Taxation, Optimization, and the January Seasonal Effect
(and Other Essays in Taxation and Finance)

by

Theodore S. Sims

A.B., Columbia College, 1967
J.D., University of Chicago, 1970

Submitted to the Department of Economics
in Partial Fulfillment
of the Requirements for the Degree of

Doctor of Philosophy
in Economics

At the
Massachusetts Institute of Technology
April 1995

© 1995 Theodore S. Sims. All rights reserved.

The author hereby grants to MIT permission to reproduce and distribute publicly paper
and electronic copies of this thesis document in whole or in part.

Signature of Author

Department of Economics
April 15, 1995

Certified by

James M. Poterba
Professor of Economics
Thesis Advisor

Certified by

Peter A. Diamond
Paul A. Samuelson Professor of Economics
Thesis Advisor

Accepted by

Richard S. Eckaus
Professor of Economics
Chair, Departmental Committee on Graduate Studies

Massachusetts Institute of Technology

Jun 29 1995
TAXATION, OPTIMIZATION, AND THE JANUARY SEASONAL EFFECT
(AND OTHER ESSAYS IN TAXATION AND FINANCE)

by

Theodore S. Sims

Submitted to the Department of Economics on April 15, 1995,
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy

Essays One and Two study the relationship between tax-motivated year-end
loss selling of securities and the January "seasonal effect".

Using 1964-1990 NYSE/AMEX daily returns for a 35-day period that spans
the turn of the year, Essay One documents a year-end depression, most evident among
small-capitalization stocks, where the January effect is also most pronounced. That
depression is negatively related to the January effect. Using a new measure of tax-loss
potential, I show also that year-end short and long-term loss potential is positively related
to the January effect, and that long-term loss potential is negatively related to year-end
returns. These findings are consistent with the tax-loss selling hypothesis, and generally
(LITWP#7.)

Essay Two models the problem of a uniformly taxed investor, endowed with
a known gain that will be taxed at the end of the year, who must decide whether to sell
a security at an offsetting loss early in the year, at the end of the year, or both. In the
face of uncertainty about the year-end price of the loss stock, I show that there exists an
interval of prices at which the investor optimally waits before taking the loss. The result-
ing prediction of a year-end rise in volume among loss stocks differs from Constantinid-
es, and is consistent with the findings of Badrinath and Lewellen (1991), Lakonishok &
Smidt (1984, 1986) and Dyl (1977). (LITWP#8.)

Essay Three applies Samuelson’s (1964) theorem on asset valuation under an
income tax with economic depreciation to two problems other than depreciation. First,
I modify Samuelson’s formulation to provide for the explicit representation of costs. In
the solution, economic "depreciation" permits costs to be deducted only to the extent that
they do not exceed current gross revenue, suggesting a "natural" capitalization rule. I
show also that tax-invariant recovery of future costs always has the property of requiring
a zero apparent rate of tax on funds held to satisfy such costs. Second, I show that the
"cancellation of indebtedness" rule properly measures a defaulting debtor’s income, but
only if, as the invariance theorem suggests, interest on the liability has been allowed as
a deduction as it accrues.

Thesis Advisor: James M. Poterba
Title: Professor of Economics
PREFACE:

OF MIDDLE AGE AND
MID-LIFE CRISSES

Sims' First and Second
Definitions of Middle Age:

1. That point in life when you first start thinking about the money in the keogh plan as "your money".

2. That point in life when the reason you think you can't do tomorrow something you just did yesterday is that you can't.

On pursuing a Ph.D. in economics at M.I.T
in (early) middle age:

It was less fun than buying a Porsche Turbo-Carrera, a good deal more work, and a lot more expensive.
ACKNOWLEDGEMENTS

I dedicate this to the memory of my father, Eugene Paul Sims, whose intellectual aspirations were always somehow thwarted, and at whose death in early September 1985 this enterprise began. I returned from the funeral, two weeks into the semester, and went to lengths to secure late registration in an introductory calculus course (I had flunked the subject twice in college) I could just as easily have audited. I had no real idea where I was headed. It was to be three years before I entered M.I.T. I dedicate it in equal measure to my mother, Deborah Shwayder Sims, for whom (like her mother before her) no sacrifice was too great where her children’s education was concerned, and who has lived to see this to its end.

After that, there are so many -- some, even, whom I have never met -- who contributed to this unlikely mid-life misadventure that it is hard to know quite where to start, and it will be difficult to stop. Five people stand out, however, as those without whom it never would have gotten off the ground, much less made it to the end. They are Daniel I. Halperin, Peter A. Diamond, James M. Poterba, Stephen W. Salant, and Linda T. M. Bui. As a lawyer, Danny is the closest to a mentor I have ever had. An indirect conversation between him and Peter (mediated, I am told, by Steven Salop, whom I still have yet to meet) sometime during the Spring of 1988, led Peter to think it might be interesting to have me enroll as a special student at M.I.T. From that point onwards through to generals, Peter discreetly engineered my progress, always giving me whatever slack I needed, never giving me quite enough rope to truly hang myself.

I have fond recollections of many conversations with Professor Diamond, but my favorite was over lunch in his office (sandwiches from the Kendall Square Au Bon Pain) in early June of 1989, when I had just qualified to retake examinations in three of the four first-year theory courses (although, I must add, by a mere point in 14.124, occasioned, I continue to insist, by having wasted precious time trying to solve an insoluble dynamic optimization problem that the instructions to the exam had nevertheless directed me to "Solve . . ."). Peter first urged me to consider matriculating at a different school. He then told me that I was welcome, if I wished, to remain another year as a special student at M.I.T. He shared with me his apprehension that I was underestimating the prospects, if I were to become a degree candidate at M.I.T., that I would fail generals. (I assured him that I wasn’t.) Finally, he informed me that if, despite all this, I wanted to proceed, he would support my application for admission. ("Affirmative action for the middle-aged" is how I recall his having put it.) Not more than five hours later he called my apartment on Beacon Street to say, if memory serves me correctly, that he "just ran into another member of the Admissions Committee, we constituted a quorum, and we decided, ‘you’re admitted’." He added that he had checked with financial aid, and that, if I did matriculate, the Department had some Olin Foundation money with which they would waive my tuition the coming year. So, having sounded all his cautionary notes, Peter left me with a choice: remain as a special student and pay $15,000 in tuition, or matriculate at an out-of-pocket cost of $0. I had been looney enough to start, but was not quite crazy enough to ignore that package of incentives.
I passed generals, with the class with which I started, then returned a year later to study econometrics and complete my minors. Thenceforth Jim Poterba (the other member of that Admissions Committee quorum) took command. He was, I think, initially skeptic about the project that I wanted to pursue. But when I showed him some pictures based on preliminary data, he took an interest. In all events he was a genuinely model dissertation advisor. I tried to take care with what I sent him. Whatever I did send him, however, elicited written comments within the month. They were detailed, insightful, realistic, constructive, supportive, and, when necessary, blunt. ("Pretentiously overwritten" is one phrase I especially recall.) I could not have asked for better or for more.

Throughout the entire process -- coursework, generals, dissertation -- Steve Salant, though physically in Ann Arbor, was always there. It has been a matter of extraordinary good fortune for me. He is my friend of longest standing, is to my eye a really fine economist, and was interested in and encouraged my decision to study economics. But he had no professional obligation to my work. He has three young children and an otherwise very busy life. Even so, he was always willing to talk by telephone, often at great length, at most any time that we were both awake. It is possible, though I judge substantially less likely, that I would have finished had he not been around.

Other members of the faculty played important roles for me, especially Jerry Hausman, Dan McFadden and Jean-Luc Vila, as well as (although they will not know it) Robert Solow and Martin Weitzman. It was a privilege to sit through Professor Solow's measured musings on the practice of economics, illustrated with five (in lieu of five dozen) papers. As for Marty Weitzman, I will never forget talking with him about my dismal performance in 14.123 (although I did get an A+ on the retake). He was utterly unable to discuss the test with me. He just kept staring in disbelief, punctuating his stare with (several) repetitions of, or variations on, the question "Why on earth are you putting yourself through all this . . . ?!?!" It was the middle of first year. I often wondered that very thing myself.

Early first year, Peter Diamond observed off-hand that M.I.T came as something of a shock to most first-year students, as they were all "accustomed to being first". I let pass without comment what I took to be hyperbole, then proceeded to learn in short order that Peter wasn't really being all that hyperbolic. At least at the beginning they, like Marty Weitzman, often seemed to look at me in way that wondered "What's he doing here?" A few -- Alan Gerber was, I believe, the first -- had the temerity to ask. But they were also immediately accepting, and proved always willing to lend a hand to a struggling colleague. We were in the soup together, and it made for a remarkably cooperative atmosphere. I now count many of them as real friends. I must mention, especially, Rabi Abraham, Yacine Ait-Sahalia, Ben Cohen, Michael J. Chapman (with whom I struggled through first year, and participated in two B league intramural hockey championship efforts, including an especially delicious defeat of Sloan in the A league semifinals in 1990), Howard Chang, Glenn Ellison, Gary Englehart, Sara Fisher-Ellison,
Alan Gerber, Michael Gibson, Charlie Hadlock (yes, Charlie), Dan Kessler, Mark McClellan, Walter Novaes, Hi'ary Sigman, and Luigi Zingales. Camaraderie entirely aside, I learned much from each of them.

Where my dissertation is concerned, special words of thanks are due to Jake A. Wetzel, who was gracious and patient while constructing and reconstructing the raw material for the initial data sets for Essay One; and Minor Sachlis of the Finance Department of George Washington University, who took a real fancy to the project, and was unstinting in sharing with me both his familiarity with the CRSP files and his programming expertise. Equally deep appreciation is due to the above-mentioned Steve Salant, who took an equally deep interest in the problem at the core of Essay Two; and whose instinct for the simplest possible way to get at the core of a problem is a virtue of his modelling I have frequently observed, even if I have not internalized the knack myself.

Others helped out in a variety of ways. In addition to those already mentioned, thanks are due to Seymour Fiekowsky, David Garlock, Irving Glick, David Liu, Deborah Minehart, Mitchell Petersen, Michael Singer, and Jeff Strnad, for helpful comments on portions of my dissertation; to Gary King, without whose cheerful attentiveness so many things in the M.I.T economics department would quickly be derailed; to Jerome Barron, the George Washington University Law School Dean who approved my starting down this road; to his successor, Jack Friedenthal, whose indulgence permitted me to continue to the end; and to Rod French, George Washington’s Vice President for Academic Affairs, whose blessing it all required. Financial support from the John M. Olin Foundation and the Dean’s Fund of the George Washington University Law School is very gratefully acknowledged.

I have yet to mention the three most special participants in this entire venture. First there are the cats, Safid (the fat white cat) and Lavendar (the little grey rat). They are litter-mates, born in 1980, who consented to travelling down this road with me, but only after tearing up Judge Jim Halpem’s hand, as he tried to trap and cage them for the plane ride from Washington to Boston, which they successfully negotiated only after an unscheduled over-night layover in Eastern Airlines baggage claim at LaGuardia. They arrived little the worse for wear, and remained faithful the entire first year, recognizing (Linda tells me) my return home each day by the sound of the car in the alley between Beacon and Marlborough Streets, observing from a perch atop the desk until I had disappeared from sight, then decamping to await me by the door. Safid did not make it to the end. In December 1988 he developed a mysterious ailment that Wayne Shapiro and the dedicated staff at Angell Memorial Hospital could not forever keep in check. Still, weak though he became, he continued to descend late each night from a nest among the forest of (never unpacked) cartons at the foot of the bed to perch atop my chest, his nose an inch from mine, his motor running, lulling me to sleep. He died in the summer of 1989. We buried him on Nantucket. Lavendar made it back to Washington, and has supervised all aspects of my dissertation, despite a serious bout with cancer, which, with wonderful
veterinary care -- from Wes Bales, Lisa Fulton, and David Saylor -- she somehow managed to fend off. As I write, daylight is just breaking in Washington, D.C., and she is sitting on a chair in the dining room (her fourth most favorite spot in all the world), grooming in preparation for the day. It's a minor miracle that she's still here.

That leaves Linda T. M. Bui, the unwisest of them all, for picking on someone 20 years her senior. We met the first day of classes, and have become steadily closer ever since. Together, we have been through examinations, econometrics papers, more examinations, general examinations, dissertations, the job market, constant shuttling between Washington and Boston, the death of two beloved cats, the gradual but ineluctable decline of yet another, and the vicissitudes of life in general. None of it seemed easy. Always, however, it has been worth the while. We shall remain together. But, difficult to believe though it may be, this, my peculiar mid-life crisis and our introductory chapter, commenced on the heels of an Autumn funeral in 1985, has finally reached its end.
Theodore S. Sims

Home: 2230 California Street, N.W. Apartment 6-C West  
Work: George Washington University National Law Center  
Washington, D.C. 20008  
(202) 462-7467  
415 Burns Hall  
Washington, D.C. 20052  
(202) 994-8326

Employment

1992-Date  Professor of Law, George Washington University, Washington, D.C.
1981-1992  Associate Professor of Law, George Washington University, Washington, D.C.
1990  Visiting Professor of Law, Boston University School of Law, Boston, Massachusetts
1977-1981  Office of Tax Policy (Office of Tax Legislative Counsel), United States Treasury
1970-1971  Law Clerk, Honorable John C. Godbold, United States Court of Appeals, (then) Fifth Circuit, Montgomery, Alabama

Education

Ph.D (Economics), 1995, Massachusetts Institute of Technology

Doctoral Research

1995  "Risk is Bad, and Time is Money, So Sometimes it Pays to Wait: Optimal Tax-Motivated Selling in a Simple Two-Period Model"
1995  "January Peaks and December Valleys: Fresh Evidence on Year-End Return Abnormalities"
1995  "Two Applications of the Samuelson Invariance Theorem"

Principal Work

1994  "Environmental ‘Remediation’ Expenses and a Natural Interpretation of the Capitalization Requirement", 47 Nat’l Tax J. 703-719

1994 "Reflections on The Long Period in Taxation: The Distributional Impact of Retirement of the Civil War Debt" (M.I.T. mimeo, April 25) (joint with Daniel Kessler)

1994 "Two Propositions on Haig-Simons Taxation and Financial Distress" (M.I.T. mimeo, February 14)

1992 "Long-Term Debt, The Term Structure of Interest, and the Case for Accrual Taxation", 45 Tax Law Rev. 313-375


1983 "The Equity Return to Mutual Life Insurance Companies and their Policyholders", in House Committee on Ways and Means, Serial 98-39, 98th Cong., 1st sess., at 591 (principal author)


Other

John M. Olin Fellowship for the study of law and economics (M.I.T 1989-1991)
ALI, ABA, AEA
Editor, Florida Tax Review

April 1995
# Table of Contents

**Introduction** .................................................. 19

## Essay One

January Peaks and December Valleys:
Fresh Evidence on Year-End Return Abnormalities

A. Introduction .................................................. 23
B. Existing Empirical Work ...................................... 28
C. Data and Methods ........................................... 32
   1. Data .................................................. 33
   2. Methods ............................................... 41
D. Results ...................................................... 42
   1. The Year-End Depression in Returns .................. 42
   2. The Relationship Between the January Effect and the Year-End Depression in Returns .................. 44
E. Turn-of-the-Year Effects and Tax-Loss Potential ............ 47
   1. Methods ............................................... 47
   2. Data and Results ...................................... 49
      a. Tax-Loss Potential and the January Effect ....... 52
      b. Tax-Loss Potential and Year-End Returns .......... 54
      c. The Influence of Market Capitalization and Volatility .... 55
F. An Interpretation ........................................... 56
G. Conclusion .................................................. 59

## Essay Two

Risk is Bad, But Time is Money, So Sometimes it Pays to Wait:
Optimal Tax-Motivated Selling in a Simple Two-Period Model

A. Introduction .................................................. 63
B. Tax-Motivated Selling ...................................... 65
   1. The Problem ........................................... 65
   2. The Treatment of Capital Losses ...................... 66
   3. Plausibility of the "Optimal" Loss Selling Assumptions ........................................... 68
C. Framework .................................................... 69
   1. Assumptions ........................................... 70
   2. Transition Relationships ................................ 71
   3. Corners in the Payoffs to Waiting or Exchanging ... 74
D. An Illustration ................................................................. 75
   1. Description ............................................................... 75
   2. Results ................................................................. 79
      a. Large \( \epsilon \): Waiting in Both Periods and Exchanging in Both Periods Both Allowed ............................................ 79
      b. Small \( \epsilon \): Both Waiting and Exchanging in Both Periods Not Allowed .................................................. 82
E. A More General Formulation ........................................... 83
   1. The Problem .............................................................. 84
   2. Characteristics of the Solution ...................................... 89
   3. Optimal Tax-Loss Selling ............................................ 91
F. Implications ............................................................... 95

APPENDIX ................................................................. 97

ESSAY THREE
Two Applications of
The Samuelson Invariance Theorem

A. Introduction ............................................................... 105
B. Economic Depreciation and Invariant Asset Valuation ........ 106
C. Environmental "Remediation" Expenses and a Natural Interpretation of the Capitalization Requirement ........................................ 108
   1. The Problem .............................................................. 108
   2. An Example ............................................................. 110
   3. Costs, Explicitly Considered, and a Natural Capitalization Rule . 113
   4. Future Environmental Costs ........................................ 117
   5. Economic Depreciation and Cash-Equivalent Accounting ........ 119
D. Haig-Simons Taxation and Financial Distress ..................... 123
   1. The Treatment of Interest in Arrears and Debt in Default .... 124
   2. Interest in Default and Accrual Invariance ..................... 126
   3. Haig-Simons Neutrality of COD ................................... 130
   4. Conclusion ............................................................. 132

REFERENCES ............................................................. 139
LIST OF TABLES

Table I-1: Summary Statistics -- Year-End and January Compound Returns (NYSE/AMEX 1964-1990) .................................................. 36

Table I-2: Mean Year-End and January Returns by Market Equity and Own Standard Deviation (NYSE/AMEX 1964-1990) .......................... 38

Table I-3: Mean Difference in Average Weekly Returns: End-of-Year vs. Rest of Year (NYSE/AMEX 1964-1989) ........................................ 43

Table I-4: Estimates of Equation (1) .............................................. 45

Table I-5: TL Summary Statistics (NYSE/AMEX 1964-1989) ................. 49

Table I-6: Means of Tax-Loss Variables (NYSE/Amex 1964-1989) .......... 51

Table I-7: Estimates of Equation (4) ............................................... 53

Table I-8-A: Decile Estimates of Equation 4(a) (By Increasing Capitalization) .......................................................... 61

Table I-8-B: Decile Estimates of Equation 4(a) (By Decreasing Volatility) .......................................................... 62

Table II-1: Price/Basis Ratios For Waiting (Binomial Probability) ......... 79

Table II-2: $\pi_{(i, i)}$ and $\theta$ ......................................................... 96

Table III-1: Economic Depreciation .................................................. 135

Table III-2-A: Computation of Economic Depreciation With Final Cost .......................................................... 136

Table III-2-B: Computation of Taxable Income With Final Cost Using Economic Depreciation .......................................................... 137
LIST OF FIGURES

Figure I-1: The January Effect (1965-1990) ........................................... 34

Figure I-2: Mean Year-End and January (Week 1) Yields by (Increasing) Capitalization and (Decreasing) Volatility (1964-1990) ......................... 38

Figure I-3: Mean May-June and July (Week 1) Yields by (Increasing) Capitalization and (Decreasing) Volatility (1965-1990) ................................. 40

Figure I-4: Spline Estimates of Slope Coefficients on YE (Equation 1(b)) ........ 46

Figure I-5: Tax-Loss Variables by Decile .............................................. 51

Figure II-1: Discrete-Choice Decision Tree .......................................... 76

Figure II-2: $q$ and $\pi_{(1,1)}(q)$ ...................................................... 81

Figure II-3: $\pi_{(1,1)}(0.50)$ and $\pi_1^*$ ........................................... 94

Figure III-1: An Arrearage Function ................................................... 128
INTRODUCTION

The subject of these essays is taxation. They otherwise reflect a range of interests, some brought with me to the formal study of economics, others acquired along the way. And they proceed from different points of view. The first two are devoted to the positive study of an aspect of the influence of taxation on behavior, specifically the relationship between tax-motivated stock market activity and the January seasonal effect. Essay One suggests, more clearly than empirical work to date, that the January seasonal is importantly related to tax-motivated trading induced by an entrenched imperfection of the income tax; in the face of contrary findings of prior theoretical work, Essay Two demonstrates that it is a priori plausible to think that such a relationship exists. In contrast, Essay Three takes up what in some sense is a converse normative issue, that of designing rules that minimize the impact of income taxation on behavior.

The January seasonal is the premier stock market anomaly. It has long been thought, naively in the eyes of many, to be related to year-end loss-taking by taxable investors. To the extent that such behavior actually takes place, it is attributable to the "realization requirement" of the income tax: with some exceptions, gains and losses from the ownership of assets are not taken into income as they accrue, but only as they are realized through (generally voluntary) sale or other disposition. That feature of the law confers on investors a "timing option," that arises because taxation of asset gains may be deferred, while losses may be established and deducted at whatever time an investor happens to choose.

An issue that has dominated the study of tax-loss selling and the January effect is how investors best should exercise that option. Work by Constantinides (1983, 1984) suggests that, behaving optimally, investors should take losses as soon as they appear. If so, loss-taking should be spread throughout the year, rather than being concentrated at the end. That claim has been a source of persistent discomfort to empirical students of January returns, whose efforts have consistently suggested a connection between year-end activity and the January seasonal effect. A substantial body of evidence, beginning with Dyl (1977) and Branch (1977), documents both elevated year-end volume (Dyl) and
elevated January returns (Branch) among issues with high year-end loss-sale potential, suggesting that year-end tax-motivated selling is related to the January effect.

Essay One contributes to that line of work by studying daily returns derived from the master files of the Center for Research in Securities Prices ("CRSP") for a 35-day period that includes the last 26 trading days in each year. In those data I document, for the first time so far as I can tell, a year-end depression in returns. That depression is concentrated among securities with loss-sale potential, has properties (like being concentrated among small capitalization issues) that mirror those of the January effect, and is statistically related to the January effect itself. The evidence thus suggests both that the year-end is characterized by disproportionate loss-selling, and that year-end loss selling is related to the January effect.

In isolation, the findings of Essay One could (and probably would) be regarded as just another ad hoc set of empirical results. So, in Essay Two, I take up the theory. Constantinides' conclusions rest on a set of assumptions, acknowledged to be abstractions from the law, that, I argue, are questionable approximations of reality. In Essay Two I relax but one of the debatable assumptions -- that short-term capital losses are uniformly more valuable than long-term losses -- and take a fresh look at the problem of optimizing tax-motivated loss realization. Using a simple two-period control model, in which an investor must choose whether to realize an available loss early in the year or at its end, I find, with surprisingly few restrictions on the probability distribution of security price movements, that the investor often does better to wait. That contradicts Constantinides' claim that losses are best realized whenever they appear. More importantly, it is consistent not merely with the evidence of year-end tax-motivated selling developed in Essay One, but with the findings of virtually all empirical work on the January effect to date.

In sum, Essays One and Two are both theoretically compatible and empirically consistent with the hypothesis that investors do exercise the timing option by voluntarily realizing losses; but that they do so disproportionately at the end of the year; and that, by doing so, they are not (as has almost universally been believed) behaving in an obvi-
ously sub-optimal way. They suggest, more clearly than has hitherto been the case, that tax-induced but otherwise optimal conduct is an important determinant of the January seasonal effect.

The source of these phenomena -- the realization requirement itself -- has been a feature of the United States income tax since its inception. It nevertheless remains controversial, because it is so glaring a departure from the income tax ideal of taxing changes in net worth as they accrue. That ideal is more than a vacant abstraction. Its significance was firmly established more than 30 years ago, when Samuelson (1964) showed that if and only if changes in value are taxed as they accrue will the valuation of assets in the presence of an income tax be independent of their holders' marginal rates. Thus, under an income tax, only accrual taxation of changes in net worth (including asset gains and losses) can eliminate differential and distortionary taxation of capital income. From this perspective, on the evidence of Essays One and Two, the January seasonal can be viewed as the reflection of time-concentrated, tax-induced distortions of behavior. Stated differently, if all securities gains and losses were annually marked to market, it is a plausible conjecture that the January effect would disappear.

