen, M. (2013) "Screening for Breast Cancer with Mammorgraphy." *Cochrane Database of Systematic Reviews* (6): CD001877.

Hasenboehler, E. A., et al. (2007) "Bias Towards Publishing Positive Results in Orthopedic and General Surgery: A Patient Safety Issue?" *Patient Safety in Surgery* 1 (1): 1–6.

Hayek, F. A. (1960) *The Constitution of Liberty*. Chicago: University of Chicago Press.

Higgs, R. (1987) *Crisis and Leviathan: Critical Episodes in the Growth of American Government*. New York: Oxford University Press.

Huynh, T., et al. (2009) "Comparison of Primary Percutaneous Coronary Intervention and Fibrinolytic Therapy in ST-Segment-Elevation Myocardial Infarction." *Circulation* 119 (24): 3101–09.

Levy, H., and Meltzer, D. (2008) "The Impact of Health Insurance on Health." *Annual Review of Public Health* 29 (2008): 399–409.

Manning, W. G., et al. (1987) "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment." *American Economic Review* 77 (3): 251–77.

Meltzer, M. I. (2001) "Introduction to Health Economics for Physicians." *Lancet* 358 (9286): 993–98.

Neumann, P. J., et al. (2000) "Are Pharmaceuticals Cost-Effective? A Review of the Evidence." *Health Affairs* 19 (2): 92–109.

Newhouse, J. P. (1992) "Medical Care Costs: How Much Welfare Loss?" *Journal of Economic Perspectives* 6 (3): 3–21.

Ulrich, M. R.; Brock, D. M.; and Ziskind, A. A. (2003) "Analysis of Trends in Coronary Artery Bypass Grafting and Percutaneous Coronary Intervention Rates in Washington State from 1987 to 2001." *American Journal of Cardiology* 92 (7): 836–39.

Weintraub, W. S., et al. (2008) "Cost-Effectiveness of Percutaneous Coronary Intervention in Optimally Treated Stable Coronary Patients." *Circulation Cardiovascular Quality and Outcomes* 1 (1): 12–20.

White, C. (2007) "Health Care Spending Growth: How Different Is the United States from the Rest of the OECD?" *Health Affairs* 26 (1): 154–61.

Forecast Bias of Government Agencies

Robert Krol

Forecasts of future economic activity underlie any budget revenue projection. However, the forecasters in a government agency may face incentives or pressures that introduce forecast bias. For example, agency forecasters may be rewarded for a rosy growth forecast that allows politicians to avoid politically costly program cuts or tax increases. Similarly they may be penalized for underforecasting economic growth. Where a reward system is asymmetric, it would make sense to observe biased forecasts.

This article evaluates real GDP forecasts of the Congressional Budget Office and the Office of Management and Budget. As a basis for comparison, the Blue Chip Consensus forecast is also evaluated. Tests in previous work assumed the forecast loss function was symmetric. This implies the political costs of a high or low GDP forecast are equal, so forecasts should be unbiased.

This article differs from previous work by conducting tests assuming the forecast loss function may not symmetric. Public choice models of political decisionmaking suggest government agencies such as the CBO and OMB face pressures that are likely to result in systematically biased forecasts. In this article, a flexible loss function allows for estimation of a parameter that captures the degree and direction of any forecast asymmetry. Elliott, Komunjer, and Timmermann (2005, 2008) show that failing to account for loss function asymmetry negatively affects tests that evaluate forecast accuracy and efficiency in the use of information available to forecasters.

Cato Journal, Vol. 34, No. 1 (Winter 2014). Copyright © Cato Institute. All rights reserved.

Robert Krol is Professor of Economics at California State University, Northridge. He thanks Shirley Svorny for helpful comments.

Evidence from the existing literature examining CBO and OMB forecast performance using the standard symmetric loss function is mixed. Some studies evaluate budget forecasts while others evaluate forecasts of economic activity, such as real GDP growth. Based on these efforts, three general conclusions can be drawn. First, short-run forecasts of GDP and revenues are generally unbiased while long-run forecasts of these variables have an upward bias.[1] Second, both short- and long-run forecasts of GDP and revenues usually fail tests of information use efficiency. Researchers find that forecasters do not use available information to improve their forecasts.[2] Third, despite what are likely to be different political pressures on different agencies, most of the studies find forecast biases to be similar across agencies.[3]

Using a flexible loss function to evaluate the CBO, OMB, and Blue Chip Consensus forecasts, I find significant evidence of asymmetry in the forecast loss functions. The CBO and the Blue Chip Consensus have a downward bias in their forecasts of real GDP growth two and five years out. The CBO forecast is consistent with the private sector consensus. The OMB forecast loss function is also asymmetric. However, the OMB bias is in the opposite direction. OMB forecasters overforecast real GDP growth at the two- and five-year horizons by 5 percent and 14 percent respectively. I argue that this finding is consistent with incentives facing the two agencies.

In addition, once the asymmetry of the forecast loss function is taken into account, the traditional finding that available information is not used in the forecasts is rejected in favor of the finding that government forecasters use available information efficiently. These results illustrate the importance of taking into account loss function asymmetries when evaluating the forecast performance of government agencies that are subjected to political pressures.

[1]These studies include Kamlet, Mowery, and Su (1987), Plesko (1988), Belongia (1988), Miller (1991), Blackley and DeBoer (1993), Campbell and Ghysels (1995), McNees (1995), Auerbach (1999), Krause and Douglas (2005), and Corder (2005). Looking at Japanese data, Ashiya (2007) finds an upward bias in Japanese government forecasts of GDP. Frankel (2011a, 2011b) finds evidence of an upward bias in economic forecasts using a sample of 33 countries.

[2]Kamlet, Mowery, and Su (1987) find the forecasts of both agencies are efficient. Belongia (1988) conducts encompassing tests and finds private forecasts add information to the CBO forecast but not the executive branch forecast. This implies some inefficiency in the CBO forecast.

[3]McNees (1995) found executive branch forecasts to be less accurate than the CBO and the Federal Reserve.

This article is organized in the following manner. The first and second sections discuss testing procedures under symmetric and flexible loss functions. The third and fourth sections report the results of the tests under alternative loss functions. The fifth section articulates why loss functions would be expected to differ among the agencies in question. The article ends with a brief conclusion.

Testing Forecast Accuracy with a Symmetric Loss Function

Underlying any forecast is a loss function. Standard forecast evaluations assume the forecast loss function to be quadratic and symmetric. A feature of this type of a loss function is that the optimal forecast is the conditional expectation, with the implication that forecasts are unbiased (Elliott, Komunjer, and Timmermann, 2005, 2008). I conduct a standard test of forecast performance by regressing the actual growth in real GDP over j periods ($\log Y_{t+j} - \log Y_t$) on the predicted growth in real GDP over j periods ($\log \hat{Y}_{t+j} - \log Y_t$):

$$(1)\quad \log Y_{t+j} - \log Y_t = \alpha + \beta(\log \hat{Y}_{t+j} - \log Y_t) + \varepsilon_t,$$

where $\log Y_{t+j}$ and $\log \hat{Y}_{t+j}$ are the logarithm of real GDP and predicted real GDP in period t+j respectively, α and β are parameters to be estimated, and ε_t is the error term, which should be uncorrelated for horizons beyond $j - 1$.[4] Under the unbiased forecast hypothesis, I test the joint null hypothesis that the parameter estimates are $\alpha = 0$ and $\beta = 1$. Rejecting the null hypothesis implies the forecasts are biased.

The second standard test examines if forecasters use available information efficiently. Past information about the economy should be uncorrelated with forecast errors. For this test, the forecast error (μ_t) is regressed on information, such as past forecast errors (μ_{t-i}), available at the time of the forecast:

$$(2)\quad \mu_t = \nu + \tau_1 \mu_{t-1} + \tau_2 \mu_{t-2} + \xi_t,$$

where ν, τ_1, and τ_2 are parameters to be estimated, ξ_t is a white noise error term, and μ_{t-i} are past forecast errors. The joint null hypothesis tested in this case is $\nu = \tau_1 = \tau_2 = 0$. Rejecting the null hypothesis means past forecast errors could be used to reduce the current forecast error. If this is the case, researchers conclude that available information is not being used efficiently.

[4] See Mincer (1969).

Testing Forecast Accuracy with an Asymmetric Loss Function

Elliott, Komunjer, and Timmermann (2005, 2008) develop a flexible loss function that provides an alternative method for evaluating forecasts. This approach allows the researcher to estimate a loss function parameter to determine the extent and direction of any asymmetry in the forecast loss function. As they show, ignoring asymmetry can bias forecast evaluation tests. Under certain conditions, a biased forecast can be optimal. If a low real economic growth forecast turns out to be politically more costly, an upward bias in the forecast is rational. This approach also provides an alternative test for how well forecasters use information available at the time of the forecast. Without accounting for purposeful bias in the forecast, we cannot effectively test whether forecasters use available information efficiently. Using this methodology, Elliott, Komunjer, and Timmermann (2005) find IMF and OECD budget deficit forecasts are optimal once the asymmetry is taken into account. Capistrán-Carmona (2008) uses this method to analyze the Federal Reserve's inflation forecast. In contrast to previous work, Capistrán-Carmona finds the Federal Reserve's forecasts to be optimal once the asymmetry of the forecast loss function is taken into account. Krol (2013) applies this approach to evaluate revenue forecasts for California. He finds a downward bias that implies optimistic revenue forecasts are politically costly. Also, forecasters use available information efficiently in contrast to much of the previous work in this area.

This article applies this approach to evaluate real GDP forecasts made each year by the Congressional Budget Office and the Office of Management and Budget. For a comparison, the Blue Chip Consensus forecast is also evaluated.

Equation 3 represents the flexible loss function used in this article:

$$(3)\quad L(\mu_{t+j}, \varphi) = [\varphi + (1-2\varphi)\ 1_{(\mu_{t+j<0})}]\ \left|\mu_{t+j}\right|^{p},$$

where $L(\mu_{t+j}, \varphi)$ is the loss function that depends on the forecast error and asymmetry parameter φ, and $1_{(\mu_{t+j<0})}$ is an indicator variable that takes on a value of one when the forecast error μ_{t+j}, is negative and zero otherwise. In order to identify φ, the parameter p is set equal to two making the loss function quadratic (Capistrán-Carmona 2008).

The relative cost of over- or underprediction can be calculated by $\varphi/(1-\varphi)$ (see Capistrán-Carmona 2008). For example, if $\varphi = 0.6$,

then underpredicting real GDP is one and a half times more costly than overforecasting real GDP growth. When the asymmetry parameter of the loss function φ is equal to 0.5, the loss function is symmetric. When $\varphi > 0.5$, underprediction is more costly than overpredicting real GDP growth. When $\varphi < 0.5$, overprediction is more costly than underpredicting real GDP growth.

The orthogonality condition of the optimal forecast under a flexible loss function and the estimate of φ are derived by assuming the forecasters minimize the expected loss function conditional on the information set available at the time of the forecast. The orthogonality condition is

$$(4)\quad E[\omega_t (\mu_{t+j} - (1 - 2\varphi) \left| \mu_{t+j} \right|)] = 0.$$

When this condition holds, the forecasts are optimal. In Equation 4, ω_t is a subset of information available to forecasters at the time of the forecast and $(\mu_{t+j} - (1 - 2\varphi) \left| \mu_{t+j} \right|$ is the generalized forecast error, the actual forecast error adjusted for the degree of asymmetry and the absolute size of the forecast error. When the loss function is asymmetric, the orthogonality condition implies the generalized forecast error rather than the actual forecast error is independent of the information subset. Tests based on the actual forecast errors suffer from an omitted variable problem, resulting in biased coefficients and standard errors.

A Generalized Method of Moments estimator is used to get consistent estimates of φ (Hansen, 1982). When more than one variable from the information set is used as an instrumental variable in estimation, the model is overidentified and a J-test can be used to test the orthogonality condition.

Empirical Results with a Symmetric Loss Function

Regressions 1 and 2 are used to examine CBO, OMB, and Blue Chip two- and five-year GDP forecasts published at the beginning of each year from 1976 to 2008.[5] Regression 1 tests the null hypothesis that the forecast is unbiased. Regression 2 examines if information

[5]Data on real GDP and forecasts come from CBO (2010). The crude oil price is the August and September value for West Texas Intermediate deflated by the CPI in that month. The August and September ten-year Treasury bond rate is the interest rate. The first and second quarter annualized growth rates are included in some versions of regression 2. These data come from FRED2 at the Federal Reserve Bank of St. Louis.

available at the time of the forecast is incorporated in the forecast. Since these forecasts are made during the fourth quarter of each year, I chose lags to ensure the data would be available at the time the forecast was made. For example, I use August and September crude oil prices.

Table 1, Panel A, reports the results on the unbiased forecast hypothesis. The unbiased forecast hypothesis is rejected for the five-year forecasts but not at the two-year horizon. This is similar to previous findings. The standard errors correct for the moving average property of the error term using the Newey and West (1987) approach. However, the Q-statistic still rejects white noise for the five-year forecasts.

Table 1, Panel B, reports results on how efficiently forecasters used available information. The first test includes a constant term and two lagged forecast errors. The second test includes a constant, two lagged real oil prices, two lagged ten-year Treasury bond interest rates, and two lagged real GDP annualized growth rates. Real oil prices represent an important supply shock. The Treasury bond rate captures general credit market conditions. Changes in real oil prices and bond rates both influence future real GDP growth. Lagged real GDP growth rates capture the recent performance of the variable forecasted.

P-values testing the joint significance of the impact of these alternative sets of variables on the forecast error are reported. Ninety-two percent of the tests reject the joint hypothesis that $\nu = \tau_1 = \tau_2 = 0$. These test results suggest the forecasts are not optimal, or that the loss functions are not symmetric.

Empirical Results with an Asymmetric Loss Function

Table 2 reports GMM estimates of φ, the asymmetry parameter, p-values associated with the J-test of the orthogonality condition of Equation 4 and the test statistic for the null hypothesis, $\varphi = 0.5$.

The OMB value for φ is significantly greater than 0.5 for all estimates. This implies OMB forecasters view underforecasting real GDP growth to be more costly than overforecasting it. In contrast, the CBO and Blue Chip Consensus values of φ are significantly less than 0.5 for all estimates. CBO forecasts are conservative and consistent with private sector projections. OMB forecasters

TABLE 1
Test Results

Panel A: Test Results for Unbiased Forecasts Assuming a Symmetric Loss Function

Forecast	α	β	P-value (1)	P-value (2)
CBO				
2 Years	0.694	0.776	.76	.07
	(.466)	(.01)		
5 Years	3.43	−0.154	.03	.00
	(.015)	(.765)		
OMB				
2 Years	1.48	.456	.13	.20
	(.112)	(.107)		
5 Years	3.43	−0.138	.02	.00
	(.006)	(.736)		
Blue Chip				
2 Years	0.844	0.784	.66	.39
	(.471)	(.047)		
5 Years	3.61	−0.217	.07	.00
	(.032)	(.720)		

Notes: Coefficient p-values are reported in parentheses. P-value (1) tests the joint hypothesis that $\alpha = 0$ and $\beta = 1$. P-value (2) tests if the regression residuals using a Q-statistic are white noise. The sample period is 1976 to 2008.

Panel B: P-Values for Tests of Information Efficiency Assuming a Symmetric Loss Function

	CBO2YR	OMB2YR	BCHIP2YR	CBO5YR	OMB5YR	BCHIP5YR
1	.004	.028	.000	.000	.000	.000
2	.015	.051	.007	.184	.012	.047

Notes: Row 1 includes a constant and two lagged forecast errors in the regression. Row 2 includes a constant, the September and August values of real oil prices (West Texas Intermediate deflated by the CPI in that month), similar lagged values of the ten-year Treasury bond interest rate, and the annualized real GDP growth rate for the first and second quarters. All variables come from the year preceding the budget. The sample period is 1976 to 2008.

TABLE 2
GMM Estimates of ϕ and Orthogonality Tests

	CBO2YR	OMB2YR	BCHIP2YR	CBO5YR	OMB5YR	BCHIP5YR
One						
φ	.432	.589	.231	.384	.797	.129
S.E.	.018	.018	.017	.040	.026	.022
J-Test	.044	.043	.061	.131	.197	.232
φ =.5	−3.78*	4.94*	−15.82*	−2.90*	11.42*	−16.86*
Two						
φ	.459	.540	.356	.390	.687	.371
S.E.	.015	.013	.014	.033	.020	.034
J-Test	.375	.461	.354	.624	.583	.644
φ =.5	−2.73*	3.08*	−10.29*	−3.33*	9.35*	−3.79*

Notes: The asymmetry parameter is φ, S.E. is the standard error of φ, J-Test is the p-value for Hansen's orthogonality test, and φ=.5 is the test statistic for testing the null hypothesis φ = 0.5.

* indicates rejection of the null hypothesis at the 1 percent level.

Each estimate is based on an alternative set of instrumental variables. One includes a constant and two lagged forecast errors in the regression. Two includes a constant, the September and August values of real oil prices (West Texas Intermediate deflated by the CPI in that month), similar lagged values of the ten-year Treasury bond interest rate, and the annualized real GDP growth rate for the first and second quarters. All variables come from the year preceding the budget. The sample period is 1976 to 2008.

produce a more optimistic picture of the country's economic future compared to the CBO and private forecasters.[6]

A potential complication for interpretation of the results is that the OMB forecast assumes the president's policies will be approved. The CBO forecast assumes current policies remain in place. This could

[6]Auerbach (1999) found both agencies made optimistic revenue forecasts during the 1986–93 period and pessimistic forecasts during the 1993–99 period. This suggests possible instability in the estimates. His sample and methodology differ significantly from this article. Also, his approach assumes a symmetric loss function. These sample periods are too short to estimate the model used in this article. To investigate for possible instability, I constructed a dummy variable for the 1993–99 period and reestimated the model. It had no impact on my results.

bias the OMB forecast upward compared to the CBO if the administration has a rosy perspective on the economic impact of its policies. Penner (2002) argues that the general uncertainties associated with making the forecast far outweigh any difference in policy assumptions.

Seventy-five percent of the forecasts fail to reject the orthogonality condition, indicating the forecasts are optimal. Unlike the results that assumed a symmetric loss function, forecasters appear to use available information efficiently once the asymmetry of the loss function is taken into account.

Why Do Agency Forecast Loss Functions Differ?

This section discusses why the forecast loss functions are asymmetric (leading to biased forecasts) and why they might differ between government agencies. Before discussing reasons for government agency forecast bias, it is worth examining the private sector performance. In a comprehensive evaluation, Batchelor (2007) finds evidence of a bias in private sector forecasts of real GDP and inflation in the G7 countries. Laster, Bennett, and Geoum (1999) and Lamont (2002) discuss rational reasons why even private sector forecasters may bias their forecasts. They argue that forecasts may depend on factors other than just statistical accuracy.

In Laster, Bennett and Geoum (1999), the forecaster's wage depends on accuracy and firm publicity. Because most forecasters are not very accurate and because it is difficult to evaluate forecasts very well in real time, forecasters simply want to do better than competitors in a given period. As a result, forecasters maximize their wage and firm publicity by biasing their forecast away from the consensus forecast. They find evidence to support their model. Independent forecasters who benefit most from favorable publicity make the most extreme forecasts. Industries that require accuracy, like banking, are closer to the consensus and less extreme.

Lamont (2002) also argues that forecasts depend on more than statistical accuracy. Other factors influencing the forecast are wages, profits, marketability, and shock value. The incentive structure rewards reputation that takes time to develop, so they manipulate their forecasts in an attempt to build reputation. If your reputation and wage depend only on accuracy, then the forecast would equal the true expectation. However, if your reputation and wage depend on your ability relative to other forecasters, you might move your

forecast away from the consensus. In this model, forecasters cannot develop a reputation by making forecasts similar to the consensus. Once again, the forecast may be biased.

Because it takes time to build a reputation, as forecasters gain experience, the uncertainty about their ability falls and their reputation is enhanced. In this case, Lamont argues that how one does relative to other forecasts become less important and forecasts begin to differ more from the consensus, again biasing the forecast. He finds evidence supporting the idea that as forecasters age and establish a reputation, they begin to make more extreme forecasts and lose accuracy. Finally, Lim (2001) provides evidence of an upward bias in analysts' forecasts of corporate earnings in return for information.[7]

Government forecasts are also likely to depend on factors other than just statistical accuracy.[8] The economic outlook of the individuals responsible for the forecast and the views of politicians who control the agency can be expected to impact an agency forecast. Politicians may have considerable influence over forecasters because they control the agency's budget as well as promotions and salaries. Politicians may have the power to appoint agency directors and are a valuable source for a job referral when political parties change and agency personnel are looking for work. We would expect government officials to reward forecasters who produce a projection that makes it easier to carry out their program. In this case, a forecaster's wage or an agency's budget will be a function of both accuracy and the extent to which the forecast accommodates the preferences of the politicians who oversee the agency. This view helps us to understand why a government forecast might be biased in a particular direction, but does not help in understanding why the CBO and OMB loss functions and forecast biases differ.

Krause and Douglas (2005) argue that institutional design determines the degree to which forecasts are influenced by political motives. They argue that the less politically insulated an agency is, the more likely it will be influenced by political motives, potentially

[7]Clatworthy, Peel, and Pope (2012) find that analysts' optimal earnings forecasts are biased under asymmetric loss functions even if actual earnings are symmetric. This reflects the fact that rewards and penalties of forecast errors are not symmetric.

[8] See Mueller (1997, 2003) and Rourke (1992) for a detailed discussion of bureaucratic behavior. For an application and evidence, see Svorny and Marcal (2002).

biasing forecasts of the economy. However, they do not find evidence to support this hypothesis in their own research, perhaps because they used a symmetric loss function. To explain their findings, they suggest that the professional credibility and reputation of a forecaster may offset at least some of the political pressure to slant a forecast in a particular direction.

Surely, economists working for the CBO and OMB benefit personally from unbiased forecasts which enhance their reputation and professional credibility, partially offsetting the political pressure to bias a forecast. A good forecasting performance can lead to lucrative private sector jobs. On the other hand, there are likely to be costs associated ignoring political pressures for a biased forecast. In addition to a reduction in agency funding and staff, economists who fail to respond to political pressures may simply be ignored and have little influence in the budgetary process.

The CBO and OMB are interesting agencies to study as their institutional designs differ. The OMB, as part of the executive branch, is controlled directly by the president and is likely to face significant pressure to bias its forecast. In contrast, the CBO, which reports to Congress rather than an individual or single party, is more independent. The CBO is accountable to members of both political parties who have different political goals. By design, the CBO budget is independent of congressional budget committees (Krause and Douglas 2005). Given the greater institutional independence of the CBO compared to the OMB, the costs associated with more objective forecasts should be lower, resulting in less optimistic forecasts.

Former OMB and CBO director Rudolph Penner (2002) argues that the CBO does not want to differ from the consensus outlook. According to Penner, large deviations from the consensus would make the CBO look partisan. Also, having a forecast that aligns with the consensus makes it easier to defend it before Congress. Furthermore, Penner points out that outside advisors contribute to the CBO forecast, which is likely to move the forecast in the direction of the consensus. Frankel (2011b) makes the more general argument that outside input can temper overly optimistic outlooks and limit the influence of politics.

The results in this article support these ideas. First, the OMB loss function suggests a low real GDP forecast is more costly to an administration than a rosy outlook. In a sense, the forecast is biased in a direction—upward—that helps the administration avoid politically

costly spending cuts or tax increases. Second, the greater independence from political pressure of the CBO and its desire to produce forecasts consistent with the private sector seems to hold. Both the CBO and the Blue Chip Consensus forecasts of real GDP growth have a similar downward bias.

Conclusion

This article evaluates the accuracy of the CBO, OMB, and Blue Chip Consensus forecasts of real GDP growth. Assuming a symmetric loss function, the unbiased forecast hypothesis is rejected for the five-year forecast, but not the two-year forecast. For the two- and five-year horizons, information efficiency is usually rejected. However, tests for loss function asymmetry suggest these results are unreliable. The proper loss function in this case is a flexible loss function.

Estimates under a flexible loss function suggest that each agency's loss function is asymmetric. These estimates indicate a significant upward bias in the OMB forecast. This is interpreted to mean executive branch political pressure influences the forecast. In contrast, both the CBO and Blue Chip forecasts have a downward bias. The CBO economic outlook is consistent with the private sector forecast. In contrast to previous work, once the asymmetry of the loss function is taken into account, government forecasters appear to use information on the economy efficiently in arriving at their GDP forecasts.

These results differ from most of the literature on government forecast evaluation. By addressing the issue of intentional forecast bias, they highlight the roll political pressure and institutional design may play in economic forecasts.

References

Ashiya, M. (2007) "Forecast Accuracy of the Japanese Government: Its Year-ahead GDP Forecast Is too Optimistic." *Japan and the World Economy* 19(1): 68–85.

Auerbach, A. (1999) "On the Performance and Use of Government Revenue Forecasts." *National Tax Journal* 52(4): 767–82.

Batchelor, R. (2007) "Bias in Macroeconomic Forecasts." *International Journal of Forecasting* 23(2): 189–203.

Belongia, M. T. (1988) "Are Economic Forecasts by Government Agencies Biased?" Federal Reserve Bank of St. Louis *Review* (November-December): 15–23.

Blackley, P. R., and DeBoer, L. (1993) "Bias in OMB's Economic Forecasts and Budget Proposals." *Public Choice* 76(3): 215–32.

Capistrán-Carmona, C. (2008) "Bias in Federal Reserve Inflation Forecasts: Is the Federal Reserve Irrational or Just Cautious?" *Journal of Monetary Economics* 55(8): 1415–27.