Samuelson's invariance theorem itself is something of a gem. It is one of the most influential pieces ever bearing on policy discussions of the income tax. Even so, the breadth of its implications are neither as widely understood nor as well appreciated as they deserve to be. Formulated to obtain a theoretically ideal system of depreciation, it more generally is the key to (income) tax neutral asset valuation. In Essay Three I use that theorem to study two income tax problems that are still imperfectly understood. The first is whether it is proper to allow business outlays (sometimes called "reverse" investments) that will be made in the future to be "depreciated," by allowing some portion of their cost to be deducted now. A decade ago, after extensive debate, the

---

1 I would judge the others to consist of Brown (1948), who showed that expensing the acquisition cost of long-lived assets is equivalent to exempting them from tax; Mirlees (1971) on progressivity; and, of course, Robert Murray Haig (1921) and Henry Simons (1938).
answer Congress gave was "no". With a simple, explicit modification of Samuelson’s original formulation, I show generally that the better answer would be "yes".

The second question might seem a bit more far afield. It concerns the "cancellation of indebtedness" (or "COD") rule, under which debt is taken into income at the time it is forgiven. The COD rule is widely regarded as an ad hoc palliative that taxes a defaulting debtor inadequately, because it taxes him later than when the loan proceeds were originally received. But it has been the object of little theoretical attention. By the simple specialization of an earlier result, also due to Samuelson (1937), I show that the COD rule does produce correct (ex post) accrual taxation, provided that, as required by the invariance theorem itself, we allow debtors to deduct interest on indebtedness as it accrues. That, however, is a far cry from what we actually do. Debtors using the cash method of accounting are permitted to deduct interest only when it is paid. And, by reason of the statutory limitations that have flourished in the last 20 years, much interest is not deductible at all. The application of the invariance theorem to the treatment of debt cancellation thus sheds critical light on, and calls into question the wisdom of, the byzantine (and, I personally believe, seriously misguided) collection of special limitations that we have in recent years deployed to govern the income taxation of debt.

Working backwards, then, what follows is about taxation and behavior. It concludes by studying how we might design rules that minimize the influence of an income tax on the decisions of economic actors. It begins, however, by studying the kinds of things that happen when we do not.

---

ESSAY ONE

January Peaks and December Valleys:
Fresh Evidence on Year-End Return Abnormalities

A. Introduction

The turn-of-the-year spike in U.S. securities markets known as the January "seasonal effect" is probably the most intensively studied market regularity. First detected about 20 years ago, it was relatively quickly documented by Keim (1983), Reinganum (1983), and Roll (1983) to be most pronounced among small capitalization stocks and concentrated at the beginning of the month. The possible explanation that has been most extensively studied -- that the January seasonal is a backwash from "price pressure" induced by year-end tax-motivated selling of loss securities -- is simultaneously regarded as the most naive.

The grounds for skepticism are several. They include the conviction that, in efficient markets, abnormal sales activity unrelated to new information should have no effect on price; and, even if it did, that the resulting disequilibrium would be recognized for what it was and quickly arbitraged away. In practice, moreover, as noted by many, including Reinganum (1983) and Brown, Keim, Kleidon & Marsh (1983), if "the price pressure argument were true, we should expect to observe a general decline in prices of small firms in December." The latter observe, however, that there is better evidence of December increases than of December declines in the prices of issues among which the January effect tends to be pronounced. In the same vein, Roll (1983) observed that, even after excluding the last trading day from December returns, there is no evidence of systematic price declines during that month; more recently, Lakonishok and Smidt (1988) have documented an end-of-December rally using the historical series of the DJIA.

---

3 For this proposition Brown, Keim, Kleidon & Marsh rely on unpublished work by Keim (1982). They claim also that the logic of the "price pressure" argument would imply an April depression in prices (and resulting "spring seasonal") as investors were forced to liquidate securities to pay income tax bills. That particular claim seems to overlook the existence of wage withholding and the requirement of making quarterly estimated tax payments, both of which tend to spread the collection of taxes over the entire calendar year.
Perhaps most importantly, the notion that investors should for tax-motivated reasons unload loss stocks in December has lacked a theoretical foundation. The theory that exists has been quite influential, and it runs just the other way. As developed by Constantinides (1983, 1984), tax-motivated selling, when carried out optimally, should not be concentrated at the end of the year. It should be (or, at least until 1986, should have been) pursued throughout the year, in the interests of maximizing taxable investors' stocks of short-term losses,4 which Constantinides modelled as uniformly more valuable than long-term losses, and as usable essentially without limitation. The continuous realization of short-term losses implies that nothing special should occur at the end of the year, leaving nothing of a tax-motivated nature from which prices might be expected to recover in January. To think otherwise, in Constantinides' (1984) words, is to "assume irrationality or ignorance on behalf of investors."

Nevertheless, there has been a steady accumulation of evidence that seems consistent with the tax-loss selling hypothesis. This paper extends that line of work. I depart from prior research, and take up the study of returns derived from daily data in the CRSP master files for 26 trading days at the close of each calendar year. In those data I document, for the first time so far as I am aware, that a year-end depression in returns does indeed exist. That depression presumably has been obscured in December monthly data in part by the fact (first noticed by Roll) that the onset of the January effect typically precedes by one trading day the actual turn of the year,5 producing an upturn on the last day of December that offsets the depression earlier in that month.

Furthermore, the year-end depression is not widely dispersed. To the contrary, it exhibits properties that are in many respects the mirror-image of those exhibited

4 A loss on a "capital asset" that is "realized" through disposition before the expiration of the short-term "holding period", the length of which has ranged between six months and one year during the past 30 years, is characterized for tax purposes as a "short-term capital loss"; thereafter it becomes a "long-term capital loss." Internal Revenue Code §§ 1222, 1223.

5 Lakonishok and Smidt (1988) also find historical evidence of a tendency of the DJIA to rise during the final week in December. As noted in Section B, Roll's finding is not free of controversy.
by returns during the first week of January. It is deepest among, but not entirely confined to, small-capitalization issues, where the January effect is most pronounced. I also document similar (and only slightly less striking) relationships between total volatility (as measured by own-variance) and both these phenomena: the year-end depression tends generally to be deeper and the January effect to be more pronounced for high-volatility than for low volatility stocks. I confirm the relationships suggested by these mirror-image properties using ordinary least squares regressions of the January effect on end-of-the-year returns. Coefficient estimates are uniformly of the expected sign and significant, and the strength of the estimates declines, ultimately to noise, with either increasing market capitalization or decreasing volatility.

Extending the work of others, principally Chan (1986), Reinganum (1983) and Roll (1983), I then delve further into the tax-loss selling hypothesis by investigating directly the relationship between year-end loss-sale potential and both the January effect and the year-end depression in returns. In contrast with prior work I run least squares regressions of the January effect on a newly devised measure of loss-sale potential, controlling directly for market capitalization. The January effect is unambiguously increasing in loss-sale potential, with both the statistical strength of the relationship and regressions' explanatory power declining almost monotonically with market capitalization and increasing with own volatility. Similar tests disclose comparable relationships (but with attenuated explanatory power) between tax-loss potential and the year-end depression in returns.

A notable feature of this evidence is that long-term year-end loss selling potential is as significantly related as short-term potential to the strength of the January effect. That is consistent with findings by Chan (1986), De Bondt and Thaler (1985, 1987), and others, and inconsistent with investor pursuit of "optimal" tax-loss selling along the lines formalized by Constantinides. In light of that fact, both Chan and De Bondt and Thaler infer that evidence of the significance for January returns of poor performance in the distant past, by "its mere presence, . . . contradicts rational tax-loss selling as an explanation of the January effect." De Bondt and Thaler (1987).
I come to the question with different prior beliefs. Constantinides’ models incorporate a series of assumptions, acknowledged to be abstractions from the law, that contribute to the predictions the models produce. In particular, as I have suggested elsewhere (Sims 1992) and has occasionally been recognized by others (e.g., Lakonishok and Smidt (1987)), the maintained assumption that short-term capital losses are, or ever were, uniformly or even importantly, more valuable than long-term losses is a drastic and debatable simplification. Until 1986, the first $3,000 of short-term capital losses was, for taxable individual investors, as valuable as the first $6,000 of long-term capital losses. Above those modest limitations, however, either species of capital loss was (and remains) deductible without limitation against both long and short-term capital gain. The capital gain netting rules of section 1222 of the United States Internal Revenue Code do require an initial matching of short-and long-term losses and gains; but, in the final analysis, all capital losses are deductible against all capital gains. What is more, the ultimate effect of any particular (short or long term) loss realization cannot be determined with certainty until all capital gains and losses (both short and long term) are accounted for at the end of each year. In fact, given the netting rules, the marginal consequence of realizing a long-term loss can easily be to reduce an investor’s (fully taxable) short-term capital gain; and the marginal consequence of realizing a short-term loss can just as easily be to reduce the investor’s taxable "net" -- *i.e.*, favorably taxed, long-term -- "capital gain". To this one must add the obvious fact that the realization of gains is both costly and by and large voluntary, and that investors have every incentive, conditional on a decision to realize a gain, to ensure its realization as a long-term capital

---

6 Under §§ 1211(a)-(b) and 1222 of the Internal Revenue Code, $3,000 of losses realized by individuals (but not by corporations) could be deducted against ordinary income, but it took $2 of long-term capital loss to produce $1 of deduction. That distinction was repealed in 1986. In contrast, both long-term and short-term capital losses are deductible against both long-term and short-term capital gains, although long-term losses must first be applied against long-term gains, and conversely for short-term losses.

7 For a recent example of the sort of complications these rules can produce in analyzing the tax consequences of particular events, see Strnad (1994).
gain. In the final analysis, modelling assumptions more plausible than Constantinides' would seem to be that most realized capital losses are deducted against realized capital gains, that most realized capital gains are voluntarily realized long-term capital gains, and hence, at least on average, that short and long-term capital losses tend to be equally valuable.

With those assumptions, much of the foundation for the belief that investors should hasten to stockpile short-term losses disappears, and with it the grounds for interpreting evidence of a relationship between long-term loss sale potential and the January effect as contradicting the tax-loss selling hypothesis. That leaves us to consider anew how investors should, in the face of uncertainty about future movement of prices, optimally deal with unrealized losses. In Essay Two, where I relax only Constantinides' assumption of differential value of short and long-term losses, I show that an investor, faced with a choice about whether to realize a loss that is available now when waiting until the end of the year might allow it to disappear, will frequently do better to wait. That model has testable implications, many consistent with empirical work to date, most directly for year-end volume. Here, however, in the context of a study of loss-sale potential, the January effect, and year-end returns, I take evidence of the statistical significance of long-term loss sale potential as at least indirectly supporting the prediction of elevated, tax-motivated year-end sales of securities at a loss, and as suggesting that existing theoretical descriptions of optimal tax-loss selling offer an account that does not comport well with investor behavior.

The balance of this Essay is organized as follows. Section B reviews in more detail the literature on tax-motivated selling and the January effect. Section C deals with data and methods: it specifies the tests to be conducted of the relationship between year-end returns and the strength of the January effect, and the construction of the data used both to identify the year-end depression in returns and in carrying out those tests.

* Or at least to realize it at the lowest possible rate, which would lead one to have expected accelerated realization of long-term capital gains before capital gains tax rates went up as the result of the Tax Reform Act of 1986, as documented in Bolster, Lindsey, and Mitrusi (1989).
Section D presents those results. Section E then investigates the relationship between tax-loss selling potential and both year-end and January returns. It describes the construction of the measure of tax-loss potential to be used, specifies the tests, and reports the additional results. Section F proposes an interpretation of the findings.

B. Existing Empirical Work

The January effect appears first to have been noticed in U.S. data by Rozell and Kinney (1976), and its existence has since been documented early in the century. Brown, Keim, Kleidon & Marsh (1983) report evidence⁹ that the phenomenon existed as early as the 1930’s, while Jones, Lee and Apenerbrink (1991) detect abnormal January returns prior to 1920. Keim (1983) was the first to document a connection between the January effect and the abnormal returns earned by small capitalization issues (the "small-firm effect"), showing that much of the small firm premium is attributable to January returns. Keim (1983), Reinganum (1983) and Roll (1983) all found the January effect to be concentrated predominantly in the first few trading days of the month, a finding routinely corroborated in subsequent work. Roll found also that the January effect actually commences on the last trading day in December, although, on the basis of intraday trading data, Griffiths and White (1993) have recently suggested that this particular finding may be attributable to Roll’s use of close-to-close price comparisons in computing daily returns.

Although work on the determinants of the January effect has involved a variety of methodologies, much of it can be broadly divided into studies of year-end volume, on the one hand, and studies of turn-of-the-year returns. The first line of inquiry originates with Dyl (1977), who, in a sample of 100 securities over a ten-year period, found evidence of an inverse relationship between appreciation during the year and year-end trading volume, with depreciated securities exhibiting abnormally high December volume and conversely. More recent work includes Lakonishok and Smidt (1984, 1986), who, while finding generally that past winners (losers) tend to exhibit abnormally

⁹ Attributed to unpublished work by Keim (1982).
elevated (depressed) volume, found also that the differences narrowed significantly during December, consistent with elevated year-end volume in losers. In studying the year-end impact of the rise in capital gains tax rates enacted in 1986, Bolster, Lindsey and Mitrusi (1989) likewise document elevated year-end volume in "losers".

The earliest work to establish a relationship between loss sale potential and turn-of-the-year returns appears to be Branch (1977), who in 1965-1974 data found evidence of elevated January returns among securities that had achieved their lows during the closing week of the preceding year. Roll (1983) and Reinganum (1983) substantiated such relationships in more detail. Roll used cross-sectional regressions of returns on individual securities (in 1962-1979 NYSE/Amex data) to document significant, negative relationships between annual returns (computed excluding the first and last five trading days in the year) and the ensuing January effect. He interpreted these findings as consistent with tax-loss selling as an explanation of the January effect, relating that interpretation to the predominance of the January effect among small-capitalization firms with the conjecture that such firms exhibited higher volatility and were therefore more likely to be candidates for tax-loss selling.

Also using 1962-1979 data, Reinganum stratified securities annually into deciles by market capitalization, and ranked each security by the ratio of its closing price to high price (denoted its "PTL") as a way of capturing its potential for year-end sale at a loss. Portfolio tabulations by PTL quartile disclosed that the "losers" were disproportionately small cap stocks; while comparisons of January returns across portfolios, after being stratified by PTL, showed the January effect to be most pronounced among "losers," regardless of capitalization. Reinganum interpreted his findings as consistent with the conclusion that both the January effect and much of the small firm effect were related to tax-motivated selling.

Reinganum's findings were extended by Chan (1986), who subdivided each of the four smallest MV portfolios constituted annually by CRSP into sub-portfolios by PTL. Each security was ranked by PTL for each of the two preceding years. Treating
securities in the highest PTL quartile as "loss" securities,\textsuperscript{10} Chan assigned each security to a sub-portfolio by whether it was a "loser" in both preceding years, a "winner" in both, or a winner in one but a loser in the other. This four-way classification was used to estimate the relationship between mean January returns and loss-sale potential, distinguishing between short and long-term loss potential. Chan found that, as measured by PTL, both long-term and short-term loss potential were positively related to January mean returns. In light of Constantinides' theoretical work, however, and in contrast with both Reinganum and Roll, Chan interpreted his findings as \textit{contradicting} the tax-loss selling hypothesis.

In pursuit of a different quarry -- the tendency of investors to "overreact" to current information -- De Bondt and Thaler (1985, 1987) develop evidence that is strikingly consistent with that of Chan. Portfolios of securities that had substantially underperformed the market during a five-year base period tended to outperform the market for several subsequent years. The premium performance in succeeding years was concentrated almost entirely in January. (See especially De Bondt and Thaler (1985), Figure 3.) Moreover, their data seem depict a persistent pattern of fourth-quarter declines, accompanied by abnormal returns in January, and could easily be interpreted as suggesting a year-end selloff, findings for which De Bondt and Thaler themselves could offer "no satisfactory explanation . . ., rational or otherwise". De Bondt and Thaler (1987, at 579.)

An alternative explanation of the January effect, also consistent with its being pronounced among issues that had suffered recent losses, is the "window-dressing" hypothesis of Haugen & Lakanishok (1987), who attribute it to professional money managers "cleaning out" losers towards the end of the year in an effort to look more like the crowd, after which they take more risky positions at the beginning of the year.

\textsuperscript{10} By construction, Reinganum's PTL is actually \textit{decreasing} in loss potential, while the statistic developed for this study (described in Section E) is \textit{increasing} in loss potential. In the interest of consistency I will use the term "highest," with respect to any measure of loss potential (however constructed), to denote those securities having the \textit{greatest} potential for producing losses.
Historical data have been used to distinguish between the two competing accounts, by studying differences between the January effect before and after tax-loss selling first became advantageous with the enactment of the War Revenue Act of 1917. Using different data, Shultz (1985) found, but Jones, Pearce, & Wilson (1987) ("JPW") did not, that the appearance of the January effect seemed to coincide with that enactment. Jones, Lee, and Apenbrink (1991) ("JLA") surmised that the conflict was attributable to differences in data sets and methods of computing January returns. Using a refinement of the JPW data, JLA studied returns during a nine-day period (beginning on the last trading day in December) for 1899-1928. They document a significant increase in turn-of-the-year returns commencing in January 1918, an increase that is pronounced among both small capitalization stocks and issues with high tax-loss potential. Consequently, JLA conclude that tax-loss selling provides a better explanation of the January effect than the "window-dressing" hypothesis.

In sum, much of the evidence is consistent with what one naively would expect if tax-loss selling is an important contributor to the January effect, especially if investors act on the assumption that long and short-term losses are of roughly equivalent value, and do not pursue loss-selling strategies of the sort envisioned by Constantinides. Both long and short-term loss-sale potential appear related to the strength of the January effect, and there is evidence relating loss-sale potential to elevated year-end volume. More recently, Ritter (1988) has suggested that there is a year-end depression in the ratio of buying to selling by individual investors that reverses at the beginning of the year. And Griffiths and White (1993) suggest that, in United States data, trades until late in the final trading of the year tend to be seller-initiated, while after the turn of the year they tend to be buyer-initiated; and that, while a similar shift is discernible in Canadian data, it occurs five days before the end of the calendar year, which coincides with the point at which, at least for purposes of taxing securities gains and losses, the Canadian tax year ends.

Griffiths and White are atypical in their attention to year-end daily data. Indeed, despite the long-standing availability of daily data, despite its extensive use in
studying returns during January, and despite evidence that the January effect commences at the end of December, thereby influencing the appearance of December monthly returns, almost no attention has been devoted to date to the return content of year-end daily data.

Given the evidence of elevated year-end volume in losers, that inattention is surprising. If the January effect is related to tax-loss selling, and if markets are not quite as efficient as we might like,\(^{11}\) or, indeed, if the demand curve for securities is not entirely flat, we would not be entirely surprised to find a depression in year-end returns, associated with elevated year-end volume, in securities that are candidates for year-end sale at a loss. Nor would we be surprised by a negative relationship between that depression and the January effect. What is more, if losses are realized primarily to offset capital gains, it should not be important, as the work by both Chan and De Bondt and Thaler already suggests, whether the loss is a long-term or short-term loss. Finally, one might also expect both year-end tax-loss selling and the strength of the January effect to be influenced by the overall performance of the market, although it is not clear a priori in which direction such an influence would cut. A strong market would produce fewer year-end losers to be sold at a loss, so that the January effect might be weaker following a year in which the market had been strong. On the other hand, if a strong market is positively correlated with the realization of gains, it would strengthen the incentive to clear out losers at the end of the year.

In all events, it is with the return content of year-end daily data that the work reported here begins.

C. Data and Methods

My principal object of study is a set of daily returns for a 35-day period that spans the turn of the year, with the principal objective of investigating (1) whether anything remarkable does occur late in November and in December, and (2) the relationship

---

\(^{11}\) Or if the inefficiencies cannot cheaply be identified and arbitraged, as suggested by, e.g., Bhardwaj and Brooks (1992).
(if any) between year-end returns and the January effect, by estimating variants of the basic equation:

\[ JE_{i,t+1} = \alpha + \beta YE_{i,t} + \theta_1 TL_{i,t} + \theta_2 TL_{i,t-1} + \delta_1 MV_{i,t} + \delta_2 Mk_{i,t} + \epsilon_{i,t} \]

where \( JE_{i,t+1} \) denotes January returns for security \( i \) in year \( t+1 \), \( YE_{i,t} \) is the security's return at the end of year \( t \), \( TL_{i,j} \) are measures of loss sale potential determined using its volume-weighted price history for years \( j = t \) and \( t - 1 \), and \( MV_{i,t} \) and \( Mk_{i,t} \) are controls for the market value of security \( i \) at the close of year \( t \) and the performance of the market in year \( t \), respectively.\(^{12}\)

1. Data

The 35 days of daily returns were drawn from the CRSP NYSE/Amex master file, and consist of the last 26 trading days in each calendar year and the first 9 trading days in the following January, covering the period 1964-65--1989-90.\(^{13}\) The mean yields for each day in the sample, pooled across securities and years, are depicted in Figure I-1.

\(^{12}\) MKT\(_i\) is the return on the value-weighted daily market index reported by CRSP, compounded in each year for a period spanning the 10th to the 27th-to-last trading day in the year.

\(^{13}\) The decision to open the sample with 1964, even though the first full year of CRSP daily returns is 1963, was dictated by the necessity of constructing statistics for each year under study using daily data from the preceding year as well.
Figure I-1 clearly shows, as Roll previously observed, that the "January" effect begins with the last trading day of the preceding year (Day 26 in the sample). Consequently, in the work reported here, the final trading day in each year was treated as part of the January effect.\textsuperscript{14} Because the January effect is so concentrated into the last trading day in December and the first four trading days in the year, daily returns at the beginning of the year were compounded into two, five-trading day "weeks" in "January"\textsuperscript{15} (hereafter denoted $W_{0i}$, $i = 1, 2$). The twenty-five days of returns from November and December -- excluding the last trading day in each year -- were compounded into a single year-end return variable, denoted $YE$.

\textsuperscript{14} The considerations that initially guided the construction of the data are discussed in more detail in Sims (1992). The original design was settled on in ignorance of Roll's (1983) finding and prior to Griffiths and White (1993), and was not influenced by a prior belief that the last trading day in the year should be treated as part of the January effect. When that preliminary finding emerged, however, the decision was made to carry out subsequent work in light of that finding. Griffiths and White (1993) conclude that trading activity before the close of the final trading day of the year is predominately seller-initiated, suggesting that much of the abnormal return on the final trading day occurs late in that day.

\textsuperscript{15} The term "trading weeks" is used loosely: although the data set was constructed by compounding returns into five-day periods, those periods in general do not correspond with calendar trading weeks.
Two versions of the data were constructed, one using daily raw returns drawn directly from the CRSP master files, and a second consisting of excess returns, computed as the residual difference between the security’s daily raw return and a return fitted using a Sharpe-Lintner capital asset pricing model, or

\[ RE_{ikt} = R_{ikt} - \left[ Rf_{kt} + \beta_{it-1} (Rm_{kt} - Rf_{kt}) \right], \]

where \( R_{ikt} \) denotes the CRSP raw return of security \( i \) on day \( k \) in year \( t \). The market return \( (Rm_{kt}) \) was taken to be the daily return on the value-weighted market index contained in the CRSP files; the risk-free rate \( (Rf_{kt}) \) was the daily equivalent of the annual yield on 90-day T-bills drawn from a series obtained from the Federal Reserve; and each security’s beta \( (\beta_{it-1}) \) is a Scholes-Williams beta, computed by CRSP using data from the immediately preceding year.\(^{16}\)

The raw returns data set contains 53,958 observations on turn-of-the-year returns. For each year from 1964-65 to 1989-90, the sample included every security in the NYSE/Amex file (1) for which a valid standard deviation, beta, and market capitalization statistic was reported by CRSP that year,\(^{17}\) and (2) for which, during that year, not more than five days of daily raw returns were missing from the 35-day turn-of-the-year interval under study. The excess returns data set contains 50,806 observations on turn-of-the-year returns drawn from the same period, satisfying the two criteria above, and (3) for which (as dictated by the construction of the excess returns specified in equation (2)) CRSP was able to compute a beta for the immediately preceding year.\(^{18}\)

Table I-1 summarizes the means, standard deviations and maximum and minimum values of the returns in each sample for the year-end variable \((YE)\), and for each

\(^{16}\) The CRSP betas are computed using data for the entire year, excluding the first and last two trading days in each year. The choice of \( \beta_{it-1} \) was dictated by the desire to avoid the use of betas constructed from data that overlapped the periods of daily returns under study.

\(^{17}\) A security was rejected if the CRSP-reported value for its standard deviation or market capitalization was non-positive, or if its beta was less than -5.

\(^{18}\) The size discrepancy in the data sets stems from this requirement.
January trading week (the $WJ_i$), pooled across both securities and years. Panel A (Panel B) of Table I-1 contains the summary statistics on the excess (raw) returns data. As with Figure I-1, Table I-1 reflects the concentration of the January effect into the first trading week in the year: the mean value for the second January trading week ($WJ_2$) is substantially closer to 1. Consistent with the observations of e.g., De Bondt and Thaler (1985), virtually all results were robust to the specification of the data. Consequently, apart from Table I-1, only the excess returns results are reported (with differences footnoted where appropriate).

**Table I-1**

**Summary Statistics**

**Year-End and January Compound Returns**

(NYSE/AMEX 1964-1990)

<table>
<thead>
<tr>
<th>Variable</th>
<th>A. Excess Returns</th>
<th>B. Raw Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 50,806)</td>
<td>(n = 53,958)</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Std.Dev.</td>
</tr>
<tr>
<td>$YE^{19}$</td>
<td>0.9937</td>
<td>0.1201</td>
</tr>
<tr>
<td>$WJ_1$</td>
<td>1.0306</td>
<td>0.09187</td>
</tr>
<tr>
<td>$WJ_2$</td>
<td>1.0089</td>
<td>0.0674</td>
</tr>
</tbody>
</table>

Refinements of the sample means are presented in Table I-2. In Panel A, the samples were stratified into decile by (increasing) market equity (MV1-MV10) for each year, and each decile was then pooled across years.\(^{20}\) The monotonic decline in the $WJ_i$ with market capitalization corroborates again that small-cap stocks participate dispropor-

---

\(^{19}\) Although not reported, returns for the five trading "weeks" in $YE$ showed modest ($\approx -0.10$) negative autocorrelation at lag 1, and insubstantial ($< -0.02$) but persistently negative autocorrelation at reported other lags.

\(^{20}\) The portfolio assignments used to construct Table I-2 differ from those reported by CRSP, which employ the applicable statistic (standard deviation or capitalization) for the immediately **preceding** year. Here, each year’s observations were reassigned to a new portfolio using the current year CRSP statistic.
tionately in the January effect. What is noticeable and novel is the *depression* in year-end returns, likewise most pronounced among small cap stocks, attenuating nearly monotonically as capitalization rises. Graphs of the mean yields for *YE* and *WJ*, -- the entries in Table I-2-A minus one -- in Panel A of Figure I-2, tell the story best.
Table I-2
Mean Year-End and January Returns
by Market Equity and Own Standard Deviation
(NYSE/AMEX 1964-1990)

<table>
<thead>
<tr>
<th>Decile</th>
<th>A. Increasing Capitalization</th>
<th>B. Decreasing Volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YE(^*) 25 days 5 days 5 days</td>
<td>YE(^*) 25 days 5 days 5 days</td>
</tr>
<tr>
<td>MV1</td>
<td>(5067) 0.9393 1.0965 1.0265</td>
<td>(5091) 0.9636 1.1024 1.0210</td>
</tr>
<tr>
<td>MV2</td>
<td>(5082) 0.9736 1.0596 1.0149</td>
<td>(5080) 0.9837 1.0536 1.0123</td>
</tr>
<tr>
<td>MV3</td>
<td>(5081) 0.9886 1.0447 1.0097</td>
<td>(5085) 0.9897 1.0352 1.0089</td>
</tr>
<tr>
<td>MV4</td>
<td>(5081) 0.9988 1.0326 1.0113</td>
<td>(5076) 0.9984 1.0288 1.0092</td>
</tr>
<tr>
<td>MV5</td>
<td>(5086) 1.0056 1.0253 1.0101</td>
<td>(5078) 1.0011 1.0212 1.0086</td>
</tr>
<tr>
<td>MV6</td>
<td>(5077) 1.0046 1.0201 1.0063</td>
<td>(5086) 1.0006 1.0166 1.0079</td>
</tr>
<tr>
<td>MV7</td>
<td>(5077) 1.0082 1.0146 1.0055</td>
<td>(5081) 1.0003 1.0140 1.0063</td>
</tr>
<tr>
<td>MV8</td>
<td>(5084) 1.0062 1.0093 1.0029</td>
<td>(5080) 1.0011 1.0105 1.0050</td>
</tr>
<tr>
<td>MV9</td>
<td>(5080) 1.0057 1.0058 1.0022</td>
<td>(5083) 1.0004 1.0115 1.0045</td>
</tr>
<tr>
<td>MV10</td>
<td>(5091) 1.0066 0.9978 0.9992</td>
<td>(5066) 0.9983 1.0119 1.0048</td>
</tr>
</tbody>
</table>

Figure I-2
Mean Year-End and January (Week 1) Yields
by (Increasing) Capitalization
and (Decreasing) Volatility

A. By Capitalization

B. By Volatility

* The last trading day in the year is excluded from YE returns and included in \( W_{J1} \) returns.
** Unequal because portfolio assignments were made by decile separately for each year.
One feature of Table I-2-A requiring mention is that the depression in year-end returns among small cap stocks was to some extent the product of construction. The portfolio assignments are by market capitalization as of the end of the current year (rather than, as reported by CRSP, for the close of the preceding year). Security that perform badly at the close of the year will tend naturally to drift into lower MV portfolios, thereby sharpening the picture of a year-end depression in returns among small-capitalization stocks. But, without a relationship between year-end performance and the January effect, that fact would do nothing to explain the rise in the January effect as market capitalization declines. Furthermore, the same qualitative pattern, attenuated only slightly in degree, was apparent when the data were stratified by CRSP (prior-year) MV portfolio assignments.

Only slightly less striking are the second panels in Table I-2 and Figure I-2, where the data have been stratified by sample standard deviation as reported by CRSP. (For ease of comparison, portfolio assignments are decreasing in standard deviation.) Consistent with Roll's conjecture that high volatility might help to explain the strength of the January effect among small-cap stocks, roughly the same pattern appears in Panel B of Table I-2 as in Panel A. Both the January effect and the year-end depression tend to be pronounced among high-volatility stocks. One departure from this pattern can be observed in the least volatile decile (SD10). There, interestingly, a depression can also be observed, a pattern that will persist in subsequent findings.

Table I-2 suggests that a year-end depression, not hitherto documented so far as I am aware, is associated with the January effect, a depression that is most con-

---

21 See footnote 20.

22 Spearman rank correlations between standard deviation and capitalization in the data were -.60, while simple correlations between portfolio assignments by capitalization and own-volatility were 0.61. The reversal in sign is attributable to the fact that portfolio assignments are increasing in capitalization but decreasing in standard deviation.

The "depression" in mean year-end yields among high-variance stocks was attenuated in the raw returns -- the means were everywhere higher in those data -- but, even there, a pattern of increasing year-end yields with decreasing volatility appeared.
spicuous among small capitalization and high volatility stocks. The phenomenon is sufficiently striking to induce concern that it somehow is an artifact of construction, although, except as noted above,23 there is no obvious reason for thinking that is so. Still, as a simple diagnostic check, both data sets were redrawn in their entirety around the beginning of July. For that specification, computations identical to those reflected in Figure I-2 for the turn-of-the-year (excess returns) data lead to Figure I-3 (scaled identically to Figure I-2). The depression appears to be characteristic of the turn of the year. It naturally invites two questions. First, how real is the apparent year-end depression? And second, how significant is its relationship to the January effect?

Figure I-3
Mean May-June and July (Week 1) Yields by (Increasing) Capitalization and (Decreasing) Volatility
NYSE/Amex Excess Returns (1965-1990)

A. By Capitalization

A. By Volatility

---

23 See notes 20-21 above and the accompanying discussion.
2. Methods

To investigate the first question I compare average weekly returns at the end of each year with comparably-computed average returns during the balance of that year, excluding the 35-day period under study. Thus, for each observation in the sample, a five-day (geometric) average year-end return was computed as the 5th root of $YE$. For that observation, the compound return for the balance of the year -- running from the 10th\textsuperscript{24} through the 27th-to-last trading day in that year, typically 217 trading days -- was raised to $5/n$ (where $n$ was the number of days during that period the stock actually traded). The resulting variable -- the difference between average weekly year-end returns and average weekly returns during the rest of the year ($DIFWRT$) -- was used to test the significance of the year-end depression.

I investigate the second question -- the relationship between year-end returns and the January effect -- by estimating a version of equation (1) from which the tax-loss variables were deleted. That equation, which (as noted by Roll) can be regarded as testing a trading rule under which securities are selected for year-end purchase on the basis of their (negative) performance during (in the context of this study) November and December, assumes a linear relationship between the year-end depression and the ensuing January effect. It is not obvious, however, that any such relationship ought to be strictly linear; in particular, it might very well not hold for securities whose year-end returns were in the vicinity of unity (or more). Consequently, two other specifications, in which the slope coefficients on $YE$ were permitted to vary with the level of the year-end depression in returns, were also estimated. The first included a quadratic term in $YE$; the second fitted the slope coefficients on $YE$ as a spline. Thus, I also estimated:

$JE_{it-1} = \alpha + \beta_1 YE_{it} + \beta_2 YE_{it}^2 + \delta_1 MV_{it} + \delta_2 MKT_t + \epsilon_{it},$

\textsuperscript{24} Excluding the (normally elevated) returns from the first nine trading days in January should tend on the whole to depress weekly returns averaged over the balance of the year; hence, it should tend to bias this procedure against finding significance to the (relative) depression in year-end returns.
(1b) \[ JE_{it+1} = \alpha + \sum_{j=1}^{5} \gamma_j S_{jit} + \delta_1 MV_{it} + \delta_2 MKT_t + \epsilon_{it}, \]

where the $S_{jit}$ are given by

\[ S_{jit} = [\text{Max}\{\text{Min}(YE_{it}, k_j), k_{j-1}\} - k_{j-1}], \quad k = 1, \ldots, 5, \text{ with} \]

(1c) \[ k_0 = 0; \quad k_1 = 0.925; \quad k_2 = 0.975; \quad k_3 = 1.025; \quad k_4 = 1.075. \]

I initially estimated each specification separately for the first ($JE = WJ_1$) and second ($JE = WJ_2$) trading weeks in January.

D. Results

1. The Year-End Depression in Returns

The means of the DIFWRT are given in Table I-3, together with t-statistics and p-values for the hypothesis that the means are 0. The data as stratified by market equity and standard deviation are reported in the left and right-hand panels, respectively. For the sample as a whole the mean is significantly less than zero -- year-end returns are comparatively depressed -- and the breakdown by either market capitalization or standard deviation indicates that the depression is significant for the smallest capitalization (or highest volatility) 30 percent of the market, with the most of the balance statistically indistinguishable from zero.\(^{25}\) Consistent with the implication of Table I-2-B, however, the least volatile decile of the market also exhibits a statistically significant year-end depression.

\(^{25}\) The same pattern, with the statistical significance of the depression extending into even higher capitalization deciles, appeared when stocks were assigned to portfolios by market capitalization as of the close of the preceding year. See the discussion accompanying notes 20-21 above. The pattern in Table I-3 was likewise found in the raw returns data, though in attenuated form, with the overall sample mean exceeding zero.
Overall, the pattern of depressed year-end returns depicted in Figure I-2 seems real. Excluding the final trading day in the year, year-end returns exhibit statistically significant depressions among small capitalization and high-volatility stocks, consistent with the hypothesis that elevated year-end volume, shown by others to be prominent among losers, reflects abnormal year-end selling pressure of the sort inferentially documented by Griffiths and White (1993). The upturn on the last trading day in the year has probably obscured this depression to students of December monthly data.

Table I-3
Mean Difference in Average Weekly Returns: End-of-Year vs. Rest of Year (NYSE/AMEX 1964-1989)

<table>
<thead>
<tr>
<th>Decile(^{26})</th>
<th>(Increasing) Market Value</th>
<th>(Decreasing) Volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean(^{27})</td>
<td>t = P &gt;</td>
</tr>
<tr>
<td>SMPL</td>
<td>-.00121</td>
<td>-11.41</td>
</tr>
<tr>
<td>ME1</td>
<td>-.00884</td>
<td>-17.50</td>
</tr>
<tr>
<td>ME2</td>
<td>-.00407</td>
<td>-9.96</td>
</tr>
<tr>
<td>ME3</td>
<td>-.00193</td>
<td>-5.13</td>
</tr>
<tr>
<td>ME4</td>
<td>-.00042</td>
<td>-1.24</td>
</tr>
<tr>
<td>ME5</td>
<td>.00068</td>
<td>2.09</td>
</tr>
<tr>
<td>ME6</td>
<td>.00028</td>
<td>0.94</td>
</tr>
<tr>
<td>ME7</td>
<td>.00091</td>
<td>3.33</td>
</tr>
<tr>
<td>ME8</td>
<td>.00046</td>
<td>1.80</td>
</tr>
<tr>
<td>ME9</td>
<td>.00045</td>
<td>1.99</td>
</tr>
</tbody>
</table>

\(^{26}\) Decile population sizes are given in Table I-2.

\(^{27}\) The mean is for the difference between the geometric five-day average return over the last 25 trading days in the year (excluding the very last day) and the comparable average for the preceding (typically) 217 trading days during the year.
2. The Relationship Between the January Effect and the Year-End Depression in Returns

Next I explore the relationship between the year-end depression and the January effect by estimating Equations (1)-(1b). Results for all specifications are presented in Panels A and B of Table I-4, which employ the first and second trading weeks in January, respectively, as the dependent variables.

The finding that the January effect is decreasing in market equity \( \hat{\delta}_1 < 0 \) is consistent with previous work. The estimates of \( \hat{\delta}_2 \) \( < 0 \) consistently imply that the January effect is decreasing in the overall performance of the market during the preceding year. As could be expected, given the concentration of the January effect in the first few days of the year, the results in Panel B, while not without interest, are generally weak. (Henceforth I therefore report only the results for Week 1.) In the Week 1 estimates, presented in Panel A, the relationship between year-end returns and the January effect is unmistakable.
Table I-4: Estimates of Equation (1)

Estimates of the relation between compound returns for the last 25 days (excluding the last day) in the year and returns for the first week in January, controlling for year-end market equity and the market, NYSE/AMEX, 1964-1989 (White std. errors). ** denotes 99% significance:

\[ JE_{it+1} = \alpha + \beta_1 YE_{it} + \beta_2 YE_{it}^2 + \sum_{j=1}^{5} \gamma_j S_{ji} + \delta_1 ME_{it} + \delta_2 Mkt_{t} + \epsilon_{it}, \]

(Eq. 1 includes only the linear term in YE; Eq. 1(a) also includes the quadratic term; Equation 1(b) excludes both but includes the spline variables \( S_{ji} \), as specified in Equation (1c) in text.)

| Coeff. | Eq. 1 | Eq. 1a | Eq. 1b | P > |t| | Eq. 1 | Eq. 1a | Eq. 1b | P > |t| |
|--------|-------|-------|-------|-----|-----|-------|-------|-------|-----|-----|
| Const. | 1.351 | 1.858 | 1.804 | 0.000 | 1.045 | 1.148 | 1.099 | 0.000 |
|        | (0.009)* | (0.059)* | (0.033)* |      | (.005)* | (0.021)* | (0.018)* |
| \( \beta_i / \gamma_i \) | -0.239 | -1.246 | -0.782 | 0.000 | -0.023 | -0.228 | -0.087 |      |
|        | (0.007)* | (0.118)* | (0.037)* |      | (.005)* | (0.039)* | (0.021)* |
| \( \beta_2 \) | --- | 0.479 | --- |      | --- | 0.0976 | --- |      |
|        |      | (0.057)* |      |      |      | (0.018)* |      |      |
| \( \gamma_2 \) | --- | --- | 0.036 | 0.440 | --- | --- | -0.026 | 0.000 |
|        |      |      | (0.046) |      |      |      | (0.030) |      |
| \( \gamma_3 \) | --- | --- | -0.199 | 0.000 | --- | --- | -0.024 | 0.383 |
|        |      |      | (0.026)* |      |      |      | (0.022) |      |
| \( \gamma_4 \) | --- | --- | -0.128 | 0.000 | --- | --- | -0.055 | 0.269 |
|        |      |      | (0.025)* |      |      |      | (0.025) |      |
| \( \gamma_5 \) | --- | --- | 0.008 | 0.407 | --- | --- | 0.034 | 0.029 |
|        |      |      | (0.010) |      |      |      | (0.012)* |      |
| \( \delta_1 \) | -0.00003 | -0.00002 | -0.00002 | 0.000 | -1e-05 | -8e-06 | -8e-06 | 0.004 |
|        | (2e-6)* | (2e-6)* | (1.7e-6)* |      | (9e-07)* | (8e-07)* | (1e-06)* |      |
| \( \delta_2 \) | -0.075 | -0.063 | -0.053 | 0.000 | -0.011 | -0.009 | -0.008 | 0.000 |
|        | (0.003)* | (0.003)* | (0.003)* |      | (0.002)* | (0.002)* | (0.002)* |      |
| \( \bar{R}^2 \) | .128 | .184 | .212 |      | 0.004 | 0.008 | 0.008 | 0.000 |
| \( F(k-1, 50,806-k) \) | 2493 | 2860 | 1951 |      | 63 | 102 | 55 |      |

a. \( JE \) are compound returns for the last day in December and the first 4 days in January; \( YE \) are compound returns for the last 25 days (excluding the last day) in the year.
b. \( P > |t| \) are reported for the coefficient estimates in Equation (1b).
c. Coefficient estimate for \( \beta_i \) is reported in cols. 1, 2, 5, and 6; the estimate for \( \gamma_i \) is in cols. 3 and 7.
d. Market equity was included in the regressions in $100,000,000.
e. \( F(k-1, 50,806-k) \) is for the joint hypothesis that all slope coefficients are zero. \( P(> F) \) was 0.00 in all cases.
The main finding -- there is an inverse relationship between year-end returns and the January effect -- is evident when only a linear term in year-end returns is included on the right-hand side (equation 1). The fit improves markedly with the inclusion of a quadratic term (equation 1(a)). The signs on the linear and quadratic coefficient estimates in (1(a)) suggest that this negative relationship becomes less pronounced (less negative) as the year-end depression declines. (The derivative of the quadratic estimate implies a minimum in the vicinity of $YE = 1.3$.\footnote{The derivative equals zero at $YE = -\hat{\beta}_1/2\hat{\beta}_2$.}) Estimating the slope coefficients as a spline (equation 1(b)), however, clearly produces the best fit. The variations in the spline slope coefficients are generally consistent with the estimated quadratic derivative. For convenience, the spline has been plotted in Figure I-4. Clearly, the depth of the year-end depression is related in an important way to the strength of the January effect.

Figure I-4
Spline Estimates* of Slope Coefficients on $YE$
(Equation 1(b))

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure.png}
\caption{January Effect/Year-End Returns 1964/1989 NYSE/AMEX Excess Returns}
\end{figure}

\begin{itemize}
\item a. The $\hat{y}$ value in the plot has been adjusted by the products of (1) $\hat{\beta}_i$, and the sample means of $MV$, and (2) $\hat{\beta}_2$ and the sample mean of $Mkt$, so that the plotted partial regression function passes through the sample mean.
\end{itemize}
Given the number of observations, the risk of type I error is in principle high. The actual estimates nevertheless seem reassuring. For several of the coefficients the hypothesis of a zero slope is not rejected; and where it is, most importantly with respect to the estimates of $\hat{\gamma}_1$, $\hat{\gamma}_3$, and $\hat{\gamma}_4$, the t-statistics of 21.3, 7.4 and 5.0, respectively, are comfortably far from 1.96 (and, indeed, from 2.576). The slope coefficients for $0.975 < YE < 1.075$, while negative and significant, are much flatter than the estimate of $\hat{\gamma}_1$, suggesting that much of the regression's explanatory power is to be found among securities with very depressed year-end returns ($YE < 0.925$). But, as suggested by Table I-2 and Figure I-2, that is precisely where the January effect is most pronounced.

E. Turn-of-the-Year Effects and Tax-Loss Potential

The findings of a year-end depression and its relationship to the January effect are suggestive. By themselves, however, they furnish no evidence of a connection between loss-sale potential or tax-motivated selling and either year-end or January returns. It is to those issues that I now turn. In this section I describe the construction of a newly-devised measure of tax-loss potential, which I then use to estimate the influence of loss potential on both the January effect and year-end returns.

1. Methods

Prior work has used relatively simple measures of tax-loss potential, most notably Reinganum's "PTL," the quotient of a security's closing price and its annual high. With the recent strengthening of the CRSP daily data, it is now possible to weight each security's daily price history by its daily volume, permitting a comparison of its closing price with a measure of the amounts acquired at each daily price during the year. But there are other shortcomings to PTL. In particular, while it provides information about loss potential, its asymmetry renders essentially indecipherable any information about unrealized security gains.

Accordingly, after determining the volume-weighted average daily (closing) price (VWP) for each security in the sample, for the 10th through 27th-to-last trading
days in each year, and using the security's closing price \( (CP) \) on the final day of that period, I calculated, as a measure of the tax-loss potential for each security \( i \) based on its price history in year \( t \):

\[
TLP_{it} = \frac{VWP_{it} - CP_{it}}{VWP_{it}}.
\]

For each security a second statistic, \( TLP_{it-1} \), was computed, using \( VWP_{it-1} \) in lieu of \( VWP_{it} \) but again comparing it to \( CP_{it} \), intended to capture the security's tax-loss potential by reference to its price history during the second preceding year. Based on a ten and one-half month price history, \( TLP_{it} \) is a reasonably good measure of short-term loss sale potential. Since anything in excess of the short-term holding period qualifies as long-term, however, \( TLP_{it-1} \) is a somewhat less accurate measure of long-term loss potential.

By construction, \( TL \) is unbounded below, bounded above by 1, and, in contrast with PTL, increasing in tax-loss potential. Zero divides losers from winners, a feature that it was anticipated would be useful in assessing the stability across winners and losers of any relationship between tax-loss potential and returns at the turn of the year. In practice it proved expeditious, in carrying out the estimation, simply to include as separate right-hand side variables the positive and negative parts of the \( TL \).

I estimated two basic equations. Lacking a clearcut model of the precise relationship between tax-loss potential and turn-of-the-year returns, I estimate a simple linear relationship between tax-loss potential and January returns, controlling for market equity, the performance of the market, and the influence of year-end returns, or

\[
VWP_{it} = \frac{\sum_d CP_{idt} Vol_{idt}}{\sum_d Vol_{idt}}.
\]

---

29 That is, loss-sale potential was determined in each year by excluding (1) the first nine trading days in January, included in the 35-day period under study for the preceding year, and (2) the year-end trading days included in the 35-day study period for the current year. Using the remaining days \( (d) \) and denoting daily volume by \( Vol_{idt} \), the statistic \( VWP_{it} \) is

\[
VWP_{it} = \frac{\sum_d CP_{idt} Vol_{idt}}{\sum_d Vol_{idt}}.
\]
\[ JE_{it+1} = \alpha_i + \sum_{j=1}^{5} \gamma_j S_{j_{it}} + \theta_{1p} TLPos_{it} + \theta_{1n} TLNeg_{it} + \theta_{2p} TLPos_{it-1} + \theta_{2n} TLNeg_{it-1} + \delta_1 ME_{it} + \delta_2 MKT_{it} + \epsilon_{it}, \]

(4a)

where \( TLPos_{it,j} \) are the positive values of \( TL_{it,j} \) -- the losers -- with the negative values set to zero, and conversely for \( TLNeg_{it,j} \) (in each case for \( j = 0, 1 \)). Similarly, I estimate the relationship between loss-sale potential and the year-end depression using

\[ YE_{it} = \alpha_i + \theta_{1p} TLPos_{it} + \theta_{1n} TLNeg_{it} + \theta_{2p} TLPos_{it-1} + \theta_{2n} TLNeg_{it-1} + \delta_1 ME_{it} + \delta_2 MKT_{it} + \epsilon_{it}. \]

(4b)

2. Data and Results

Summary statistics for the \( TL \) are presented in Table I-5. As is to be expected, there is far more dispersion in the \( TL_{t+1} \).