Campbell, B. and Ghysels, E. (1995) "Federal Budget Projections: A Nonparametric Assessment of Bias and Efficiency." *Review of Economics and Statistics* 77(1): 17–31.

Clatworthy, M. A.; Peel, D. A.; and Pope, P. F. (2012) "Are Analysts' Loss Functions Asymmetric?" *Journal of Forecasting* 31(8): 736–56.

Congressional Budget Office (2010) "CBO's Economic Forecasting Record: 2010 Update." Washington: Congressional Budget Office.

Corder, J. K. (2005) "Managing Uncertainty: The Bias and Efficiency of Federal Macroeconomic Forecasts." *Journal of Public Administration Research and Theory* 15(1): 55–70.

Elliott, G.; Komunjer, I.; and Timmermann, A. (2005) "Estimation and Testing of Forecast Rationality under Flexible Loss." *Review of Economic Studies* 72(4): 1107–25.

_____ (2008) "Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?" *Journal of the European Economic Association* 6(1): 122–57.

Frankel, J. (2011a) "Over-optimism in Forecasts by Official Budget Agencies and Its Implications." *Oxford Review of Economic Policy* 27(4): 536–62.

_____ (2011b) "A Solution to Fiscal Procyclicality: The Structural Budget Institutions Pioneered by Chile." *Journal EconomiaChilena* 14(2): 39–78.

Hansen, L. P. (1982) "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 59(4): 1029–54.

Kamlet, M. S.; Mowery, D. C.; and Su, T-T. (1987) "Whom Do You Trust? An Analysis of Executive and Congressional Economic Forecasts." *Journal of Policy Analysis and Management* 6(3): 365–84.

Krause, G. A., and Douglas, J. W. (2005) "Institutional Design versus Reputational Effects on Bureaucratic Performance: Evidence from U.S. Government Macroeconomic and Fiscal Projections." *Journal of Public Administration Research and Theory* 15(2): 281–306.

Krol, R. (2013) "Evaluating State Revenue Forecasting under a Flexible Loss Function." *International Journal of Forecasting* 29(2): 282–89.

Lamont, O. A. (2002) "Macroeconomic Forecasts and Microeconomic Forecasters." *Journal of Economic Behavior and Organization* 48(3): 265–80.

Laster, D.; Bennett, P.; and Geoum, I. S. (1999) "Rational Bias in Macroeconomic Forecasts." *Quarterly Journal of Economics* 114(1): 293–318.

Lim, T. (2001) "Rationality and Analysts' Forecast Bias," *Journal of Finance* 56 (1): 369–85.

McNees, S. K. (1995) "An Assessment of the Official Economic Forecasts." *New England Economic Review* (July/August): 14–23.

Miller, S. M. (1991) "Forecasting Federal Budget Deficits: How Reliable Are US Congressional Budget Office Projections?" *Applied Economics* 23(12): 1789–99.

Mincer, J. (1969) *Economic Forecasts and Expectations Analyses of Forecasting, Behavior, and Performance.* New York: Columbia University Press.

Mueller, D. C. (1997) *Perspectives on Public Choice: A Handbook*. Cambridge: Cambridge University Press.

_____ (2003) Public Choice III. Cambridge: Cambridge University Press.

Newey, W. K., and West, K. D. (1987) "A Simple, Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3): 703–08.

Penner, R. G. (2002) "Dealing with Uncertain Budget Forecasts." *Public Budgeting and Finance* 19(2): 193–206.

Plesko, G. A. (1988) "The Accuracy of Government Forecasts and Budget Projections." *National Tax Journal* 41(4): 483–501.

Rourke, F. E. (1992) "Responsiveness and Neutral Competence in American Bureaucracy." *Public Administration Review* 52(6): 539–46.

Svorny, S., and Marcal, L. (2002) "The Allocation of Federal Funds to Promote Bureaucratic Objectives: An Empirical Test." *Contemporary Economic Policy* 20(3): 209–20.

A Fresh Look at Climate Change

Paul Ballonoff

Recently *The Economist* (2013a), a prominent journalistic advocate of strong policies to control CO_2 emissions, expressed their puzzlement on the absence of warming over the last 15 years. They observed that this flat period of global average temperature occurred despite that CO_2 emissions from human sources continued at an increased rate. The total human-produced CO_2 emissions in that period of flat temperatures represent a quarter of all such emissions ever produced. The standard climate models, such as those used by the United Nation's International Panel on Climate Change (UN IPCC), anticipated that such massive CO_2 increases should have caused continuing increases in average global temperatures. *The Economist* noted that observed global average temperature is now at the lowest end of the predicted range, and that if the present trend continues, the actual temperatures will soon be below even the lowest forecasts. Most recently, Fyfe, Gillett, and Zwiers (2013) demonstrated that the current climate models have experienced a systematic failure—a finding very similar to Knappenberger and Michaels (2013).

Given the large difference of observed data from the forecasts that underlie much current policy, it is timely to ask if the climate debates are addressing the right questions. Comparison of forecasts to observations is the right way to start asking. If the forecasts used to set policy are not accurate, then policies based on those forecasts warrant

---

Cato Journal, Vol. 34, No. 1 (Winter 2014). Copyright © Cato Institute. All rights reserved.

Paul Ballonoff is an independent consultant in international energy development. He thanks Ken Wentzel and Patrick Michaels for their generous comments and suggestions in preparing this article.

review. This is important for all of the purposes for which climate policies may be set, but this article concentrates on country development policies related to energy, especially electricity. Those policies are critical, because it is widely accepted that more than one billion people have no access to reliable electric grid power and therefore must turn to other sources for heat and light (Ballonoff 2013). The cost to provide that electricity, and also meet the continued and expanding needs of developed and developing countries, is estimated in the trillions of dollars. Our understanding of climate change and how it interacts with continued expansion of use of energy resources thus has a profound effect on assuring such huge capital cost is invested in the most effective way.

The Status of Climate Science

The foundation of the modern climate change discussion is the accurate observation that human activity has significantly increased the atmospheric concentration of CO_2, and that such activity is continuing (Tans 2009). Increased CO_2 concentration, especially when amplified by predicted feedback effects thus also is assumed to predict increasing global average atmospheric temperature. Depending on the degree of warming expected, other serious and mainly undesired effects are predicted. As *The Economist* (2013a) observed, the average global temperature did rise on average over the previous century. Following a 25-year cooling trend post-World War II, temperatures increased at an especially strong rate in the quarter century ending in 1997. The trend of that warming period, the correlation with increased CO_2, and the fact of human activity causing that CO_2 increase apparently supported use of projection models extending that trend to future years. Such projections were the basis for the UN's 1997 IPCC analysis on which much current policy is based. It is thus at least ironic that 1997 was also the last year in which such measured global average temperature increase took place.

One of the key features of the IPCC forecast, and greenhouse effect forecasts generally, is the expected feedback loops. One of those is that the presumed drier and hotter conditions on the ground would cause expanded desertification and deforestation. A distinct kind of greenhouse effect is also predicted from increased CO_2 concentration—namely, the aerial fertilization effect, which is that plants grow better in an atmosphere of higher CO_2. Many analysts, such as the IPCC, clearly thought the greater effect would be from

heating, not plant growth. One must assume this was an intentional judgment, as the IPCC was aware of the CO_2 aerial fertilization effect from its 1995 Second Assessment Report, which contained empirical evidence of increased greening in enhanced CO_2 environments (Reilly 2002: 19). In contrast, climate analysts such as those with the Cato Center for the Study of Science have argued since 1999 that atmospheric temperature is much less sensitive to increased concentration of CO_2 (Michaels 1999b).

While in fact heating has not occurred as the IPCC forecasted, greatly increased global biomass is indeed demonstrated. Well documented evidence shows that concurrently with the increased CO_2 levels, extensive, large, and continuing increase in biomass is taking place globally—reducing deserts, turning grasslands to savannas, savannas to forests, and expanding existing forests (Idso 2012). That survey covered 400 peer-reviewed empirical studies, many of which included surveys of dozens to hundreds of sources. Comprehensive study of global and regional relative greening and browning using NOAA data showed that shorter-term trends in specific locations may reflect either greening or browning, and also noted that the rapid pace of greening of the Sahel is due in part to the end of the drought in that region. Nevertheless, in nearly all regions and globally, the overall effect in recent decades is decidedly toward greening (de Jong et al. 2012). This result is also the opposite of what the IPCC expected.

Global greening in response to increased CO_2 concentrations was clearly predicted by a controlled experiment of the U.S. Water Conservation Laboratory conducted from 1987 through 2005 (Idso 1991).[1] In that study, half of a group of genetically identical trees were grown in natural conditions and the other half in the same conditions but in an atmosphere of enhanced CO_2 concentration. By 1991 the Agricultural Research Service (ARS) reported that the trees in the enhanced CO_2 environment contained more than 2.8 times more sequestered carbon than the natural environment trees (i.e., were 2.8 times larger). By 2005, when the experiment was ended, the total additional growth of the enhanced CO_2 trees was 85 percent more than that of the natural-condition trees, both in woody mass and in fruit.

One reason for expanded growth even into dry environments is a seldom remarked propensity that CO_2 induced growth due to aerial

[1]See USDA, "Long-term Sour Orange Tree CO_2 Enrichment Project" (www.ars.usda.gov/Main/docs.htm?docid=9723).

fertilization also greatly increases a plant's efficiency of use of water. The ARS further documented this effect in a 2011 study, citing the extensive literature demonstrating that enhanced CO_2 environments "impact growth through improved plant water relations" (Prior et al. 2011). Similar results, both as to aerial fertilization effect and increased efficiency of water use, were found by the joint study of the USDA and the U.S. Department of Energy on the effects of CO_2 on agricultural production in the United States (Reilly 2002). In that study, the effect of forecasted increased CO_2 concentration, together with the increased warming forecasted, was shown to cause up to 80 percent increases in agricultural productivity, and *decreased* use of water since the growth would occur faster and with more efficient water use by plants. While different crops were forecasted to respond differently, most crops were positively affected, with a range from 10 percent reduction in yield up to 80 percent increase. Even considering the complex interactions with market conditions, the overall effect was certainly found to be favorable.

Using demonstrated experimental data, the 1991 ARS study also predicted effects of further or even greatly enhanced atmospheric CO_2 concentrations, such as from the expected large increase that might come (and subsequently did come and is continuing) especially from developing and newly industrializing countries. Comparing demonstrated warming to that date to the evidence, the ARS study concluded:

> If past is prologue to the future, how much more CO_2 induced warming is likely to occur? Very little. . . . The warming yet to be faced cannot be much more than what has already occurred. . . . A doubling of current emissions, for example, would lead to an atmospheric CO_2 content on the order of 700 ppm, which would probably be climatically acceptable, but only if the earth's forests are not decimated in the meantime [Idso 1991: 964–65].

The 1991 study noted that expanded forested areas would allow even greater atmospheric CO_2 concentrations. To assure the measured results were accurate and a reasonable basis on which to infer the effect of global-scale CO_2 concentration, the ARS also published results of eight additional distinct empirical studies of natural processes, each of which independently verified that the measured results found by direct experiment were a reasonable basis for such

extrapolation (Idso 1998). The effects were recently further verified by models whose results were compared to empirical data on Australian and other arid regions. Modeling water use by plants in enhanced CO_2 environment, the study predicted the effect on plant growth in dry regions and verified the result empirically compared to actual measurements over a 30-year period (AGU 2013). The data verified the prediction both in the direction and in the quantity of effect observed: Enhanced CO_2 improves water use by plants and reduces, not increases, dry regions by making them greener.

Thus, evidence to date implies that the view that global temperature is far less sensitive to CO_2 than many fear, is likely correct. Simultaneously, demonstrated experimental evidence on plant growth predicted exactly what the now extensive empirical literature shows: Enhanced CO_2 is associated with greatly increased biomass production, even in dry climates. The extent of increased CO_2 sequestration both in soil and in biomass associated with increased atmospheric concentration has also been documented (Pan et al. 2011). Those results, while not what the IPCC predicted, do not imply we should have no concerns about climate policy.

The Status of Climate Policy

Clearly, there is an imperative to rethink climate policy. The issues are real. Even the 1991 ARS study that predicted stable temperatures despite large increases in CO_2 argued for avoiding deforestation and allowing expanded forest coverage. That same study also shows why crafting a proper policy must be based on evidence.

A common belief that deforestation was taking place widely and destructively apparently motivated the ARS discussion in 1991. Massive uncontrolled forest fires in Indonesia, parts of the Amazon Basin, and Southeast Asia, as well as overgrazing and firewood collection in sub-Saharan Africa reinforced that impression, and clearly demonstrated a need for improved forestry management. Climate policies were never intended as a substitute for good basic resource stewardship; done well, they enhance it. But unfortunately, certain climate policies have caused the opposite of good stewardship. Simply allowing markets to determine the result would have been far better.

First, increased biomass is taking place exactly as predicted by the more complete CO_2 analysis. However, localized deforestation exists. The climate change literature clearly documents that over 50 percent of all wood use is for energy production, and over half of all

deforestation is due to clearing to expand biofuel production (Klenk et al. 2012). This is an unintended result of policies that subsidize biofuels to reduce CO_2 emissions from use of coal or oil for electricity generation, or oil for vehicle fuel. The next largest source of deforestation is for harvesting forest products (charcoal and wood) for household heating and cooking fuels. Such products are demanded especially by populations that do not have access to reliable grid electricity (Ballonoff 2013). A classic demonstration of this fact is the Google Earth view of the island of Hispaniola, which shows a green and vibrant Dominican Republic that has relatively reliable grid electricity next to a brown and deforested Haiti that does not.

As *The Economist* (2013b) also noted, unintended consequences are not limited to the developing world. The largest and most rapidly growing source of renewable energy in Europe—wood—is highly favored by subsidies intended to promote renewable sources. Wood can be used in power plants designed for coal with minimal modification of the plants, and it can be dispatched as a fuel like coal. In sharp contrast to other renewables, wood can be used for base-load power, backup, and grid regulation. However, paying high subsidies to substitute wood for coal seems counterproductive: aerial fertilization is absorbing the added CO_2 and thus avoiding predicted warming.

The absence of reliable, lower-cost grid power has motivated consumers of forest products for energy to be among the cash buyers of kerosene (Ballonoff 2013). However, kerosene is a major source of indoor and outdoor air pollution. The desire to replace kerosene with cleaner sources is not affected by whether CO_2 sequestration arguments are accurate. Both a large share of deforestation and the expanding use of noxious kerosene are important (and unintended) results from carbon policies, as well as from other practices that prevent operation of lower-cost grids. Projects that might rapidly expand comparatively lower-cost grid energy and improve economic development through traditional fuels are often opposed due to their carbon impacts. Counterintuitively, expanding competently operated electricity grids—even if operated to some extent from carbon fuels—may better preserve forests and have more favorable impacts on carbon. The clearly demonstrated evidence of the actual effect of higher atmospheric CO_2 concentration ought to significantly change the way such issues are analyzed and policies set.

Cost of energy efficiency compared to cost of capital for new capacity, and to the cost for use of existing capacity, is critical even if

the climate change CO_2 damage arguments are wrong. Private capital, which is the only source which can cover the massive capital requirements forecasted to meet global requirements for energy, requires that the capital is applied effectively (Ballonoff 2013). Comparative cost is critical to such analysis. All of such judgments are enhanced when the use is also shown to be the least cost and thus most easily repaid. If warming is not the threat expected, CO_2-based constraints on form of capital expansion for energy generation may be counterproductive. Thus, carbon-based fuels, which often allow lower-cost capital, and which frequently are operated from larger-scale generation attached to relatively extensive transmission and distribution grids, remain serious candidates for least-cost planning. A recent study by the World Bank found that "scenarios, based on realistic unit costs, also show that for a majority . . . decentralized power supply is unlikely to be cheaper than grid supplies any time soon" (Deichmann et al. 2010). Thus, when capital is used more effectively for generation, as may occur in central grids using larger-scale thermal sources (assuming the grid is competently, and not corruptly, operated) this also frees capital for expanding investments in efficiency. Climate policies that have often sought to force the limited capital available in developing economies for electricity development into the highest-cost and least reliable generation sources, have also made effective economic development of those least-developed economies much more difficult.

Many clean-energy generation policies, including distributed generation, use of renewable generation, use of natural gas, and applications of clean coal, have their own relative merits, for reasons quite independent of CO_2 issues. Improved air quality such as lower nitrox, lower sulfur, lower soot, and fewer noxious fumes, from substituting those for less clean fuels, are not affected by whether CO_2 sequestration is effective. Distributed generation also may have benefits for efficiency of operation of power systems, provided the associated reliability issues are properly managed and paid for. The ability to manage power system reliability when additional renewables are added is enhanced if more traditional generation, which can be more flexibly dispatched, can be considered as a source of providing reliability and system backup. Also, renewable generation is often desired for a different aspect of energy security: avoidance of foreign supply and avoidance of price volatility for imported fuels. Especially when combined with the commonly used device of a fixed price for

a long term, which is more feasible when the principal costs are capital, not fuels, renewables can provide more stable prices for future electricity generation. None of those factors are removed by the status of CO_2 sequestration arguments. Indeed, if demand for traditional fuels returns, then the associated market conditions which allow their price volatility may increase, and may thus induce increased demand for renewables, precisely to avoid such enhanced price risks in adjacent markets. All of those effects, however, will result from the normal operations of market economics and from least-cost system reliability planning and operation.

CO_2 itself is also a potential resource of economic value, both for the fertilization effect and for other uses. A creative use of the massive amounts of CO_2 expected from direct and indirect coal liquefaction and subsequent conversion to Fischer-Tropsch liquid fuels, which have no entrained heavy metals and therefore command a large premium in the marketplace, is to pipe the CO_2 for reinjection in oil fields to facilitate enhanced oil recovery. For example, the new owners of the Great Plains Coal Gasification Project in North Dakota financed and constructed a 200-mile pipeline to Canada and sold CO_2 for use in enhanced oil recovery. Similar projects are being considered in both the developed and developing world. CO_2 capture for use in commercial greenhouses to grow biomass products—whether for agriculture, medical products, flower industry, or others—is already established technology. Such projects can and should be considered as part of the economics of developing-country electrification projects, as one means of expanding the economic value of those projects. The argument is similar to that for cogeneration, or for district heating systems, namely, use of a by-product from a thermal plant for a joint output—heat, in the case of cogeneration, and extracted CO_2 for the aerial fertilization or liquid injection source.

While expanded use of renewables can be a desirable policy for the reasons just summarized, expanded renewable generation also affects operations required for assuring power system reliability, and especially for assuring adequate supply is available when demanded. The U.S. National Energy Technology Laboratory has demonstrated that additional cycling of thermal plants needed to compensate for nondispatchability of renewable generation also damages those thermal plants and induces additional system costs (US DOE 2012). In developing countries, such effects exacerbate the condition of often already poorly maintained grid generation capacity. But very

often only thermal units can provide the needed added dispatchable reliability. This is also true even in developed countries that have advanced renewable generation programs. Thus, in late 2012, the German government, having realized that massive increases in wind generation were making the transmission grid less stable, that nuclear was no longer a politically acceptable option for base load, and that added wind was also not a base-load resource, quietly began immediate construction of 23,000 MW of new coal capacity. The funds for that came in part from Germany's green energy surcharges on consumers. In general, compensating for renewable energy generation profile characteristics raises both the capital and operational cost of providing reliable power.

Since developing countries have often been more focused on effects of actions on people in the shorter term, many such countries are already more sensitive to concrete actions with real effects, than on more abstract climate arguments. Some, such as the recent Vietnamese National Green Growth Strategy, have already focused on practical policies and their real economic costs and consequences, including effects on clean environments for reasons other than beliefs about carbon, such as practical effects of distributed renewable generation, or the economics of liquefied CO_2 irrespective of temperature arguments. Such practical policies may be unaffected by beliefs about CO_2 sequestration effect on global temperatures.

An enhanced CO_2 atmosphere also portends changes in agriculture. A naive expectation is that more green means better growth for existing products, expansion of productive agricultural environments, faster growing forests and thus better product harvests, reduced desertification, and other—mainly beneficial—economic consequences. Such predominantly beneficial effects on growth due to the aerial fertilization effect are exactly what a special panel of the U.S. Department of Agriculture found when they examined the effects of increased CO_2 on U.S. agriculture (Reilly 2002). But details of effects on specific species and specific environments may differ. And, like the cited USDA study, nearly all such analysis in the past several decades has assumed that continued warming was taking place. The analyses thus also typically assume presence of environmental harms due to warming, but which warming is not occurring. Analysis of effects of enhanced CO_2 concentrations without assumed warming, as well as careful examination of how aerial fertilization effects interact with developing-country agricultural as well as other economic aid programs seems called for.

On that line, a tantalizing hint of an additional greenhouse feedback effect is suggested by the conclusion of de Jong et al. (2012: 653) on the relative distributions of global greening and browning: "In general greening prevails in all land cover classes. . . . The strongest indication for this was found in croplands and the weakest in needleleaf forests." But croplands of course are also intensely managed. And in general, human care of cropland precisely because it is market driven, requires the most intense and responsible management of both short term and longer term (meaning also, capital) risks. In a forecast made in May 1999, Ballonoff (2000) predicted that there would be no Y2K failure of the North American electricity grid on midnight of December 31 of that year. Using the power grid as an example, he reasoned that human responsible management in systems of diverse ownership in a market system are a significant part of making markets more reliable and allowing them to better adapt even to potential major operational risks. The fact that private ownership and management demonstrably has such effects has hardly been accounted for in the climate policy literature. Even if warming were a real risk, the case that only government action forcing changes in investment policies is the required solution has not been demonstrated. Indeed, the 2002 USDA study found that markets will adapt to such changes in their normal course (Reilly 2002).

Many other arguments have been made on the consequences of the expected effects of heating due to higher CO_2 concentrations. I leave for others to debate those issues, simply noting that in doing so one hopes they consider both the heating issues and the enhanced CO_2 issues, and recognize that either could have outcomes not realized by simple projections of trends. More accurate prediction and more careful attention to empirical evidence will help ensure better use of scarce energy resources.

Conclusion

Climate science, and especially understanding the interaction of human activity with climate, remains one of the key scientific challenges of our time. Humans have profoundly affected vast swaths of Australia, Eurasia, and the Americas, at least in the immediate term, by agricultural and other practices. But understanding long-term effects is not so easily guided by seemingly simple projections from short-term observations. For example, it is widely accepted that in

the 50,000 years that humans have occupied the continent of Australia, the flora and fauna have been changed profoundly (Gammage 2012). Yet, even with a century of very detailed research, it is not known for certain that the seemingly obvious inference that use of fire for clearing land over that long period has had any long-term effect on the natural climate of Australia (Smith 2013, Mulvaney and Kamminga 1999: 60–62).

It is therefore not speculation that humans can have large-scale effects on the environment. It is critical we take the risks seriously. Global climate has a profound effect on human viability. Geologically, we are at the warm cusp of an interglacial period. The period of human recorded history has occurred within a period of warming generally. Glaciation historically has occurred rather rapidly on geological time scales. The risk of severe cooling does not seem imminent, nor does the risk of severe human-induced warming.

If we can find scientifically demonstrated ways to regulate the global thermostat, we certainly want to know what those might be. The extravagant claims made by many in the climate change community have not advanced that effort, and may have contributed a widening mistrust of use of science for determining policy. Real science is not simply the application of ad hoc models to predict pending disasters; it also compares the results of predictions to actual events. The technical community that has produced false predictions of global warming, by failing to compare predictions to subsequent actual events, adds an unfortunate chapter to a long history of abuse of the appearance of science for political purposes (Michaels 1999a, 1999b).

The empirically demonstrated evidence on water use by plants in an enhanced CO_2 environment is the opposite of the commonly claimed effect from models that look only at assumed increased heating due to CO_2 increases. Empirically, CO_2 has recently been associated with warming only until increased green growth set in. That increased growth however continues so long as the extra CO_2 is present. Despite reluctant rhetoric, other climate modelers recently studying the process have also created models that show higher CO_2 concentration increases biomass. (Cox et al. 2013, Huntingford et al. 2013). But like the IPCC, many such authors seem to regard the model, not the reality against which compared, as the primary evidence. That attitude is unique in the physical or biological sciences, where reliability of prediction is judged by correspondence to empirical evidence. Reflecting a similar error, much climate policy

relies heavily on projecting assumed trends. NASA, for example, has recently displayed the results of an entire set of models that *assume* continued warming and then predict its effects.[2] But the prediction is meaningful only if the future warming exists. Trend data are only reliable for forecasts if the underlying conditions assumed remain constant and are a relatively complete description of the underlying real processes. Climate trend models have not fully accounted for the ability of plants to use water more efficiently at higher CO_2 concentrations and have underrated the capacity for aerial fertilization to sharply improve sequestration via plant growth. Had they done so, like the 1991 ARS study or the 2012 Australian analysis, they may have predicted temperature and other effects more accurately.

The misuse of modeling as a surrogate for science, which superficially allows advocacy to claim science without looking at actual evidence, has not been unique to climate warming. The new ice age foreseen in the 1950s to early 1970s did not visit us. The U.S. National Center for Atmospheric Research, one of the more prominent prophets of the new ice age, later switched to prophecy of global warming—presumably for political rather than scientific purposes. The forecasted population explosion and exhaustion of physical resources did not carry the earth past a presumed inherent carrying capacity by the early 1990s, foretold in well-known studies led by the Club of Rome and the American Association for Advancement of Science (Hardin, Lyons, and Edelson 1973; Meadows, Singer, and Perlman 1973; Meadows et al. 1974), and which were criticized even at the time (Cole et al. 1973). A 2004 update (Meadows et al. 2004) to the forecast changed the dates but not the methods, and did not improve the forecast. Instead of explosive growth, world population growth slowed, itself unpredicted by all but one theory (Ballonoff 1998). Moreover, as mineral prices were falling in real terms, efficiency of use increased, and absolute remaining known resources have generally grown in both relative and absolute terms. This is especially true for energy resources: known reserves have grown, despite that total use has far outpaced forecasts, and real price (as opposed to nominal dollar price) has fallen, not exploded to the predicted heights forecasted by the U.S. Department of Energy and others (Ballonoff and Moss 1991, Ballonoff 1997). The technology-driven mechanisms of expanding reserves also characterize the current expansion of

[2]See www.nasa.gov/topics/earth/features/wetter-wet.html.

energy reserves through fracking. What all of these examples show is that models alone, without comparison of their results to actual evidence and without embodying what is known from experimentally demonstrated behavior of nature, are not science, and may be extremely misleading foundations on which to base policy.

This article has focused on the empirical effects of climate policy related to effects of atmospheric CO_2 concentration on energy development policy, especially for electricity. Some policies are not affected by the evidence on CO_2 concentration. Energy efficiency remains a compelling goal in all climate scenarios because it leads to the most effective use of all energy capital investment and is readily achieved by normal market forces. Renewable generation remains a desired option for certain purposes of energy security, as well as for aspects of grid supply (assuming the grids are competently operated, reliable, and paid for), for potentially reducing grid losses, and as a substitute for other fuels. Such purposes, as well as programs such as "sustainable landscapes" that encourage preservation and expansion of green areas, seem justified by the demonstrated empirical effects of CO_2.

But many of the documented effects of current climate policies show a counterproductive effect on development. Efforts to reduce CO_2 emissions by subsidizing biofuels, including subsidizing wood itself as an electricity generation fuel, appear instead to be the principal cause of deforestation. Policies to avoid carbon fuels may be inhibiting development of more economically efficient central grids and degrading the operation of existing grids, thus making more difficult the task of serving the underserved with reliable and low-cost electricity. That result, in turn, paradoxically causes expanded use of hydrocarbons in the form of kerosene, with its own soot and air pollution effects, as kerosene is available where grids are not. Forest products are also harvested to substitute for unavailable reliable electric power and, as a result, contribute to deforestation. The empirically demonstrated ability of global greening to absorb greatly enhanced CO_2 concentrations and mitigate warming would seem to make policies to avoid carbon fuels in developing countries unnecessary. The demonstrated natural sequester of CO_2 in plants as atmospheric CO_2 concentration increases seems to obviate the need for foreign development capital projects to artificially sequester CO_2. In sum, the rather extensive funds dedicated to such uses by multilateral and national sources of development capital would be more effective in meeting development goals if used to increase least-cost reliable supply.

It is apparent that the demonstrated science of the direct and indirect effects due to increased CO_2 concentration is rather different from that expected by many. Past climate policy has very often been based principally on models that have not been borne out by experience. Models alone are not science; models merely reflect the assumptions embedded in them. In climate models, and climate policy generally, those assumptions have apparently not reflected demonstrated evidence. Climate policy should reflect what experimental and empirical evidence show to be true.

References

American Geophysical Union (AGU) (2013) "Elevated Carbon Dioxide Making Arid Regions Greener." AGU Release No. 13–24 (31 May). Available at www.agu.org/news/press/pr_archives/2013/2013-24.shtml.

Ballonoff, P. A. (1997) *Energy: Ending the Never-Ending Crisis.* Washington: Cato Institute.

_____ (1998) "Liberalism, Libertarianism, Free Choice." In R. Trappl (ed.) *Cybernetics and Systems '98*. Vol. 1: 431–35. Vienna: Austrian Society for Cybernetic Studies.

_____ (2000) "Melting Snow: Run-off from the Heights of Systems Forecasts." In R. Trappl (ed.) *Proceedings of the European Meetings of Cybernetics and Systems Research*, 445–49. Vienna: Austrian Society for Cybernetic Studies.

_____ (2013) "Providing Access to Electricity for the Unserved: A Free-Market Solution." *Cato Journal* 33 (1): 29–46.

Ballonoff, P. A., and Moss, D. (1991) "Natural Gas Price Forecasting: One Explanation for Upward Bias." Energy Exploration and Exploitation 9 (3): 107–21.

Cole, H. S. D.; Freeman, C.; Jahoda, M.; and Pavitt, K. L., eds. (1973) *Thinking about the Future: A Critique of the Limits to Growth*. London: Chatto and Windus for Sussex University Press.

Cox, P. M., et al. (2013) "Sensitivity of Tropical Carbon to Climate Change Constrained by Carbon Dioxide Variability." *Nature* 494 (February): 341–44.

Deichmann, U., et al. (2010) "The Economics of Renewable Energy Expansion in Sub-Saharan Africa." World Bank Policy Paper No. 5193 (January) 2010.

de Jong, R., et al. (2012) "Trend Changes in Global Greening and Browning: Contribution of Short-Term Trend to Longer-Term Change." *Global Change Biology* 18: 642–55.

Fyfe, J.; Gillett, N. P.; and Zwiers, F. W. (2013) "Overestimated Global Warming over the Past 20 Years." *Nature Climate Change* 3 (September): 767–69.

Gammage, B. (2012) *The Biggest Estate on Earth, How Aborigines Made Australia.* New York: Cambridge University Press.

Hardin, G. J.; Lyons, R.; and Edelson, E. (1973) *The Tragedy of the Commons Revisited.* Washington: American Association for the Advancement of Science.

Huntingford, C., et al. (2013) "Simulated Resilience of Tropical Rainforests to CO2-Induced Climate Change." *Nature Geoscience Letters* 6 (March): 268–73.

Idso, C. D. (2012) "Terrestrial Biosphere: How Is It Responding to Rising Atmospheric CO_2 and Warmer Temperatures?" Center for the Study of Carbon Dioxide and Global Change (5 December). Available at www.co2science.org/education/reports/greening/TheStateofEarthsTerrestrialBiosphere.pdf.

Idso, S. B. (1991) "The Aerial Fertilization Effect of CO_2 and Its Implications for Global Carbon Cycling and Greenhouse Warming." *Bulletin of the American Meteorological Society* 72 (7): 962–65.

_____ (1998) "CO_2-Induced Global Warming." *Climate Research* 10: 69–82.

Klenk, N. L., et al. (2012) *Deforestation, Forest Management and Governance.* Chichester, UK: John Wiley.

Knappenberger, P. C., and Michaels, P. J. (2013) "Policy Implications of Climate Models on the Verge of Failure." Paper delivered at the American Geophysical Union Science Policy Conference, Washington, D.C., June 24–26.

Meadows, D. H.; Meadows, D. L.; Randers, J.; and Behrens III, W. W. (1974) *The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind.* New York: Universe Books.

Meadows, D. H.; Randers, J.; and Meadows, D. L. (2004) *Limits to Growth: The 30-Year Update.* White River Junction, Vt.: Chelsea Green.

Meadows, D. L.; Singer, S. F.; and Perlman, D. (1973) *The Limits to Growth: A Debate*. Washington: American Association for Advancement of Science.

Michaels, P. J. (1999a) "The Consequences of Kyoto." Cato Policy Analysis No. 307. Available at www.cato.org/pubs/pas/pa-307.html.

_____ (1999b) "The Long Hot Year: Latest Science Debunks Global Warming Hysteria." Cato Policy Analysis No. 329. Available at www.cato.org/pubs/pas/pa-329.pdf.

Mulvaney, J., and Kamminga, J. (1999) *Prehistory of Australia.* Washington: Smithsonian Institution Press.

Pan, Y., et al. (2011) "A Large and Persistent Carbon Sink in the World's Forests." *Science* 333: 988–93.

Prior, S. A., et al. (2011) "A Review of Elevated Atmospheric CO_2 Effects on Plant Growth and Water Relations: Implications for Horticulture." *Horticultural Science* 46 (2): 159–62.

Reilly, J. M., ed. (2002) *Agriculture: The Potential Consequences of Climate Variability and Climate Change for the Unites States.* A Report of the National Agricultural Assessment Group for the USDA Global Change Research Program. New York: Cambridge University Press.

Schneider, S. H., and Mesirow, L. E. (1976) *The Genesis Strategy.* New York: Plenum.

Smith, M. (2013) *The Archaeology of Australia's Deserts*. New York: Cambridge University Press.

Tans, P. (2009) "An Accounting of the Observed Increase in Oceanic and Atmospheric CO_2 and an Outlook for the Future." *Oceanography* 22 (4): 26–35.

The Economist (2013a) "Wood: the Fuel of the Future." (6 April). Available at www.economist.com/news/business/21575771-environmental-lunacy-europe-fuel-future?frsc=dg%7Cc.

_____ (2013b) "Climate Science: A Sensitive Matter." (30 May). Available at www.economist.com/news/science-and-technology/21574461-climate-may-be-heating-up-less-response-greenhouse-gas-emissions.

U.S. Department of Agriculture (USDA) "Long-Term Sour Orange Tree CO2 Enrichment Project." Available at www.ars.usda.gov/Main/docs.htm?docid=9723.

U.S. Department of Energy (US DOE) (2012) "Impact of Load Following on Power Plant Cost and Performance: Literature Review and Industry Interviews." US DOE National Energy Technology Laboratory Document DOE/NETL-2013/1592 (October).

Contingent Liability, Capital Requirements, and Financial Reform

Joshua R. Hendrickson

A bank is considered insolvent when its liabilities (deposits) exceed the value of its assets (reserves, loans, and securities). If assets exceed liabilities, any losses experienced on the asset side of the bank balance sheet result in a corresponding loss in the bank's capital. Insolvency occurs only in the event of losses exceeding the value of capital. All else equal, a bank with more capital is at lower risk of insolvency because the value of the bank's capital fluctuates with the value of assets.

Understanding the basic analytics of a consolidated bank balance sheet provides important context for calls for financial reform in the wake of the recent financial crisis. For example, recent discussion of financial reform focuses on the role of the mixture of debt and equity finance in banking. It has been argued that banks hold an insufficient amount of capital (Miles, Yang, and Marcheggiano 2012). Put differently, the claim is that banks finance too much activity with debt than with equity. As a result, some have called for imposing higher capital requirements (Admati and Hellwig 2013).

While it is true that banks that hold more capital are at lower risk of insolvency, the logic behind calls for higher capital requirements is flawed. The flaw in this argument is that it mistakes the means for the end. The objective of banking reform is conceivably to reduce the

---

Cato Journal, Vol. 34, No. 1 (Winter 2014). Copyright © Cato Institute. All rights reserved.

Joshua R. Hendrickson is Assistant Professor of Economics at the University of Mississippi. He thanks Harlan Holt for many useful conversations about capital requirements as well as Jim Dorn and an anonymous referee for thoughtful comments on a previous draft. The usual caveat applies.

risk of insolvency among banks and other financial firms. Higher levels of capital are a means by which this can be achieved because it insulates depositors from losses, but it does not address the underlying causes that lead to insolvency.

An alternative solution is to give banks an incentive to be more prudent. For example, from the Civil War until the New Deal, nationally chartered banks had double liability. Similarly, even state chartered banks had some degree of contingent liability, in some cases more stringent than federal law. In addition, many banks outside the United States had similar liability structures. Under contingent liability, bank shareholders were subject not only to losses from the initial investment but also to losses suffered by depositors. Given that bank managers and members of the board of directors were often large shareholders of the bank, in some cases required to be by law, contingent liability gave banks the incentive to be more prudent with lending by aligning the interests of the shareholders with the depositors.

One might be tempted to argue that altering bank incentives and imposing capital requirements are likely to result in the same outcome with respect to the level of capital held by banks. However, even if this were true, the means by which this outcome is achieved is fundamentally different and has important implications for bank behavior both in lending standards and in the event of asset losses. Historical evidence suggests that contingent liability reduced bank risk taking by giving bank managers and shareholders the incentive to do so. This article argues that successful banking reform would give banks the incentive to take on less risk rather than imposing higher capital requirements.

Double Liability System

Under current federal law, U.S. banks are limited liability corporations. However, this has not always been the case. The National Banking Act of 1864 established double liability for bank shareholders:

> The shareholders [of every national banking association] shall be held individually responsible, equally and ratably, and not one for another, for all contracts, debts, and engagements of such association, to the extent of the amount of their stock therein, at the par value thereof, in addition to the amount invested in such shares [Sec. 12; see also U.S. Revised Statutes Sec. 5151 (1875) 12 U.S.C., Sec. 63].

This law held bank shareholders responsible for their initial investment in the bank as well as an amount equal to the par value of the shares in the event of insolvency in order to repay depositors. Under limited liability, shareholder losses are limited to the value of the initial investment in the event of insolvency. A regime of contingent liability is therefore unique in the sense that it requires shareholders to compensate depositors for losses out of their personal wealth if the remaining assets of the bank are insufficient to cover liabilities.

The law applied only to nationally chartered banks, but 35 states imposed double liability on shareholders. Other states imposed even more stringent laws. For example, Colorado imposed triple liability on shareholders whereas California adopted a system of unlimited liability (Vincens 1957).

Some form of contingent liability for banks was not unique to the United States. In the 19th century, many banks in the United Kingdom were subject to multiple rather than limited liability. Canadian banks were also subject to double liability until 1934, which coincided with the creation of the Bank of Canada.

The imposition of double liability might seem odd to contemporary legal and economic scholars because limited liability is now the standard practice. Nonetheless, as Evans and Quigley (1995) argue, broader liability structures, including unlimited liability, have the potential to overcome information asymmetries between creditors and shareholders. The purpose of imposing double liability on a bank's shareholders is to align the incentives of managers, directors, and other shareholders with the interests of depositors. In contrast, under a limited liability system, bank shareholders can push some of the losses onto depositors, and there is an incentive for the bank to invest in riskier assets in an attempt to earn a larger profit.

Historically, the incentives of bankers were often directly changed as a result of the fact that many states had laws that required board members to purchase a minimum amount of equity. In addition, federal law required bank directors to own at least $1,000 worth of the bank's stock (Mitchener and Richardson 2013). Under a contingent liability regime, the ability of the shareholders to pass along losses to depositors is limited. A bank with unlimited liability, like those in California and Scotland, could not pass along any of the losses to depositors. Contingent liability structures therefore internalize, at least to some degree, the losses borne by depositors in the event of insolvency.

The era of double liability in the United States was largely a success. Macey and Miller (1992), show that despite the difficulty and costliness of collecting assessments from shareholders, over 50 percent of such collections were received from 1865 to 1934. While this might not seem like a high success rate, it is important to remember that many of the shareholders of banks during this era were bank managers and board members who often faced corresponding issues of personal solvency. Assessing losses as a percentage of total liabilities is more indicative of the relative success of double liability. According to Macey and Miller (1992), depositor losses as a percentage of total liabilities were only 0.044 percent from 1865 to 1934. Even during the period from 1930 to 1934, when bank failures were more common, losses amounted to only 0.072 percent of total liabilities. In addition, the authors find that voluntary liquidations significantly outnumbered forced liquidations due to insolvency.

Contingent liability also reduced risk taking among banks. Grossman (2001) finds that banks in states with contingent liability had lower failure rates, higher capital ratios, and higher liquidity ratios than banks in states with limited liability. This evidence, however, is weaker for the 1920s.

Mitchener and Richardson (2013) find stronger evidence that contingent liability reduced risk taking among banks. The authors use differences in the dates of both the adoption of and departure from contingent liability regimes across states to examine the changes in risk-taking behavior through the early 20th century. They find that double (or multiple) liability reduced leverage ratios. Banks with double liability also maintained a larger share of retained earnings as a percentage of loans relative to banks with limited liability. The higher percentage of retained earnings meant that banks were in better position to sustain significant declines in the value of their assets. Finally, the authors attribute the increase in bank leverage after the New Deal to the fact that double liability was replaced by limited liability and federal deposit insurance.

Despite this relative success, the system of double liability in the United States ended in 1933 with amendments to the National Banking Act and the Federal Reserve Act. In particular, these amendments removed double liability from shares issued prior to June 1933. In 1935, further amendments, which took effect in 1937, eliminated double liability for all shares outstanding regardless of the issue date. In conjunction with the changes made to the liability

regime in 1933, the U.S. government created the Federal Deposit Insurance Corporation in an effort to insure depositors against losses. Vincens (1957) attributes this policy shift to the substantial cost associated with collections from shareholders as a result of both the number of bank failures and the severity of the Great Depression. Macey and Miller (1992) similarly note that the shift from double to limited liability was due to the dispersion of shareholders, the corresponding detachment of decisionmaking of ordinary shareholders, and the fact that many shareholders during the period from 1930 to 1934 were personally insolvent and thus unable to pay the assessments.

The shift from double liability to limited liability in the United States was not a shift in the preferences of depositors for limited liability; it was one that was imposed by the political process. Thus, an interesting question is: What contractual arrangements would banks and depositors agree to in the absence of legal and political forces? The Scottish experience with contingent liability is particularly useful in this context.

At the beginning of the 19th century in Scotland, the three largest and most prominent banks were chartered with limited liability. However, over the subsequent half-century, a significant number of unlimited liability joint stock banks emerged. As Evans and Quigley (1995: 505) note, "By the end of the free banking period in 1844, they had surpassed the limited liability firms as the dominant element in the Scottish banking system" (see also White 1984). The failure of the City of Glasgow Bank in 1878, however, represented a critical juncture in the Scottish banking system as it significantly called into question the desirability of unlimited liability. While Scottish banks were inclined to eliminate unlimited liability, there remained concern "about the stability of the banking system if some of the risk assumed by shareholders was simply transferred onto depositors" (Evans and Quigley 1995: 508). Rather than adopt limited liability, however, the Scottish banks adopted multiple liability. The change to multiple liability meant that shareholders were still responsible for the losses to depositors, but that there was an upper bound on this liability. Also, depositors had an incentive not only to monitor the wealth of shareholders, as was the case under unlimited liability, but also to monitor the bank regarding issues related to solvency.

The decision of the Scottish banks to offer multiple liability rather than limited liability therefore offers a potential comparison of the desirability of each structure in a market environment. Put differently,

since banks with multiple liability competed alongside those with limited liability, it is possible to evaluate the preferences of depositors for one structure relative to the other. Evans and Quigley (1995) present evidence that suggests that the market share of the limited liability banks declined after the banks with unlimited liability changed to multiple liability. In fact, deposits in the banks with multiple liability grew at a rate over twice as high as the deposits in the limited liability banks. The authors also note that the state-chartered banks petitioned the government to amend their charter such that the banks would have both Treasury oversight and multiple liability in order to gain a competitive advantage. Evans and Quigley (1995) argue that the change in market share and the desire of the chartered banks to amend their charter represent evidence that banks with multiple liability were preferred to the chartered, limited liability banks in the context of market competition. This provides strong evidence of depositor preferences for some form of contingent liability.

Incentives or Rules?

If one accepts the premise that banks hold too little capital, then it is natural to ask why such circumstances exist and to what extent policy can mitigate this inefficiency. Much of the analysis pertaining to why banks hold an insufficient amount of capital emphasizes the favorable treatment of debt relative to equity in the corporate tax structure (see Auerbach 2002; Graham 2003; Desai, Foley, and Hines 2004; Cheng and Green 2008; and Weichrieder and Klautke 2008). Similarly, bank bailouts by governments can provide an increased incentive toward leverage and the purchase of risky assets. The same can be said about government-provided deposit insurance.

If all that is preventing banks from holding the optimal level of capital is the tax system, deposit insurance, and government bailouts, it would seem that the correct policy would be to eliminate the favorable tax treatment of debt, reduce or eliminate the deposit insurance, and end the process by which large banks are bailed out by the government. Regardless of the desirability of those policies, they are unlikely to resolve the shortage of capital. As noted, there is a clear difference in the level of capital held before and after the shift in policy in 1933. As such, limited liability plays a significant role in the amount of capital that banks desire to hold, even when other factors are constant.

Instituting capital requirements is a much more politically feasible policy than any of the above options. Nonetheless, the emphasis on capital requirements is misguided. While higher capital requirements reduce the risk of insolvency in the context of a balance-sheet exercise, it is altogether unclear that those requirements would do much to make banks more prudent. For example, it is possible that banks would increase exposure to risk in an attempt to earn the same level of profit that they would have under their preferred mix of debt and equity. This is especially true if there is an expectation of government bailouts in the event of insolvency.

More important, however, is the fact that the imposition of capital requirements has adverse consequences in the event that banks suffer losses. Consider the following example of two banks holding the same level of capital. The first bank, which will be called Bank A, is holding the level of capital because of the decisions made by the managers. The second bank, hereafter Bank B, is holding the particular level of capital because of the imposition of capital requirements. Now suppose that each bank suffers a loss of the same size, which is assumed to be less than the value of its capital. For both banks, capital declines. In the case of Bank A the managers of the bank have the ability to determine when to raise more capital. However, in the case of Bank B, the bank is forced to increase capital in order to maintain a level consistent with the capital requirements. This might be particularly difficult for the bank to do if the loss suffered by the bank is particularly large or if such losses are widespread in the banking system.

The shift toward capital requirements also puts strong demands on bank regulators. Mitchener and Richardson (2013: 23) note:

> Capital requirements . . . place demands on regulators to verify balance sheet particulars with regularity, and then report these publicly to achieve market discipline. Executing this task, however, is complicated by reporting standards (marking to market versus book value) and the opacity of many types of assets. Banks have become increasingly adept at satisfying regulatory capital by shifting assets "off the balance sheet."

Capital requirements exacerbate the shift in the burden of risk management to the regulator as opposed to the bank and its shareholders. In addition, capital requirements provide banks with an

incentive to circumvent the intentions of the regulation while remaining officially compliant.

Finally, much of the analysis that pertains to why banks hold an insufficient amount of capital examines the choice of the mix between debt and equity as though it were solely the decision of the bank. In reality the observed mixture of debt and equity is the equilibrium outcome of the interaction between banks and their liability holders. This distinction is important because the equity and debt of a bank yield different services for the liability holders of banks. Debt issued by the bank in the form of deposits also serves as a medium of exchange whereas bank equity does not. As a result, under certain circumstances banks might issue more debt relative to equity because the former is preferred by liability holders.[1] In this case, it is possible that the imposition of capital requirements is welfare-reducing, or at least that optimal levels of capital have been overstated.

By contrast, requiring that bank shareholders are subject to contingent liability provides an incentive for banks to internalize any potential losses to depositors since bank shareholders are responsible for those losses. Realigning the incentives of shareholders to be consistent with those of depositors has a number of advantages relative to capital requirements. For example, the mix between debt and equity for a bank with contingent liability is chosen by the bank. Banks with a riskier portfolio of assets might decide to finance a greater share of their activity through equity rather than debt. Correspondingly banks with less risky portfolios might choose a smaller fraction of equity finance.

Capital requirements are unlikely to be risk-adjusted. With limited liability, banks do not have an incentive to internalize losses to depositors in the event of insolvency and are therefore likely to choose a riskier portfolio of assets. Compliance with capital requirements in this instance provides a bank with the appearance of propriety even if the bank is at greater risk of losses and insolvency. In addition, even if capital requirements are risk-adjusted, this adjustment would be at the discretion of regulators rather than bank

[1]Hendrickson and Holt (2013) show that when there is a shortage of transaction assets, bank liability holders strictly prefer deposits to equity. This might explain why Macey and Miller (1992) find evidence that banks with double liability often held less capital than those with limited liability. For more on asset shortages, see Caballero (2006).

managers and shareholders. In the context of limited liability, it is possible that regulation could improve on the allocation of bank resources in the event that banks take on too much risk since shareholders do not have an incentive to internalize depositor losses. However, shareholders with contingent liability are likely to have better assessments of the risk in comparison with regulators since the shareholders would stand to lose some amount of their personal wealth in the event of a bank failure.

This point is especially important given the nature of regulation. It is possible, for example, that a system in which bank shareholders have limited liability and regulators impose risk-adjusted capital requirements could result in an efficient use of resources. This statement, however, relies on two critically important assumptions. First, the ability of bank regulators to promote an efficient allocation of resources assumes that regulators are guided solely by the interests of promoting solvency in the banking system and ignores the political economy aspect of bank reform and regulation.[2] In addition, a system of limited liability with risk-adjusted capital requirements requires a particular sort of specialized knowledge that may not be possessed or even obtained by regulators.

Even if there were a significant decline in the value of its assets, a bank with contingent liability would be permitted to have a lower level of capital at its own discretion. Banks with contingent liability would not be forced to raise capital in the wake of large and significant losses on the asset side of the balance sheet. Nevertheless, bank shareholders would still have an incentive to ensure that the bank take the necessary steps to prevent such large losses and lower levels of capital from increasing the risk of insolvency.

Contingent liability also gives bank shareholders the incentive to be proactive in the event of large and significant losses. If shareholders believe that the bank is at greater risk of insolvency, there is an incentive to voluntarily liquidate assets rather than risk personal wealth in the event of a forced liquidation. This incentive is clear from historical evidence. As noted, Macey and Miller (1992) document the fact that voluntary liquidations significantly outnumbered forced liquidations during the period in which U.S. bank shareholders were subject to double liability.

[2]For a discussion of the political economy aspect of regulation, with particular attention to the recent financial crisis in the United States, see Johnson and Kwak (2010).

Overall, contingent liability provides banks with more flexibility in decisionmaking and in dealing with declines in asset values while also providing bank shareholders with better incentives to monitor risk than do capital requirements. Imposing higher capital requirements would continue the three-quarter-century-long trend of shifting the burden of assessing risk from the bank and its shareholders to regulators.