<table>
<thead>
<tr>
<th>Table I-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>TL Summary Statistics</td>
</tr>
<tr>
<td>(NYSE/AMEX 1964-1989)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>( TL_t )</td>
</tr>
<tr>
<td>( TL_{t+1} )</td>
</tr>
</tbody>
</table>

In Panels A and B of Table I-6 the means of the \( TL \) are reported by market capitalization and volatility. In the former, essentially a variation on Reinganum’s (1983) results, the means of the \( TL_t \) decrease monotonically (reflecting decreasing loss-sale potential) with capitalization, while the \( TL_{t+1} \) decrease sharply in the smallest deciles and
then tend to flatten out. What Panel B adds is that this pattern is nearly as sharply in evidence when securities are ranked by volatility. On average, high (own) volatility securities tend to exhibit a pronounced potential for year-end sale at a loss. The patterns are plotted in Figure I-5.
Table I-6
Means of Tax-Loss Variables
(NYSE/Amex 1964-1989)

<table>
<thead>
<tr>
<th>A. By Capitalization</th>
<th>B. By Volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decile</td>
<td>(n=)'</td>
</tr>
<tr>
<td>MV1</td>
<td>(5067)</td>
</tr>
<tr>
<td>MV2</td>
<td>(5082)</td>
</tr>
<tr>
<td>MV3</td>
<td>(5081)</td>
</tr>
<tr>
<td>MV4</td>
<td>(5081)</td>
</tr>
<tr>
<td>MV5</td>
<td>(5086)</td>
</tr>
<tr>
<td>MV6</td>
<td>(5077)</td>
</tr>
<tr>
<td>MV7</td>
<td>(5077)</td>
</tr>
<tr>
<td>MV8</td>
<td>(5084)</td>
</tr>
<tr>
<td>MV9</td>
<td>(5080)</td>
</tr>
<tr>
<td>MV10</td>
<td>(5091)</td>
</tr>
</tbody>
</table>

Figure I-5
Tax-Loss Variables by Decile

A. By Capitalization

B. By Volatility

* Unequal because portfolio assignments were made by decile separately for each year.
3. Results
   
a. Tax-Loss Potential and the January Effect

   Estimates of equations (4a) and (4b) are presented in Panels A and B of Table I-7. To facilitate interpretation of the estimates, two versions of (4a), in which \( W_{ij} \) is the dependent variable, are reported, from the second of which the spline variables (the \( S_{jm} \)) were deleted. Also reproduced in Panel A are the results from estimating equation (1b), as originally reported in column 3 of Table IV. In Panel B, \( YE \) is the dependent variable.

   Looking first at the complete estimate of equation (4a) in col. 1, the coefficient (\( \hat{\theta}_{1p} \)) on \( TLP_{0.5} \) -- short-term losers -- is positive and significant, confirming directly that loss-sale potential is related to the strength of the January effect. While the estimate on \( TLP_{0.5} \) (\( \hat{\theta}_{1p} \)) is also positive and significant, it is substantially closer to zero, and an F-test rejects the hypothesis \( \theta_{1p} = \theta_{1n} \). The estimates on the long-term tax-loss variables (\( \hat{\theta}_{2p} \) and \( \hat{\theta}_{2n} \)) present a slightly more complicated picture. The coefficient on the long-term losers (\( \hat{\theta}_{2p} \)) is also positive and significant. That finding is consistent with the evidence of Chan (1986) and De Bondt and Thaler (1985), and implies that the January effect is increasing in potential for year-end sale at a long-term loss. It is inconsistent with the hypothesis that investors engage in optimal (short-term) tax-loss trading as envisioned by Constantinides. Less expected, however, is the significant, negative estimate of (\( \hat{\theta}_{2n} \)), the coefficient on the long-term winners. It implies that the January effect is increasing in a winner’s potential for sale at a long-term gain. One could imagine plausible explanations, but the finding is probably a spurious artifact of size: when equation (4a) was reestimated separately by decile (typically \( n < 5100 \)), the hypothesis of a zero slope on \( \hat{\theta}_{2n} \) was rarely rejected at any conventional level of significance.
Table 1-7: Estimates of Equation (4)

Estimates of the relations (a) between returns during the first week in January and (1) tax-loss potential and (2) compound returns for the last 25 days (excluding the last day) in the year, and (b) between year-end returns and tax-loss potential, in each case controlling for year-end market equity and the market in the prior year; NYSE/AMEX, 1964-1989 (White std. errors). * denotes 99% significance.

(a): \[ JE_{it} = a_i + \sum_{j=1}^{5} \gamma_j S_{it} + \theta_1 TLPot_{it} + \theta_2 TLPot_{it-1} + \theta_3 TLPot_{it-2} + \delta_1 ME_{it} + \delta_2 MKT_{it} + \epsilon_{it} \]

(b): \[ YE_{it} = a_i + \theta_1 TLPot_{it} + \theta_2 TLPot_{it-1} + \theta_3 TLPot_{it-2} + \delta_1 ME_{it} + \delta_2 MKT_{it} + \epsilon_{it} \]

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>A. ( WJ_1 )</th>
<th>B. ( YE )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(4a)</td>
<td>(4a)</td>
</tr>
<tr>
<td></td>
<td>(No Spline)</td>
<td>(Tab.IV)</td>
</tr>
<tr>
<td>Coeff.:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Const.</td>
<td>1.635</td>
<td>1.016</td>
</tr>
<tr>
<td></td>
<td>(0.031)*</td>
<td>(0.004)*</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>-0.678</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(0.036)*</td>
<td></td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>0.111</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td></td>
</tr>
<tr>
<td>( \gamma_3 )</td>
<td>-0.180</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(0.025)*</td>
<td></td>
</tr>
<tr>
<td>( \gamma_4 )</td>
<td>-0.169</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(0.025)*</td>
<td></td>
</tr>
<tr>
<td>( \gamma_5 )</td>
<td>0.024</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>0.101</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>(0.005)*</td>
<td>(0.006)*</td>
</tr>
<tr>
<td>( \theta_2 )</td>
<td>0.021</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.005)*</td>
<td>(0.006)</td>
</tr>
<tr>
<td>( \theta_3 )</td>
<td>0.043</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>(0.003)*</td>
<td>(0.003)*</td>
</tr>
<tr>
<td>( \theta_4 )</td>
<td>-0.004</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.001)*</td>
<td>(0.002)*</td>
</tr>
<tr>
<td>( \delta_1 )</td>
<td>-0.0002</td>
<td>-0.0002</td>
</tr>
<tr>
<td></td>
<td>(1.5e-6)</td>
<td>(0.00001)*</td>
</tr>
<tr>
<td>( \delta_2 )</td>
<td>-0.003</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)*</td>
</tr>
<tr>
<td>( \hat{R}^2 )</td>
<td>0.256</td>
<td>0.134</td>
</tr>
<tr>
<td>( F(1,4) )</td>
<td>1589</td>
<td>1306</td>
</tr>
</tbody>
</table>

a. \( JE \) are returns for the last day in December and the first 4 days in January; \( YE \) are returns for the last 25 days (excluding the last day) in the year; the \( S_i \) are as specified in equation (1c) in text.

b. Market equity was included in the regressions in $100,000,000.

c. \( \text{Fk} - 1, 50,806 - k \) is for the joint hypothesis that all slope coefficients are zero. \( P(>F) \) was 0.00 in all cases.
It is informative to compare the estimates in column 1 of Table I-7 with the separately estimated tax-loss variables and spline coefficients reported in columns 2 and 3. While the full estimate clearly produces the best fit, both the magnitude and significance of each set of estimates is essentially unaltered when the other variables are omitted: the variables of interest in this study -- the $\gamma$s and $\hat{\theta}_{1p}$ and $\hat{\theta}_{2p}$ -- exhibit remarkable stability. And, overall, at least as measured by $\bar{R}^2$, each set of variables has considerable but overlapping explanatory power for performance during the first part of January. Tax loss potential, especially short-term potential but significantly long-term potential, is clearly related to the January effect. Even controlling for the influence of tax-loss potential, however, year-end returns have substantial explanatory power of their own.

b. Tax-Loss Potential and Year-End Returns

The proposition that tax-loss potential and year-end returns are less than collinear for the January effect is borne out by the estimate of equation (4b), reported in Panel B of Table I-7. The relationship between tax-loss potential and the depression in year-end returns seems markedly less strong than that between loss-sale potential and the January effect itself. The coefficient estimate on the short-term loss variable is statistically indistinguishable from zero. The estimate of the long-term tax-loss variable ($\hat{\theta}_{1p}$), on the other hand, is quite far from zero and significant. In fact, the point estimate on $\hat{\theta}_{2p}$, when the dependent variable is year-end returns, is of about the same magnitude as (but opposite in sign from) the estimate on the short-term tax loss measure when the January effect is on the right hand side in the full estimate of equation (4a). This suggests, even more strongly than the estimate of equation (4a) alone, that long-term loss selling is an especially intensive year-end phenomenon. This inference is corroborated by the fact that, when the spline variables are deleted from equation (4a) -- as reported in column 2 of Table I-7 -- the point estimate on the short-term loss variable ($\hat{\theta}_{1p}$) is essentially unchanged, while the magnitude of the estimated coefficient on the long-term variable ($\hat{\theta}_{2p}$) approximately doubles. To some extent, then, year-end returns and long-
term loss sale potential appear to contain the same information about ensuing January returns.

c. The Influence of Market Capitalization and Volatility

To explore how these relationships vary with the characteristics of the securities involved, I reestimated equation (4a) separately for each decile of the market by both market equity and own volatility. These estimates are reported in Tables I-8-A and B (contained in Appendix A). The patterns are generally consistent. One prominent feature is that the explanatory power of the regressions is quite high among the smallest stocks, and then decreases monotonically in capitalization (Table I-8-A.) The slope coefficients ($\gamma_i$) on low year-end return stocks ($YE < .925$) are not far from (although statistically distinguishable from) -1.0 in the smallest deciles and then rise nearly monotonically, with virtually all the spline coefficients indistinguishable from zero among the largest capitalization stocks. With the coefficients on the tax loss variables the picture is somewhat different. The estimate $\hat{\delta}_{ip} > 0$ persisted through about half the market: short-term loss potential was positively and significantly related to the January effect. More interesting, however, the estimated effect of long-term loss-sale potential ($\hat{\theta}_{2p}$) was positively related (with 99% confidence) to the January effect in all MV deciles. The general picture is of a strong relationship between a deep depression in year-end returns and the January effect, most pronounced among small-capitalization stocks, but with persistent significance of long-term loss-sale potential. That these findings are not simply an artifact of small standard errors is implied by the degeneration of the most of the individual coefficient estimates to insignificance in the MV9-MV10 deciles.

A somewhat less distinct pattern emerges when the market is stratified by variance (Table I-8-B). There, the best fit, and the most markedly negative spline coefficients on the low year-end return stocks, again tends to be found among in most volatile deciles. More generally, however, both the explanatory power of the regressions and the significant, negative coefficient estimates on $\gamma_i$ remain evident among even the less volatile issues. Indeed, although the fit varies somewhat, quite similar qualitative
characteristics are shared by all the regressions for portfolios SD3-SD9. Interestingly, though, the basic pattern exhibited by the most volatile deciles -- negative, significant estimates of the spline coefficients and a large, positive coefficient on the long-term loss variable -- is almost as pronounced in the least volatile (SD10) decile. One possible story that is consistent with this finding would be of a disposition to year-end profit-taking in favorably taxed long-term winners, together with the realization of offsetting losses. Given the premise that year-end profit-taking in long-term winners does occur, the concomitant realization of offsetting long-term losses would not be especially surprising.

Altogether, Tables I-7 and I-8 portray clear relationships between tax-loss potential and the year-end depression and the strength of the January effect, most pronounced among issues that experienced particularly deep depressions in year-end returns. The breakdowns by portfolio suggest that both low capitalization and high volatility play a role in these relationships, but that market capitalization is the more important consideration.

F. An Interpretation

The estimate of Equation (4a), with the spline variables deleted, indicates clearly that year-end loss-sale potential has explanatory power for returns at the beginning of January, consistent with the tax-loss selling hypothesis. When compared to the estimate of equation (1a) it appears, however, to have less explanatory power than do the year-end returns themselves. From this one might infer that the relationship between the year-end depression and the January effect, documented in Section D, while partly attributable to year-end tax-motivated selling, reflects other influences too. The estimate of Equation (4b), on the other hand, indicates that loss-sale potential -- especially short-term loss sale potential -- does less well at explaining the year-end depression in returns than at explaining the January effect itself. This suggests that loss-sale potential influences the January effect, in some measure at least, outside the confines of the year-end period selected for this study.
I suggest the following reconciliation of these findings. It is amply clear from the estimates of equations (4a) and (4b), which confirm and extend findings by De Bondt and Thaler (1985, 1987) and Chan (1986), that long-term loss sale potential is significantly related both to the January effect and to the depression in returns that I have identified during the last 25 trading days of the year. Such findings, hitherto regarded as atheoretical and mystifying, I do not regard as implausible at all. Costly stockpiling of short-term losses that may not be usable for years is not altogether obviously an optimizing pursuit. Some losses unavoidably (or, at least, optimally) will become long-term. The (equally costly) decision to realize long-term losses is best made when they are most likely -- ideally, a.s. -- to be used; and (because of discounting) most cheaply pursued as late as possible in the game.

Just how late is optimally late is not clear. The estimate of equation (4b) suggests that, for at least some long-term losses, the last month of the year is soon enough. That the process may begin earlier is made plausible by the hypothesis of Ritter (1988) that individual investors temporarily "park" the proceeds of (relatively) late-in-the-year loss sales pending still later re-entry into the market.\textsuperscript{30} That the pace of loss realization by individual investors rises during the year has been documented in a recent study (using 1971-1979 data) by Badrinath and Lewellen (1991). But perhaps the most striking evidence, disbelieve it though they may, is De Bondt and Thaler's (1985) Figure 3: the picture of a persistent but gradually decaying fourth-quarter sell-off and January rebound in long-term losers seems unmistakable. It is corroborated in this study by

\textsuperscript{30} Constantinides (1984) suggests, to the contrary, that "rational" individual investors who unloaded loss stocks near the end of the year would simultaneously purchase other loss stocks then being disposed of by different investors, thereby eliminating any (possibly) depressing effect on market prices. This suggestion would appear to require either (a) an extreme belief in the real-time ability of investors to determine which issues are experiencing year-end declines because of tax-motivated selling and which are declining in response to new, adverse information, or (b) that investors holding securities with unrealized losses engage in a search for other investors holding different but risk-compatible loss securities with whom they can effectively "swap," a search that would itself be relatively costly.
estimates (the $\hat{\theta}_n$) that imply that long-term loss sale potential is significantly related to both the year-end depression and the January effect.

Short-term loss sale potential, on the other hand, seems unmistakably related to the January effect but without explanatory force for returns during the final 25 days of the year. It is undeniable that short-term losses can (or at least could before 1986) in some instances be more valuable than long-term losses at the margin. And, especially given prevailing professional wisdom, there surely must have been many who have acted as though they were. So it is plausible to think that there has been at least some disposition in practice to "take" losses that materialize quickly in newly-acquired ("disappointing"?) securities, and there is (pre-1986) evidence in Badrinath and Lewellen (1991) that investors behave(d) marginally in that way. Hence, at least when compared with long-term losses, the looming transition from short-term to long-term status in loss securities may have sharpened investors' incentives to accelerate the realization of some short-term losses, reducing the concentration of short-term losses in the final days of the year. Such a process could easily account for the reduced explanatory power of the tax-loss variables generally for the year-end depression than for the January effect; and, more particularly, could account for the insignificant coefficient estimate on the short-term tax-loss variable in the estimate of equation (4b).

As for the explanatory force of year-end returns for the January effect beyond what is also accounted for by the tax-loss variables alone, I can only suggest -- it is simply a version of Roll's (1983) original conjecture and it will still sit uneasily with many -- that the "pressure" created by an accelerating disposition to sell losers as the year-end draws near, accumulated over a longer period than the more-or-less arbitrarily selected 25 days used in carrying out this study, creates a disequilibrium condition from which the affected securities are unable to recover until the pressure is "relieved" by the cessation of year-end loss taking at the very end of the year, as documented by Griffiths and White (1993). In any market other than the market for securities that would be an unremarkable account. Perhaps it is a plausible explanation here.
G. Conclusion

The findings of this study are amply consistent with prior work on the relationship between year-end tax motivated selling and the January seasonal. But the return content of year-end daily data adds a new dimension to that work. The absence of a year-end depression in December monthly data has long been thought to undermine the view that year-end tax-loss selling gives rise to the January effect. The year-end depression documented here, and its evident relationship to January returns, thus tend strongly to substantiate the belief that tax-induced year-end behavior is an important determinant of the striking phenomena of early January. They fall short, however, of telling an entire story. In the face of Constantinides’ claim that year-end loss selling is irrational, the work reported here could easily be regarded as just more in the way of ad hoc empirical results. To deal with that objection, I turn to the theory in Essay Two.
### Table I-8-A
(Decile Estimates of Equation 4(a) -- By Increasing Capitalization)
(See legend and notes to Table I-7)

<table>
<thead>
<tr>
<th>Coeff.</th>
<th>MV1</th>
<th>MV2</th>
<th>MV3</th>
<th>MV4</th>
<th>MV5</th>
<th>MV6</th>
<th>MV7</th>
<th>MV8</th>
<th>MV9</th>
<th>MV10</th>
</tr>
</thead>
<tbody>
<tr>
<td>n =</td>
<td>5.067</td>
<td>5.082</td>
<td>5.081</td>
<td>5.081</td>
<td>5.086</td>
<td>5.077</td>
<td>5.077</td>
<td>5.084</td>
<td>5.080</td>
<td>5.091</td>
</tr>
<tr>
<td>Const.</td>
<td>1.661</td>
<td>1.678</td>
<td>1.642</td>
<td>1.420</td>
<td>1.278</td>
<td>1.312</td>
<td>1.198</td>
<td>1.161</td>
<td>1.124</td>
<td>0.933</td>
</tr>
<tr>
<td>(\gamma_1)</td>
<td>-0.719</td>
<td>-0.703</td>
<td>-0.674</td>
<td>-0.416</td>
<td>-0.230</td>
<td>-0.276</td>
<td>-0.167</td>
<td>-0.135</td>
<td>-0.111</td>
<td>0.057</td>
</tr>
<tr>
<td>(\gamma_2)</td>
<td>0.505</td>
<td>0.240</td>
<td>0.233</td>
<td>-0.148</td>
<td>-0.248</td>
<td>-0.105</td>
<td>-0.173</td>
<td>-0.187</td>
<td>-0.100</td>
<td>-0.000</td>
</tr>
<tr>
<td>(\gamma_3)</td>
<td>-0.335</td>
<td>-0.065</td>
<td>-0.256</td>
<td>-0.100</td>
<td>-0.012</td>
<td>-0.105</td>
<td>-0.113</td>
<td>-0.110</td>
<td>-0.106</td>
<td>-0.076</td>
</tr>
<tr>
<td>(\gamma_4)</td>
<td>0.007</td>
<td>-0.316</td>
<td>-0.079</td>
<td>-0.146</td>
<td>-0.310</td>
<td>-0.188</td>
<td>-0.199</td>
<td>-0.252</td>
<td>-0.224</td>
<td>-0.168</td>
</tr>
<tr>
<td>(\gamma_5)</td>
<td>-0.115</td>
<td>-0.024</td>
<td>-0.062</td>
<td>-0.037</td>
<td>-0.018</td>
<td>-0.019</td>
<td>-0.034</td>
<td>0.007</td>
<td>-0.024</td>
<td>0.034</td>
</tr>
<tr>
<td>(\theta_{1\gamma})</td>
<td>0.176</td>
<td>0.168</td>
<td>0.110</td>
<td>0.058</td>
<td>0.030</td>
<td>0.025</td>
<td>0.006</td>
<td>-0.005</td>
<td>-0.019</td>
<td>-0.009</td>
</tr>
<tr>
<td>(\theta_{1\eta})</td>
<td>0.008</td>
<td>0.020</td>
<td>0.013</td>
<td>0.021</td>
<td>0.047</td>
<td>0.033</td>
<td>0.033</td>
<td>0.038</td>
<td>0.055</td>
<td>0.051</td>
</tr>
<tr>
<td>(\theta_{2\gamma})</td>
<td>0.037</td>
<td>0.021</td>
<td>0.045</td>
<td>0.047</td>
<td>0.038</td>
<td>0.042</td>
<td>0.042</td>
<td>0.039</td>
<td>0.025</td>
<td>0.017</td>
</tr>
<tr>
<td>(\theta_{2\eta})</td>
<td>-0.011</td>
<td>-0.005</td>
<td>-0.007</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.001</td>
<td>0.001</td>
<td>0.003</td>
<td>-0.000</td>
<td>0.004</td>
</tr>
<tr>
<td>(\delta_1)</td>
<td>-0.034</td>
<td>-0.013</td>
<td>-0.004</td>
<td>-0.003</td>
<td>-0.001</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td>(\delta_2)</td>
<td>0.028</td>
<td>-0.001</td>
<td>0.000</td>
<td>-0.000</td>
<td>-0.026</td>
<td>-0.029</td>
<td>-0.015</td>
<td>-0.015</td>
<td>-0.002</td>
<td>0.018</td>
</tr>
<tr>
<td>(\bar{R}^2)</td>
<td>0.285</td>
<td>0.257</td>
<td>0.240</td>
<td>0.179</td>
<td>0.136</td>
<td>0.138</td>
<td>0.103</td>
<td>0.096</td>
<td>0.055</td>
<td>0.029</td>
</tr>
<tr>
<td>F(10, n-11)</td>
<td>185</td>
<td>169</td>
<td>147</td>
<td>101</td>
<td>74</td>
<td>75</td>
<td>54</td>
<td>50</td>
<td>28</td>
<td>15</td>
</tr>
<tr>
<td>Coeff.</td>
<td>SD1</td>
<td>SD2</td>
<td>SD3</td>
<td>SD4</td>
<td>SD5</td>
<td>SD6</td>
<td>SD7</td>
<td>SD8</td>
<td>SD9</td>
<td>SD10</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>n =</td>
<td>5,091</td>
<td>5,080</td>
<td>5,085</td>
<td>5,076</td>
<td>5,078</td>
<td>5,086</td>
<td>5,081</td>
<td>5,080</td>
<td>5,083</td>
<td>5,066</td>
</tr>
<tr>
<td>Const.</td>
<td>1.765</td>
<td>1.659</td>
<td>1.400</td>
<td>1.424</td>
<td>1.331</td>
<td>1.380</td>
<td>1.321</td>
<td>1.320</td>
<td>1.414</td>
<td>1.324</td>
</tr>
<tr>
<td></td>
<td>(0.060)*</td>
<td>(0.084)*</td>
<td>(0.049)*</td>
<td>(0.054)*</td>
<td>(0.051)*</td>
<td>(0.072)*</td>
<td>(0.063)*</td>
<td>(0.069)*</td>
<td>(0.073)*</td>
<td>(0.101)*</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.750</td>
<td>-0.650</td>
<td>-0.373</td>
<td>-0.399</td>
<td>-0.317</td>
<td>-0.401</td>
<td>-0.320</td>
<td>-0.327</td>
<td>-0.426</td>
<td>-0.321</td>
</tr>
<tr>
<td></td>
<td>(0.068)*</td>
<td>(0.097)*</td>
<td>(0.055)*</td>
<td>(0.060)*</td>
<td>(0.057)*</td>
<td>(0.082)*</td>
<td>(0.070)*</td>
<td>(0.075)*</td>
<td>(0.080)*</td>
<td>(0.110)*</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.500</td>
<td>0.288</td>
<td>-0.127</td>
<td>-0.047</td>
<td>-0.031</td>
<td>0.045</td>
<td>-0.198</td>
<td>-0.102</td>
<td>-0.136</td>
<td>-0.265</td>
</tr>
<tr>
<td></td>
<td>(0.224)</td>
<td>(0.175)</td>
<td>(0.099)</td>
<td>(0.088)</td>
<td>(0.076)</td>
<td>(0.085)</td>
<td>(0.065)*</td>
<td>(0.060)</td>
<td>(0.055)</td>
<td>(0.052)*</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>-0.168</td>
<td>-0.295</td>
<td>-0.094</td>
<td>-0.125</td>
<td>-0.178</td>
<td>-0.211</td>
<td>-0.044</td>
<td>-0.140</td>
<td>-0.075</td>
<td>-0.111</td>
</tr>
<tr>
<td></td>
<td>(0.263)</td>
<td>(0.132)</td>
<td>(0.097)</td>
<td>(0.082)</td>
<td>(0.067)*</td>
<td>(0.063)*</td>
<td>(0.055)</td>
<td>(0.045)*</td>
<td>(0.040)</td>
<td>(0.031)*</td>
</tr>
<tr>
<td>$\gamma_4$</td>
<td>-0.460</td>
<td>-0.043</td>
<td>-0.212</td>
<td>-0.126</td>
<td>-0.243</td>
<td>-0.190</td>
<td>-0.207</td>
<td>-0.146</td>
<td>-0.186</td>
<td>-0.132</td>
</tr>
<tr>
<td></td>
<td>(0.206)</td>
<td>(0.116)</td>
<td>(0.087)</td>
<td>(0.073)</td>
<td>(0.067)*</td>
<td>(0.061)*</td>
<td>(0.060)*</td>
<td>(0.054)*</td>
<td>(0.045)*</td>
<td>(0.039)*</td>
</tr>
<tr>
<td>$\gamma_5$</td>
<td>-0.045</td>
<td>-0.056</td>
<td>-0.046</td>
<td>-0.057</td>
<td>-0.027</td>
<td>-0.040</td>
<td>-0.040</td>
<td>-0.017</td>
<td>-0.038</td>
<td>-0.058</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.023)</td>
<td>(0.026)</td>
<td>(0.019)*</td>
<td>(0.036)</td>
<td>(0.029)</td>
<td>(0.037)</td>
<td>(0.049)</td>
<td>(0.033)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>$\theta_{ip}$</td>
<td>0.166</td>
<td>0.120</td>
<td>0.064</td>
<td>0.038</td>
<td>0.033</td>
<td>0.030</td>
<td>0.018</td>
<td>0.006</td>
<td>0.016</td>
<td>-0.022</td>
</tr>
<tr>
<td></td>
<td>(0.018)*</td>
<td>(0.018)*</td>
<td>(0.012)*</td>
<td>(0.011)*</td>
<td>(0.010)*</td>
<td>(0.010)*</td>
<td>(0.010)*</td>
<td>(0.009)</td>
<td>(0.010)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>$\theta_{in}$</td>
<td>0.008</td>
<td>0.013</td>
<td>0.039</td>
<td>0.064</td>
<td>0.045</td>
<td>0.026</td>
<td>0.057</td>
<td>0.036</td>
<td>0.046</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.013)</td>
<td>(0.012)*</td>
<td>(0.011)*</td>
<td>(0.010)*</td>
<td>(0.010)*</td>
<td>(0.011)*</td>
<td>(0.010)*</td>
<td>(0.010)*</td>
<td>(0.010)*</td>
</tr>
<tr>
<td>$\theta_{ip}$</td>
<td>0.026</td>
<td>0.027</td>
<td>0.031</td>
<td>0.038</td>
<td>0.035</td>
<td>0.052</td>
<td>0.042</td>
<td>0.050</td>
<td>0.050</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>(0.009)*</td>
<td>(0.008)*</td>
<td>(0.007)*</td>
<td>(0.007)*</td>
<td>(0.006)*</td>
<td>(0.007)*</td>
<td>(0.007)*</td>
<td>(0.007)*</td>
<td>(0.009)*</td>
<td>(0.008)*</td>
</tr>
<tr>
<td>$\theta_{in}$</td>
<td>0.001</td>
<td>-0.000</td>
<td>-0.004</td>
<td>-0.005</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.004</td>
<td>0.003</td>
<td>-0.002</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$\delta_{i}$</td>
<td>-0.001</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(0.000)*</td>
<td>(0.000)*</td>
<td>(0.000)*</td>
<td>(0.000)*</td>
<td>(0.000)*</td>
<td>(0.000)*</td>
<td>(0.000)*</td>
<td>(0.000)*</td>
<td>(0.000)*</td>
<td>(0.000)*</td>
</tr>
<tr>
<td>$\delta_{c}$</td>
<td>-0.051</td>
<td>-0.045</td>
<td>-0.029</td>
<td>-0.031</td>
<td>-0.016</td>
<td>0.003</td>
<td>-0.005</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(0.019)*</td>
<td>(0.011)*</td>
<td>(0.009)*</td>
<td>(0.008)*</td>
<td>(0.008)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$\hat{R}$^{2}</td>
<td>0.266</td>
<td>0.228</td>
<td>0.167</td>
<td>0.168</td>
<td>0.136</td>
<td>0.147</td>
<td>0.122</td>
<td>0.122</td>
<td>0.147</td>
<td>0.185</td>
</tr>
<tr>
<td>F(10, n-11)</td>
<td>168</td>
<td>137</td>
<td>94</td>
<td>94</td>
<td>74</td>
<td>81</td>
<td>65</td>
<td>65</td>
<td>81</td>
<td>106</td>
</tr>
</tbody>
</table>
ESSAY TWO