Marketability and Transferability

While theory and evidence suggest that a contingent liability regime provides better incentives for banks than capital requirements, the main criticism of contingent liability has been with the marketability and tradability of shares.[3] Legal scholars have argued that shares with unlimited liability shift the distribution of risk to wealthier shareholders (Halpern, Trebilcock, and Turnbull 1980).[4] In the event of bankruptcy, if a number of shareholders are insolvent as well, the burden of repayment for liabilities would shift to the wealthiest of the remaining shareholders. This is potentially problematic because of the implications for asset pricing. A shareholder that is wealthy relative to other shareholders would value the stock at a price below that of the other shareholders. Symmetrically, shareholders with little wealth relative to other shareholders would have a higher valuation for the stock. Standard asset pricing theory suggests that the price of a stock should be equal to the present discounted value of its future dividends. If the relative wealth of the shareholders affects the valuation of individual shareholders, then it might be difficult to ascertain a common market price. It is therefore argued that regimes of unlimited liability "create a significant measure of uncertainty in the valuation of securities and threaten the existence of organized securities markets" (Halpern, Trebilcock, and Turnbull 1980: 147).

Woodward (1985) also raises concerns about the transferability of shares in the absence of limited liability. She argues that under

[3]This view dates back at least to Walter Bagehot's writing in *The Economist* and the *Saturday Review*, who argued that unlimited liability joint stock banks that existed at the time of his writing would ultimately have shareholders with few assets. For more on Bagehot's view, see Hickson and Turner (2003).

[4]See also Easterbrook and Fischel (1985) and Grundfest (1992), who argue that minimum capital requirements are attempts to reduce the social costs that result from limited liability regimes.

contingent liability, if only the current shareholders were subject to liability, then the wealthiest shareholders would sell shares if there was a threat of bankruptcy. Those willing to buy shares (assuming symmetric information) would be those with too little wealth to be pursued in the event of a bankruptcy. In this case, the contingent liability regime would become a de facto limited liability regime. It would seem that unlimited liability regimes would require the limitation of transferability of shares to prevent this outcome.

Concerns about the marketability and transferability of shares, however, are largely unfounded. For example, much of the criticism of contingent liability regimes assumes that such a regime would be one of unlimited liability and that the liability would be joint and several. Put differently, this assumption implies that in the event of bankruptcy the difference between liabilities and assets would be assessed to shareholders in proportion to their holdings. In the event that some shareholders were insolvent or otherwise unable to meet this obligation, their assessments would be transferred to wealthier shareholders. This increases the costs associated with holding shares because a shareholder would now need to have information about the wealth of fellow shareholders in order to determine the liability associated with owning shares. All else equal, this characteristic would certainly reduce the marketability of shares. As Hansmann and Kraakman (1991) argue, however, unlimited liability does not imply that the liability be joint and several. In fact, the contingent liability regime that existed in the United States was not joint and several, but rather shareholder assessments were determined by the value of their shares determined by the receiver in the event of bankruptcy.

Much of the concern surrounding the marketability and the transferability of shares in companies with contingent liability is based on theoretical work. In particular, the arguments described earlier suggest that when shareholders are subject to unlimited liability, the market for the firm's shares will be less liquid, ownership will be more concentrated, and there will be evidence of higher risk reflected in share prices. The existing empirical evidence casts doubt on those concerns.

Grossman (1995), for example, examines the experience of American Express during the 1950s. American Express was initially chartered in 1850 as an unlimited liability joint stock company. American Express did not become a limited liability corporation until 1965. Grossman examines the experience of American Express in the

1950s because it provides an example of a firm in which shareholders were subject to unlimited liability at a time when the vast majority of other firms' shareholders were subject to limited liability. The evidence shows that shares of American Express were dispersed among 25,000 shareholders.[5] Shares in American Express were also listed in the financial press among other actively traded stocks, which provides indirect evidence that shares were not illiquid relative to shares of firms subject to limited liability.[6] In addition, using a capital asset pricing model, Grossman fails to find evidence that American Express shares were more risky than the overall market. This evidence casts doubt on the hypothesis that shares subject to pro rata unlimited liability would be subject to limited marketability and transferability.

The experience of unlimited liability joint stock banks in Ireland in the 19th century provides further evidence against the hypothesis that shares with unlimited liability are subject to limited marketability and transferability. The argument made by Halpern, Trebilcock, and Turnbull (1980) and others is that firms with unlimited liability would have share prices that were functions of both the expected income of the firm and the wealth of the shareholders and that this characteristic would prevent the marketability of shares. The inability to determine a common market price implies that wealthier individuals would pay lower prices for shares. Using detailed information from the archives of the Ulster Banking Company in Ireland during the 19th century, Hickson and Turner (2003) fail to find any evidence that wealth had an effect on the price paid by shareholders. In subsequent work, Hickson, Turner, and McCann (2005) show that a liquid market existed for the shares of Ulster Banking Company and that there was no identifiable change in liquidity after the bank became a limited liability corporation in 1883.

Similarly, the experience of Irish banks provides evidence against Woodward's (1985) critique that unlimited liability would create an incentive for wealthy shareholders to sell shares when the bank was threatened with bankruptcy. While the idea that wealthy shareholders would like to escape their liability is reasonable, shareholders of

[5]As Grossman (1995) notes, the number of shareholders implies that on average each shareholder owned just over 80 shares.

[6]American Express shares were traded over the counter and volume was not publicly recorded. Indirect evidence is therefore needed to assess liquidity.

Irish banks were subject to post-sale extended liability in which shareholders were subject to assessments three years from the sale of shares.[7] Shareholders that foresaw an impending bankruptcy were not capable of avoiding the liability.

Conclusion

In the wake of the recent financial crisis, advocates of policy reform have emphasized the imposition of greater capital requirements as a way to prevent bank insolvency. The intuition behind this recommendation is that capital provides a buffer to depositors in the event of significant declines in the value of assets on a bank's balance sheet. All else equal, a higher level of capital (a greater provision of equity finance) reduces the risk of insolvency and protects depositors from losses.

While the logic of the advocacy of greater capital requirements in reducing insolvency is not at issue, there is reason to believe that meaningful banking reform requires a much different approach. For example, banks with greater capital requirements might be at a reduced risk of insolvency, but the imposition of such requirements does not necessarily alter the incentives of the bank. Two banks with balance sheets of the same size, holding the same amount of capital, might have significantly different risk profiles on the asset side of the balance sheet. Nonetheless, each bank would be compliant with regulation. In addition, with capital requirements banks would be forced to raise capital in the aftermath of significant declines in the value of assets on the bank's balance sheet, a time when banks are likely to find such actions most difficult.

An alternative to capital requirements is to provide banks with an incentive to internalize the losses faced by depositors in the risk of insolvency. One way to alter bank incentives is to impose some form of contingent liability, in which bank shareholders would not only lose the value of their initial investment in the event of insolvency but would also be subject to compensating depositors for any losses. This is in stark contrast to the limited liability of bank shareholders under present law. Under the present system, bank shareholders have no responsibility to compensate depositors. Contingent liability

[7] This was amended in the late 19th century to limit the post-sale liability to one year from the sale.

therefore causes bank shareholders to internalize the costs to depositors of insolvency. As a result, contingent liability realigns the incentives of bank shareholders to be cognizant of the preferences and concerns of depositors, which results in less risky behavior on the part of the bank.

Historical evidence suggests that contingent liability regimes have more desirable characteristics than limited liability regimes. Evidence from the United States shows that banks with contingent liability took on less risk and less leverage than their limited liability counterparts. There is also evidence in the United States that contingent liability led to voluntary liquidations that seemingly reduced the number of insolvencies among banks during this period. In addition, the successes of contingent liability do not appear to be confined to the United States. The available evidence on Scotland during the 19th century lends credence to the view that banks with contingent liability were preferred to those with unlimited liability during a time at which these banks competed with one another for market share.

Nonetheless, the main criticism of contingent liability regimes is that they limit the marketability and transferability of shares and therefore impede investment and economic growth. Although this claim has been subject to much debate, arguments against contingent liability regimes are largely theoretical. The empirical evidence on the subject is limited because unlimited liability regimes largely existed prior to the emergence of organized financial markets or in cases in which there is little or no measure of comparison. Existing empirical evidence, however, shows that the theoretical concerns surrounding contingent liability regimes are largely unfounded. This evidence provides further support for the claim that contingent liability regimes are preferable to limited liability regimes with capital requirements.

A system of contingent liability is theoretically preferable to a regime of limited liability on the grounds that it provides better incentives for banks. In addition, the historical evidence suggests that regimes of contingent liability have many preferable characteristics relative to those of limited liability. Taken as a whole, it should be clear that contingent liability provides a preferable alternative to the present regime of limited liability and that meaningful banking reform should seek to realign the incentives of banks rather than merely imposing higher capital requirements.

References

Admati, A., and Hellwig, M. (2013) *The Bankers' New Clothes.* Princeton, N.J.: Princeton University Press.

Auerbach, A. (2002) "Taxation and Corporate Financial Policy." *American Economic Review* 92: 67–78.

Caballero, R. (2006) "On the Macroeconomics of Asset Shortages." In A. Beyer and L. Reichlin (eds.) *The Role of Money: Money and Monetary Policy in the Twenty-First Century.* Frankfurt, Germany: European Central Bank.

Cheng, Y., and Green, C. (2008) "Taxes and Capital Structure: A Study of European Companies." *The Manchester School* 76 (S1): 85–115.

Desai, A.; Foley, C. F.; and Hines, J. (2004) "A Multinational Perspective on Capital Structure Choice and Internal Capital Markets." *Journal of Finance* 59 (6): 2451–3487.

Easterbrook, F. H., and Fischel, D. R. (1985) "Limited Liability and the Corporation." *University of Chicago Law Review* 52: 89–117.

Evans, L. T., and Quigley, N. C. (1995) "Shareholder Liability Regimes, Principle-Agent Relationships, and Banking Industry Performance." *Journal of Law and Economics* 38 (2): 497–520.

Graham, J. (2003) "Taxes and Corporate Finance: A Review." *Review of Financial Studies* 16 (4): 1075–1129.

Grossman, P. Z. (1995) "The Market for Shares of Companies with Unlimited Liability: The Case of American Express." *Journal of Legal Studies* 24 (1): 63–85.

Grossman, R. (2001) "Double Liability and Bank Risk Taking." *Journal of Money, Credit and Banking* 33 (2): 143–59.

Grundfest, J. A. (1992) "The Limited Future of Unlimited Liability: A Capital Markets Perspective." *Yale Law Journal* 102: 387–425.

Halpern, P.; Trebilcock, M.; and Turnbull, S. (1980) "An Economic Analysis of Limited Liability in Corporation Law." *University of Toronto Law Journal* 117: 117–50.

Hansmann, H., and Kraakman, R. (1991) "Toward Unlimited Shareholder Liability for Corporate Torts." *Yale Law Journal* 100: 1879–1934.

Hendrickson, J. R., and Holt, H. (2013) "Why Don't Banks Hold More Capital?" Working Paper, University of Mississippi.

Hickson, C. R., and Turner, J. D. (2003) "The Trading of Unlimited Liability Bank Shares in Nineteenth-Century Ireland: The Bagehot Hypothesis." *Journal of Economic History* 63 (4): 931–58.

Hickson, C. R.; Turner, J. D.; and McCann, C. (2005) "Much Ado about Nothing: The Limitation of Liability and the Market for 19th Century Irish Bank Stock." *Explorations in Economic History* 42 (3): 459–76.

Johnson, S., and Kwak, J. (2010) *13 Bankers: The Wall Street Takeover and the Next Financial Meltdown.* New York: Vintage Books.

Macey, J. R., and Miller, G. P. (1992) "Double Liability of Bank Shareholders: History and Implications." *Wake Forest Law Review* 27: 31–62.

Mitchener, K. J., and Richardson, G. (2013) "Does 'Skin in the Game' Reduce Risk Taking? Leverage, Liability, and the Long-Run Consequences of New Deal Banking Reform." NBER Working Paper No. 18895.

Miles, D.; Yang, J.; and Marcheggiano, G. (2012) "Optimal Bank Capital." *Economic Journal* 123 (567): 1–37.

Vincens, J. R. (1957) "On the Demise of Double Liability of Bank Shareholders." *The Business Lawyer* 12 (3): 275–79.

Weichrieder, A., and Klautke, T. (2008) "Taxes and the Efficiency Costs of Capital Distortions." CESifo Working Paper No. 2431.

White, L. H. (1984) *Free Banking in Britain.* Cambridge: Cambridge University Press.

Woodward, S. (1985) "Limited Liability in the Theory of the Firm." *Journal of Institutional and Theoretical Economics* 141: 601–11.

The Explicit Costs of Government Deposit Insurance

Thomas L. Hogan and William J. Luther

The Diamond-Dybvig (DD) model is often cited as a theoretical justification for government deposit insurance. In the model, rational agents find it in their interest to withdraw their bank deposits if they suspect other depositors plan to do likewise. When a sufficient number of agents are expected to liquidate their accounts, a bank run ensues. Guaranteeing deposits through a system of government-administered deposit insurance removes the temptation to run on the bank and thereby precludes the need to ever use the deposit insurance. As Thomas Sargent makes clear, deposit insurance enters the model as a costless solution:

> The good news in the Diamond-Dybvig model . . . is that if you put in government-supplied deposit insurance, that knocks out the bad equilibrium. People don't initiate bank runs because they trust that their deposits are safely insured. And a great thing is that it ends up not costing the government anything to offer the deposit insurance! It's just good all the way around [Rolnick 2010: 31].

Diamond and Dybvig (1983: 44) conclude, "Government deposit insurance can improve on the best allocations that private markets provide."

Cato Journal, Vol. 34, No. 1 (Winter 2014). Copyright © Cato Institute. All rights reserved.

Thomas L. Hogan is Assistant Professor of Finance and a member of the Manuel H. Johnson Center for Political Economy at Troy University. William J. Luther is Assistant Professor of Economics at Kenyon College. They thank the Cato Institute and Mercatus Center at George Mason University for supporting this research.

In practice, however, government-provided deposit insurance is not a costless solution. It is frequently invoked to cover the losses of failed banks. In the United States, deposit insurance provided by the Federal Deposit Insurance Corporation (FDIC) is administered as a "rainy-day fund." Each member bank pays a premium based on its risk rating and on the amount of insurable deposits held by the bank. Premiums are held in a deposit insurance fund (DIF) administered by the FDIC. When a member bank becomes insolvent, debts to its depositors are paid out of the DIF and premiums for all banks increase until the fund is restored. Annual premiums reflect the cost of operating and administering the system and recent losses due to bank failures. The cost of guaranteeing deposits through such a system is decidedly nonzero.

The costs of providing insurance are sufficiently high as to warrant their inclusion in any reasonable model of deposit insurance. If the corresponding benefits of deposit insurance were sufficiently large and the alternative means of acquiring these benefits were either nonexistent or sufficiently costly, then ignoring the real-world costs of government-provided deposit insurance is perhaps appropriate. However, we do not believe this is the case. The benefits are not so large that one need not be concerned with costs, and there are potentially superior alternatives to government-provided deposit insurance. If one is to consider alternatives by engaging in comparative institutional analysis, a better understanding of the costs of government deposit insurance is required.

In what follows, we explore the explicit costs of government-provided deposit insurance. We focus on the FDIC as a specific example. First, we review the DD model and show how FDIC deposit insurance differs from the model in several key respects. Second, we discuss the history of the FDIC, paying particular attention to how the maximum amount covered, number of bank failures, and cost of managing the deposit insurance fund have changed over time. Third, we briefly discuss private deposit insurance and other risk-constraining mechanisms as alternatives to government-provided insurance. Finally, we offer some concluding remarks.

Theory of Deposit Insurance

The DD model demonstrates that although banks can reduce individual risk by acting as financial intermediaries, they create

systemic risk in the potential for bank runs. Under certain conditions, financial contagion can cause all banks, even solvent ones, to be run upon simultaneously. Diamond and Dybvig (1983) propose that government insurance can costlessly limit the risk of bank runs by guaranteeing the values of customer deposits. Here we review the DD model to demonstrate how deposit insurance under FDIC differs from that posited by Diamond and Dybvig.

Banks in the DD model provide insurance against some form of uncertainty about the future. Suppose there is a group of agents with only one type of opportunity for production over three periods 0, 1, and 2. For each unit of capital invested in the production process in period 0, any agent can earn a return of $R > 1$ in period 2 or withdraw his original investment of 1 unit in period 1. Many analogies have been used for this scenario: planting corn that grows in the future but provides a meager yield if harvested early (Selgin 1993); a business project where investors' time horizon is uncertain (Diamond 2007); a real estate fund that may be relinquished early at a discount or held to maturity (Sebastian and Tyrell 2006). The asymmetry of future payments in each of these cases makes their payoff patterns suboptimal. Because the agents invested in these technologies tend to be (or at least are assumed to be) risk averse, they would prefer to accept a reduction of their high potential payment in period 2 in return for a small increase in their potential payment in period 1.

A bank can be created to reduce the cost of uncertainty by smoothing the potential future payoffs. Agents that invest in the bank receive a deposit contract that allows them to choose between future payoffs of return r_1 in period 1 or r_2 in period 2 where $1 < r_1 < r_2 < R$. However, the bank uses the same production technology as agents in the economy. In period 1 it has only 1 unit of capital per deposit contract but has promised each agent $r_1 > 1$. If too many agents redeem their deposits in period 1, the bank will not have sufficient capital to fulfill its obligations, and the bank will go into default. Once it becomes known that the bank may default, all agents have an incentive to redeem their deposits immediately in period 1 since no capital will be left in period 2. This flood of simultaneous redemptions constitutes a bank run. The danger of bank runs is most poignant when consumer preferences are unknown. Because each agent fears that the others may withdraw early, bank runs become a

self-fulfilling prophecy: any indication that there *may* be a bank run can itself *cause* a bank run.[1]

Diamond and Dybvig (1983) propose that a system of government-provided deposit insurance can mitigate the danger of bank runs. The authors assume that the government has an advantage over private banks because it can enact its desired policies ex post once the optimal allocation of resources is known. "In particular, it can tax those agents who withdrew 'early' in period T = 1." By contrast, "a private insurance company is constrained by its reserves in the scale of unconditional guarantees which it can offer" (Diamond and Dybvig 1983: 413). They conclude that "this asymmetry allows a potential benefit from government intervention" (p. 414). Once it is known that the government will redistribute any undeserved gains from bank runs, agents no longer have an incentive to run on the bank. Hence, the government's commitment to providing deposit insurance precludes the possibility of a bank run and guarantees that deposit insurance payouts will never be necessary. In this way, government deposit insurance becomes a costless solution to the problem of bank runs.

Other works extend the DD model to examine the optimality of deposit insurance under a variety of assumptions. Dowd (1988), Wallace (1990), Selgin (1993), and Green and Lin (2000) propose alternative measures, such as proper capital allocation and suspending deposit redemptions to improve upon government deposit insurance. Peck and Shell (2003) show even those optimal contracts may be subject to runs. Others study the effects of signaling and information on the potential for runs (e.g., Samartin 2003; Andolato, Nosal, and Wallace 2007). However, each employs a model applicable only under a strict set of assumptions, with little consensus as to which is the most useful representation of deposit insurance in practice.

Diamond and Dybvig (1983) acknowledge the implementation of deposit insurance is likely to be suboptimal. They note that their model produces "a very strong result (which may be too strong) about the optimality of government deposit insurance" (p. 414). The costs of actual deposit insurance deviate from the DD model in several ways. Taxes assessed on banks to fund deposit insurance will cause the provision of insurance to be suboptimal because there are real costs to assessing and collecting taxes. Diamond and Dybvig

[1] Diamond and Dybvig (1983: 410) argue that a bank run may be caused by any "commonly observable random variable in the economy . . . even sunspots."

(1983: 415) state that "if a nonoptimal tax must be imposed, then when t is stochastic there will be some tax distortions and resource costs associated with government deposit insurance."[2] Furthermore, to the extent that failures resulting from bank runs are indistinguishable from other types of failure, government-provided deposit insurance is more likely to be under- or oversupplied.

Government deposit insurance departs markedly from that proffered in Diamond and Dybvig (1983). To further illustrate the differences between deposit insurance in theory and practice, and to more clearly understand the actual costs, we examine deposit insurance offered under the FDIC.[3]

FDIC Deposit Insurance

The FDIC was established by the Banking Act of 1933 primarily in response to widespread bank failures in that year. Prior to the FDIC, deposit insurance was provided at the state level. However, rural bank failures during the economic downturn of 1921 and the decade of crop failures that followed proved too much for these funds to handle, and all had ceased operations by 1930. Despite initially opposing federal deposit insurance, President Franklin D. Roosevelt signed the bill into law on June 16, 1933. By January of the following year the program was up and running.[4]

The history of the FDIC has been presented in much greater detail elsewhere.[5] Since we are ultimately concerned with the cost of deposit insurance in practice, our aim is limited to expositing the

[2] Diamond and Dybvig (1983: 416) claim that "so long as the government can impose some tax to finance the insurance, no matter how distortionary, there will be no runs and the distorting tax need never be imposed."

[3] Considering a single case of government-provided deposit insurance allows for greater depth of analysis. However, it provides no assurance that the case at hand is representative. Our decision to focus on the FDIC reflects our familiarity with the case and the ease with which data can be obtained. As a result, the reader is cautioned to consider the generalizability of the case at hand.

[4] Initially, the FDIC was designed to start on July 1, 1934. A late amendment proposed by Senator Arthur Vandenburg created the Temporary Federal Deposit Insurance Fund, which would operate in the interim from January 1, 1934 to July 1, 1934. Extensions approved on June 16, 1934, and June 28, 1935, postponed the transition to the permanent corporation to September 1, 1935.

[5] For a more detailed account, the reader is advised to consult FDIC (1984), Calomiris and White (1994), Kaufman (2002), Bradley (2000), and Kroszner and Melick (2008).

size, scope, and function of the FDIC over time. Specifically, we consider the maximum amount covered, number of bank failures, and the cost of managing the deposit insurance fund since the FDIC was established.[6]

FDIC Coverage

Both the real amounts covered by FDIC deposit insurance and the percentage of bank deposits insured by the FDIC have consistently increased over time. When the deposit insurance program kicked off in 1934, the maximum amount insured per depositor was set at $2,500. Depositors holding $2,500 or less in a member bank would have the entire balance of their accounts "backed by the full faith and credit of the United States government," which promised to pay 100 cents on the dollar should the bank become unable to meet the demands of the depositor.[7] Depositors holding more than $2,500 would receive 100 cents on the dollar for the first $2,500 held in eligible accounts, but no coverage was extended to the remaining balance.

Bradley (2000: 7) maintains that the $2,500 maximum was established for two reasons. First, it put banks on a level playing field with the Postal Savings System (PSS). Deposits held in the PSS were limited to $2,500, and since the PSS was a government program, those accounts were already backed by the full faith and credit of the U.S. government. Establishing the maximum amount covered under the FDIC at $2,500 merely extended the coverage already offered on PSS accounts to deposits held at banks. Second, the $2,500 maximum served as a compromise between bankers and depositors. Still struggling from the financial crisis, bankers feared they would be unable to afford assessments for complete coverage. Depositors, on the other hand, preferred complete coverage. Since roughly 97 percent of depositors in 1933 held less than $2,500 in accounts, limiting coverage to $2,500 per depositor allowed for lower assessments while providing most depositors with complete coverage.

The maximum amount insured per depositor is legally established in nominal terms. As a result, it must be continually adjusted to keep

[6] These data are taken from the FDIC's *2010 Annual Report* available at www.fdic.gov/about/strategic/report/2010annualreport/index_pdf.html.

[7] As described in Title IX of the Competitive Equality Banking Act of 1987.

TABLE 1
Nominal Maximum Individual FDIC Coverage Amounts, by Period

Time Period	Nominal Amount Covered by FDIC
1934–35	$2,500
1935–50	$5,000
1950–66	$10,000
1966–69	$15,000
1969–74	$20,000
1974–80	$40,000
1980–2008	$100,000
2008–Present	$250,000

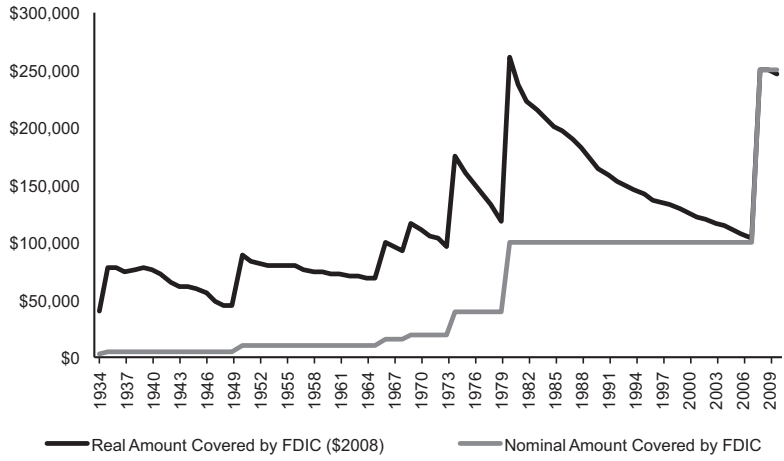
Source: Federal Deposit Insurance Corporation, *2010 Annual Report*.

up with inflation if coverage is to be held constant through time. Since 1934, the maximum amount covered under FDIC has been increased seven times. Historical coverage amounts are presented in Table 1.

The infrequency with which the maximum amount insured per depositor has been adjusted means that the real (inflation-adjusted) amount covered by the FDIC varied significantly over the history of the program. Since there are positive costs associate with changing the nominal maximum, it does not necessarily follow that infrequent adjustments are inefficient.[8] Nonetheless, it is important to consider the real maximum amount insured per depositor. Controlling for inflation allows one to see how much of a depositor's wealth is actually protected by deposit insurance at a particular point in time, as well as how coverage levels have changed through time. Understanding the extent to which accounts are protected by deposit insurance will be useful in assessing the historical cost of providing insurance under the FDIC. The inflation-adjusted maximum amount insured per depositor over time is depicted as a solid black

[8] In a 1980 speech before Congress, for example, then-FDIC Chairman Irving H. Sprague estimated that increasing the maximum amount covered would result in up to $750,000 in direct costs to the FDIC (Bradley 2000: 19).

FIGURE 1
Real and Nominal Maximum FDIC Coverage Amounts, 1934–2010



Source: Federal Deposit Insurance Corporation, *2010 Annual Report*.

line in Figure 1 alongside the solid grey line representing the nominal amount covered.

Since its inception, the real scope of federal deposit insurance (reported in constant 2008 dollars) has increased by roughly 514 percent, from $40,168 in 1934 to $246,706 in 2010. Steady inflation from 1935 to 1965 and infrequent adjustments (only once over the period, in 1950) resulted in a relatively large range of coverage, from a low of $44,669 in 1948 to a high of $89,337 in 1950. The average for the 30-year period totaled $70,974. The 15-year period that followed saw several increases in FDIC coverage (in 1966, 1969, 1974, and 1980).[9] From 1965 to 1980, the real amount covered per depositor increased by roughly 282 percent. The largest increase in the period occurred in 1980, when the real amount covered was more than doubled from $118,624 to $261,290. In 1980, the real amount

[9] The amount of public unit deposits covered was also increased in 1974. Whereas the maximum covered for nonpublic unit accounts was upped from (nominal) $20,000 to $40,000, the maximum for public unit deposits increased from $20,000 to $100,000. We report only the increase for nonpublic unit deposits in Table 1 and Figure 1.

covered per depositor was greater than any other year in the history of the FDIC.

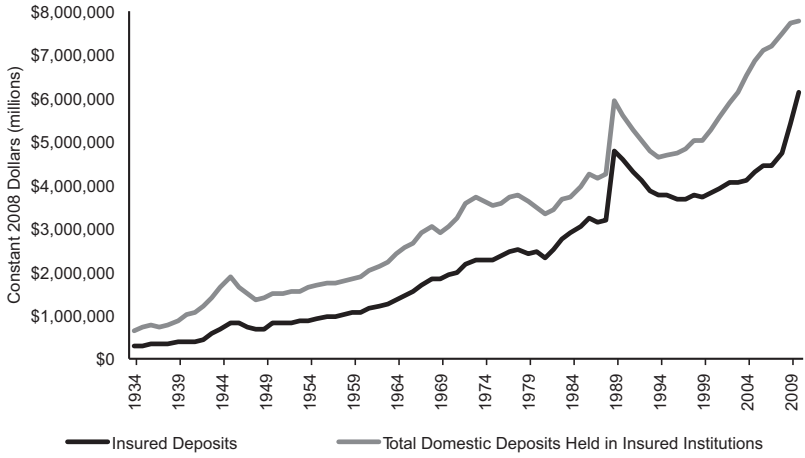
Inflation eroded the real scope of deposit insurance from 1980 to 2007. However, over the entire period, coverage amounts remained well above the 1934 to 1979 inflation-adjusted average of $85,720. In 2008, the real amount covered by the FDIC increased by 141 percent, from $103,840 to $250,000.[10] As such, it is the single largest real increase in the history of the program. As of 2010, the real amount covered by the FDIC totaled $246,706. Although large by historical standards, the real coverage amount today is comparable to that offered in 1980.

Although the maximum amount covered under FDIC has outpaced inflation on average, the real increase in deposit insurance might merely reflect the increase in income over the same period. As incomes increase, depositors become wealthier and the maximum amount covered must also increase if the proportion of the average depositor's wealth protected under FDIC is to be maintained. However, this is not generally the case. From 1934 to 1965, real coverage per depositor increased at an annualized rate of 1.68 percent from $40,168 to $68,350, whereas real GDP per capita increased at an annualized rate of 3.52 percent from $8,391 to $25,406. Over the 16-year period from 1965 through 1980, real GDP per capita grew at an annualized rate of 1.46 percent while real coverage per depositor managed an astounding 8.74 percent. The maximum amount covered fell relative to GDP from 1980 to 2007. Even still, real coverage has outpaced real GDP per capita over the history of the program, growing at 2.39 and 2.25 percent per year, respectively.

As population and income per capita have increased over the last 75 years, the amount of funds held in member institutions and the amount insured under the FDIC have also increased. Insured and total domestic deposits held in member institutions, measured in millions of 2008 dollars, are depicted in Figure 2 as solid black and grey lines respectively. Total deposits grew at an average

[10] The statutory increase from $100,000 to $250,000 in September 2008 was originally set to revert back to $100,000 in January 2010. The temporary increase was then extended to last until January 2014. The signing of the Dodd-Frank Wall Street Reform and Consumer Protection Act on July 21, 2010, made the temporary increase permanent.

FIGURE 2
Insured and Total Domestic Deposits, 1934–2010



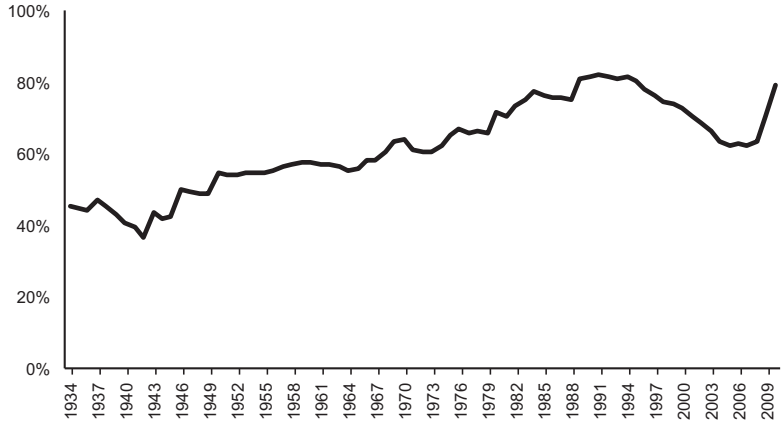
Source: Federal Deposit Insurance Corporation, *2010 Annual Report*.

annual rate of 3.29 percent, from $643,660 million in 1934 to $7,783,791 million in 2010. Over the same period, insured deposits increased by 4.04 percent per year, from $290,418 million to $6,139,149 million.

Also observable in Figure 2, the proportion of total deposits insured under the FDIC has increased over the history of the program. To make this even clearer, we present the proportion of total domestic deposits held in insured institutions covered by the FDIC in Figure 3. At its inception in 1934, 45.12 percent of domestic deposits held in member institutions were insured. The proportion then fell to an all-time low of 36.54 percent in 1943, after which it increased steadily to a high of 82.05 percent in 1991. In 2010, the FDIC covered 78.87 percent of domestic deposits held in insured institutions.

In general, the above data confirm that federal deposit insurance coverage has increased since 1934. Upward adjustments to the nominal maximum amount insured per depositor have outpaced both inflation and growth in GDP per capita. As a result, the proportion of deposit balances covered by the FDIC has grown as well.

FIGURE 3
Insured Deposits as a Percentage of Total Deposits, 1934–2010



Source: Federal Deposit Insurance Corporation, *2010 Annual Report*.

Bank Failures under the FDIC

As discussed earlier, deposit insurance removes the incentive for depositors to run on solvent banks. Since a run forces the bank to liquidate its long positions in order to cover withdraws on demand, one might expect there to be fewer bank failures when deposits are insured. However, it does not follow that deposit insurance will prevent all (or even most) bank failures. Banks also fail from making unsound investments in risky assets. Poorly managed banks suffer losses just like any other poorly managed firm, and sufficiently bad banks fail. Absent a carefully constructed counterfactual, the raw number of bank failures in the presence of deposit insurance provides no indication as to whether (and to what extent) the scheme is working.

Although the number of failing banks cannot be used to gauge how well government-provided deposit insurance is working, it could significantly affect the cost of the program. In practice, it is difficult to distinguish whether a bank has failed because of a run or whether depositors are running because a bank has failed. Under the FDIC, depositors are compensated regardless of what prompts the bank to fail. Hence, even if the deposit insurance scheme is working, the costs of providing deposit insurance might still exceed the benefits.

TABLE 2
Number of Bank Failures, by Decade

Decade	Failures
1930s[a]	312
1940s	99
1950s	28
1960s	43
1970s	76
1980s	1,015
1990s	490
2000s	197
2010s[b]	157

[a] Data unavailable for years 1930, 1931, 1932, 1933.

[b] Only includes data from 2010.

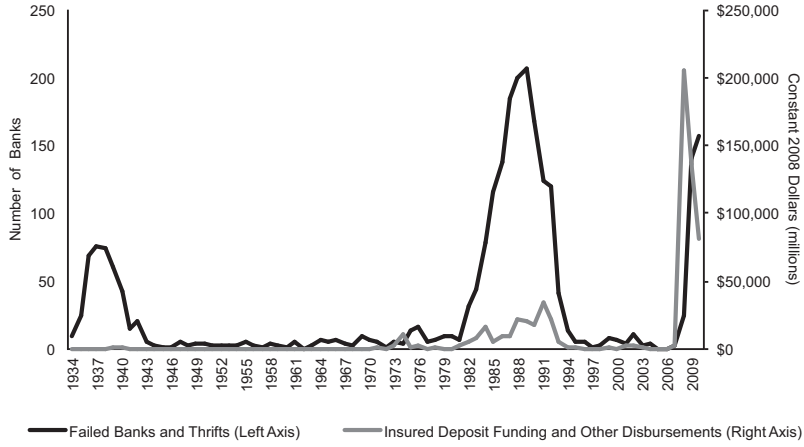
Source: Federal Deposit Insurance Corporation, *2010 Annual Report*.

The number of bank failures for each decade between 1930 and 2010 are presented in Table 2.[11] Over the period, 2,417 banks and thrifts failed—roughly 31 per year on average. Bank failures are not distributed evenly over the period, however. The 1950s and 1960s saw relatively few failures—only 28 and 43, respectively. The 1980s, in contrast, witnessed 1,015 failures—accounting for more than 40 percent of all failures between 1934 and 2010. Similarly, more failures occurred in 2010—the last year for which data are available—than in the 1950s and 1960s combined.

The black line in Figure 4 depicts the number of bank failures in each year between 1934 and 2010. Three periods stand out. During the second half of the Great Depression (1935–42), 381 banks closed their doors at an average rate of 3.97 banks per month. The average rate equaled 9.38 banks per month during the Savings and Loan Crisis (1982–94). In total, 1,464 banks closed over the period, with 206 occurring in just one year (1989)—more than any other single year in the series. Finally, in the most recent financial crisis (2008–10), 322 banks failed at an average rate of 8.94 per month.

[11] Since the series starts in 1934, data are unavailable for years 1930, 1931, 1932, and 1933.

FIGURE 4
Disbursements to Depositors of Failed and Assisted Banks under FDIC, 1934–2010



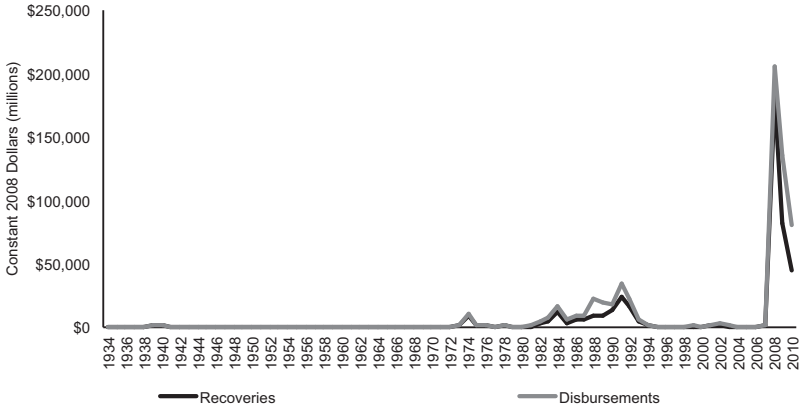
Source: Federal Deposit Insurance Corporation, *2010 Annual Report*.

Although the number of bank failures provides some context, we are ultimately concerned with the cost of bank failures on the deposit insurance program and should therefore consider the size of banks failing. One indicator of the magnitude of banks failing is the amount disbursed by the FDIC to depositors. Disbursements to depositors of failed and assisted banks by the FDIC from 1934 to 2010 are depicted by the solid grey line in Figure 4. Disbursements are measured in millions of inflation-adjusted dollars (100 = 2008) along the right vertical axis. The graph is consistent with the stylized facts of the three periods discussed above. The average bank that failed between 1935 and 1942 was relatively small. Many were rural banks, holding little on deposit. The FDIC disbursed only $4,347 million over the period—averaging $543 million per year and $11.41 million per failed bank. Failures occurring between 1982 and 1994 were generally more substantive than those in the earlier period. Disbursements totaling $177,703 million—more than 40 times the amount disbursed in the second half of the Great Depression—reflect the greater number of larger banks failing over a longer period of time. On average, the FDIC disbursed $13,669 million per year between 1982 and 1994, more than $121.38 million per failed bank. Finally, as expected, the most recent financial crisis saw significantly larger banks failing over a much shorter period. Although more banks failed between

1982 and 1994 than 2008 and 2010, the amount disbursed to depositors of failed banks more than doubled to $422,597 million in the latter period. Between 2008 and 2010, disbursements by the FDIC averaged $140,866 million per year, $1,312.41 million per failed bank.

The aforementioned disbursements are offset in part by recoveries obtained from failed and assisted banks. Recoveries indicate any value recovered from the sale of a failed or assisted bank's assets. Recoveries by the FDIC from 1934 to 2010 are depicted by a solid black line in Figure 5.[12] For reference, we include disbursements over the same period as a solid grey line. Both disbursements and recoveries are measured in millions of inflation-adjusted dollars (100 = 2008). From 1935 to 1942, $3,989 million was recovered from failed banks—$498.68 million per year on average, or $10.47 million per bank failure. As the number and size of failing banks picked up during the Savings and Loan Crisis, the amount recovered also increased. Recoveries totaled $115,076 million from 1982 to 1994, averaging $8,852.04 million per year and $78.60 million per failed bank. Finally, from 2008 to 2010, the amount recovered exceeded $339,529 million. In other words, roughly $113,176.57 million was recovered each year over the three-year period—averaging $1,054.44 million per failed bank.

FIGURE 5
Disbursements and Recoveries under FDIC, 1934–2010

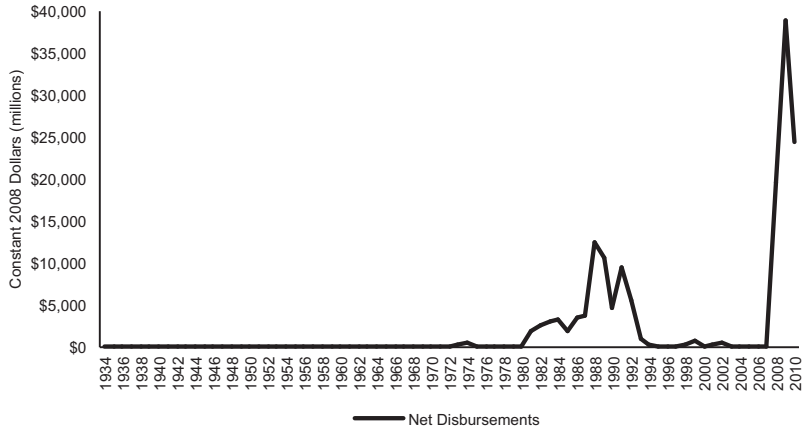


Source: Federal Deposit Insurance Corporation, *2010 Annual Report*.

[12] We include additional recoveries estimated by the FDIC but not yet realized. Years with nonzero estimated recoveries include 1991, 1992, 1998, 1999, 2001, 2002, 2003, 2007, 2008, 2009, and 2010.

In order to more clearly depict the difference between disbursements and recoveries over time, net disbursements by the FDIC between 1934 and 2010 are presented in Figure 6. Net disbursements indicate the amount disbursed to depositors of failed banks less recoveries from the sale of bank assets. Again, we consider the three periods of above average bank failures. Net disbursements averaged $44.65 million per year from 1935 to 1942, less than $1 million for each failed bank. From 1982 to 1994, net disbursements averaged $4,817.39 million per year—ranging from a high of $12,590.84 million in 1988 to a low of $260.12 million in 1994—with each bank failure producing an average net disbursement of $42.78 million. Net disbursements were significantly higher from 2008 to 2010. Roughly $27,689.01 million was disbursed each year on net—more than $257.97 million for each failed bank.[13]

FIGURE 6
Net Disbursements under FDIC, 1934–2010



Source: Federal Deposit Insurance Corporation, *2010 Annual Report*.

How much has the FDIC spent to cover the deposits of failed and assisted banks? Net disbursements by the FDIC over the history of the program exceed $152,000 million—averaging roughly $1,974 million per year. From 1934 to 2010, cumulative net disbursements grew at an average annual rate of 15.11 percent. Growth in

[13] Although net disbursements were largest in the recent financial crisis, the percentage of disbursements recovered was also higher in this period than in the S&L crisis.

cumulative net disbursements was just below trend from 1934 to 1980, averaging an annual rate of 14.80 percent. From 1980 to 1992 the average annual growth rate shot up to 30.91 percent, reflecting the wave of bank failures associated with the Savings and Loan Crisis. Cumulative net disbursements returned to trend between 1992 and 2006, averaging just 0.35 percent growth per year over the period. Then, as bank failures increased during the most recent financial crisis, cumulative net disbursements shot back up. From 2006 to 2010, cumulative net disbursements have grown at an annualized rate of 17.20 percent.

Government-provided deposit insurance removes the incentive for depositors to run on banks and thereby prevents a solvent bank from failing because an unnecessary run has rendered it illiquid. It does not prevent banks from failing for other reasons. The case of FDIC illustrates the point. From 1934 to 2010, 2,417 FDIC member banks failed. The commitment to insuring deposits at these banks resulted in substantial costs for the deposit insurance program.

One might be tempted to conclude that the costs arising from the failure of insolvent banks could be avoided by only promising to pay out in cases of bank failures caused by runs (i.e., when banks are solvent but illiquid). If deposit insurance removes the incentive for depositors to run on a bank, failures resulting in the presence of deposit insurance must stem from other factors. Hence, deposit insurance should never pay out in practice and, as in the DD model, deposit insurance becomes a costless solution. Unfortunately, this simple solution is ultimately ineffective, as it fails to take into account expectations involved in the dynamic game. If agents know in advance that deposit insurance will never pay out in practice, they can have little confidence that their deposits will be covered in the event of a run-induced failure. Although their belief that deposits will be covered in the event of a run would be sufficient to prevent such runs from occurring, they have no basis to hold such beliefs if they know deposit insurance never pays out in practice. In other words, the inability to distinguish run-induced failures from other types of failure cuts both ways: the guarantor cannot identify which depositors should be covered ex post and depositors cannot hold beliefs ex ante that the set of those ultimately denied coverage will be limited to those that should not have been covered. Indeed, paying out in cases of

known insolvency might be a costly signal necessary to induce depositors to hold the beliefs on which deposit insurance schemes rely. If this is the case, the costs associated with nonrun bank failures cannot be avoided under government-provided deposit insurance schemes.

Cost of Managing the Deposit Insurance Fund

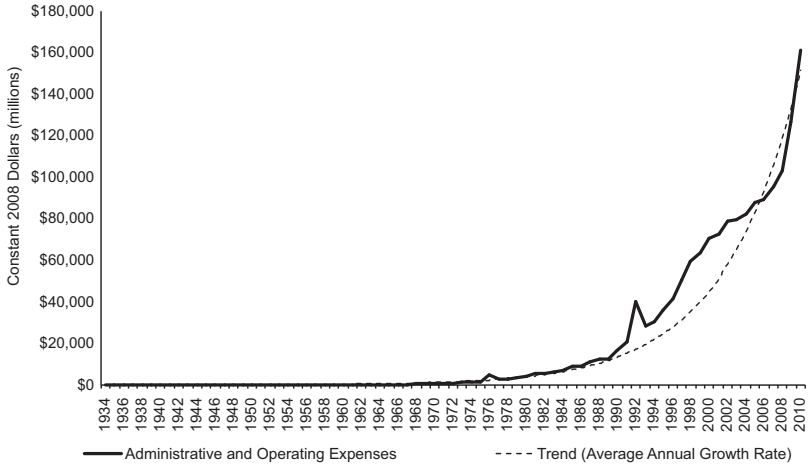
In order to cover the positive net disbursements arising from bank failures, the FDIC manages a deposit insurance fund primarily supported by fees assessed annually on member banks.[14] Managing the DIF is anything but costless. The FDIC has teams of administrators, auditors, and staff—a total of 8,150 employees in 2010. According to its annual report, the FDIC paid these employees over $1 billion in wages and salaries for the year. Furthermore, the organization owns $416 million in plant, property, and equipment with which these employees conduct the day-to-day operations of the FDIC. Although these costs are conspicuously absent in the DD model, they are essential for government-provided deposit insurance in practice.

Annual administrative and operating expenses from 1934 to 2010 are presented in Figure 7 with a solid black line.[15] Administrative and operating expenses have grown at an average annual rate of 10.23 percent since 1934. The average annual growth rate is depicted as a dashed black line in Figure 7. From 1934 to 1987, administrative and operating expenses grew roughly at trend, averaging $1,559 million each year. Administrative and operating expenses outpaced trend from 1988 to 1999, growing at an average annual rate of 13.71 percent and averaging $34,332 each year. Returning to trend in 2005, administrative and operating expenses have since grown at roughly 10.18 percent each year.

[14] The DIF was originally named the Federal Deposit Insurance Fund. In 1989, it was renamed the Bank Insurance Fund (BIF) in accordance with the Financial Institution Reform, Recovery, and Enforcement Act (FIRREA), which also created the Savings Association Investment Fund (SAIF) to be administered by the FDIC (FDIC 1998: 51). The BIF and SAIF were later merged into a singled fund, the DIF, according to the Federal Deposit Insurance Reform Act of 2005 (Pennacchi 2009: 6 n.12).

[15] Administrative and operating expenses do not include charges to failed bank receiverships managed by FDIC.

FIGURE 7
Annual Administrative and Operating Expenses, 1934–2010

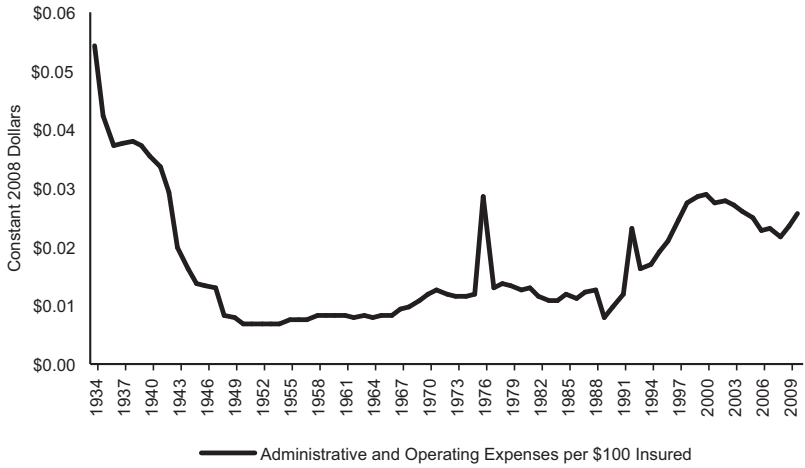


Source: Federal Deposit Insurance Corporation, *2010 Annual Report*.

The exponential increase in administrative and operating expenses corresponds to the increasing scope of the FDIC. Recall that coverage has increased significantly over the history of the program, both in terms of the number of depositors covered and the extent of coverage each depositor receives. As such, it is worth considering the cost of a constant coverage amount over time. Annual administrative and operating expenses per $100 insured between 1934 and 2010 are depicted by a solid grey line in Figure 8.

Constant coverage administrative and operating expenses fell from $0.054 per $100 insured in 1934 to $0.008 in 1948. It then leveled off, averaging $0.010 from 1949 to 1989, with the exception of 1978 when it spiked to $0.028. Constant coverage administrative and operating expenses climbed from $0.007 in 1950 to $0.029 in 2000, before falling back down to $0.023 in 2007. In 2010, administrative and operating expenses per $100 insured totaled $0.026. Over the entire period, administrative and operating expenses averaged $0.017 per $100 insured, with nearly three quarters of the observations falling between $0.007 and $0.027.

FIGURE 8
Annual Administrative and Operating Expenses per $100 Insured, 1934–2010



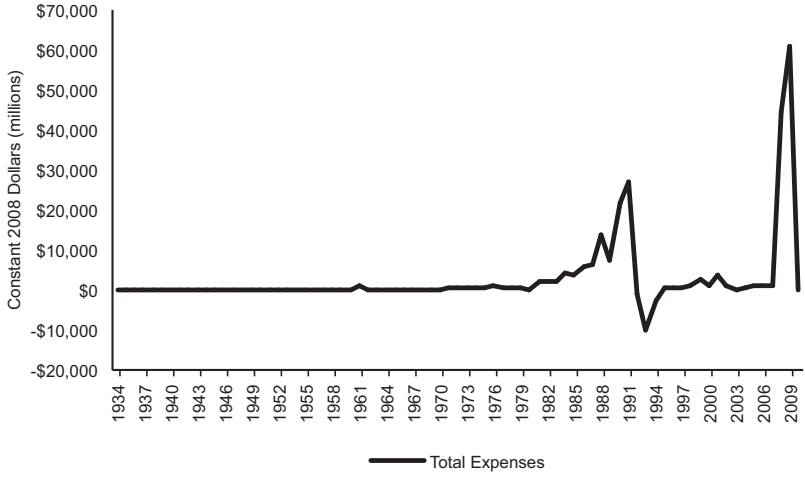
Source: Federal Deposit Insurance Corporation, *2010 Annual Report*.