Risk is Bad, But Time is Money, So Sometimes it Pays to Wait:
Optimal Tax-Motivated Selling in a Simple Two-Period Model

A. Introduction

Evidence accumulated steadily over the past 15 years\(^{31}\) consistently suggests a connection between the January "seasonal effect" in the U.S. securities markets and year-end tax-motivated sales of securities trading at a loss. The basic story is that the January effect, partly at least, is a "rebound" from a transitory disequilibrium condition that a wave of year-end loss selling creates. But the "tax-loss selling" hypothesis is viewed with suspicion. Apart from the possible market inefficiency it implies, theory, due to Constantinides (1983, 1984), suggests that, if investors behave rationally, tax-motivated loss selling should be distributed throughout the year rather than being concentrated at the end. If so, nothing remarkable ought to occur in the vicinity of the end of the year. The theory has been exceedingly influential. It is striking to find investigators, whose evidence suggests a connection between loss-sale potential and the January effect but is incompatible with the "optimal" time-invariant pattern of tax-loss selling that Constantinides predicts, choosing to disbelieve their own work. (See, most notably, De Bondt and Thaler (1985, 1987) and Chan (1986).)

This Essay revisits the theory. In Section B I argue that several of the assumptions on which Constantinides modelled the problem are questionable approximations of reality. In Sections C-E I relax one of the debatable assumptions -- that "short-term" capital losses should be modelled as uniformly more valuable than "long-term" capital losses -- and proceed to reinvestigate the problem. In Section D I use a simple decision tree to capture the basic intuition on which I wish to focus attention: even with uncertainty about future price movements, a loss that could be taken today might also be around to be realized tomorrow. With transaction costs and a positive discount rate, taking a loss today is more costly than taking the same loss tomorrow. Depending on to-

\(^{31}\) See Part B of Essay One and, especially, Ritter (1988) for surveys.
day's price, the discount and transaction costs rates, and the probability distribution of the future price movement, it will, in terms of expected cost, often be better to wait.

In Section E I formalize the intuition of Section D using a two-period control model with uncertainty. Although I do not obtain a closed-form solution, the first-order condition for this simple formulation produces a number of insights, with testable implications for year-end volume, into an investor's optimized decision whether to realize a loss for tax purposes before the end of the year. The insights have considerable intuitive appeal. The decision to wait turns on the probability of an upward price movement at year's end sufficient to dissipate the loss. The decision to sell turns symmetrically on the probability of a downward movement sufficient to induce the investor to sell at a loss once again.

Finally, I establish two propositions that, on the assumption that short-term and long-term losses are equally valuable, indirectly provide an essentially complete characterization of the solution to the problem of optimal tax-loss selling. There exists an interval of prices at which an investor could realize a loss before the end of the year but at which it is nevertheless optimal to wait. Everything other than time held constant, the lower bound of that interval rises as the end of the year draws near, implying a rise in year-end loss selling. In Section F I summarize the model's implications, addressing both the matter of consistency with existing empirical work and implications for further research.

One virtue of the approach taken here is that the insights obtained are not restricted by detailed assumptions about the probability distribution of security price movements. Little more is required than that, to a sufficient degree, and with positive probability, stock prices can go up or go down.
B. Tax-Motivated Selling

1. The Problem

The insights to be exemplified in Section D and formalized in Section E concern the problem faced by an investor who, before the "end" of his taxable year, may have (or already has) "realized" a taxable "capital" gain, and can, if he chooses, realize immediately an offsetting loss. Take the simplest case: the investor already has realized the gain. If there were no cost to taking the loss -- no commissions, no bid-ask spread, and securities sold at a loss could immediately and costlessly be reacquired -- there would not be any reason to wait. If, however, doing so were costly the investor might instead hesitate. If, in particular, the loss were certain also to be available at the very end of the year, waiting would clearly be best. With a positive discount rate and everything else constant, a given loss is most cheaply realized at the latest possible moment in time.

Uncertainty complicates the decision. If by the end of the year the loss stock went up, the loss would be smaller and therefore less valuable and might even completely disappear. So waiting is risky, since diminution of the loss could leave the gain wholly or partly (but permanently) exposed to taxation. The risks of not waiting include the already identified possibility that the loss will not go away, and hence that the costs of realizing it will have been incurred needlessly early in time. Less obvious is the

---

32 The term "realized" has technical and accepted but differing meanings in statistics and Federal taxation. Under Internal Revenue Code § 1001(a)-(b) the term "realized" denotes a gain (or loss) that has become exposed to taxation by the "sale or other disposition" of an asset, the sense in which the term is used in the sentence in text. In general it will be clear whether the term is being used to denote a taxable event or the realization of a random variable, but it seems appropriate to insert a cautionary note here.

33 For individual investors, tax is imposed on the net gain (after deducting all losses) on capital transactions for the year. Under Internal Revenue Code § 1211(b), unused losses may be carried forward for use in succeeding years, but may not be carried back and applied against gains that were exposed to taxation in previous years. Thus, once an individual taxpayer's capital gain has been exposed to taxation, it is (in the absence of a net operating loss carryback under Internal Revenue Code § 172) forever exposed to taxation. For a corporation the discontinuity is less stark. Corporate capital losses may (under § 1211(a)) be carried back and deducted against previously taxed corporate gains.
possibility, in the event of a further decline, that it might become profitable to realize an additional loss by selling again at the end of the year. If so, the investor will incur two rounds of costs to realize a loss that, by waiting until the end of the year, he could have realized in a single transaction entailing a single round of costs.

Existing work by Constantinides (1983, 1984), which predicts that it is not optimal to wait (and more generally that tax-motivated selling should be distributed throughout the year), imposes more structure on the problem by adopting one or both of two further assumptions. The first is that all losses realized in any year in fact are deductible that year. Second, and more importantly, it assumes that "short-term" capital losses are uniformly more valuable than "long-term" term capital losses. The principal question I wish to explore is whether these particular assumptions are pivotal to the finding that losses are best realized sooner than later, and to the related prediction of essentially time-invariant loss-selling. Whether that question is worthwhile pursuing turns on the plausibility of the assumptions, which in turn requires a look at the law.

2. The Treatment of Capital Losses

For income tax purposes most gains and losses on securities transactions by individual investors are treated as "capital" gains and losses. Apart from defining what it means to be a "capital" transaction, the relevant provisions of the U.S. Internal Revenue Code ("I.R.C.") (1) provide for favorable treatment of "long-term" capital gains, (2) specify the way in which capital losses are to be deducted against capital gains, and (3) provide for the "carryover" of "unused" capital losses.

a. The Taxation of Capital Gains. Favorable taxation of net (long-term) capital gains is familiar. Any "net" (long-term) "capital gain" -- net gain from the sale or exchange of a "capital asset" held for more than one year -- is taxed at a preferential

---

A capital gain or loss is defined by I.R.C. § 1222 as gain or loss from the "sale or exchange of a capital asset," in turn defined in I.R.C. § 1221. The determination of what constitutes a "capital asset" in the hands of a corporation is generally more complex than it is for individuals. It is to be emphasized that what is a "capital asset" for income tax purposes is not equivalent to any conventional economic definition of the term "capital".
rate,\textsuperscript{35} whereas gain from capital assets held for one year or less are not taxed at any favorable rate.

b. \textit{The Deductibility of Capital Losses}. The deductibility of losses from the sale of capital assets is somewhat more complex. I.R.C. § 1222 requires short-term capital losses first to be deducted ("netted") against short-term capital gains, and long-term losses against long-term gains. Any resulting "net" short-term loss may then be deducted against any remaining net long-term gain (and conversely, when that is the case). These provisions, together with lenient taxation of long-term capital gain, might be taken to suggest that short-term capital losses are more valuable than long-term losses, but that is not necessarily so. At the margin, a short-term loss may easily be deducted against a long-term gain; the converse may equally be so.\textsuperscript{36} More generally, the \textit{actual} consequence of short or long-term losses to any given taxpayer in any given year depends on the overall \textit{mix} of that taxpayer’s short and long-term transactions \textit{that year}.

A portion of any surplus of capital losses over capital gains for a year may be deducted against non-capital gain income (such as wages, dividends and interest). That deduction is limited, however, to a maximum of $3,000 per year, and it is only available (under I.R.C. § 1211) to individuals. Only to that extent (and only until 1986) were short-term losses explicitly more valuable than long-term losses: it took $2 of long-term capital loss but only $1 of short-term loss to produce $1 of deduction. In 1986 even that limited distinction disappeared. Now, $3,000 of \textit{either} long- or short-term capital loss in excess of all capital gains produces a $3,000 individual deduction.

c. \textit{Carryover of Unused Losses}. For individuals, capital losses that are "unused" because they exceed \textit{both} capital gains \textit{and} the additional $3,000 deduction may

\textsuperscript{35} The capital gains "preference" has varied over the years, and was completely eliminated in 1986. In 1990 a small capital gains preference reappeared, which has since become a rate advantage that, at its current maximum, stands at 11.6 percent.

\textsuperscript{36} For example, a $1,000 short-term loss taken by an investor with previously realized long-term but no short-term gains would be deductible against the long-term gains; and a $1,000 long-term loss realized by an investor with short-term but no long-term gains would be deductible against the short-term gains.
be "carried forward" and treated as a capital loss in the following year.\textsuperscript{37} For corporations, capital losses in excess of capital gains may be carried back and deducted against previously taxed gains in any of the three preceding taxable years, and carried forward to any of the five following years.

3. Plausibility of the "Optimal" Loss Selling Assumptions

From this summary it is immediately clear that, unless an investor consciously tailors the amount of capital losses he realizes each year to the capital gains he reasonably anticipates having that year, some portion of the losses may end up \textit{not} being currently deducted. Their value will typically then be reduced because they will be "carried over" and deducted against gains realized in some later year.\textsuperscript{38} Constantinides (1983) expressly acknowledges both the counterfactual nature of his contrary assumption and that the assumption strengthened his results.\textsuperscript{39}

Even less plausible is the \textit{joint} assumption that all losses are deductible \textit{and} that all short-term losses are more valuable than long-term losses. Even before 1986, when explicit differential treatment was eliminated, short-term losses were explicitly more valuable only to the extent of the $3,000 deduction against ordinary income for individuals. At the individual level, however, that is a quantitatively trivial feature of the taxation of capital losses, and surely insufficient to justify the costs of portfolio churning dictated by policies of realizing all short-term losses of the sort Constantinides recommends.

\textsuperscript{37} See footnote 33 above. Through the operation of the rather complex rules of I.R.C. § 1212(b), losses carried over are characterized as short-term or long-term losses in the following year. Moreover, because § 1212(b) provides simply for the carryover of unused losses to the "succeeding" taxable year, the carryover period is effectively unlimited in duration.

\textsuperscript{38} If, because of loss unavailability or sheer inattention, a \textit{corporate} taxpayer has failed to offset some prior year capital gains with losses, a loss in excess of gains in the current year may be carried back and deducted against gains that were previously taxed. See footnote 33 above.

\textsuperscript{39} See Constantinides (1984) at 76.
In contrast, both long-term and short-term capital losses are ultimately deductible without limitation against both short-term and long-term capital gains. It is therefore reasonable to suppose that most capital losses are in fact deducted against other capital gains. On that assumption the relative value of short and long-term losses is in principle unclear. In practice, however, voluntarily realized capital gains are almost invariably long-term capital gains. On the additional assumption that most capital gains are voluntarily realized gains, it follows that most capital losses -- short-term or long-term -- will be deducted against long-term capital gains. I therefore suggest that a more plausible modelling assumption is that long-term and short-term capital losses will on average be of equal value.\textsuperscript{40}

Ideally, a complete reinvestigation of the loss-selling question would alter both debatable assumptions. Constantinides's work suggests, however, that it is the assumption of differential value of short-term and long-term losses that is probably most pivotal.\textsuperscript{41} What is more, adopting an assumption of uncertain gains would require the incorporation of two stochastic processes -- one for the price process of the security to be sold at a loss, another for the materialization of gains -- into the formalization and solution of the problem. So, in the interest of both simplicity and isolating what a priori is most likely to be crucial to Constantinides's findings, the work presented here preserves the assumption that all realized losses may be currently deducted, and relaxes only the assumption of differential taxation of short-term and long-term losses.

C. Framework

Even with the assumption that short-term and long-term losses are equally valuable, an investor's decision at the beginning of the year will be influenced in an intricate way by the interplay between that decision and nature's choice of the price at the

\textsuperscript{40} For dealers, at least before 1993, more favorable ordinary loss treatment was available, but, there as well, the value of a loss had nothing to do with how long the securities have been held. Now, dealers are required to mark such securities to market annually under I.R.C. § 475.

\textsuperscript{41} In Constantinides (1983) stochastic realization of gain (modelled as a Poisson process) is allowed.
end of the year. So, before either illustrating or modelling the problem, it will be useful to make our assumptions precise; to specify the transition relationships that govern the environment the investor faces at the end of the year, depending on both nature and the investor's first period decision; and, finally, to spell out just how the two interact to give rise to the principal intricacies in the solution to the investor's problem.

1. Assumptions

Consider an investor who at some point before the end of a year (Period 1) has already realized a gain ($G$) from the sale of a capital asset. Unless offset by a loss that is realized for tax purposes by the end of the year (Period 2), the gain will be taxed at rate $\theta \in (0, 1)$. The investor holds a single, infinitely divisible share of a security, for which the investor previously paid (including any acquisition costs) $b_1$, denoted the investor's "adjusted basis" in the security for determining any subsequently realized gain or loss. I assume that $b_1 < G$, which ensures that any loss realized on the security by the end of the current year will be deductible against the current year gain. I assume, finally, that, at the beginning of Period 1, the security is trading at a price (more specifically restricted below) such that an immediate sale would produce a loss ($p_1 < b_1$). In general, the investor is free to realize a fraction $\alpha \in [0, 1]$ of any available loss in Period 1, in Period 2, or in both.

The one-way cost of realizing a loss in either period is a fraction $\gamma \in [0, 1]$ of the price at which the security is sold. If a loss is realized during Period 1, I assume (following Constantinides (1984)) that the security is immediately repurchased, at a price identical to that at which it was sold, on which an additional commission must be

---

42 That is, since any loss will be the excess of the investor's adjusted basis over the price, the maximum loss that (ignoring expenses) the investor can realize (at price 0) is just $b_1$.

43 That is, I assume that the "wash-sale" limitation of Internal Revenue Code § 1091 -- which disallows any loss from the sale of any security that is reacquired (directly or through the purchase of an option) during the 60-day period centered on the date of sale -- does not bind. This assumption is adopted both in the interests of consistency with Constantinides and because it is unlikely to affect the results. Even if one does not assume (as do many financial economists) that the limitation can easily be circumvented by "selling Ford and buying GM", its impact could
paid. All transactions costs are assumed to be deductible: the cost incurred at the time of purchase is added to the investor's ("adjusted") basis, and is therefore deducted at the time of any subsequent sale; while the cost of sale is deducted from the gross proceeds received. I assume, finally -- it is an important assumption -- that if the investor realizes a loss in Period 2 he again repurchase the stock and pays the applicable commission. The one-period discount rate is $\delta \in (0, 1]$.

I underscore two features of these assumptions. The first is that the value of a loss is independent of how long the investor held the stock: a short-term loss is assumed to be no more or less valuable than a long-term loss. The second is $b_1 < G$. Following Constantinides, this implies that the value of a loss taken in either Period is undiminished by the possibility that the loss may not be usable this year.

2. Transition Relationships

With these assumptions, any loss realized in Period 1 and carried forward to be deducted in Period 2 will be

\[
L_2 = L_2(\alpha_1) = \alpha_1 [b_1 - p_1 (1 - \gamma)],
\]

and the cost of realizing that loss (and immediately repurchasing the stock) will be $2\alpha_1 \gamma p_1$. Taking account of the tax savings from any Period 1 loss, which will not be enjoyed until Period 2, it is immediately clear that, uncertainty aside, it will be cost-saving (though not necessarily cost-minimizing) for the investor to sell in Period 1 if and

be modelled as an additional (non-deductible) cost of establishing a loss, which would be incurred irrespective of when the loss was established.

---

44 This assumption is needed to ensure that the results in either the illustration in Section D or the model of Section E are not driven by an unrealistic cost structure, in which the rate at which costs are incurred is lower if they are simply incurred later. It is also consistent with the underlying assumption of the problem that losses are being realized solely to minimize taxes, in positions that the investor would just as soon preserve. Finally, it has the virtue of allowing for the possibility that the two period version of the problem in Section E might readily be embedded in a repeated longer- horizon formulation.
only if the discounted tax savings from deducting the loss exceed the cost of realizing the loss, or

$$2\gamma p_1 - \theta \delta L_2 = p_1[2\gamma + \theta \delta (1-\gamma)] - \theta \delta b_1 < 0, \text{ or}$$

$$(2a) \quad p_1 < \frac{\theta \delta b_1}{2\gamma + \theta \delta (1-\gamma)} = p_1^* < b_1,$$

(where the last inequality follows from the assumptions $\theta < 1$ and $\delta \leq 1$). Call the price $p_1^*$ defined by (2a) the Period 1 "critical price". It is increasing in both the tax rate ($\theta$) and time ($\delta$), and decreasing in the transaction cost rate ($\gamma$). A threshold implication of $p_1^* < b_1$ is the perhaps obvious point that, with positive transaction costs -- irrespective of uncertainty -- modest losses aren't worth the taking.

Whether, at Period 1 prices below $p_1^*$, it is cost-minimizing to realize an available loss must take into account the investor's options in Period 2. If the investor were to wait, his basis for determining any Period 2 loss would simply be $b_2 = b_1$, which he would then compare to the realized Period 2 price. If, however, the investor had sold in Period 1, his basis for determining any subsequent (gain or) loss would be reduced\(^{45}\) to the price at which he reacquired the security in Period 1, augmented by the cost of reacquisition, and that adjusted basis would be compared to the Period 2 price. Generally,\(^{46}\) the Period 2 basis is

$$(1b) \quad b_2 = b_2(\alpha_1) = (1-\alpha_1)b_1 + \alpha_1 p_1(1+\gamma).$$

The investor's decision in Period 2 is governed by the relationship between his Period 2 basis and the Period 2 price, assumed generally to be given by

\(^{45}\) I show formally in Section E that $\alpha_1 > 0 \Rightarrow b_2 < b_1$.