Annual total expenses for the deposit insurance fund are presented in Figure 9. In addition to administrative and operating expenses, total expenses incudes provision for insurance losses and interest and other insurance expenses.[16] Annual expenses averaged $208 million between 1934 and 1980, ranging from $55 million in 1951 to $827 million in 1961. With the Savings and Loan Crisis underway, average total expenses increased to $5,515 million between 1981 and 1995. Variability also increased over the period, ranging from a high of $26,755 million in 1991 to a low of –$10,049 million in 1993.[17] Total expenses then receded, averaging only $1,166 million from 1996 to 2007. More recently, bank failures—which required enormous provisions for insurance losses—caused total expenses to increase markedly. Total expenses equaled $44,340 million in 2008 and $60,926 million in 2009.

[16] Funding transfers from the FSLIC resolution fund totaling $139.4 million are included in years 1989, 1990, 1991, and 1992.

[17] Negative total expenses are recorded when the DIF experiences a negative "provision for insurance losses," indicating that net disbursements throughout the year and into the future were and/or are expected to be lower than previously expected.

FIGURE 9
Annual Total Expenses for Deposit Insurance Fund, 1934–2010



Source: Federal Deposit Insurance Corporation, *2010 Annual Report*.

Cumulative total expenses have grown exponentially at an average annual rate of 9.33 percent. Prior to the Savings and Loan Crisis, the cost of the FDIC amounted to roughly $11,802 million. From 1982 to 1994, total expenses increased by an astounding $80,428 million—more than six times the cost of the FDIC from 1934 to 1981—bringing the total cost to $92,230 million. Another significant increase occurred between 2007 and 2010. Cumulative total expenses more than doubled over the period, increasing from $106,502 million in 2007 to $211,841 million in 2010.

Reviewing the history of the FDIC yields several stylized facts. First and foremost, it is quite clear that government-provided deposit insurance is costly. Indeed, administrative and operating expenses have increased over the history of the program. Part of the observed increase can be attributed to increases in coverage. More wealth is held in the banking system today than in 1934, and the FDIC protects a larger amount of each depositor's wealth. Even still, annual administrative and operating expenses per $100 insured were higher in the 1990s and 2000s than in the 1950s and 1960s. Adding to the increase in administrative and operating expenses are costs associated with bank failures. Provision for insurance losses in the Savings

and Loan Crisis from 1982 to 1994 and the more recent crisis from 2008 to 2010 have drastically increased the cost of deposit insurance. These stylized facts provide a starting point for comparing government-provided deposit insurance with potential alternatives.

Alternatives to Government-Provided Insurance

Our analysis demonstrates that government-provided deposit insurance departs markedly from the theoretical ideal conceptualized in the DD model. If alternatives were nonexistent or sufficiently costly, the shortcomings of government-provided deposit insurance in practice might be of little consequence. Therefore, it is useful to consider some potential alternatives. We briefly review private insurance, self-regulation, and other market mechanisms that might protect depositors from bank failures. Although a full comparative institutional analysis is beyond the scope of this article, the potential alternatives discussed below should provide the reader with a clearer view of the sort of comparisons we have in mind.

An obvious alternative to government-provided deposit insurance is privately provided deposit insurance. Such schemes were once common in the United States, arising naturally out of the private regional associations used to clear checks and banknotes. According to Calomiris (1990: 284–85), "Formal coinsurance arrangements among bank clearinghouse members, and less formal arrangements among other banks—especially in the branch-banking states of the antebellum South—provided many of the features of government deposit insurance."

Private and privately managed deposit insurance is not merely a thing of the past. Demirgüç-Kunt and Sobaci (2000) consider 109 countries with explicit deposit insurance programs.[18] Eleven of these countries—including Argentina, Austria, Brazil, Finland, France, Germany, Luxembourg, Norway, Switzerland, Tanzania, and the United Kingdom—have privately managed programs. Another 24 countries have deposit insurance programs that are managed jointly through some sort of public-private partnership. Only 33 countries have publicly managed deposit insurance programs like the United States.

[18] See also: Demirgüç-Kunt and Sobaci (2001), Demirgüç-Kunt, Karacaovali, and Laeven (2005).

In considering alternatives to government-provided deposit insurance, we might look to a broader set of institutions that stop bank runs and reduce the spread of panics. Clearinghouses would seem to fit the bill. Prior to the Federal Reserve, private banks used clearinghouses to coordinate activities. Clearinghouses worked primarily by allowing banks to coordinate activity when it was in the interest of the system. According to Gorton (1985: 277), "During financial panics, the clearing house united banks into an organization resembling a single firm." Acting in concert made it easier to suspend convertibility of banknotes into specie, which prevented bank runs from destroying otherwise solvent banks. As Gorton and Mullineaux (1987: 462) make clear, "The advantages of the [commercial-bank clearinghouse] organization were such that within a decade a large number of local clearinghouses were formed."

When held liable for losses, bank managers have an incentive to curtail excessive risk.[19] Historically, bank owners faced unlimited or double liability for any losses they incurred. Hence, owners had a strong incentive to monitor the bank's risk-taking activities and remove excessively risky managers. Macey and Miller (1992: 34) find that "empirical evidence substantiates the inference that double liability was an effective regulatory system" and "unlike deposit insurance, the threat of double liability appears to have induced caution on the part of bank managers in their use of depositors' funds."

Prior to FDIC deposit insurance, bank owners were more vigilant in monitoring risk. They sometimes required bank managers to post performance bonds to the equivalent to one or more years of their annual salaries, which would be forfeited in the case of the bank's failure (White 2011). Similar tools are once again being implemented today in the form of "clawback" clauses, which require bank managers to repay past bonuses. In recent cases, however, such clauses are generally triggered by ethics violations rather than performance alone (Attwood 2012, Hodgson 2012). We hope that regulators will consider these mechanisms of depositor protection as alternatives to government-provided deposit insurance.

[19] The incentives bank managers face are widely discussed in the economic literature on free banking. White (1984, 1989), Selgin (1988), and Dowd (1993) all discuss the manager's need to balance the marginal benefits and costs of the assets and liabilities under his control. Moreover, they provide evidence that such risk-management strategies have been employed successfully to reduce bank risk and enhance economic stability in Britain, Scotland, Australia, and the United States.

Conclusion

Much analysis of government-provided deposit insurance evaluates a costless means of protecting depositors. As the case of the FDIC illustrates, government deposit insurance is costly. Costs are incurred in assessing fees from banks, covering losses when banks fail, and monitoring the banking system. An understanding of the costs of government-provided deposit insurance in practice is essential for any discussion of potential alternatives.

In this article, we reviewed the performance of the FDIC and examined how the costs of federal deposit insurance have changed over time. We also used the FDIC experience as a benchmark against which potential alternatives might be compared. We found that deposit insurance in practice departs markedly from the ideal, costless system espoused by Diamond and Dybvig (1983). As such, a meaningful analysis of deposit insurance must incorporate the costs of government-provided insurance.

References

Andolatto, D.; Nosal E.; and Wallace, N. (2007) "The Role of Independence in the Green-Lin Diamond-Dybvig Model." *Journal of Economic Theory* 137: 709–15.

Attwood, M. (2012) "Banks Take Up 'Clawback' Schemes." *Financial News* (24 August). Available at: www.efinancialnews.com/story/2012-08-24/mercer-banks-pay-clawback-schemes.

Bradley, C. M. (2000) "A Historical Perspective on Deposit Insurance Coverage." *FDIC Banking Review* 13: 1–25.

Calomiris, C. W. (1990). "Is Deposit Insurance Necessary? A Historical Perspective." *Journal of Economic History*, 50 (2): 283–95.

Calomiris, C. W., and White, E. N. (1994) "The Origins of Federal Deposit Insurance." In C. Goldin and D. Libecap (eds.) *The Regulated Economy: A Historical Approach to Political Economy*, 145–88. Chicago: University of Chicago Press.

Chari, V. V. (1989) "Banking without Deposit Insurance or Bank Panics: Lessons from a Model of the U.S. National Banking System." Federal Reserve Bank of Minneapolis *Quarterly Review* (Summer 1989): 3–19.

Demirgüç-Kunt, A.; Karacaovali, B.; and Laeven, L. (2005) "Deposit Insurance around the World: A Comprehensive Database." World Bank Policy Research Working Paper No. 3628.

Demirgüç-Kunt, A., and Sobaci, T. (2000). "Deposit Insurance around the World: A Data Base." Washington: The World Bank. Available at http://siteresources.worldbank.org/INTFR/Resources/475459-1108066643741/data.pdf.

_____ (2001) "A New Development Database: Deposit Insurance around the World." *World Bank Economic Review* 15 (3): 481–90.

Diamond, D. W. (2007) "Banks and Liquidity Creation: A Simple Exposition of the Diamond-Dybvig Model." Federal Reserve Bank of Richmond *Economic Quarterly* 93: 189–200.

Diamond, D. W., and Dybvig, P. H. (1983) "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 91: 401–19.

Dowd, K. (1988) "Option Clauses and the Stability of a Laissez-faire Monetary System." *Journal of Financial Services Research* 1: 319–33.

_____ (1993) *Laissez-faire Banking*. New York: Routledge.

Federal Deposit Insurance Corporation (1984) *The First Fifty Years: A History of the FDIC 1993–1983.* Washington: FDIC.

_____ (1997) *History of the Eighties: Lessons for the Future.* Washington: FDIC.

_____ (1998) *A Brief History of Deposit Insurance in the United States.* Washington: FDIC.

_____ (2010) *2010 Annual Report*. Available at www.fdic.gov/about/strategic/report/2010annualreport/index_pdf.html.

Gilbert, R. A. (1990) "Market Discipline of Bank Risk: Theory and Practice." Federal Reserve Bank of St. Louis *Review* 72: 3–18.

Green, E. J., and Lin, P. (2000) "Diamond and Dybvig's Classic Theory of Financial Intermediation: What's Missing?" Federal Reserve Bank of Minneapolis *Quarterly Review* (Winter 2000): 3–13.

Gorton, G. (1985) "Clearinghouses and the Origin of Central Banking in the United States." *Journal of Economic History* 45: 277–83.

Gorton, G., and Mullineaux, D. J. (1987) "The Joint Production of Confidence: Endogenous Regulation and Nineteenth Century Commercial Bank Clearing Houses." *Journal of Money, Credit and Banking* 19: 457–68.

Hodgson, P. (2012) "It Is Time for Real Bank Clawbacks." *Forbes* (15 November). Available at www.forbes.com/sites/paulhodgson/2012/11/15/it-is-time-for-real-bank-clawbacks.

Kaufman, G. G. (2002) "FDIC Reform: Don't Put Tax Payers Back at Risk." Cato Institute Policy Analysis No. 432.

Kroszner, R. S., and Melick W. R. (2008) "Lessons from the U.S. Experience with Deposit Insurance." In A. Demirguc-Kunt, E. Kane, and L. Leaven (eds.) *Deposit Insurance around the World: Issues of Design and Implementation*, 188–217. Cambridge, Mass.: MIT Press.

Macey, J. R., and Miller, G. P. (1992) "Double Liability of Bank Shareholders: History and Implications." *Wake Forest Law Review* 27: 31–62.

McCulloch, J. H., and Yu, M. (1998) "Government Deposit Insurance and the Diamond-Dybvig Model." *Geneva Papers on Risk and Insurance Theory* 23: 139–49.

Peck, J., and Shell, K. (2003) "Equilibrium Bank Runs." *Journal of Political Economy* 111: 103–23.

Pennacchi, G.G. (2009) "Deposit Insurance." Paper prepared for AEI Conference on Private Markets and Public Insurance Programs (August). Available at cbpp.illinois.edu/pdf/research.

Rolnick, A. J. (2010) "Interview with Thomas Sargent." Federal Reserve Bank of Minneapolis *The Region* (September): 26–39.

Samartin, M. (2003) "Should Bank Runs Be Prevented?" *Journal of Banking and Finance* 27: 977–1000.

Sebastian, S., and Tyrell, M. (2006) "Open End Real Estate Funds: Diamond or Danger?" SSRN Working Paper (March). Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=893121.

Selgin, G. A. (1988) *The Theory of Free Banking: Money Supply under Competitive Note Issue*. Lanham, Md.: Rowman and Littlefield.

_____ (1993) "In Defense of Bank Suspension." *Journal of Financial Services Research* 7: 347–64.

Thomson, J. B. (1987) "The Use of Market Information in Pricing Deposit Insurance." *Journal of Money, Credit and Banking* 19: 528–37.

Van den Heuvel, S. J. (2008) "The Welfare Cost of Bank Capital Requirements." *Journal of Monetary Economics* 55: 289–320.

Wallace, N. (1988) "Another Attempt to Explain an Illiquid Banking System: The Diamond-Dybvig Model with Sequential Service Constraint Taken Seriously." Federal Reserve Bank of Minneapolis *Quarterly Review* (Fall 1988): 3–16.

_____ (1990) "A Banking Model in Which Partial Suspension Is Best." Federal Reserve Bank of Minneapolis *Quarterly Review* (Fall 1990): 11–23.

White, E. N. (2011) "Rethinking the Regulation of Banking: Choices or Incentives?" Conference paper prepared for the Witherspoon Institute (5–6 December).

White, L. H. (1984) *Free Banking in Britain*. Cambridge: Cambridge University Press.

_____ (1989) *Competition and Currency*. New York: New York University Press.

_____ (1999) *The Theory of Monetary Institutions.* Malden, Mass.: Blackwell.

Pathological Altruism and Pathological Regulation

Paul H. Rubin

A concept recently developed by scholars in psychology and biology is "pathological altruism." (Oakley 2013, Oakley et al. 2012). A pathological altruist is defined as "a person who sincerely engages in what he or she intends to be altruistic acts, but who harms the very person or group he or she is trying to help, often in unanticipated fashion; or harms others; or irrationally becomes a victim of his or her own altruistic actions." (Oakley, Knafo, and McGrath 2012: 4). We may relate this concept to Buchanan's Samaritan's dilemma: Buchanan's Samaritan is the altruist, and the pathology is that the recipient will be in the "no work" cell, so that the Samaritan becomes a victim of his own altruistic actions (Buchanan 1975).

So far as I know, this concept has not yet been used in economics. However, it can become an extremely useful tool for economists, and particularly public choice economists. This is because there are many public policies that are harmful in one or more of the senses above, and yet which are supported by well-meaning citizens and voters. The notion of pathological altruism is by no means a complete explanation for undesirable special interest legislation or regulation. However, it can be a valuable additional tool in the economist's toolkit for explaining such legislation. This is because if an interest group can harness the altruistic beliefs of voters, it will be more successful in obtaining the benefits it desires. Because of voters' limited

---

Cato Journal, Vol. 34, No. 1 (Winter 2014). Copyright © Cato Institute. All rights reserved.

Paul H. Rubin is Samuel Candler Dobbs Professor of Economics at Emory University.

attention and understanding, the actual act may be pathological in the sense defined above, and yet still appeal to voters. Voters are likely to pay attention to motives, not results, if there is a plausible connection between the policy and the goal.

Psychologists have developed many sources of pathological altruism and used it to explain many behaviors. Many of these are not directly relevant for economic analysis. However, one source that has been identified is simply error: "When people who feel empathy at witnessing another's misfortunes falsely believe that they caused the other's problems, or falsely believe that they have the means of to relieve the person of suffering, they have erred in their analysis of the situation" (O'Connor et al. 2012: 11). O'Connor et al. also relate this notion to guilt about inequality, where those with more wealth feel guilty with respect to those who have less (p. 16). They explain this guilt in evolutionary terms, arguing that in the evolutionary environment equality was an important social value. This is similar to arguments I have made elsewhere (Rubin 2002, 2003).

While Oakley et al. (2012) present many examples of pathological altruism for individuals in their day-to-day lives, one would expect the notion to be highly relevant for policy analysis. This is because of the standard notion of "rational ignorance." If a policy can present a plausible altruistic justification, it generally does not pay for voters to further examine this basis. Moreover, policies are extremely difficult to analyze and even if voters desired to determine their effects, they would have a good deal of difficulty doing so. Knowledge of the effects of policies is not direct, but must be teased out of the data using statistical or econometric tools, and even then there is often disagreement among experts about the effects of policies. This disagreement is fueled by the incentives of participants in political debates to find or fund experts who will espouse their views. As a result, it would not be surprising if voters erroneously support counterproductive or pathologically altruistic policies.

The point is that policies need not actually benefit the purported beneficiaries. As long as a convincing story can be told about beneficiaries, the political process may adopt the policies. The notion of pathological altruism can be added to the public choice economist's standard notion of rational ignorance to create a powerful new tool for analysis.

There is another potential benefit from applying the notion of pathological altruism to economic behavior. Scholars in fields such as

psychology and perhaps biology are generally more favorable disposed to regulation than are many economists. Many psychologists and others seem to already accept the notion of pathological altruism. One of the key papers was published in *Science* (Oakley 2013). If we economists can relate the notion of inefficient regulation to the notion of pathological altruism, we may be able to enlist some of these scholars in efforts to reform inefficient or undesirable regulation. Moreover, if we economists can point out that many regulations are a result of pathological altruistic processes, we may be able to gain allies in our attempts to eliminate inefficient regulations.

The Supply Side of Regulation

Since at least the important paper by Stigler (1971), economists studying regulation have focused on the demand side of regulation—that is, which industries have the political power to obtain the benefits of favorable regulation from the state. However, there is also a supply side. Certain forces make it easier for some industries to obtain regulation. In particular, there are characteristics of some industries that make it easier for the political process to regulate them. Understanding those characteristics will shed some light on the form of regulation for particular industries, and may also help suggest policies which can reduce inefficient regulation.

There are three fundamental political justifications for regulation: efficiency enhancement, altruism, and fear. Economists, of course, emphasize efficiency. However, efficiency does not have a large political constituency; most voters do not think in economic or efficiency terms (Rubin 2003, Caplan 2007). Indeed, even when there are efficiency justifications for some regulation (e.g., using antitrust to eliminate deadweight losses), political explanations are more likely to be in terms of altruistic redistribution—lowering prices for consumers (altruism), or taking away monopolists ill-gotten gains (fear). Since Stigler's article economists have tended to treat justifications for regulation as a detail, and to concentrate on the power of various interest groups to achieve their goals. In this article, I argue that those altruistic justifications are more than a detail and can help us understand regulation, and perhaps improve it.

Virtually all special interest legislation is justified as some form of altruism towards some party. Humans are naturally altruistic in some circumstances (Rubin 2002). We are altruistic toward members of

our own group as a result of evolutionary pressures for kin selection. We are altruistic toward those who are less fortunate as a result of outside factors (but not because of shirking). We are also altruistic in circumstances where some assistance can help a person become productive again.

In some circumstances, humans are also prone to fear of dominants (Rubin 2003). If some party can be demonized as someone who wants to take unfair advantage of us, then it is easier to justify regulation. This is facilitated by our natural zero-sum thinking, so that profits made by the demonized industry or group can be viewed as having come from exploited consumers, rather than having been earned by increased productivity (which is not well understood by untrained people). The evolutionary environment was one of forced equality, with would-be dominants being punished (Rubin 2003, Boehm 2012).

It is necessary to tell some story about deservingness or about the danger of exploitation that makes the proposed regulation fit into one of these natural categories. However, we have very little intuitive understanding of efficiency (Rubin 2003). Politically it is much easier to justify some intervention if it can be cast in terms of benefits for some well-defined deserving group, or as protection from some other group, rather than as reducing deadweight loss or increasing efficiency.

Beneficiaries of Altruism

There are three basic types of altruism in relation to interest group legislation: (1) altruism toward members of the interest group, (2) altruism toward consumers of the product supplied by the interest group, and (3) altruism toward seemingly unrelated persons. I consider each.

Altruism toward Members of the Interest Group

Altruism may be aimed at members of the interest group if a case can be made that members are deserving of sympathy. As an example of altruism directed at suppliers, consider agriculture. The farm support program is unabashedly aimed at benefitting farmers. When the programs began, it was possible to speak about the family farm and the value of the virtues generated by family farms. "Supporters of farm subsidies have argued that such programs stabilize agricultural commodity markets, aid low-income farmers, raise unduly low

returns to farm investments, aid rural development, compensate for monopoly in farm input supply and farm marketing industries, help ensure national food security, offset farm subsidies provided by other countries, and provide various other services" (Summer 2008). The beneficiaries of those policies are then visualized as farmers who through no fault of their own are experiencing low incomes and therefore need assistance. Moreover, sometimes middlemen are blamed for low farm incomes, thus making use of the anti-exploitation branch of regulatory justifications.

Tariffs are also often presented as benefiting workers in the protected industry. Humans have no good intuitive understanding of the productivity of capital, as shown by the appeal of Marxism, so capital owners are not a good target for altruism, but displaced workers are, and this is often a selling point for protectionist measures. This altruism is strengthened when it is argued that a tariff is only needed for a short time, until the domestic industry can "catch up." Moreover, in the case of tariffs and other import reducing activities, the natural tendencies of humans to aim altruism at members of their own group can be exploited to further advance the policies. This is also true of other policies, as when farmers are benefited because of alleged "unfair" foreign competition.

Consumers

Claims for consumer benefits are probably the most common altruistic justification for special interest legislation. Occupational licensing may be the most common form of protectionist legislation, and this is invariably justified as protecting consumers. Kleiner and Krueger (2009) show that licensing increases wages by about 14 percent, and that occupational licensing has been increasing, going from less than 5 percent of the U.S. labor force in the 1950s to 29 percent in 2008. Economic factors can explain much of the variation in which professions are licensed (Stigler 1971). Stigler's factors explain the demand side of special interest regulation.

The ability to convince voters that protection is beneficial to consumers is a supply-side factor. The argument is that consumers will be harmed by incompetent (unlicensed) practitioners of some profession. Harmed or potentially harmed consumers are a natural target for altruistic beliefs. The more easily harms can be perceived by consumers and the more easily these harms can be described, the greater the willingness of consumers as voters to allow such regulation. For

example, much regulation is justified in terms of health and safety. Consumers can easily perceive the harms from incompetent medical practitioners or harmful drugs and so are willing and perhaps eager for such regulation. Much regulation is the result of some systemic malfunction, as when the FDA was given increased power as a result of the thalidomide disaster. The Type 2 error, the harm from lack of practitioners, is more difficult to explain or perceive. This may be because of consumers' zero-sum bias; they may not perceive that the number of practitioners can be influenced by regulatory variables.

Health and safety regulation can be justified as protecting consumers from careless or incompetent practitioners, such as physicians. However, regulation is sometimes justified as protecting consumers from those who would exploit them. A major theme in health regulation is that pharmaceutical companies are unscrupulous and would gleefully sell harmful drugs if only they were allowed to. Similar justifications are commonly used for financial regulation, where it is argued that financial firms or certain classes of lenders (e.g., "predatory" lenders) would harm consumers if they could, and only the government can protect consumers from this harm. This form of justification plays in part on humans' zero-sum thinking (Rubin 2003)—if drug firms or financial firms are making money, then it must be harming consumers. It also explains why politicians planning to regulate some industry will first spend resources criticizing this industry.

Seemingly Unrelated Third Parties

Finally, much special interest regulation is justified as benefitting seemingly unrelated parties. One form of such regulation that is well known is Bruce Yandle's (1983) famous "Bootleggers and Baptists" example, where sellers of illegal liquor form an informal alliance with religious persons who want to maintain the illegality of liquor sales on Sundays or with ministers who want to maintain their audiences and members.

Dudley and Brito (2012: 19) provide many examples of this phenomenon:

> Modern-day stories of bootleggers and Baptists abound. Large biotechnology companies join with food safety activists to encourage stricter regulation of new foods involving genetic engineering, thus putting smaller competitors who cannot

> afford the regulatory compliance costs at a disadvantage. Tobacco companies supported legislation that would have required cigarettes to receive FDA premarketing approval, which would make it harder for new brands to enter the market. Solar power manufacturers support regulation that inhibits the production of conventional, competing sources of power (oil, coal, and gas). Food and toy companies lobby for more regulation to ensure their products' safety, thereby keeping out foreign competitors that may not be able to demonstrate that their products meet the same standards. U.S. testing laboratories argue on safety grounds against European requests to permit manufacturers of low-risk workplace electrical products to self-certify compliance with regulations rather than subject them to third-party testing. Big box retailers with vast resources are the largest supporters of minimum wage laws that raise the costs of doing business for their mom-and-pop competitors.

Of course, unions are also supporters of minimum wages which make the cost of competitors (nonunion workers) more expensive. Traditionally, minimum wages have also had a geographic component, as northern high-wage workers wanted to disadvantage southern low-wage competitors. The justification, providing higher incomes to poor persons, is a natural form of altruism. An added justification is protecting workers from greedy employers who would exploit them by paying low wages.

In addition to minimum wages, there are other examples where interest groups sell policies as benefitting the poor. Food stamps are viewed as an antipoverty program, but food stamps must be used for American food and are administered by the Department of Agriculture—traditionally the source of benefits to agricultural interests, not to lower-income people. Various builders associations are great supporters of public housing.