\(^{46}\) In general, Treasury Regulations § 1.1012-1(c) implicitly require an investor to maintain separate bases for lots of the same security acquired separately at different prices, a detail that, by averaging basis, this formulation ignores. Since, with the assumption of "large" G, however, the investor ends up optimizing at a corner in Period 2, any resulting imprecision is immaterial.
\[ \tilde{p}_2 = p_1 \cdot \bar{\varepsilon}, \quad \bar{\varepsilon} = f(\bar{\varepsilon}) \text{ on } \varepsilon < 0 < \bar{\varepsilon}. \]

Any loss taken in Period 2 produces immediate tax savings. Consequently, it is cost-minimizing for the investor to sell in Period 2 only if the Period 2 savings exceed the Period 2 cost, or

\[ 2 \gamma p_2 - \theta \left[ b_2 - p_2(1 - \gamma) \right] < 0, \quad \text{or} \]

\[ p_2 < \frac{\theta b_2}{2 \gamma + \theta (1 - \gamma)} = p^*_2(b_2). \]

The Period 2 "critical price" defined by (2b) is just the undiscounted analogue of the Period 1 critical price \( p^*_1 \) defined by (2a). Since, however, the Period 2 adjusted basis is a function of the Period 1 choice \( (\alpha_1) \), \( p^*_2(b_2) = p^*_2(b_2(\alpha_1)) \). If the investor waits in Period 1, so that \( \alpha_1 = 0 \) and \( b_2 = b_1 \), then, given \( \delta \leq 1 \) and consistent with the fact that \( p^*_1 \) is increasing in time,

\[ p^*_2(b_1) = \frac{\theta b_1}{2 \gamma + \theta (1 - \gamma)} \geq \frac{\theta \delta b_1}{2 \gamma + \theta \delta (1 - \gamma)} = p^*_1, \]

So a second implication of formalizing the problem is that, even if the price doesn't change, the critical price at which (uncertainty aside) the investor becomes indifferent about whether or not to take a loss goes up from Period 1 to Period 2. Stated differently, some losses that are not worth taking in Period 1 become worth taking in Period 2.

When necessary, I will parameterize the example in Section D with \( \gamma = 0.02 \), \( \theta = 0.28 \) and \( \delta = 0.97 \) (which could be interpreted as taking Period 1 to be the "middle" of the year and the risk-free annual rate to be 6 percent). With that parameterization, (2a) yields

\[ p^*_1 = 0.88709 \ b_1. \]

---

\(^{47}\) See footnote 44 above.
while, if the investor waits in Period 1, (2c) produces

\[ p_2^*(b_1) = 0.89059 b_1. \]

For the obvious reason, in what follows I impose the restriction \( p_1 < p_1^* \).

3. Corners in the Payoffs to Waiting or Exchanging

The principal complication to this problem stems from the fact that, regardless of whether the investor waits or exchanges in Period 1, the probability distribution of \( \bar{\epsilon} \) can introduce corners into the Period 2 consequences of the Period 1 decision. If he waits and the stock price moves up, that movement may either (1) reduce the loss or (2) cause it completely to disappear. The latter occurs if

\[ (3a) \quad p_1 + \epsilon \geq p_2^*(b_1). \]

the price at which -- when the investor waits in Period 1 -- it is not cost-saving to realize any Period 2 loss. If so, an investor who waited in Period 1 will refrain from selling again. Otherwise he will realize a (reduced) loss in Period 2.

Conversely, if the investor exchanges and the stock price moves down, that movement may or may not be sufficient to make it cost-saving to realize a loss again. It will be sufficient only if (assuming \( \alpha_1 = 1 \) and, in accordance with (1b), \( b_2 = p_1(1 + \gamma) \)),

\[ (3b) \quad p_1 - \epsilon < p_2^*(p_1(1+\gamma)). \]

Otherwise the investor who exchanged in Period 1 will choose to wait in Period 2.

Expressions (3a) and (3b) suggest generally that, at least to a point, the wider the support for the probability distribution of \( \epsilon \), the more complicated the investor's problem will be. In the general formulation in Section E, which allows for both a continuous probability distribution and a continuous decision, these considerations influence the solution in a way that is relatively opaque. So, I begin with a more straightforward
and intuitive example, in which both are discrete, that sheds light on how the presence of these corners can influence the investor's decision.

D. An Illustration

1. Description

In this Part I parameterize the price process as binomial,

\[ \tilde{p}_2 = \begin{cases} p_1 + e, & \text{prob} = q, \ (e > 0) \\ p_1 - e, & \text{prob} = 1 - q. \end{cases} \]

and construct a decision tree for the problem. In so doing I adopt as an assumption what I prove for the more general formulation of Section E: that the optimal choice in both Period 1 and Period 2 will be found at an extreme setting of the control. Thus, in Period 1 the investor chooses to wait \((W_1)\) or exchange \((X_1)\) by selling and repurchasing the stock. The stock price then moves up or down, after which the investor chooses to wait \((W_2)\) or exchange \((X_2)\) in Period 2. All eight possible sequences of moves by the investor and nature are arrayed in Figure II-1 leading to the payoffs (valued as of Period 2, with subscripts suppressed for convenience) calculated at the right of the tree.  

\[ \text{To illustrate the computation of the payoffs, using the least and most complicated possibilities, assume first that the investor chooses to wait in Period 1 \((W_1)\), the stock price moves up \((u)\), and the investor again chooses to wait \((W_2)\). Then no loss is realized in either period and the investor pays tax of } \theta G \text{ (as given by entry (1a)).} \]

Suppose, conversely, that the investor chooses \(X_1\), the stock price moves down, and the investor then chooses \(X_2\). In that event he will realize a Period 1 loss of

\[ b_1 = p_1(1 - \gamma). \]

(expression (1a), with \(\alpha_1 = 1\)), his Period 2 basis, according to (1b) (again with \(\alpha_1 = 1\)), will be adjusted to

\[ b_2 = p_1(1 + \gamma). \]

so his Period 2 loss (after the price decline) will be

\[ p_1(1 + \gamma) - (p_1 - e)(1 - \gamma). \]
In each of the uppermost four sequences the investor waits in Period 1, so he at most realizes a loss in Period 2; in each of the bottom four entries he exchanges in Period 1, so he at least realizes a loss in Period 1.

Standard use of the tree is to compute, for each possible Period 1 action, the expected cost of nature's move; and then to choose that action with the lowest expected cost in Period 1. Here, for each Period 1 choice and each move by nature, there are two possible moves between which the investor must choose optimally in Period 2. At the

The Period 1 transactions costs (valued as of Period 2) and the Period 2 transactions costs are (respectively)

$$2\gamma p_1 \delta^{-1}, \text{ and } 2\gamma(p_1 - e).$$

Thus, taking account of both transactions costs and tax savings, total costs (as given by entry (IVb)) are

$$\theta[G - b_1 + p_1(1 - \gamma) - p_1(1 + \gamma) + (p_1 - e)(1 - \gamma)] + \delta^{-1} 2\gamma p_1 + 2\gamma(p_1 - e).$$
two outer Period 2 nodes (I and IV), the least cost action will depend, in accordance with (3a) and (3b), on the choice of ε imposed on the model. Of the remaining four possibilities, however, two may be eliminated from consideration. Given \( p_1 < p_1^* < p_2^*(b_1) \) -- i.e., at price \( p_1 \) it is cost-saving to realize a loss in at least one of the two Periods -- if the investor chooses \( W_1 \) and nature chooses \( d \), it must also save money (and now be cost-minimizing) to sell in Period 2 (\( X_2 \)). So at node II the sequence \( W_1-d-W_2 \) is ruled out. Conversely, if the investor chooses \( X_1 \), his basis is adjusted downward to \( p_1(1 + \gamma) \).

But, for \( \theta < 1 \),

\[
p_1(1 + \gamma) > p_1 > \frac{\theta p_1(1 + \gamma)}{2 \gamma + \theta(1 - \gamma)} = p_2^*(p_1(1 + \gamma)),
\]

so, if nature now chooses \( u \), the stock price will exceed \( p_2^* \) and there will be no loss to realize in Period 2. Hence, at node III the sequence \( X_1-u-X_2 \) is likewise ruled out.

Without arbitrarily restricting the problem by imposing a value for \( \epsilon \), we can write the expected costs of (respectively) waiting and exchanging in Period 1 (in each case followed by optimal conduct in Period 2) for all possible epsilons as

\[
E_\epsilon[W_1(p_1, b_1)] = \theta G + q \min\{ 0, (p_1 + \epsilon)(2 \gamma + \theta[1 - \gamma]) - \theta b_1 \}
\]

(4a)

\[
+ (1 - q)[(p_1 - \epsilon)(2 \gamma + \theta[1 - \gamma])], \quad \text{and}
\]

\[
E_\epsilon[X_1(p_1, b_1)] = \theta [G - b_1 + p_1(1 - \gamma)] + \delta^{-1}2 \gamma p_1 + (1 - q) \min\{ 0, (p_1 - \epsilon)(2 \gamma + \theta[1 - \gamma]) - \theta p_1(1 + \gamma) \}.
\]

(4b)

The arguments of the Min operator in (4a) correspond to the cases in which \( \epsilon \) is (or is not) sufficiently large to satisfy the inequality (3a). (If it is, the investor, having waited in Period 1, optimally waits again in Period 2, and, when the stock price moves up, the operator picks 0 as the Period 2 cost.) Similarly, the arguments of the operator in (4b) correspond to whether \( \epsilon \) is large enough to satisfy (3b). (If that inequality is satisfied the
investor, having exchanged in Period 1, exchanges again in Period 2, and the operator picks the non-zero argument as the Period 2 cost when the stock price moves down.)

We can now investigate in what circumstances (if any) it is optimal to wait in Period 1 by imposing

\[(4c) \quad E_t[W_t(p_1, b_1)] \leq E_t[X_t(p_1, b_1)],\]

and solving the inequality for \(p_1\). There are four possible Period 1 combinations, depending on \(\epsilon\) and the inequalities (3), the solutions to which are arrayed in matrix form in Table II-1. For convenience, I have normalized \(b_1 = 1\), so that both the disturbance (\(\epsilon\)) and the Period 1 price (now denoted by \(\pi\)) above which it is optimal to wait are expressed as a percentage of \(b_1\).
Table II-1
Price/Basis Ratios For Waiting
(Binomial Probability)\(^{49}\)

<table>
<thead>
<tr>
<th>ACTION(_1)</th>
<th>Exchange ((X_i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(u-W_2)</td>
<td>(u-W_2)</td>
</tr>
<tr>
<td>(d-X_2)</td>
<td>(d-W_2)</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{Wait} & \\
(W_i) & \\
{/} \quad \text{NATURE/} \\
\text{ACTION}\_2 & \quad \text{ACTION}\_2 \\
\text{u-W}_2 & \quad (p - \epsilon < p_r (p(1+\gamma)) < p_r (b_i) < p + \epsilon) \quad (p_r (p(1+\gamma)) < p - \epsilon < p_r (b_i) < p + \epsilon) \\
\text{d-X}_2 & \quad \pi \geq \frac{q^2 \delta}{2 \gamma (1 - \delta) + \epsilon \delta (1 + \gamma)} \quad \pi \geq \frac{q^2 \delta - \epsilon (1 - q) \delta (2 \gamma + \theta (1 - \gamma))}{2 \gamma (1 - \delta (1 - q) + q \delta (1 - \gamma)}
\end{align*}
\]

2. Results

a. Large \(\epsilon\): Waiting in Both Periods and Exchanging in Both Periods Both Allowed

The northwest \((1, 1)\) entry in Table II-1 furnishes the most general insight with the least complication. It gives a ratio \((\pi)\) of Period 1 price to basis -- which for convenience I denote \(\pi_{(1,1)}\) (and similarly for the other \(\pi_{(i,j)}\)) -- above which it is optimal

\(^{49}\) The expressions in parentheses give the restrictions (in terms of the inequalities (3a) and (3b)) appropriate for each entry in the Table.
to wait in Period 1 by evaluating (4c) for the case in which both corners in the Period 2 payoffs occur: \( \epsilon \) is sufficiently large so that, if the investor waits and the price moves up, the loss completely disappears, while if the investor exchanges and the price moves down the investor exchanges again. Since \( \pi \geq \pi_{(1,1)} \) allows for the possibility that the loss may completely disappear, it skirts none of the risks of waiting.\(^{30}\)

In the Appendix I show that, for sufficiently large \( \epsilon \) (anything more than about 7.5 percent of \( b_1 \) for the assumed parameter values), \( \pi_{(1,1)} \) is an upper bound on the price at which it becomes optimal to wait. Since, however, \( \epsilon \) is chosen so that both corner limitations occur, the precise magnitude of \( \epsilon \) itself plays no further role in determining the optimal decision. The price above which it is optimal to wait is solely a function of the parameters, and the relative probabilities that the price will move up or move down.

The insight provided by evaluating \( \pi_{(1,1)} \) is striking. Using \( q = 0.50 \), the investor optimally waits if the ratio of price to basis satisfies

\[
\pi \geq \pi_{(1,1)}(0.50) = 0.810.
\]

Observe that this is less than 0.88709, the value of \( p^*_1 \), above which it does not pay to realize a loss even in the absence of uncertainty. Hence, there exists an interval of prices below \( p^*_1 \) at which it remains optimal to wait, even when uncertainty is introduced. An investor holding a stock that cost $100, trading at a loss of nearly 20 percent in the middle of the year, with mean-zero expectations about the year-end innovation in price, will clearly do better to wait.

\(^{30}\) In terms of (3a) and (3b), \( \epsilon \) satisfies:

\[
p_1 - \epsilon < p_2^*(p_1(1+\gamma)) < p_2^*(b_1) < p_1 + \epsilon.
\]

Furthermore, this case corresponds, in the more general formulation in Section E, to an assumption that the support for the probability distribution of \( \bar{\epsilon} \) includes the Period 2 critical price regardless of whether the investor waits or exchanges in Period 1, i.e., that

\[
p_1 - \bar{\epsilon} < p_2^*(p_1(1+\gamma)) < p_2^*(b_1) < p_1 + \bar{\epsilon}.
\]
Mean-zero expectations are not, however, an especially attractive assumption.

With such expectations the investor might well be inclined to take the loss and invest in another security about whose prospects he was more optimistic. And, as can easily be seen by differentiating \( \pi_{(1,1)} \), the price above which it is optimal to wait is strictly increasing in \( q \). But a second, equally striking insight is that for any \( q < 1 \) there exists some \( \pi_{(1,1)} < \pi_1^* (= p_1^* / b_1) \) above which it is optimal to wait. This can be seen by setting \( q = 1 \) to find that

\[
\pi_{(1,1)}(1) = \frac{\theta \delta}{2 \gamma + \theta \delta (1 - \gamma)} = \pi_1^*.
\]

Hence, for any \( q < 1 \) the Period 1 price above which it is optimal to wait lies below \( p_1^* \), as shown in Figure II-2:

![Figure II-2](image)

Using, for example, \( q = 0.70 \) (with \( \epsilon = 0.10 \)) -- which corresponds to a little less than 5 percent (semi-annual) positive drift\(^{\text{51}}\) -- the investor still (optimally) waits as long as

\[ E[\bar{\epsilon}] = 0.7 \cdot 0.1 - 0.3 \cdot 0.1 = 0.04. \]

so, at a price/basis ratio of about 0.85, the drift is \( 0.04 / 0.85 = 0.047 \).

\(^{51}\) That is,

\[ E[\bar{\epsilon}] = 0.7 \cdot 0.1 - 0.3 \cdot 0.1 = 0.04. \]
\[ \pi \geq \pi_{(1,1)}(0.70) = 0.852. \]

b. Small \( \epsilon \): Both Waiting and Exchanging in Both Periods Not Allowed

The other three entries in Table II-1 correspond to the cases in which one or both of the inequalities (3) is not satisfied. Of the three conditions, that least favorable to the proposition that it may be optimal to wait uses \( \pi \geq \pi_{(1, 2)} \). That condition corresponds to the case in which \( \epsilon \) is so small that, if the investor exchanges in Period 1 and the stock price moves down he does not exchange again; whereas the Period 1 price is so close to \( \pi^* \), that, if the investor waits and the price moves up, the loss nevertheless disappears. It therefore subjects the investor to the corner limitation on the payoff only when he waits in Period 1. Even there, however, there are prices below \( \pi^*_1 \) at which the investor optimally waits. This can easily be seen by noting that \( \pi_{(1, 2)} \) is increasing in \( q \) and decreasing in \( \epsilon \), and that setting \( q = 1 \) and \( \epsilon = 0 \) in \( \pi_{(1, 2)} \) yields \( \pi^*_1 \). Thus, any positive \( \epsilon \) and any smaller \( q \) produces \( \pi_{(1, 2)} < \pi^*_1 \). For example, \( q = 0.70 \) and \( \epsilon = 0.02 \) yields

\[ \pi_{(1,2)}(0.70, 0.02) = 0.8771 < 0.88709. \]

So (using the example above) the investor still waits at any price above $87.70.

The implication is this. If an investor benefits equally from losses, irrespective of duration, there is some ratio of price to basis above which, even with uncertainty, the investor does better to wait. That finding persists in Section E. First, however, one final insight -- perhaps the most striking -- can be obtained by solving the condition \( \pi \geq \pi_{(1,1)} \) for \( q \), to find that the investor optimally waits whenever

\[ q \leq \frac{2 \gamma (1 - \Theta \delta) \pi}{\Theta \delta (1 - \pi(1 + \gamma))} = \frac{\kappa \pi}{(1 - \pi(1 + \gamma))} = h(\pi). \]
The numerator of \( h(\pi) \) is just *price* (expressed as a percentage of basis), multiplied by the round-trip transaction costs rate (net of the discounted tax savings from deducting the costs); the denominator is just *loss*, multiplied by the discounted tax rate at which the investor benefits from the loss. Thus, (4d) has the natural interpretation that, if the probability that the price will go up is less than the ratio of (1) the cost of realizing the loss, to (2) the tax benefit of deducting the loss, the investor does better to wait.

That the investor’s decision should turn on the ratio \( \pi \) makes intuitive sense. When \( \pi \) is high, so is the cost of realizing the loss, while the loss itself (and resulting tax savings) is relatively small. As the price falls, the loss gets larger and the cost of realizing it declines.\(^{53}\) Thus, at some value of \( \pi \in (0, \pi^*_1) \), we would expect the investor to become indifferent between realizing a Period 1 loss and not, as we have just seen to be the case. Essentially the same condition will emerge as the key to the formulation in Section E.

E. A More General Formulation

I proceed now on the assumptions specified generally in Section C, allowing (1) the probability distribution of the Period 2 price to take the general form given by (1c), and (2) the investor to sell any fraction \( \alpha_j \in [0, 1] \) of the loss stock in Period \( j = 1, 2 \). With those assumptions I formulate the problem using a two-period control model, for which no closed form solution exists. Nevertheless, I formalize the problem and obtain conditions on the derivative that correspond to the conditions that emerged from the discretized example in Section D. Then, I establish two propositions that show that the basic insights of Section D are preserved in this more general formulation.

---

\(^{52}\) Note that \( h(\pi) \) it is increasing in \( \gamma \) and \( \pi \), and decreasing in \( \theta \) and \( \delta \). Observe also that, setting \( h(\pi) = 1 \) and solving for the price yields \( \pi = \pi^*_1 \).

\(^{53}\) In practice, this may not strictly speaking be correct, since many brokerage commission schedules have some minimum commission, in which event the schedules are affine rather than strictly linear, as I have assumed.
1. The Problem

**Period 2 Minimized Cost.** Working backwards from Period 2, if the investor sells a fraction $\alpha_2$ of his shares (which, to repeat, he is again assumed immediately to repurchase),\(^{54}\) his total Period 2 cost, taking account of any Period 2 sale *and* (using (1a)) any loss from Period 1, is

$$\alpha_2 2 \gamma p_2 + \theta \left\{ G - L_2 - \alpha_2 \left[ b_2 - p_2 (1 - \gamma) \right] \right\}$$

(5a)

$$= \alpha_2 \left\{ 2 \gamma p_2 - \theta \left[ b_2 - p_2 (1 - \gamma) \right] \right\} + \theta \left( G - L_2 \right).$$

Clearly, it is cost saving to realize a loss in Period 2 only when the first term in (5a) is negative, that is, when the loss is sufficient to produce tax savings that offset the transaction cost, based on the Period 2 asset price. As in Section D, that corresponds to the condition

$$p_2 < p_2^*(b_2),$$

given by (2b). The assumption $G > b_1 > p_1$, together with the corollary -- suggested by (1b) and justified formally below -- that $b_1 \geq b_2$, implies that, if this condition is satisfied, the investor realizes the *entire* available loss. So in Period 2 the investor minimizes cost at an extremity of the control set by choosing

$$\alpha_2 = \begin{cases} 
1, & \text{if } 2 \gamma p_2 - \theta \left[ b_2 - p_2 (1 - \gamma) \right] < 0, \\
0, & \text{otherwise}.
\end{cases}$$

(5b)

Using (5a) and (5b), define the minimized Period 2 cost, as a function of the Period 2 states $b_2$ and $L_2$ and the price $p_2$, by

$$V_2(b_2, p_2, L_2) =$$

$$\min_{\alpha_2 \in \{0, 1\}} \left\{ \alpha_2 \left\{ 2 \gamma p_2 - \theta \left[ b_2 - p_2 (1 - \gamma) \right] \right\} + \theta \left( G - L_2 \right) \right\}$$

---

\(^{54}\) See footnote 44 above.
\[ \begin{align*}
(5c) \quad & \begin{cases} 
2\gamma p_2 - \theta [b_2 - p_2 (1 - \gamma)] + \theta \{G - L_2\}, & p_2 < p_2^* \ (\alpha_2 = 1), \\
\theta \{G - L_2\}, & \text{otherwise.}
\end{cases}
\end{align*} \]

**Period 1 Problem.** In Period 1, the investor chooses \( \alpha_1 \) to minimize the sum of (1) the Period 1 cost of realizing losses, and (2) the expected minimized Period 2 cost, discounted to Period 1, or

\[
\text{Min}_{\alpha_1 \in \{0, 1\}} \{ \alpha_1 p_1 2\gamma + \delta E_1 \left[ V_2 (b_2, \tilde{p}_2, L_2) \right] \},
\]

\text{subject to}

\[
b_1 \text{ given}, \\
p_1 \text{ given}, \\
L_1 = 0,
\]

and subject to the equations ((1a)-(1c)) that define the transition from Period 1 to Period 2 in terms of the Period 1 states and \( \alpha_1 \).

As a preliminary matter, a Period 1 sale at price \( p_1 = b_1 \) would, taking account of the fact that sales commissions reduce the amount realized from the sale, itself produce an apparent "loss." It is reasonably evident that an investor will not gain by realizing a loss produced solely by the expense of the sale; even so, it will be useful to justify that claim formally before we proceed. Thus, if the investor were to sell at \( p_1 = b_1/(1 + \gamma) < b_1 \), the Period 1 loss (\( L_2 \)), and the cost and basis at the start of Period 2, would be

\[
\hat{L}_2 = \alpha_1 [b_1 - p_1 (1 - \gamma)] = \alpha_1 b_1 \frac{2\gamma}{1 + \gamma};
\]

\[
\text{cost} = \alpha_1 2\gamma p_1 = \alpha_1 b_1 \frac{2\gamma}{1 + \gamma}; \text{ and}
\]

\[
\hat{b}_2 = b_1.
\]
So the "loss" just equals the cost, and the Period 2 basis and Period 2 critical price are what they would have been without any Period 1 sale. If, then, the realized price in Period 2 exceeded \( p^*_2(b_1) \), the Period 1 cost would have been expended for no return. On the other hand, substituting \( \hat{b}_2 \) and \( \hat{L}_2 \) into the minimized Period 2 cost function (with \( \alpha_2 = 1 \)), and evaluating that function at \( p^*_2(b_1) \), produces

\[
\theta G - \alpha_1 \theta b_1 \frac{2 \gamma}{1 + \gamma},
\]

which, when added to the Period 1 cost, yields total cost of

\[
\theta G + \alpha_1 b_1 \frac{2 \gamma}{1 + \gamma} (1 - \theta) > \theta G.
\]

So incurring cost in Period 1 just to deduct that cost cannot minimize aggregate cost. If, finally, \( p_2 < p^*_2 \), that loss can always be realized in Period 2. Hence, optimally,

\[
(6a) \quad p_1 > \frac{b_1}{1 + \gamma}, \quad \alpha^*_1 = 0, \quad b_2 = b_1.
\]

Also

\[
(6b) \quad \alpha_1 > 0, \quad p_1 < \frac{b_1}{1 + \gamma}, \quad b_2 < b_1.
\]

Returning to (6), expanding the expectation operator, differentiating with respect to \( \alpha_1 \), and using the transition relationships (1a)-(1c), yields

\[
\frac{\partial}{\partial \alpha_1} \left\{ \alpha_1 p_1 2 \gamma + \delta \int V_2(b_2, \tilde{p}_2, \tilde{L}_2) f(\tilde{e}) d\tilde{e} \right\}
\]

\[
= p_1 2 \gamma + \frac{\partial}{\partial \alpha_1} \left\{ \delta \int V_2(b_2(\alpha_1), p_1 + \tilde{e}, L_2(\alpha_1)) f(\tilde{e}) d\tilde{e} \right\}
\]

86
\[ p_1 2 \gamma + \delta \int \left\{ \left[ p_1 (1 + \gamma) - b_1 \right] \frac{\partial V_2(\cdot)}{\partial b_2} + \left[ b_1 - p_1 (1 - \gamma) \right] \frac{\partial V_2(\cdot)}{\partial L_2} \right\} f(\bar{e}) d\bar{e} \]

(7)

\[ p_1 2 \gamma - \theta \delta \left[ b_1 - p_1 (1 - \gamma) \right] + \delta \int \left\{ \left[ p_1 (1 + \gamma) - b_1 \right] \frac{\partial V_2(\cdot)}{\partial b_2} \right\} f(\bar{e}) d\bar{e}, \]

where the last line follows from the fact that, independent of \( \alpha_1 \) and \( \bar{e} \),

(8a) \[ \frac{\partial V_2}{\partial L_2} = -\theta. \]

The first term in (7) is the increase in Period 1 commission expense attributable to selling additional stock. The second is the discounted value of the resulting reduction in Period 2 tax liability (\( \theta \delta \), multiplied by the Period 1 loss). The third is the discounted value of the expected increase in Period 2 tax liability, attributable to the fact that sale at a loss in Period 1 reduces the shareholder’s basis (in accordance with (1b), with \( p_1 < b_i/(1 + \gamma) \), as justified by (6b)), and hence his potential for realizing losses in Period 2. With \( \alpha_1 \) buried in the expectation operator, we cannot solve explicitly for the optimal choice in Period 1. Dynamic programming offers a possible avenue to a complete solution; but, in contrast with a standard dynamic programming problem, we wish to allow for a possibly non-stationary sequence of optimal decisions within each year. Nevertheless, we can obtain an essentially complete characterization of the solution from the first-order condition for this simple two-period model.

Note, first, that for \( \alpha_2 = 0 \) (corresponding with \( p_2 \geq p^*_2 \)), the derivative of the value function (5c) with respect to \( b_2 \) vanishes, as does the integrand in (7). For \( \alpha_2 = 1 \) (or \( \bar{p}_2 < p^*_2 \)), on the other hand

(8b) \[ \frac{\partial V_2}{\partial b_2} = -\theta, \]
again independent of $\alpha_1$ and $\tilde{\epsilon}$. Using these observations we can rewrite (7) in the following equivalent ways:

\begin{align}
(7a) & \quad p_1^2 \gamma (1 - \theta \delta) - \theta \delta \left[ b_1 - p_1 (1 + \gamma) \right] \left[ 1 - \text{Prob}(\tilde{p}_2 < p_2^*) \right] \\
(7b) & \quad = p_1^2 \gamma (1 - \theta \delta) - \theta \delta \left[ b_1 - p_1 (1 + \gamma) \right] \text{Prob}(\tilde{p}_2 \geq p_2^*) \\
(7c) & \quad = p_1^2 \gamma (1 - \theta \delta) - \theta \delta \left[ b_1 - p_1 (1 + \gamma) \right] \text{Prob}(a_2 = 0).
\end{align}

For $\gamma = 0$, the derivative is (as we would expect) always non-positive: waiting gains nothing, since at zero cost there is nothing to lose by realizing an available loss now. For strictly positive $\gamma$, on the other hand, the sign of the derivative depends (using (7b)) on the weight attached to the probability that the realized Period 2 price will equal the price at which it saves nothing to realize any loss in Period 2. With sufficiently high probability the derivative is negative, and increasing $\alpha_1$ reduces total expected cost, since it is that much more likely that a loss that could be realized today will not be available tomorrow. With the probability weight in (7b) set to one, however, we still can require that the derivative be positive and solve for $p_1$, in which event we obtain $p_1^*$ (as given by (2a)), the price at which it pays to wait irrespective of the Period 2 price.

The probability weight in (7) must be handled with care. Recall that the critical price ($p_2^*$) is a function of the Period 2 basis (see (2b)), which in turn is a function of the Period 1 control (see (1b)). Hence the relationship between the $p_2$ and $p_2^*$ in (7b), and the sign of the derivative, is itself a function of $\alpha_1$. We can, however, make that relationship more explicit by rewriting the derivative (the details are in the Appendix) as
\[ g(\alpha_1, \pi) = \pi 2 \gamma (1 - \theta \delta) - \theta \delta \left[ 1 - \pi (1 + \gamma) \right] \text{Prob} \left( \tilde{\epsilon} \geq \epsilon^*(\alpha_1, \pi) \right), \]

where

\[ \pi = \frac{p_1}{b_1}, \quad \text{and} \quad \epsilon^*(\alpha_1, \pi) = \frac{\theta \{ (1 - \alpha_1) + \alpha_1 \pi (1 + \gamma) \}}{2 \gamma + \theta (1 - \gamma)} - \pi. \]

In (9), as in the simpler formulation of Section D, the optimal choice of \( \alpha_1 \) depends on the parameters of the problem and on the ratio of price to basis (\( \pi \))\(^55\), which also determines the ratio of Period 1 price to realizable loss. Henceforth, set \( \pi_i = p_i/b_1 \), and \( \pi_i^* = p_i^*/b_1 \).

2. Characteristics of the Solution

In the Appendix, I show that \( g(\alpha_1, \pi) \) is strictly decreasing in \( \alpha_1 \). Consequently, the problem (6) is strictly concave in \( \alpha_1 \), and (as in Period 2) the Period 1 solution will be found at a corner of the control set. In light of that fact, and to see what more specifically guides the choice, I investigate the sign of \( g(\alpha_1, \pi) \) for \( \gamma > 0 \), with \( \alpha_1 \) set to both 0 and 1. First, setting \( \alpha_1 = 0 \) in (9), imposing \( g(0, \pi) > 0 \), and solving for the probability weight, we find the derivative is positive when

\[
\text{Prob} \left( \tilde{\epsilon} \geq \epsilon^*(0, \pi) \right) = \text{Prob} \left( \tilde{\epsilon} \geq \frac{\theta}{2 \gamma + \theta (1 - \gamma)} - \pi \right) < h(\pi),
\]

where \( h(\pi) \) is as given in (4d). The first term in \( \epsilon^*(0, \pi) \) is just \( \pi_2^*(b_1) \). Thus, the weight is for the probability of drawing an \( \epsilon \) sufficient to move the Period 2 price from \( \pi_1 \) to \( \pi_2^*(b_1) \). Observe, moreover, that (10a) is an exact counterpart of the condition \( \pi \geq \pi_{(1, 1)} \) from Section D, as rewritten in (4d), correcting only for differences in formulation. There, \( \epsilon > 0 \) was assumed to produce \( p_2 \geq p_2^*(b_1) \), so the investor optimally wait-
ed if \( q \leq h(\pi) \). Here, \( h(\pi) \) is the probability weight associated with all values of \( \epsilon \) sufficient to produce \( \pi_2 \geq \pi^*_2(b_1) \). Otherwise the conditions are identical. As in Section D, the investor optimally chooses \( \alpha_1 = 0 \) and waits in Period 1 if, conditional on waiting, the probability of waiting again in Period 2 does not exceed the ratio of transaction costs (net of discounted tax savings) to the discounted value of taxes saved. Thus, a probability bound of 0.50 again corresponds to \( \pi = 0.810 \), now with \( \epsilon^*(0, \pi) = 0.0806 \) (given \( \pi^*_2(b_1) = 0.8906 \)). Here, then, for an asset trading at \( p_1 = \pi b_1 = 0.810 b_1 \), the derivative of the investor's problem evaluated at \( \alpha_1 = 0 \) is positive if the probability of an 8.1 percent price rise is no more than 0.50.

Next, setting \( \alpha_1 = 1 \) in (9) and requiring \( g(1, \pi) < 0 \), we can solve again for the probability weight, and rewrite the result in terms of the complementary probabilities, to find that the derivative is negative at \( \alpha_1 = 1 \) when

\[
(10b) \quad \text{Prob}(\bar{\epsilon} < \epsilon^*(1, \pi)) = \text{Prob}\left(\bar{\epsilon} < \pi \frac{-2\gamma(1-\theta)}{\gamma + \theta(1-\gamma)}\right) \leq 1 - h(\pi).
\]

Notice that, just as \( \pi + \epsilon^*(0, \pi) = \pi^*_2(b_1) = \pi^*_2(\alpha_1 = 0) \),

\[
\pi + \epsilon^*(1, \pi) = \frac{\pi(1+\gamma)\theta}{\gamma + \theta(1-\gamma)} = \pi^*_2(p_1(1+\gamma)) = \pi^*_2(\alpha_1 = 1).
\]

Thus, the critical movement of \( \bar{\epsilon} \) is now to the point at which, having already realized a loss in Period 1, it is optimal to realize an additional loss in Period 2, thereby duplicating the investor's cost. If that probability is sufficiently high it pays to reduce \( \alpha_1 \) to less than 1. Otherwise, the investor optimally chooses \( \alpha_1 = 1 \).

In this formulation, then, the derivative very well may be positive at \( \pi_1 < \pi^*_1 \) and \( \alpha_1 = 0 \), in which event the investor does better to wait. More generally, however, the optimal Period 1 choice may depend on both the probabilities (1) conditional on waiting in Period 1, that the price will rise to the point at which the investor waits again in Period 2, and (2) conditional on exchanging in Period 1, that the price then falls to the point at which an additional loss is realized in Period 2. If the first is sufficiently high,
$g(0, \pi) < 0$ and (since $g(\alpha_1, \pi)$ is decreasing in $\alpha_1$) the investor chooses $\alpha_1 = 1$. If the second is sufficiently low, $g(1, \pi) \geq 0$ and the investor chooses $\alpha_1 = 0$.

3. Optimal Tax-Loss Selling

That leaves the important question unanswered. One can easily imagine probability distributions that produce $g(0, \pi) > 0 > g(1, \pi)$. For the binomial distribution of Section D, the condition $\pi = \pi_{(2, 2)}$ in Table 1 (for both sufficiently small $\epsilon$ and a choice of $\pi$ that ensure that $\pi_{(2, 2)}$ is the appropriate condition) is an obvious case in point.\(^{56}\) There the investor waits at that (and any higher) ratio of price to basis; below that the investor does best to exchange. In the more general case now under study, however, it is not obvious that the same conclusion will hold. Instead, the optimal choice generally cannot be determined without positing a distribution for $\bar{\epsilon}$ and evaluating the objective function (6) at the extreme settings of the control (although concavity in $\alpha_1$ ensures that the optimum will be found at one of those extremes).

The example of Section D does suggest, however, that for some probability distributions for $\bar{\epsilon}$, there exists some price above which it is optimal to wait. In this part I begin by showing that a simple condition on the support of $f(\bar{\epsilon})$ ensures generally that this will be so. Specifically, I show that if the support of $\bar{\epsilon}$, when the Period 1 price/basis ratio is $\pi^*_1$, includes $\epsilon < 0$ sufficient so that, if the investor sold in Period 1 at $\pi^*_1$, he could draw an $\epsilon$ that would induce him to sell again in Period 2, there exists some $\pi_i < \pi^*_1$ above which it is optimal to wait. That is,

\begin{itemize}
  \item \textbf{Theorem 1.} Suppose that $f(\bar{\epsilon})$ is continuous and strictly positive on $\bar{\epsilon} < \bar{\epsilon} < 0$, and that $\epsilon^*(1, \pi) \in (\epsilon, 0)$ for all $\epsilon^*$ to be considered. If

  $$\text{Prob}(\bar{\epsilon} < \epsilon^*(1, \pi_1)) = \beta > 0,$$

\end{itemize}

\(^{56}\) For that case, $\text{Prob}(\bar{\epsilon} \geq \epsilon^*(0, \pi)) = 0 < h(\pi) \forall \pi > 0$; and $\text{Prob}(\bar{\epsilon} < \epsilon^*(1, \pi)) = 0 \leq 1 - h(\pi)$. 

91
\[ \exists \hat{\pi} < \pi^*_1 \text{ and } h(\hat{\pi}) \text{ satisfying} \]
\[ \text{Prob}(\hat{\epsilon} < \epsilon^*(1, \hat{\pi})) \geq 1 - h(\hat{\pi}). \]

**Proof.** The proof is in the Appendix.

**Comment.** The requirements of continuity and strict positivity on the density are adopted solely for convenience. The important condition is \( \epsilon^*(1, \pi^*_1) \in (\epsilon, 0) \). That ensures that there is an open interval of positive probability in the lower tail of the probability distribution for \( \hat{\epsilon} \) that includes \( \epsilon^* \) if the investor were to sell in Period 1 at \( \pi^*_1 \). With that condition, Theorem 1 establishes that we can find a price strictly less than \( \pi^*_1 \) at which, if the investor exchanges in Period 1, the probability of drawing an \( \epsilon \) that will induce a Period 2 exchange exceeds the probability bound given by (10b). The import of that fact is given by

**Corollary 1.** If Theorem 1 holds, then

\[ (a) \ g(1, \hat{\pi}) > 0, \text{ and} \]
\[ (b) \ g(\alpha_1, \hat{\pi}) > 0 \ \forall \alpha_1 \in [0, 1]. \]

**Proof.** Part (a) follows from Theorem 1 and (10b). Given (a), Part (b) then follows from the fact that \( g(\alpha_1, \pi) \) is decreasing in \( \alpha_1 \).

Theorem 1 and its Corollary establish that positive probability of the event \( \hat{\epsilon} < \epsilon^*(1, \pi^*_1) \) is sufficient to guarantee the existence of Period 1 prices \( \pi < \pi^*_1 \) at which the derivative of the Problem (6), evaluated at \( \alpha_1 = 1 \), is positive, and therefore of Period 1 prices at which, in the face of uncertainty about the price in Period 2, it is nevertheless optimal to choose \( \alpha_1 = 0 \). When Theorem 1 holds, we can also establish that there exists a unique Period 1 price, \( \pi^* \), at which the investor is indifferent between taking the loss available in Period 1 and waiting until Period 2.
THEOREM 2. Suppose that Theorem 1 holds. If

\[ \frac{\partial \Pr(a_1, \pi)}{\partial \pi} < \frac{2\gamma(1-\theta\delta) + \theta \delta (1+\gamma) \Pr(a_1, \pi)}{\theta \delta [1-\pi(1+\gamma)]}, \]

where \( \Pr(a_1, \pi) \) is the probability weight in (9), then

(a) there exists a unique \( \pi^* \in (0, \hat{\pi}) \) at which the investor is indifferent between \( a_1 = 0 \) and \( a_1 = 1 \), and

(b) \( \pi^* > \pi_1^* = 0, \) and \( \pi^* < \pi_1^* = 1. \)

PROOF. The proof is in the Appendix.

COMMENT. The condition is of a technical nature. Obtained by requiring that the derivative of (9) with respect to \( \pi \) be positive and solving for the derivative of the probability weight, it is essentially a condition on the slope of the distribution function of \( \hat{\varepsilon} \) that will be satisfied in practice by probability distributions typically assumed for security price processes. It insures that \( g(a_1, \pi) \) is strictly increasing in \( \pi \).

What Theorem 2 adds to Theorem 1 is simply that there exists some unique ratio of price to basis \( \pi^* \), with \( \pi^* < \hat{\pi} < \pi_1^* \), below which the investor does better to exchange. At or above \( \pi^* \), however, the investor does better to wait. With these simple requirements the results obtained in Section D for the binomial distribution continue more generally to hold. There is an interval of prices above which, in the face of uncertainty, the investor generally does better to wait. That is the important conclusion of the paper.

Although the model is formulated with only two periods, it is implicitly parameterized by the continuous passage of time, captured by \( \delta \) in the first-order condition (9).

A consequence of Theorem 2 is

COROLLARY 2. If Theorem 1 and the condition to Theorem 2 hold, then, prior to the arrival of Period 2, \( \pi^* \) is increasing in time.

PROOF. Differentiate (9) with respect to \( \delta \) to obtain

\[ \frac{\partial g(a_1, \pi)}{\partial \delta} = -n \pi 2 \gamma \theta - n \theta [1 - \pi(1+\gamma)] \text{Prob} \left[ \hat{\varepsilon} > \varepsilon^*(a_1, \pi) \right] < 0, \]
for $\pi < \pi^* < 1/(1 + \gamma)$. Since $\delta' > \delta = g(\alpha^*_1, \pi^*, \delta') < g(\alpha^*_1, \pi^*, \delta) = 0$, the fact that $g(\alpha_1, \pi)$ is increasing in $\pi$ requires $\pi'' > \pi'$ for $g(\alpha''_1, \pi'', \delta') = 0$. \[ \square \]

With the assumption that investors acquired the loss security at prices distributed between 0 and the security's (recent) high price, Corollary 2 implies that, as the distance between Period 1 and the end of the year shrinks, volume in the loss stock should rise. That is simply because, as $\pi^*$ rises with time, the number of investors holding that security at $\pi < \pi^*$ (and who therefore find it optimal to sell) should likewise rise, and the number who continue to wait should decline. Note, however, that even for $\delta$ arbitrarily close to 1, and $\pi^*_1$ arbitrarily close to $\pi^*_2(b_1)$, it remains the case, under the conditions of Theorem 1, that $\hat{\pi} < \pi^*_1 = \pi^*_2$. Consequently, if Theorem 1 holds, then, regardless of how close Period 1 and Period 2 are in time, there will still be a discontinuous movement, between Period 1 and Period 2, in the price/basis ratio below which it is optimal to exchange (as illustrated in Figure II-3, using $\pi_{(1, 1)}$ from the binomial example of Section D). The natural implication is that, at least where securities with loss sale potential are concerned, something discernible ought to occur in the vicinity of the end of the year.

Figure II-3

$\pi_{(1, 1)}(0.5)$ and $\pi$:

($\theta = 0.28; \gamma = 0.02$)
F. Implications

I argued in Section B that for tax purposes it is more plausible to believe a priori that long-term and short-term losses -- "capital" or "ordinary" -- on average are of equal value to investors than that short-term capital losses are systematically and substantially more valuable. What Sections D and E establish is that, without the assumption of differential value, the theoretical foundation for the belief that investors optimally should liquidate losers whenever they materialize disappears. With (I maintain) the more plausible assumption of approximately equivalent value of losses of different duration, Sections D and E demonstrate, to the contrary, that an investor will often do better to wait.

That prediction compels a fresh look at the evidence on the January effect. Although often suffused with a discomfort attributable to Constantinides’ theoretical work, most of that evidence in fact is consistent with the proposition that year-end tax-motivated selling bears importantly on the January effect. The model developed here suggests that findings that the January effect is related to both long and short-term year-end loss sale potential (as in Chan (1986)), or of a year-end depression in issues in which the January effect is pronounced (as documented in Essay One), may be quite consistent with optimizing behavior. Even more directly to the point is that vein of the January effect literature, beginning with Dyl (1977) and including Lakonishok and Smidt (1986), that documents a year-end rise in volume among securities with loss sale potential, of the sort that this paper predicts. Particularly striking is the recent evidence of Badrinath and Lewellen (1991), who studied the conduct of an actual sample of investors (rather than, as is typical, the CRSP data) during the period 1971-1979, in which they documented a monotonic rise over the course of the year in sales of securities with loss sale potential, just as Corollary 2 predicts.

The model developed here suggests that, in getting to the bottom of the January effect, even more careful attention to year-end daily data would be fruitful. Indeed, the model contains one readily testable implication that is substantially more precise than simply that volume in securities with loss-sale potential should rise as the end of the year draws near. In Table II-2 I have arrayed the ratio $\pi_{(1,1)}$ from Table 1,
using a plausible range of post-war discount rates ($\delta$) and (for convenience) a 50 percent probability that the price will move up, for each long-term capital gain tax rate ($\theta$) that has prevailed since 1962.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>0.350</th>
<th>0.280</th>
<th>0.250</th>
<th>0.200</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9850</td>
<td>0.8532</td>
<td>0.8130</td>
<td>0.7906</td>
<td>0.7429</td>
</tr>
<tr>
<td>0.9000</td>
<td>0.8506</td>
<td>0.8100</td>
<td>0.7875</td>
<td>0.7394</td>
</tr>
<tr>
<td>0.9500</td>
<td>0.8470</td>
<td>0.8060</td>
<td>0.7832</td>
<td>0.7347</td>
</tr>
<tr>
<td>0.9300</td>
<td>0.8433</td>
<td>0.8018</td>
<td>0.7788</td>
<td>0.7299</td>
</tr>
</tbody>
</table>

It can easily be seen that while $\pi_{(1, 1)}$ is not very sensitive to $\delta$, it is quite sensitive to the capital gains rate, declining monotonically with $\theta$. As noted in connection with Corollary 2, if investors’ cost bases are dispersed, at lower values of $\pi_{(6, \beta)}$ fewer investors should find it cost-minimizing to take losses immediately and more should be content to wait until the end of the year. Consequently, Table II-2 suggests that not only should we expect to find elevated year-end volume in losers, generally; but, more specifically, that we should expect to find that phenomenon to be discernibly more pronounced in years when the long-term capital gain rate was low. Testing that relationship stands out as a logical next step.
APPENDIX

1. The domain of $\pi_{(1,1)}$ (\(\equiv p_{(1,1)}/b_1\)). For convenience, let

\[ h = 2 \gamma + \theta(1 - \gamma). \]

Then for $\epsilon$ to satisfy both (3a) and (3b), we must have

\[ \begin{align*}
(A1a) & \quad \epsilon \geq p_2^*(b_1) - p_1 = \frac{\theta b_1}{h} - p_1, \quad \text{and} \\
(A1b) & \quad \epsilon > p_1 - p_2^*(p_1(1 + \gamma)) = p_1 - \frac{\theta(p_1(1+\gamma))}{h}.
\end{align*} \]

Equating the right-hand sides of (A1a) and (A1b) and solving for $p_1$ yields

\[ \begin{align*}
(A2a) & \quad p_{(1,1)} = \frac{b_1 \theta}{2h - \theta(1 + \gamma)} \approx 0.8159 \, b_1.
\end{align*} \]

Using (A1a) or (A1b) yields

\[ \begin{align*}
(A2b) & \quad \epsilon_{(1,1)} = \frac{b_1 \theta(h - \theta(1 + \gamma))}{2h - \theta(1 + \gamma)} \approx 0.0747 \, b_1.
\end{align*} \]

2. $\pi_{(1,1)}$ as an upper bound on the price/basis ratio above which it pays to wait.