Various environmental and green movements have spawned many "bootlegger-Baptist" alliances, in addition to those mentioned above. European food interests promulgated the label of "Franken foods" and lobbied for bans on genetically modified foods that competed with their more primitive technologies. Diaper delivery services claimed that disposable diapers clogged landfills. Some conservatives denigrate the notion of global warming because they believe it is a liberal ploy to generally increase government control of the economy. Of course, advocates of the dangers of global warming in conjunction

with producers of alternative energy sources demonize traditional energy producing firms. Municipalities impose bottle fees, ostensibly to reduce landfills usage, but also to increase revenues. Ackerman and Hassler (1981) documented the alliance between environmentalists and producers of high-sulfur coal. Ethanol was originally sold as an environmental benefit but actually benefits corn producers. In honor of ethanol, probably the most expensive such policy, I will call these "Green-Corn" alliances.

These are all examples of what I have called altruistic benefits for seemingly unrelated third parties. In many cases the third party does not actually gain, and most of these examples are socially inefficient. Therefore, these are all examples of pathological altruism.

Fear of Dominants

In addition to altruism, humans also have less desirable motivations. In particular, we are sometimes driven by fear of outsiders, or of excessive dominance by elites. Both of these emotions are based on our evolutionary history, in which there was more conflict than we like to admit. In some circumstances, voters can be made to fear some interest group, and this can be a justification for regulation. In recent times, President Obama has demonized the financial services industry in obtaining additional regulation of this industry. He has also recently attacked universities, and advocated pathologically altruistic policies that will lead to further increases in tuition.

Form of Benefits

If some special interest benefit is justified through an appeal to altruism, then this appeal determines the form of the benefit. Thus, the political justification for regulation is important in part because this justification determines the form of benefits. This explains why few benefits are in the form of direct cash transfers.

A major exception is agriculture. Here, the claim is that farmers' incomes are too low, and so a direct transfer of income through increasing the prices of agricultural goods and buying up of surpluses is possible. To the extent that low-wage workers are the ostensible beneficiary of minimum wage laws, a direct increase in wages is also a feasible policy. However, note that in both cases the transfers are paid for by consumers who are forced to pay higher prices, not directly by taxation.

Regulation aimed at protecting consumers can take two forms, depending on the underlying purpose. If the purpose is to "protect" consumers from incompetent practitioners, then simple entry limitation can be successful. This may be in the form of testing or of educational requirements (or both) for entry into the industry. Once these are in place, there is little need for further regulation, and practitioner groups are sometimes criticized for inadequate policing of incompetent members who have gotten through the licensing process.

If the justification is to protect consumers from "exploitive" practitioners, then the process is much more complex. This explains why agencies such as the FDA and the SEC have developed large bureaucracies and complex sets of rules. Currently regulations are being written under the Dodd-Frank Act aimed at protecting consumers from evil financial firms; this followed a political effort to demonize Wall Street.

Of course, once these rules are in place, then members of the enforcing bureaucracy themselves become an interest group with an interest in continuing and increasing regulation (Niskanen 1971). Because these bureaucrats are the only people who understand the regulations, they are in a powerful position to lobby for them and explain why they are needed. Moreover, since they understand the workings of the bureaucracy and the rules enforced by the agencies they are in a position to leave the agency and make money in the private sector, helping firms deal with the regulatory thicket. (Disclosure: At one point in my career, after I left the FTC, I went to work for a consulting firm whose main business was dealing with the FTC regulatory process.) These former bureaucrats (lawyers, economists, and others) who assist firms in dealing with the regulatory agencies are then themselves another special interest group with an interest in maintaining the regulatory regime.

The most interesting cases are the seemingly unrelated third party cases. Since there are many forms of such third parties, there are many types of benefits generated. One form of benefit is to use the power of the government to increase the direct cost of a rival's product, as do minimum wage laws. More common is using the power of the government to harass some rival, as by excessive regulation or an outright ban on the competing products (Salop and Scheffman 1983). This applies to many Green-Corn interferences, such as the Franken foods and diaper examples. Another example is mandated

purchases of some product (e.g., ethanol as a gasoline supplement, ostensibly for environmental reasons but actually to benefit corn growers). Regulation itself can be a method of increasing rivals' costs. Large firms have a comparative advantage in dealing with regulation since it imposes a fixed cost and thus disproportionately burdens smaller firms. For example, many large firms are advocates of increased privacy regulation, in part because this will make entry into Internet businesses more difficult (Rubin and Lenard 2011).

Policy Implications

If we realize that the altruistic justifications for regulation are an important component of the regulation itself and that much of this altruism is pathological, then there are possible arguments that can be made to reduce inefficient regulation. While cost-benefit analysis is an important tool for economists analyzing regulation, analysis of the political justifications for regulation may also be useful. If the goal of regulation is something other than efficiency, then showing that the regulation is inefficient may be politically irrelevant. If we want to criticize regulation, then it might be more effective to criticize it in terms of the stated purpose of the regulation. I consider each type.

Members of the Interest Group

If some regulation is specifically aimed at increasing incomes of the members of the group itself, then showing that it is inefficient is irrelevant. There are, however, arguments that can still be made. One is that members of the benefitted group are not particularly deserving. For example, showing that farmers who benefit from price supports are large corporate farms, rather than small family farms, is a useful (although not necessarily successful) tactic. Similarly, showing that a program harms another deserving group (e.g., families with children who consume milk) may be a useful tactic. It may also pay to borrow a trick from the pro-regulatory toolbox and demonize beneficiaries. "Big farms want to profit by making your children's milk more expensive." Of course, this creates the risk of demonization which could cause an overreaction and could lead to equally inefficient regulation on the other side. Another alternative is to calculate the cost to society of the benefits. Thus, it may be useful to calculate the cost per job saved of a tariff. For example, it has been calculated that a tariff on Chinese tires cost $900,000 per U.S. job saved

(Hufbauer and Lowery 2012). If numbers such of these can be widely promulgated and understood, there may be some pressure to reduce regulation.

Consumers

Politically, it is difficult to argue against regulation aimed at protecting consumers, even if it is pathological and inefficient. Consumers place a high value on safety and are not willing to listen to arguments that a particular policy does not really add to safety. It may be helpful to show that policies are actually counterproductive. For example, Peltzman (1973, 1975) showed that drug safety regulation and auto safety regulation actually increased risk, and Rubin and Shepherd (2007) have shown that modern American tort law actually leads to increased deaths. Nonetheless, we still have inefficient drug and auto regulation, indicating that this approach may not be successful. However, modern tort law has been curtailed, in part perhaps by research showing its costs and inefficiencies. It may also be possible to show that there are better (i.e., more efficient) alternatives for achieving a particular goal. Thus, economists argue that a carbon tax is a more efficient way of reducing carbon consumption than the patchwork of laws and regulations now on the books.

Seemingly Unrelated Third Parties

This class of regulations will be particularly difficult to attack because the third parties are often politically powerful. For example, no one wants to attack Yandle's Baptist ministers. In today's world, attacking Greens will not get one very far. Thus, the politics of limiting inefficient Baptist-bootlegger or Green-Corn regulation will be delicate.

Perhaps the most politically relevant way of reducing this form of regulation is to point out the connections between the Baptists and the bootleggers. The secret to successful regulation of this sort is to hide the links. In Yandle's original example, the role of the bootleggers is hidden; the Baptists are the front men. Similarly, the corn producers are behind the scenes, and the environmentalists are visible. If citizens can be made aware of the behind-the-scenes manipulation, then there may be some possibility of defeating the regulation.

Conclusion

Scholars in psychology and biology have recently begun to understand that many policies sold as being altruistic are in fact harmful and "pathological." In this article, I extend the notion of pathological altruism to economic regulation, and show that this concept can be applied to much inefficient regulation. This approach may be useful for several reasons, both theoretical and in terms of policy. First, it will help us understand some of the bases for regulation. Second, it may enable economists to enlist assistance from other scholars in attacking undesirable legislation. Third, by exploring the inefficient justifications for regulation, it may be possible to create politically useful counterarguments.

References

Ackerman, B., and Hassler, W. T. (1981) *Clean Coal/Dirty Air: or How the Clean Air Act Became a Multibillion-Dollar Bail-Out for High-Sulfur Coal Producers.* New Haven: Yale University Press.

Boehm, C. (2012) *Moral Origins: The Evolution of Virtue, Altruism, and Shame.* New York: Basic Books.

Buchanan, J. M. (1975) "The Samaritan's Dilemma." In E. Phelps (ed.) *Altruism, Morality and Economic Theory.* New York: Russell Sage Foundation.

Caplan, B. (2007) *The Myth of the Rational Voter: Why Democracies Choose Bad Policies.* Princeton, N.J.: Princeton University Press.

Dudley, S. E., and Brito, J. (2012) *Regulation: A Primer.* 2nd ed. Mercatus Center, George Mason University, and Regulatory Studies Center, George Washington University.

Hufbauer, G. C., and Lowery, S. (2012) "U.S. Tariffs: Saving Few Jobs at High Cost." Pearson Institute for International Economics, Policy Brief 12–9. Available at www.iie.com/publications/interstitial.cfm?ResearchID=2095.

Kleiner, M. M., and Krueger, A. B. (2009) "Analyzing the Extent and Influence of Occupational Licensing on the Labor Market." NBER Working Paper No. 14979.

Niskanen, W. A. (1971) *Bureaucracy and Representative Government.* Chicago: Aldine.

Oakley, B. A. (2013) "Concepts and Implications of Altruism Bias and Pathological Altruism." *Proceedings of the National Academy of Science* (June).

Oakley, B. A.; Knafo, A.; Madhavan, G.; and Wilson, D. S., eds. (2012) *Pathological Altruism*. New York: Oxford University Press.

Oakley, B. A.; Knafo, A.; and McGrath, M. (2012) "Pathological Altruism: An Introduction." In Oakley et al. (2012: chap. 1).

O'Connor, L. E.; Berry, J. V.; Lewis, T. B.; and Silver, D. J. (2012) "Empathy Based Pathogenic Guilt, Pathological Altruism, and Psychopathology." In Oakley et al. (2012: chap. 2).

Peltzman, S.(1973) "An Evaluation of Consumer Protection Legislation: The 1962 Drug Amendments." *Journal of Political Economy* 81 (5): 1049–91.

_____ (1975) "The Effects of Automobile Safety Regulation." *Journal of Political Economy* 83 (4): 677–726.

Rubin, P. H. (2002) *Darwinian Politics: The Evolutionary Origin of Freedom.* New Brunswick, N.J.: Rutgers University Press.

_____ (2003) "Folk Economics." *Southern Economic Journal* 70 (1): 157–71.

Rubin, P. H., and Lenard, T. M. (2011) "The Federal Trade Commission Penalizes Google for Being Successful." *Forbes* (30 June).

Rubin, P. H., and Shepherd, J. (2007) "Tort Reform and Accidental Deaths." *Journal of Law and Economics* 50 (2): 221–38.

Salop, S. C., and Scheffman, D. T. (1983) "Raising Rivals' Costs." *American Economic Review* 73 (2): 267–71.

Stigler, G. (1971) "The Theory of Economic Regulation." *Bell Journal of Economics* 2 (1): 3–21.

Sumner, D. A. (2008) "Agricultural Subsidy Programs." *The Concise Encyclopedia of Economics.* Library of Economics and Liberty (24 August). Available at http://econlib.org/library/Enc/AgriculturalSubsidyPrograms.html.

Yandle, B. (1983) "Bootleggers and Baptists: The Education of a Regulatory Economist." *Regulation*7 (3): 12–16.

Book Reviews

Free Market Revolution: How Ayn Rand's Ideas Can End Big Government
Yaron Brook and Don Watkins
New York: Palgrave Macmillan, 2013, 251 pp.

In *Free Market Revolution: How Ayn Rand's Ideas Can End Big Government*, Yaron Brook, executive director of the Ayn Rand Institute (ARI), and Don Watkins, a fellow at ARI, give a full-throated and spirited defense of Rand's arguments for freedom, self-actualization, and the just society. The book is a clear explanation of objectivism that weaves in timely and accurate policy discussions, such as the chapter on health care, that buttress the overall point.

Watkins and Brook are as uncompromising as Rand. "Only Rand's morality of rational selfishness," they write, "can resolve the contradiction at the root of the founding and provide the idealism, the consistency, and the intellectual clarity necessary to end Big Government." Such directness is often needed when defending free markets, but it is not always the best strategy. In this case, that uncompromising nature condemns the book to be praised by those who already agree and derided by those who do not.

Free-market advocates often say that we're on a precipice—a choice looms before us: totalitarianism or freedom. *Free Market Revolution* effectively argues that it is time for a radical reassessment of the American character. Between looming debts, fiscal instability, and a brave new world of centralized health care, between the Tea Party and Occupy Wall Street, there is something in the air.

Cato Journal, Vol. 34, No. 1 (Winter 2014). Copyright © Cato Institute. All rights reserved.

For Brook and Watkins, the only answer to our problems, as is clear from the title, is the ideas of Ayn Rand. While free-market luminaries such as F. A. Hayek and Milton Friedman have ostensibly pushed for a freer society, the authors argue that they have failed to create a moral justification for capitalism. As such, "today's alleged free market champions are powerless to stop Big Government." "To do that," they say, "you need Ayn Rand."

Full disclosure: I am not an objectivist. I have considered myself a libertarian since I was a teenager, but, unlike most teenage freedom-lovers, I did not find my way to Rand's books until my mid-20s. By that time I had read most of the libertarian canon, and I did not need to be converted. When I first read Rand I was struck by her uncompromising positions, and I admired her Aristotelian/Nietzschean approach to the philosophy of freedom. I also admired her understanding that freedom needs a moral defense, even if I believe her defense to be lacking. I see her approach as a unique way to appeal to some people and one of the many ways that libertarians can advocate freedom.

But many objectivists disagree with me, and it seems Brook and Watkins are two of them. To them, Rand's way is not just one way to advocate for freedom, it is the only way. They are exasperated by lukewarm defenses of freedom, apologies for "selfishness," and defenses of "self-sacrifice" that they see as being little better than full-blown collectivism.

In constructing their case for full-blown and unapologetic capitalism, Brook and Watkins effectively argue that profit is not evil, that businessmen are not parasites, and that wealth does not simply pop into existence; wealth is created and earned. Rand herself was at her best when defending these truths, which are fundamental to a well-grounded defense of capitalism.

Brook and Watkins are less effective, however, in defending against—indeed, even addressing—arguments against capitalism that are more common and, to non-objectivists, more convincing. Most people do not embrace the epistemological and metaphysical framework erected by Rand and endorsed by Brook and Watkins. Most people do not see a direct connection between "A=A" and a robust defense of free markets. This is true whether or not objectivist philosophy is correct.

Unfortunately, many objectivists regard non-objectivists as being confused and having cluttered minds rather than having cogent objections to the philosophy. Some of this probably comes from

Rand's tendency to be dismissive of those who disagreed with her—see, for example, her flippantly brushing aside Ludwig Wittgenstein's seminal work on concept formation as being the product of a "mind out of focus." Thankfully, Brook and Watkins avoid this singularly unattractive objectivist tendency.

Yet, like many objectivists, Brook and Watkins still describe the world in terms of diametrically opposed intellectual camps, those who are with us and those who are against us. In reality, our opponents are far more marginal. Most people in the Western world believe in capitalism. They acknowledge that wealth is created and earned, that businessmen can be admirable, and that free markets have done incredible things for human flourishing.

But for most people the idea that profits are an unalloyed virtue is a strange one, if not preposterous. Brook and Watkins rightly criticize the negative view of profits as misguided and lament its widespread acceptance. Hollywood, the mainstream media, and popular culture in general treat the word "profit" as synonymous with "morally problematic," if not downright evil.

Yet Brook and Watkins do not adequately address the concerns of those who are not so much "anti-business" but rather not 100 percent "pro-business." Such people believe in profits but not at the expense of the environment, workers' rights, the well-being of the poor, and human flourishing. They do not believe that businessmen are categorically evil, rather they believe that the negative tendencies of business should be countered by positive government regulations. Similarly, they believe that the negative tendencies of government should be countered by positive effects of businesses. Such mixed pro-business/pro-government people may be wrong about the facts, but they are not categorically putting their foot down on either the government or business side.

Nevertheless, Brook and Watkins, like Rand, treat the world as if it were neatly divided into the "altruists" (the arch enemies of objectivists) and the "selfish" (the objectivist heroes). It's the black hats versus the white hats, and the only relevant question is which side you are on.

If only it were that simple. Unfortunately, most destructive regulation and encroaching state activity comes from people choosing on the margins—that is, those who want one more regulation to curb what they perceive as unjust business practices but who don't want to destroy business.

Brook and Watkins probably agree with me that most government growth occurs due to people choosing government policies on the margins, and they would likely argue that the only way to stop a death by 1,000 cuts is to articulate an unyielding moral position that all cuts are wrong. I disagree that this rhetorical strategy alone is the best way to stop the cuts. I also believe that 500 cuts are better than 1,000.

Now, it may be true, and I grant that it often is, that an argument allowing 500 cuts may be philosophically indistinct from one that allows for 1,000, or even a million. Brook and Watkins ask, "Isn't the only difference between socialism and the entitlement state one of degree?" Perhaps, but it is a range of degrees upon which much human suffering can be mapped.

As a libertarian, I read Rand, as well as Brook and Watkins, with a smile and a fist raised in solidarity. They make free-market champions feel like Nietzschean heroes fighting against a recalcitrant, if not willfully obtuse, prevailing narrative. They make businessmen feel like Superman and regulators into Lex Luthors. They ask you to own your life and to stop apologizing for it.

But for non-libertarians who read objectivists, I fear the opposite is true. Those who are unfamiliar with the broad range of arguments supporting capitalism and the empirical evidence of freedom's benefits are likely to charitably describe Brook and Watkins's views as lacking nuance. Uncharitably, they would accuse them of Pollyannaish naiveté, if not downright malice.

I'm sure Brook and Watkins don't care what those who oppose capitalism think of them, but perhaps they would if objectivist arguments are convincing fewer people than they are pushing away.

This is why it is important to be ecumenical in allowing "acceptable" arguments for liberty. Whereas I welcome objectivists into the libertarian tent as part of a multimembered consortium to rollback encroaching government, I fear that many objectivists would not give me the same courtesy.

Yet, despite my misgivings, *Free Market Revolution* is a good book for convincing some types of people and inspiring others already convinced. We need more than objectivism, however, if we're going to win the fight against Big Government.

Trevor Burrus
Cato Institute

Political Bubbles: Financial Crises and the Failure of American Democracy
Nolan McCarty, Keith Poole, and Howard Rosenthal
Princeton, N.J.: Princeton University Press, 2013, 368 pp.

The majority of books on the recent financial crisis tend to be written either by economics/finance experts or by journalists. While the journalistic accounts occasionally focus on political actors, it is usually in the manner of "bad people doing bad things" rather than with a theoretical framework. The economic accounts, with some exception, rarely incorporate the politics of finance. It is this vacuum that *Political Bubbles* attempts to fill.

The authors are three prominent political science professors whose work will be familiar to many *Cato Journal* readers. Poole and Rosenthal's previous joint works have made significant contributions to the Public Choice literature on economic legislation. All three authors can rightly be called pioneers in the modern academic literature on ideology. Readers of Poole and Rosenthal's *Congress: A Political-Economic History of Roll Call Voting* (1997) and *Ideology and Congress* (2007) will recognize much of that work here. Chapters 2, 3, and 4 of *Political Bubbles* are largely an introduction to the authors' previous work. Those already familiar with this work can either skim or skip these chapters.

The framework outlined in these early chapters focuses on what the authors call the "Three I's": ideology, institutions, and interests. That framework is used to explain the development of "political bubbles," which are defined as "a set of policy biases that foster and amplify the market behaviors that generate financial crises." It should be clear from this definition that the authors begin with the premise that financial crises are the result of markets. By relying on this definition, the authors rule out the possibility that government itself can be a generator of financial crises. From this starting point, the authors ask which ideologies are likely to constrain government from controlling financial markets. Not surprisingly, the authors quickly conclude that free-market conservatism is the "belief structure most conducive to supporting political bubbles." The rest of the book is spent trying to show how Congress became more free-market oriented and adopted deregulatory policies that contributed to the recent bubble.

The foundation of the analysis is a spatial model of voting, which the authors developed in previous works. Based on this model,

members of the House and Senate are assigned an "ideology score." The scores are calculated upon actual roll call votes, where "yeas" and "nays" represent different ends of a two-dimensional scale. The authors impose these vote scores on the traditional "liberal" and "conservative" positions. The authors claim these scores are consistent predictors of voting behavior. Positions that do not fit into this framework, such as a libertarian one, are shoehorned by excluding noneconomic roll call votes. False predictions that are out of the sample, such as Senator Russ Feingold's vote against Dodd-Frank, are simply explained away in an ad hoc fashion.

These scores were previously used in the three authors' 2006 book *Polarized America*. With a title like that you can perhaps guess that the authors blame political gridlock for much of the cause and "lack" of response to the crisis. As someone who spent the years leading up to the crisis on the staff of the Senate Banking Committee, I can certainly say that gridlock, or rather needing to reach 60 votes in the Senate, was a significant obstacle to reforming Fannie Mae and Freddie Mac, as well as hampering attempts in 2006 to bring more competition to the credit-rating agencies. The authors' thesis is consistent with my experience but for different reasons than the authors suggest, which I will return to. The remaining chapters essentially apply the authors' previous work to the particulars of the recent financial crisis.

Despite chapters on the role of interests and institutions, the bulk of *Political Bubbles* is dedicated to ideology. A brief chapter only illustrates that (1) campaign contributions from the financial services industry have been increasing, and (2) a handful of members of Congress have a concentration of financial services employment in their districts. We are also shown that members of the relevant congressional committees receive more contributions from finance than members not on those committees. None of this is really new or shocking, and the authors do not demonstrate the importance of these findings. As is often the case, the mere suggestion of money is supposed to be damning. Those requiring a higher level of proof will be left disappointed.

The discussion on institutions is only slightly more in-depth, which is particularly surprising given this is a book by three political scientists. After a very cursory overview of features such as the filibuster, the presidential veto, and the review function of the courts, the authors conclude that our system requires supermajorities and that those

supermajorities can shift dramatically with relatively small changes in the makeup of Congress—what the authors label "pivot points."

As an example of a pivot point, they cite the special election of Massachusetts Senator Scott Brown. I agree with the authors that Brown's election shifted the balance of ideology within the Senate. I disagree about Brown's impact. Much is made of Brown's objection to a prefunded resolution mechanism in Dodd-Frank's Title II, which would have assessed an insurance fee on financial institutions (something like a FDIC for nondeposit creditors). Senator Brown viewed it as a tax, and in order to gain his support the fund was dropped. Instead, Dodd-Frank has a postfunding mechanism. The authors paint this as a huge ideology swing.

Which brings me to my biggest criticism of *Political Bubbles*: relatively minor, even trivial, policy differences are presented as massive ideological canyons. Because of the institutions not mentioned by the authors, such as agenda control by the committees, Scott Brown's election had a relatively minor impact on Dodd-Frank. That Dodd-Frank does not fix the causes of the crisis has little to do with Scott Brown and everything to do with the disinterest of the relevant committee chairs in addressing those issues.

Unfortunately, the book is marred in several places by occasional outright factual errors. For instance, the authors claim that Fannie Mae and Freddie Mac were not allowed to acquire risky subprime mortgages (p. 131). As I've explained elsewhere (Cato Institute Briefing Paper No. 120, March 7, 2011), this is false, as is clear from the both the regulatory and statutory language governing their allowable purchases. As I also document, Fannie Mae and Freddie Mac did purchase subprime loans. The authors also claim that Fannie Mae is in receivership, which is also incorrect (both are in conservatorship). Finally, they also claim that the Community Reinvestment Act (CRA) prohibited banks from redlining but did not require banks to make risky loans. Again the authors simply have their facts wrong. There's no prohibition on redlining in CRA, and the 1995 regulatory changes to CRA implementation encouraged banks to lower underwriting standards. While many academics have an allergic reaction to actually reading statutes and regulations, that is no excuse for not at least consulting a legal expert in this area. Such simple errors undermine confidence in the authors' knowledge of financial regulation.

Beyond the handful of objective factual errors, the authors make a number of assertions that are at least debatable (and in my opinion

outright false). For instance, they dismiss objections to the Dodd-Frank Act as a "bailout bill" as "dubious." An objective reading of Dodd-Frank would conclude otherwise, but there is little evidence that the authors actually have read Dodd-Frank. Of greater relevance to the authors' framework is their use of ideology in describing individual politicians. For example, the former chair of the House Financial Services Committee, Mike Oxley (of Sarbanes-Oxley fame), is described as a "free market conservative advocate." However, there is no credible definition of free-market advocate that would include Oxley.

This is where *Political Bubbles* falls most flat. In constraining their definition of "ideology" to a one-dimensional measure, the authors cannot distinguish actual free-market advocates from crony capitalists. In fact, they explicitly say they see little difference, claiming that free-market advocates and crony capitalists have bonded (p. 270). One could as easily take the case of Fannie Mae and claim that progressives and crony capitalists bonded, but such would undermine the authors' thesis that activist government is the solution.

Despite the brief chapter on interests, *Political Bubbles* downplays the role of special interests in the financial services industry, attributing the vast majority of legislative changes to "ideology." In my seven years as a Senate staffer working on these issues, I found that almost all of it is driven by interests with very little role played by ideology.

Political Bubbles contains a number of policy recommendations. Some of these are related to the arguments advanced by the authors, such as filibuster reform; others, such as limiting bank activities, feel tacked on. The financial rules they suggest mirror conventional wisdom and seem to be included largely for that reason. Their more interesting recommendations, such as set rules that account for political risk, are useful considerations. Overall, the recommendations are a mixed bag—something for everyone to agree with and object to, but little in the way of convincing argument.

I've quipped in many presentations that one cause of the financial crisis is that "democracy loves a bubble." By this I mean that the people generally enjoy the appearance of expanding wealth that accompanies a bubble. Few politicians can successfully lean against the wind. With that in mind, I greatly looked forward to *Political Bubbles*. I was sadly disappointed that the role of democracy is never really addressed, nor is the authors' theory really applicable beyond the recent crisis. The authors are ill informed on issues of finance and

do not consider alternative theories, even within political science, as an explanation of the crisis. Despite these flaws, *Political Bubbles* is an important book. I would include it in any seminar on the crisis, as it injects an analysis into the discussion that is all too often missing. I would also hope that *Political Bubbles* spurs other scholars to plow this field.

Mark A. Calabria
Cato Institute

The Bet: Paul Ehrlich, Julian Simon, and Our Gamble over Earth's Future
Paul Sabin
New Haven, Conn.: Yale University Press, 2013, 304 pp.

Yale historian Paul Sabin's *The Bet: Paul Ehrlich, Julian Simon, and Our Gamble over Earth's Future* is worth a read because of its detailed tour through the world of environmental doomsaying. Yet, in the end, I was profoundly disappointed, consigning this book to my very large Cassandra File because Sabin endorses that doomsaying as expressed by dreaded global warming.

The Bet is about a very public wager between economist Julian Simon and the serial apocalypse predictor Paul Ehrlich (cosigned with John Holdren, President Obama's science advisor). Ehrlich bet that the price of five key metals would rise between 1981 and 1990, and Simon bet they would decline (in constant dollars). Ehrlich was spectacularly wrong, but nonetheless continued, and continues, to enjoy a substantial public presence despite virtually none of his predictions coming true.

Simon, who died in 1998, was very poorly compensated and couldn't even get the University of Maryland to give him a secretary. He found it increasingly hard to publish in the academic literature, and he was confounded by the durability of the environmental apocalypse meme. In 1995, he wrote in the *San Francisco Chronicle*: "After 25 years of the doomsayers being proven entirely wrong, their credibility and influence waxes ever greater." Indeed. Ehrlich has been showered with goody-goody prizes around the planet. He's still a staple on the dinosaur media. Being fundamentally wrong has been very, very good to him, and being right bought Simon virtually nothing.

Sabin is far too kind to Ehrlich. One of the few criticisms comes at the end: "Most fundamentally, human history over the past forty years has not conformed to Paul Ehrlich's predictions." Sabin cannot bring himself to say what can be said with more economy of words, namely, "Ehrlich's predictions were wrong."

Further, he comes perilously close to blaming Simon for another end of the world—that is, the mother of all environmental scares, global warming. Sabin contends that "the most pernicious current reflection of Ehrlich and Simon's clash is the ongoing political impasse over climate change." That's the kind of fatuous academic blather we've grown to expect from the global warming crowd. Blame Julian Simon's bet. Heck, why not trot out the Koch Brothers while you're at it? Blame anyone but the dons of environmental science, who have every interest to hype every issue.

What a spectacular misreading of recent history. The reason that the Senate didn't pass the House's 2009 cap-and-trade bill isn't because of Julian Simon. It's because Senate staffers read polls. Three days after the House passed that infamous bill, Scott Rasmussen's generic congressional ballot switched from Democrat to Republican. Given that the index is a three-day running average, this was as close to a single cause-and-effect poll shift as it ever gets.

Did Julian Simon cause this? Sixty-five seats in the House went from Democrat to Republican in the 2010 mid-term, as did control of the chamber. Virtually every close race was lost by a Democrat who voted for cap-and-trade. In the Senate, every close contest was won by a Democrat. Both houses voted for the oxymoronic Affordable Care Act, but only the one voted for cap-and-trade.

While trotting out his glib explanation for the failure of his preferred policy, Sabin bemoans the generic failure of environmental hype. "Conservative commentators," he notes, "have warned of 'apocalypse fatigue.'" Fact check: According to wordspy.com, I coined the term, and I'm no conservative. As I said on CNN's Crossfire (February 10, 1992),

> I think the problem, Larry, is that we keep on seeing this science by press release with these apocalyptic pronouncements. If we were in California, we'd probably say the people are getting apocalypse fatigue and each one of these things has to be hyped more and more and more, and you know that.

Along with many other folks, I am indeed tired of environmental apocalypse. Born in 1950, I believe I am living through the death throes of my eighth apocalypse—global warming—and witnessing the birth of the ninth, the erroneously named "ocean acidification." And Paul Ehrlich championed every single one, which are, in order: the population bomb, the limits to growth, mass extinction, the world food crisis, global cooling, acid rain, ozone depletion, and global warming.

If Ehrlich didn't live in Stanford's protective bubble, he might realize that he is just another—very vocal—prophet of doom and that the chances of his being generally wrong (not just in his bet with Simon) were and are extremely high. He somehow doesn't get that the intelligentsia have spent millennia getting paid for warning about "The End," and that their institutions, be it the Church, the millennial cult, or academia, thrive on the largess tendered to them out of fear of the reigning apocalypse fad. Sabin should provide an explanation for the numbing repetition of this process. In fact, it is a logical consequence of the way we do environmental science today.

Virtually all such science is funded by a monopoly provider—the federal government. Given that science issues literally compete with each other for finite resources, there is every incentive to portray your issue as vitally important. Al Gore called for fighting global warming to be "the central organizing principle for world civilization."

The incentive structure in academic science is perilously intertwined with this type of exaggeration. Those who call out people or issues are likely to be thrown out of the academy. Those who research the apocalypse du jour gain massive funding and receive honors and lifetime employment. Woe be to the researcher who submits a manuscript debunking any of the hype, as it is likely to be vigorously and negatively reviewed. Praise be to the one who writes the "it's even worse than we thought" paper, which will be guaranteed a light review, as well as newspaper coverage, and the researcher will get permanent first-class upgrades. It is odd that Sabin, as a historian, avoids examining the root cause of what he feels is a "gamble for the earth's future." No hype there.

Despite its flaws and lacunae, I still recommend Sabin's book—probably not for someone like me who's lived through it all, but for students and young professionals to grasp the serial ineptitude of ecological doomsaying. Those new to the field will also note the bone-shattering intolerance of any criticism of Cassandra. Those who don't

know the environmental hype that elected Jimmy Carter are likely to repeat history. They almost did in 2000.

When speaking of President Carter, who may correctly be viewed as the gray eminence of global gloom, Sabin exhibits his innate reluctance to criticize greens, even when they are (as they usually are) profoundly wrong. How else to explain the following passage summing up Carter's energy ineptitude:

> Carter's belief [was] that demand for oil would outstrip production in the early 1980s. If action wasn't taken, he said in one speech, the world's entire proven reserves of oil could be gone by 1990. Viewed through a later lens, Carter's fears about looming shortages appear exaggerated.

"Appear"? How about "were"?

For me, the take-home lesson in *The Bet* is that a guy can make a pretty good living by shorting apocalypse futures.

Patrick J. Michaels
Cato Institute

The Great Rent Wars: New York 1917–1929
Robert M. Fogelson
New Haven: Yale University Press, 2013, 512 pp.

Economic shocks in an unregulated textbook world are managed through the price system. During gluts, prices fall and the least efficient firms lose wealth and exit the market. The result is that supply falls and demand increases. Eventually a new equilibrium is reached in which prices increase toward marginal cost and risk-adjusted returns to firms equal the cost of capital. During shortages, prices rise, existing firms receive rents, and new firms enter the market. The result is that supply increases and demand falls. Eventually a new equilibrium is reached in which prices decrease toward marginal cost and risk-adjusted returns to firms fall to equal the cost of capital.

In *The Great Rent Wars: New York 1917–1929*, Robert M. Fogelson expands my one paragraph, antiseptic, economist's account to over 400 pages and describes in great detail housing-market shocks in New York City during and following World War I. First some stylized facts. From 1903 to 1916, NYC experienced an unprecedented

expansion of rental housing supply. Four hundred thousand units were built, and, by 1916, 40 percent of all apartments were built after 1903. During this time, even though the population grew by a million and 40,000 old tenements were demolished, the vacancy rate grew and rents fell.

From 1917 thorough the late 1920s, the expansion of supply ended. Initially WWI regulations stopped residential construction. After the war, everyone expected a construction boom, but it did not occur because materials and labor costs rose faster than tenants' willingness to pay. Construction costs rose 50 percent from 1913 to 1918. By 1920 the vacancy rate was 0.3 percent. Rents rose dramatically, and tenants resisted with rent strikes. Landlords responded with eviction attempts.

The remainder of the book describes how the state legislature wrestled with these events. The conflicts and arguments from 100 years ago about the pros and cons of market intervention are very similar to what we hear today. Some argued for tax incentives and reduced down payments to increase housing supply. Others proposed to increase supply by modifying the Tenement House Act of 1901 to make converting single-family houses into apartments cheaper. Fire departments and tenement reformers predictably opposed such modifications because they would result in unacceptable (for them) fire-safety risks. Future mayor and then congressman LaGuardia thought public housing was the answer. Real estate developers opposed LaGuardia, arguing that they couldn't compete against taxpayer-subsidized housing and that public housing was unconstitutional.

In 1919, the New York state legislature altered laws governing property rights and the regulation of housing. Tenants were given the right to 40 days notice of lease termination. In addition, the 1901 tenement law was amended to allow single-family homes to be converted more easily into apartments. The banking law was amended to reduce down payments from 60 percent of property value to 40 percent.

In 1920, the legislature passed a series of bills giving courts the right to allow tenants to pay their existing rent if the courts deemed a new rent unreasonable. Landlords couldn't just evict a tenant because they were "undesirable" and couldn't cut off services to tenants who were paying below-market rents. New buildings were exempt from these laws. The attempt to control economic rents

(through price controls for existing units) while preserving market forces at the margin (through exemption of new construction) is remarkably similar to the policies enacted by Congress in 1973 and 1975 in reaction to the oil price shock of 1973—that is, putting price controls on "old oil" while keeping world market prices for "new oil" and imports.

Property owners argued that the new laws were unconstitutional. In April 1921, the U.S. Supreme Court upheld Congress's emergency rent control for the District of Columbia and New York state's emergency rent laws. Justice Oliver Wendell Holmes, writing for the majority, said that if the legislature has the power to regulate building heights (a 1909 decision involving Boston, *Welch v. Swasey*), it also has the power to regulate building rents. As described by Fogelson, the dissent from Justice Joseph McKenna reads very much like a Cato amicus brief today:

> "Why is it the solicitude of the police power of the state of New York to keep from competition an apartment in the City of New York?" To say that it is to supply homes to the homeless "does not satisfy" because all the laws do is keep one tenant in and another out. This they do by withdrawing "the dominion of the property from its owner, superseding the contracts he confidently made under the law then existing and subjecting them to the fiat of subsequent law." "If such an exercise of government is legal, what exercise of government is illegal? Houses are necessary, but other things are as necessary. May they too be taken from the direction of their owners and disposed of by the government?"

The remaining chapters of the book explain in great detail the difficulties faced by the courts that implemented the emergency rent laws. How were judges supposed to figure out a reasonable rent when the landlords said they were not earning a sufficient return and the tenant said they were not able to pay more? What was a "fair rent" under abnormal conditions? The system jammed up under the strain of these issues, and, by August 1921, 10,000 tenant-landlord cases were scheduled for trial in the Bronx second district court alone.

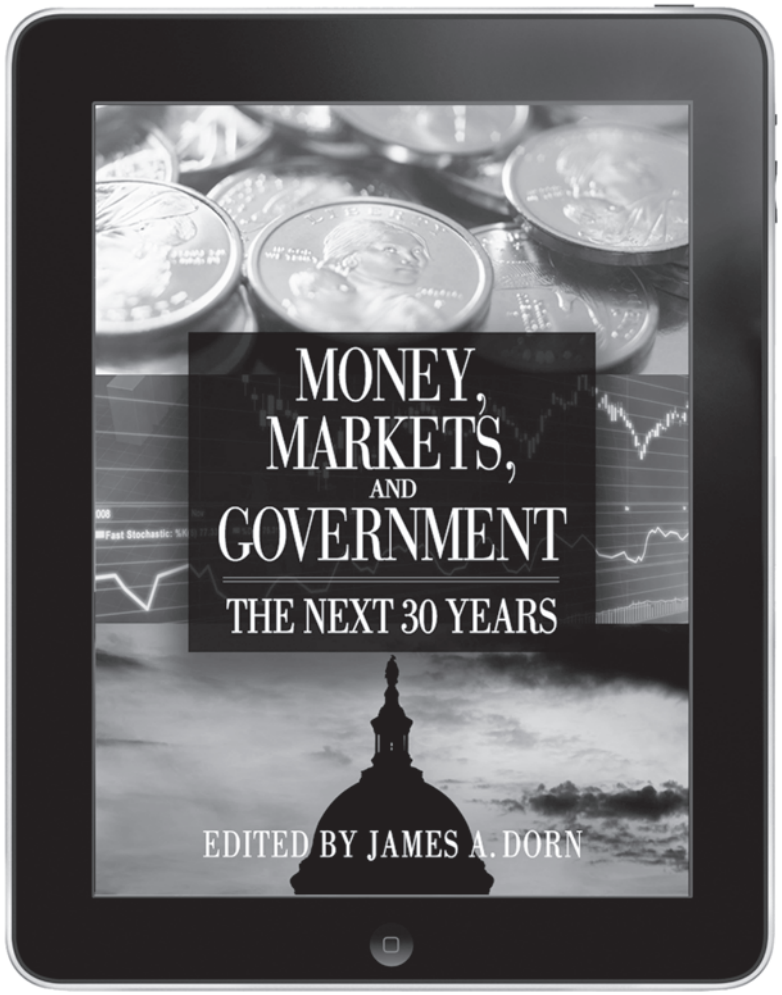
Judges responded by implementing the judicial equivalent of public utility rate regulation. In August 1921, a New York Court of Appeals ruled that a fair return was 10 percent of market value or less and net income was gross rent minus expenses including depreciation.

The emergency statutes were extended in 1922 and 1924. But the beginning of the end also started in 1924 when the U.S. Supreme Court struck down the District of Columbia emergency rent control system because the Court considered the emergency to be over. In 1926, New York's laws were extended one more year to apartments that rented for less than $20 a room per month. In 1927, the extension was for apartments that rented for less than $15 a room. The laws expired June 1, 1929.

This book reinforces in great detail some central lessons of applied microeconomics. First, never intervene in particular markets to remedy what are essentially distributional issues rather than market failures. The administrative nightmare of implementing the emergency rent laws in New York was directly analogous to the complexity of oil price controls in the 1970s. Second, the intellectual arguments for and against market intervention are remarkably constant over time. During some eras market arguments win while during others intervention arguments prevail. Shortages of necessities such as housing and energy, even if they are caused by other perverse policies, often result in the acceptance of interventionist policies. In all, Fogelson's book is an occasionally interesting, if a little too exhaustive, history of the predictable effects of microeconomic intervention.

Peter Van Doren
Cato Institute

NEW EBOOK FROM THE CATO INSTITUTE



The 2008–2009 financial crisis and Great Recession have vastly increased the power and scope of the Federal Reserve and radically changed the financial landscape. This ebook, an edited volume of papers presented at the Cato Institute's 30th Annual Monetary Conference, examines those changes and considers how the links between money, markets, and government may evolve in the future.

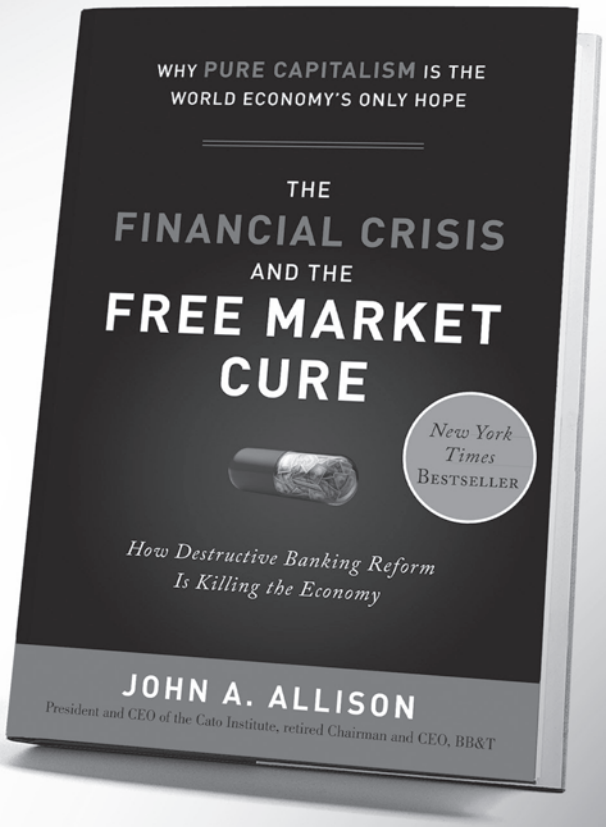
Topics covered include how the choice of monetary regimes affects economic freedom and prosperity, the policy steps needed to avoid future financial crises, the limits of monetary policy, the lessons from the Eurozone debt crisis, and China's path toward capital freedom.

CATO INSTITUTE | $3.99 AT CATO.ORG/STORE AND OTHER EBOOK RETAILERS.

The #1 *Wall Street Journal* bestseller *Forbes* calls "one of the most important books of the year."

"A call to arms for a nation on the brink."
—The Blaze

"Authoritative analysis of the problems of the American economy."
—The Weekly Standard



SCAN HERE TO BUY NOW!

Mc Graw Hill Education

Learn More. Do More.

FOLLOW US: BusinessBlog

MHPROFESSIONAL.COM

McGill Law Journal
Revue de droit de McGill

Bilingual. Bijuridical.
Student Run. Peer Reviewed.

See why the *McGill Law Journal* is among the best in the world.
Subscribe today!

To Subscribe: Visit: http://lawjournal.mcgill.ca / E-mail: journal.law@mcgill.ca / Phone: 1 514 398 7397

•

•

, & &

& ,

- . 33 83 71 . ** 3

*. . 3 ** * 33 3 1

D CJ CEG

E G D	*- B DIEKHAN*	*' EG AG*	*EC ' BI D*
GBH D	*I G EBB G*	*EJGA*	*IEG KH DHED*
EN AA D	*DD- GEL*	*EHJ - JGK A*	*JHHB E N*
K	*E D- EGI G*	*J N G*	*. EBH C B*
GHE G I DH	*K D GHI K*	*C H G A*	*EB OB*
C GH	*D GL K*	*H L I*	*B C G G - JGO*
LGD BD	*BMD G - EINB*	*G IC D*	*B D E DHED*

HEJI I H B GAE AIAD G JJAE II G EJAA JAE
- IJ ID HHH HJ I I BAE AEIA J E HAL

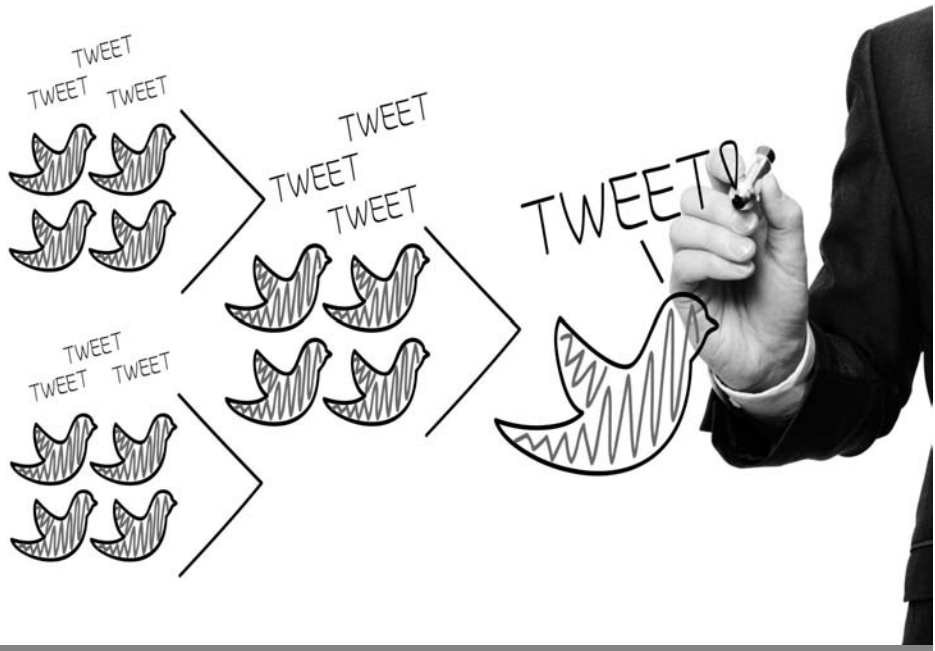
- HE I E J 3AE

Education is important to everyone.

Join in the discussion...

Follow us on Twitter **@EducationNext**

Also find us on **Facebook, Pinterest, and Google+**



EducationNext

A JOURNAL OF OPINION AND RESEARCH

Educationnext·org

One of the Most Cited Journals in International Relations

International Security

Steven E. Miller, Editor in Chief
Sean M. Lynn-Jones and Owen R. Coté Jr., Editors
Diane J. McCree, Managing Editor

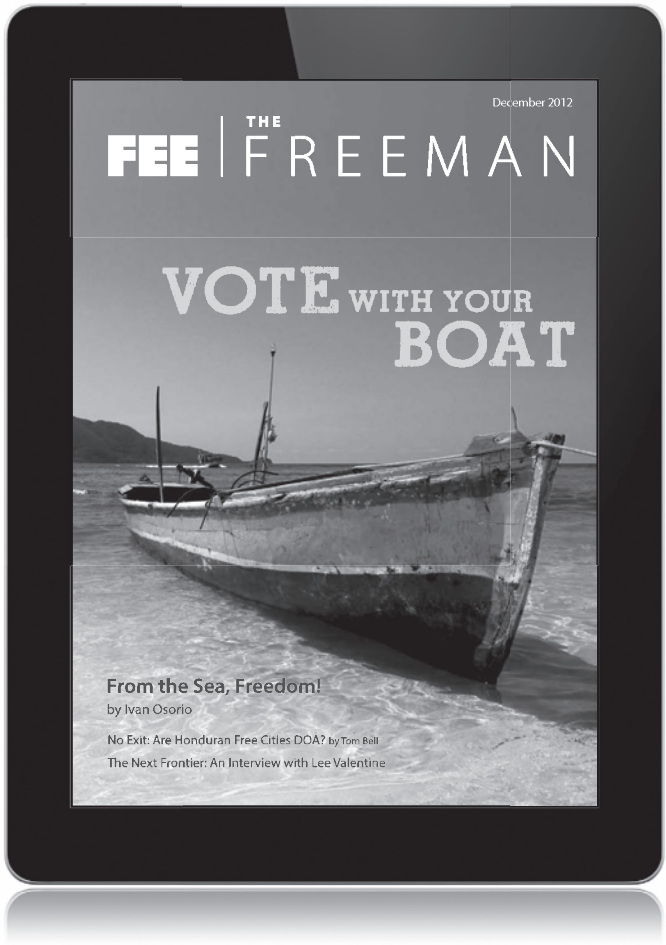
Samuel P. Huntington, Harvard University

John Lewis Gaddis, Yale University

mitpressjournals.org/is

Now Available

$2.49 a month for Kindle readers



On the Kindle app for your smartphone, computer, or tablet.

Cato Institute

Founded in 1977, the Cato Institute is a public policy research foundation dedicated to broadening the parameters of policy debate to allow consideration of more options that are consistent with the principles of limited government, individual liberty, and peace. To that end, the Institute strives to achieve greater involvement of the intelligent, concerned lay public in questions of policy and the proper role of government.

The Institute is named for Cato's Letters, libertarian pamphlets that were widely read in the American Colonies in the early 18th century and played a major role in laying the philosophical foundation for the American Revolution.

Despite the achievement of the nation's Founders, today virtually no aspect of life is free from government encroachment. A pervasive intolerance for individual rights is shown by government's arbitrary intrusions into private economic transactions and its disregard for civil liberties. And while freedom around the globe has notably increased in the past several decades, many countries have moved in the opposite direction, and most governments still do not respect or safeguard the wide range of civil and economic liberties.

To address those issues, the Cato Institute undertakes an extensive publications program on the complete spectrum of policy issues. Books, monographs, and shorter studies are commissioned to examine the federal budget, Social Security, regulation, military spending, international trade, and myriad other issues. Major policy conferences are held throughout the year, from which papers are published thrice yearly in the Cato Journal. The Institute also publishes the quarterly magazine Regulation.

In order to maintain its independence, the Cato Institute accepts no government funding. Contributions are received from foundations, corporations, and individuals, and other revenue is generated from the sale of publications. The Institute is a nonprofit, tax-exempt, educational foundation under Section 501(c)3 of the Internal Revenue Code.

CATO INSTITUTE
1000 Massachusetts Ave., N.W.
Washington, D.C. 20001
www.cato.org

8^{00}

THE FOREIGN AID ILLUSION

The idea that global poverty could be eliminated if only rich people or rich countries were to give more money to poor people or to poor countries, however appealing, is wrong. Its undying popular appeal seems to be based on the commonsense notion that, if I am poor, and you give me money, I am less poor. But foreign aid doesn't work that way; the belief that it does is what I call the "aid illusion," itself a barrier to better policies. . . .

The key to understanding why aid does not work lies in the relationship between aid and politics. Political and legal institutions play a central role in creating the environment needed to nurture prosperity and economic growth. . . . But large inflows of foreign aid change local politics for the worse and undercut the institutions needed to foster long-run growth.

—Angus Deaton

"The Great Escape," *Foreign* Policy (16 September 2013).

CATO INSTITUTE
www.cato.org

the CATO JOURNAL

VOLUME 34 NUMBER 1 • PUBLISHED BY THE CATO INSTITUTE • WINTER 2014