I now show, as claimed in Part IV, that

\[ \epsilon > \epsilon_{(1,1)} = p_{(1,1)} < p_{(1,1)}. \]

Proof. Observe first that for $\Delta > 0$ and $p_1 = p_{(1,1)} - \Delta$,

\[ \begin{align*}
& \quad \left[ p_{(1,1)} - \Delta \right] - p_2^*(p_{(1,1)} - \Delta)(1 + \gamma) \\
& = \left[ p_{(1,1)} - p_2^*(p_{(1,1)}(1 + \gamma)) \right] - \left[ \Delta - p_2^*(\Delta(1 + \gamma)) \right] \\
& < p_{(1,1)} - p_2^*(p_{(1,1)}(1 + \gamma)) < \epsilon_{(1,1)},
\end{align*} \]
(using \( p_2^*(b) < b \) and (A1b)) so, for \( \varepsilon > \varepsilon_{(1,1)} \) and \( p_1 < p_{(1,1)} \),

\[
p_1 - \varepsilon < p_2^*(p_1(1 + \gamma))
\]

i.e., the domain of \( p_{(1,2)} \) and \( p_{(2,2)} \) is empty. The domain of \( p_{(2,1)} \), on the other hand, is non-empty only for

\[
p_1 < p_2^*(b_1) - \varepsilon = \frac{\theta b_1}{h} - \varepsilon.
\]

But for

\[
\varepsilon < \frac{\theta b_1}{h} - p_{(2,1)},
\]

using the definition of \( p_{(2,1)} \) in Table 1,

\[
p_{(2,1)} = \frac{\varepsilon \delta h}{2\gamma(1 - \theta \delta - q \delta[1 - \theta])}
\]

\[
< \left( \frac{\theta b_1}{h} - p_{(2,1)} \right) \left( \frac{q \delta h}{2\gamma(1 - \theta \delta - q \delta[1 - \theta])} \right)
\]

\[
= (\theta b_1 - h p_{(2,1)}) \left( \frac{q \delta h}{2\gamma(1 - \theta \delta - q \delta[1 - \theta])} \right),
\]

or (using the definition of \( h \))

\[
p_{(2,1)} < \frac{q \theta \delta b_1}{2\gamma(1 - \theta \delta - q \delta[1 - \theta]) + q \delta h}
\]

\[
= \frac{q \theta \delta b_1}{2\gamma(1 - \theta \delta) + \theta \delta q(1 + \gamma)}
\]

\[= p_{(1,1)} \]

3. *Curvature of the derivative of (6).* We can rewrite the derivative (as given in (7b)) to make more explicit its dependence on \( \alpha_1 \) by noting that

\[
\hat{p}_2 \geq p_2^*
\]

\[\Leftrightarrow\]

(A3a)

\[p_1 + \check{\varepsilon} \geq p_2^*
\]
\[ \hat{\epsilon} \geq p_2^* - p_1. \]

Consequently, we can replace \( p_2^* \) by its definition (2b), replace \( b_2 \) in the definition of \( p_2^* \) using the transition relationship (1b), and then use (A3) to rewrite (7b) entirely in terms of the states observable at Period 1, the Period 1 decision, and \( \hat{\epsilon} \):

\[ (A3b) \quad np_1 2\gamma (1 - \theta \delta) - n\theta \delta \left[ b_1 - p_1 (1 + \gamma) \right] \text{Prob} \left( \hat{\epsilon} \geq \frac{\theta (b_1 (1 - \alpha_1) + \alpha_1 p_1 (1 + \gamma))}{2\gamma + \theta (1 - \gamma)} - p_1 \right). \]

Replacing \( p_1 \) by \( \pi \) and \( b_1 \) by 1 (and appropriately recalling \( \epsilon \)) produces \( g(\alpha_1, \pi) \) in (9).

I now show that:

\[ \frac{\partial g(\alpha_1, \pi)}{\partial \alpha_1} < 0. \]

**Proof.** For \( \gamma > 0 \), the sign of (9) depends on the weight attached to the probability of drawing an \( \epsilon \) sufficient to move the Period 2 price from \( \pi \) to at least the critical price \( \pi_2^* \) (= \( p_2^*/b_1 \)). Given \( \pi \), however, both \( \pi_2^* \) and \( \epsilon^* \) in the probability weight depend on \( \alpha_1 \). The first term in \( \epsilon^*(\alpha_1, \pi) \) -- the price/basis ratio \( \pi_2^* \), written as a function of \( \alpha_1 \) and \( \pi \) -- is positive. From (6a), we know it is not worthwhile to realize a loss in Period 1 if \( \pi \geq b_1/(1 + \gamma) \) (with \( b_1 \) normalized to 1). For \( \pi < 1/(1 + \gamma) \), on the other hand, the derivative of \( \epsilon^* \) with respect to \( \alpha_1 \) is

\[ \frac{\partial \epsilon^*}{\partial \alpha_1} = \frac{\theta (\pi (1 + \gamma) - 1)}{2\gamma + \theta (1 - \gamma)} < 0, \]

so \( \epsilon^* \) is strictly decreasing in \( \alpha_1 \). Consequently, if we fix the support for the probability distribution of \( \hat{\epsilon} \), the value that \( \hat{\epsilon} \) must exceed in Period 2 decreases in \( \alpha_1 \), and the probability of drawing an \( \epsilon \) that exceeds that value increases. Hence, the probability weight on the negative term in \( g(\alpha_1, \pi) \) is increasing in \( \alpha_1 \).
4. Proof of

THEOREM 1. Suppose that \( f(\varepsilon) \) is continuous and strictly positive on \( \varepsilon < \bar{\varepsilon} < 0 \), and that \( \varepsilon^*(1, \pi) \in (\varepsilon, 0) \) for all \( \varepsilon^* \) to be considered. If

\[
\text{Prob}(\bar{\varepsilon} < \varepsilon^*(1, \pi^*_1)) = \beta > 0,
\]

then

\[
\exists \hat{n} < \pi^*_1 \text{ and } h(\hat{n}) \text{ satisfying}
\]

\[
\text{Prob}(\bar{\varepsilon} < \varepsilon^*(1, \hat{n})) > 1 - h(\hat{n}).
\]

PROOF. Since \( h(\pi^*_i) = 1 \), \( g(1, \pi^*_i) > 0 \) (since \( \text{Prob}(\bar{\varepsilon} < \varepsilon^*(1, \pi^*_i)) > 0 \)). Note from (9) (with \( \theta \in (0, 1) \)) that

\[
\frac{\partial \varepsilon^*(\alpha_1, \pi)}{\partial \pi} = \frac{\alpha_1 \theta (1 + \gamma)}{2 \gamma + \theta (1 - \gamma)} - 1 > 0
\]

for \( \alpha_1 > \frac{2 \gamma + \theta (1 - \gamma)}{\theta (1 + \gamma)} > 1 \),

so for \( \alpha_1 = 1 \), \( \varepsilon^*(1, \pi) < 0 \) is decreasing in \( \pi \) and its absolute value is increasing. Consequently, \( |\varepsilon^*(1, \pi)| < |\varepsilon^*(1, \pi^*_i)| \) for \( \pi < \pi^*_i \). Since \( h(\pi^*_i) = 1 \) and \( h(\pi) \) is continuous for \( \pi < \pi^*_i < 1/(1 + \gamma) \), by the intermediate value theorem we can, by choosing \( \pi \) sufficiently close to \( \pi^*_i \), make \( h(\pi) \) arbitrarily close to 1, and \( 1 - h(\pi) \) arbitrarily close to zero. In particular, then, for any \( \beta > 0 \) we can (using the definition of \( h(\pi) \) in 4(a)) choose \( \hat{n} \) to satisfy

\[
1 - h(\hat{n}) = 1 - \frac{\kappa \hat{n}}{1 - \hat{n}(1 + \gamma)} = \beta
\]

or

\[
\hat{n} = \frac{1 - \beta}{\kappa + (1 - \beta)(1 + \gamma)}
\]

Note that

\[
h(\hat{n}) = 1 - \beta < 1,
\]
so, by construction, \( \hat{\pi} < \pi^* \), which implies \( |\epsilon^*(1, \hat{\pi})| < |\epsilon^*(1, \pi^*)| \), or, equivalently, \( \epsilon^*(1, \hat{\pi}) > \epsilon^*(1, \pi^*) \). Hence \( \hat{\pi} + \epsilon^*(1, \hat{\pi}) > \hat{\pi} + \epsilon^*(1, \pi^*) \).

Denoting by \( \text{Prob}_t \) the probability at price \( \hat{\pi} \), the assumptions about \( f(\bar{\varepsilon}) \) imply that

\[
\text{Prob}_t \left( \hat{\pi} + \epsilon^*(1, \pi^*_1) < \hat{\pi}, \hat{\pi} + \epsilon^*(1, \pi^*_1) \right)
\]

\[
= \text{Prob}_t \left( \hat{\pi} + \epsilon^*(1, \pi^*_1) < \hat{\pi}, \hat{\pi} + \epsilon^*(1, \pi^*_1) \right)
\]

\[
= \text{Prob}_t \left( \epsilon^*(1, \pi^*_1) < \hat{\pi} < \epsilon^*(1, \pi^*_1) \right) > 0.
\]

Now assume that

\[
\text{Prob}_t (\bar{\varepsilon} < \epsilon^*(1, \pi^*)) \leq 1 - h(\hat{\pi}) = \beta.
\]

Then

\[
\beta = \text{Prob}_t (\bar{\varepsilon} < \epsilon^*(1, \pi^*)) < \beta
\]

leads to a contradiction, from which it follows that

\[
\text{Prob}_t (\bar{\varepsilon} < \epsilon^*(1, \pi^*)) > 1 - h(\hat{\pi})
\]

5. Proof of

\[
\text{THEOREM 2. Suppose that Theorem 1 holds. If}
\]

\[
\frac{\partial \text{Pr}(\alpha_1, \pi)}{\partial \pi} < \frac{2\gamma(1 - \theta \delta) + \theta \delta(1 + \gamma) \text{Pr}(\alpha_1, \pi)}{\theta \delta[1 - \pi(1 + \gamma)]},
\]

where \( \text{Pr}(\alpha_1, \pi) \) is the probability weight in (9), then

(a) there exists a unique \( \pi^* \in (0, \hat{\pi}) \) at which the investor is indifferent between \( \alpha_1 = 0 \) and \( \alpha_1 = 1 \), and

(b) \( \pi > \pi^* \Rightarrow \alpha_1^* = 0, \) and \( \pi < \pi^* \Rightarrow \alpha_1^* = 1. \)
PROOF. (a) Denote the objective function (6), written in terms of $\pi$, by $c(\alpha_1, \pi)$. Expression (1c) means that $\pi > 0$ has positive probability at any price $\pi$, so that, using (9), $g(\alpha_1, 0) < 0$ for any choice of $\alpha_i$. From that fact, the corollary to Theorem 1, and the intermediate value theorem, there exists some $\pi_{01} \in (0, \pi)$ satisfying

$$g(1, \pi_{01}) = 0.$$ 

(Given the condition of the theorem, $\pi_{01}$ is unique.) Since $g(\alpha_1, \pi)$ is decreasing in $\alpha_1$, it follows that

$$g(\alpha_1, \pi_{01}) > 0, \quad \forall \alpha_1 < 1,$$

and therefore that

$$c(1, \pi_{01}) - c(0, \pi_{01}) > 0.$$ 

By similar reasoning, there exists a unique $\pi_{00}$, strictly less than $\pi_{01}$ given the condition of the theorem, satisfying

$$g(0, \pi_{00}) = 0,$$

$$g(\alpha_1, \pi_{00}) < 0, \quad \forall \alpha_1 > 0, \text{ and}$$

$$c(1, \pi_{00}) - c(0, \pi_{00}) < 0.$$ 

Define $C(\pi)$ by

$$C(\pi) \equiv c(1, \pi) - c(0, \pi).$$

Note that

$$C(\pi_{00}) < 0,$$

$$C(\pi_{01}) > 0,$$

and that

$$\frac{\partial}{\partial \alpha_1} \left( \frac{\partial c(\alpha_1, \pi)}{\partial \pi} \right) = \frac{\partial}{\partial \pi} \left( \frac{\partial c(\alpha_1, \pi)}{\partial \alpha_1} \right) = \frac{\partial g(\alpha_1, \pi)}{\partial \pi} > 0,$$

(where the inequality follows from the condition of the theorem), so that
\[
\frac{dC(\pi)}{d\pi} = \frac{\partial c(1, \pi)}{\partial \pi} - \frac{\partial c(0, \pi)}{\partial \pi} > 0.
\]

Hence, again by the intermediate value theorem, there exists a unique \( \pi^* \in (\pi_{00}, \pi_{01}) \) at which

\[
C(\alpha, \pi^*) = c(1, \pi^*) - c(0, \pi^*) = 0.
\]

(b). From Rolle’s theorem it now follows that there is some \( \alpha^*_1 \in (0, 1) \) satisfying

\[
\frac{\partial c(\alpha^*_1, \pi^*)}{\partial \alpha_1} = g(\alpha^*_1, \pi^*) = 0,
\]

so that

\[
\alpha_1 < \alpha^*_1 \Rightarrow g(\alpha_1, \pi^*) > 0,
\]

\[
\alpha_1 > \alpha^*_1 \Rightarrow g(\alpha_1, \pi^*) < 0.
\]

Finally, since \( C(\pi) \) is strictly increasing in \( \pi \),

\[
\pi > \pi^* \Rightarrow C(\pi) > 0, \text{ while}
\]

\[
\pi < \pi^* \Rightarrow C(\pi) < 0,
\]

so the investor’s optimal choice is \( \alpha^*_1 = 0 \) for \( \pi > \pi^* \), and conversely for \( \pi < \pi^* \).
ESSAY THREE

Two Applications of
The Samuelson Invariance Theorem

A. Introduction

In a brief paper published 30 years ago, Paul Samuelson (1964) established a simple proposition that has been extraordinarily influential in policy discussions respecting the Federal income tax: to insure that the valuation of a productive asset is invariant to the tax rate of the owner, depreciation should be economic, in the sense that the deductions periodically allowable for depreciation of an asset should mirror the actual decline in its value. The attractive implications of Samuelson’s theorem are that an income tax need not affect after-tax rates of return to differing investments in a (differentially) distortionary fashion; and, moreover, that, with economic depreciation, after-tax rates of return to investments will decline in proportion to each investor’s marginal tax rate, thereby implementing the progressive effect of however graduated a rate structure Congress has prescribed. The "invariance theorem" (which has been shown by Fane (1987) to hold under uncertainty) thus essentially defines proper asset depreciation under a Haig-Simons (or "accretion") type tax.

But, as Samuelson noted, the import of the invariance result is more far-reaching. It was established by valuing an abstract representation of a stream of receipts, so in principle it applies to virtually any "investment", including not merely assets that conventionally are regarded as "depreciable," but also to outlays made post-production rather than in advance, as well as to wasting financial assets such as debt. Indeed, Samuelson’s theorem can be regarded as defining the ideal treatment of accretions (positive or negative) to wealth -- accumulation, generally -- under a Haig-Simons tax.57

57 A version of Part C of this Essay appeared as "Environmental ‘Remediation’ Expenses and a Natural Interpretation of the Capitalization Requirement", 47 National Tax Journal 703-719 (© 1994, reprinted with permission). The corresponding ideal for a consumption-type tax was implicitly established by Brown (1948), who showed that expensing asset costs is, with some assumptions, equivalent to exempting asset yields from tax. Under a "cash-flow" consumption tax, asset values are also effectively tax-rate invariant, since the rate on capital income is uniformly zero, as noted by, for example, Andrews (1974), Warren (1975) and Strnad (1992).
Congress has generally been slow to appreciate the breadth of Samuelson's insight. This paper considers two cases in point. The first is a series of enactments of a bit more than a decade ago, in which Congress prescribed the income tax treatment of a wide range of so-called "future costs," such as nuclear power plant decommissioning expenses. The treatments prescribed were generally consistent with one another. I show that, when judged by the criterion of Samuelson's paper, they are also consistently wrong. The second is the income tax treatment of debt, as to which the patchwork of recent congressional enactments defies any ready categorization. The aspect taken up here is the taxation of "cancellation of debt" (COD). Little attention has been paid to whether that rule is substantively sound. I show that it would be correct, at least if we treated interest generally as Samuelson's theorem prescribes, something that we do not, however, currently do.

In Section B I review Samuelson's theorem itself. In Section C, motivated by recently renewed controversy over the treatment of expenses of environmental remediation and restoration, I specialize it to the taxation of future costs. In Section D I take up the COD rule.

B. Economic Depreciation and Invariant Asset Valuation

Samuelson began by representing with the Fisher integral the value \( V(t) \) of an asset that produces a continuous net revenue stream \( N(t) \), known with certainty in advance, in an environment in which the pre-tax interest rate is \( y^8 \),

\[
V(t) = e^{yt} \int_{s=t}^{T} N(s) e^{-ys} ds, \tag{1}
\]

the time derivative of which is given by:

\[
V'(t) = yV(t) - N(t). \tag{2}
\]

I do not here assume the existence of any hypothetical "no-tax" world. I interpret (1) as the value of the asset to a tax-exempt investor in a taxable environment.
The taxable counterpart to (1), with tax imposed at rate $z$ and depreciation allowed in computing taxable income at a rate (possibly) dependent on the marginal tax rate, and therefore denoted $D(t, z)$, is:

$$V(t, z) = e^{y(1-z)t} \int_{s=t}^{T} \left\{ N(s) - z \left[ N(s) - D(s, z) \right] \right\} e^{-y(1-z)s} ds.$$

Samuelson showed that (3) is tax-rate invariant if and only if:

$$D(t, z) = D(t)$$

$$= N(t) - yV(t)$$

$$= -V'(t),$$

that is, only if the depreciation allowance is equal to the change in the asset’s value, given by the negative of its time derivative $V'(t)$.

The decomposition of the depreciation schedule (4) merits careful attention. It says that, for an investor who receives an (includible) net receipt $N(t)$ at time $t$, a principal element of the depreciation allowance is a deduction equal to the amount of the receipt. The intuition underlying this feature of the schedule is that two things occur simultaneously at time $t$: one is that the investor is enriched by the amount of the receipt; the other, however, is that the stream of remaining receipts to which ownership of the asset entitles them declines by an exactly offsetting amount. Hence, her net worth is unaffected by the mere conversion into cash of her claim to the receipt at time $t$.

The significance of the second term of (4), $-yV(t)$, becomes clear if (4) is substituted into (3) to obtain the actual tax rate-invariant expression for $V(t)$:

$$V(t, z) = V(t) = e^{y(1-z)t} \int_{s=t}^{T} \left\{ N(s) - zyV(s) \right\} e^{-y(1-z)s} ds.$$

The important feature of (5) is that, when economic depreciation is allowed, the only thing that is taxed is the instantaneous accrual of yield (at rate $y$) to the net value ($V(t)$)
of the asset itself. Thus, while most of the attention has been devoted to the allowance of economic depreciation, the equally fundamental implication of allowing economic depreciation is that it leads to the taxation of just \( yV(t) \). It is in this sense that economic depreciation implements a pure accrual tax.\(^{59}\)

C. Environmental "Remediation" Expenses and a Natural Interpretation of the Capitalization Requirement

1. The Problem

The income taxation of outlays for environmental restoration (or "remediation") has become a hot topic again. It elicited extensive attention a decade ago, when the focus was on the treatment of such "future costs" as the expenses of surface mine reclamation and nuclear power plant decommissioning, now explicitly covered by Internal Revenue Code (I.R.C.) §§ 468 and 468A. Although not currently controversial, surface mining affords a simple illustration. As a condition of mining coal by surface methods, a coal mine operator is unconditionally required by state law to restore its site, when no longer actively mined, to a reasonable facsimile of the condition the land was in before mining operations commenced. Based on its experience, it can estimate with reasonably accuracy (1) when it will cease its mining operations, and (2) how much it will cost to reclaim the land.

\(^{59}\) This is just a generalization of the more familiar observation that, for an asset with constant productivity (a "one-hoss shay"), "economic" depreciation produces a pattern of asset "depreciation" identical to the schedule of principal amortization for a level-payment fully amortized conventional mortgage loan. See Chirelstein (1994) § 6.08(d). In the latter instance it is familiar, of course, that the "proper" amount to be includible in the income of the lender (the "investor") is just the product of the outstanding "principal balance" and the instrument's "yield-to-maturity," as now provided for in I.R.C. §§ 1272(a). That is precisely what (5) produces in the way of taxable income to the holder of an asset subject to economic depreciation.

It is somewhat more realistic to expect that we can achieve approximately accurate accrual taxation of debt than that we can of depreciable assets. For a careful study of the application of Samuelson's theorem to depreciation when revenue profiles are known only statistically in advance, and of the difficulties in making the theorem operational when uncertainty is introduced, see Strnad (1991).
For tax purposes, how should the anticipated expenses be taxed? There is no dispute about whether they are deductible. The controversy is about when, and the basic possibilities are two. One is to allow a at the time they are actually paid. Equivalently, one could discount the entire future cost to its present value, using an after-tax discount rate, and allow a deduction in that discounted amount now, but no further deduction at any other time. I will refer to these approaches as "cash" (or "cash-equivalent") accounting. The alternative is some form of "accrual." In the literature the accrual approaches are described in a variety of ways, and justified by a variety of analogies, but they are similar in both provenance and operational content. Operationally, they all allow the present value of the liability -- determined using a pre-tax discount rate -- to be deducted when the liability first becomes fixed and reasonably quantifiable, and thereafter to allow the balance of the liability to be deducted over time as it accrues. And they all claim kinship to Samuelson's paper. The efforts 10 years ago (and since) have not produced an agreement on how this problem should properly be solved. As the result of the Tax Reform Act of 1984, however, the basic issues have been by and large resolved. The statute now leans heavily in the direction of cash-equivalent accounting.

---

60 Some will object to the implications of this terminology, which I adopt in the interests of convenience. Actually, any stream of deductions with the property that their aggregate present values, consistently computed using an after-tax discount rate, just equals the future outlay, should produce what I have called a "cash-equivalent" outcome. Compare Bailey (1974). Devotees of this approach include, in particular, Fiekowsky (1984) and Cunningham (1985), as well as current I.R.C. §§ 468 and 468A.

61 A useful summary can be found in Cunningham (1985), who, however, disagrees with them all.


63 Section 461(h) largely dictates "cash." Sections 468 and 468A, although appearing to allow some form of advance accrual, actually operate to allow deductions whose aggregate present values, computed using an after-tax discount rate, are equal to the present value of a deduction for the expenses when they are actually paid. See note 1.
But Samuelson's theorem speaks not merely to the problem of asset "depreciation" conventionally understood, but more generally to the deductibility of expenses and, in principle at least, to the equally fundamental matter of "capitalization". In what follows, then, I introduce into Samuelson's original formulation a modification that takes explicit account, not merely of the "depreciation" of an initial investment, but also of subsequently incurred expenses. This treatment has the virtue of clarifying (at least in principle) the confusing distinction between the treatment of "ordinary and necessary" future expenses and future "capital" costs. It also helps to identify more precisely the methodological divide between proponents of accrual and those (presumably including Congress) who claim that cash equivalent accounting is theoretically correct. In particular, it suggests that there will always be tax invariant solutions to the problem of future costs; that they all can be accounted for within the framework of Samuelson's solution to the problem of "economic" depreciation; but that they all have the property of apparently requiring a zero rate of tax.

2. An Example

Before proceeding further, it is worth turning to a simple illustration, used by Sunley (1984), and elaborated on in detail by Fieckowsky (1984) and Cunningham (1985), in exploring the taxation of a "negative salvage value" asset, that requires an extraordinary outlay at the close of the asset's service life. Consider first, however, the asset in the absence of the final outlay, assuming that it has five years of constant productivity, generating annual net income of $1,000, in an environment with a 10 percent pre-tax discount rate and a 30 percent tax rate. If acquired for $3,791, the asset's internal rate of return equals the pre-tax discount rate (and its net present value is zero).

Standard "ideal" treatment of that asset, if purchased for $3,791, is illustrated in Table III-1 (contained in the Appendix). Column (3) of Table III-1 contains the present value of each $1,000 receipt, discounted to Time 0 (the beginning of Period 1) at 10 percent, while Column (4) computes the sum of the present values of all remaining payments, first at Time 0, and then as of the end of each of Periods 1-4. The standard
way of determining "economic" depreciation of this asset is to compute the successive changes in the values reported in Column (4), which I report in Column (5).\textsuperscript{64}

The alternative (and equivalent) procedure implied by expression (4) is to subtract from each periodic payment in Column (2) the product of (a) the 10 percent internal rate of return to the asset and (b) the asset's remaining value as of the close of the preceding period, as reported in the preceding line of Column (4). The reader can easily verify that this produces the same schedule of depreciation as the "standard" method.\textsuperscript{65} In Columns (6)-(8) I compute taxable income, tax, and after-tax cash flow from the asset, and discount the latter to Time 0 at an after-tax discount rate of 7 percent, verifying that the asset value is unaffected by the tax. The important thing to note, however, as suggested by (5) and as the reader can again easily verify, is that the amount ultimately taxed to the holder is just the product of (a) the 10 percent internal rate of return and (b) the asset's remaining value as of the close of the preceding period.

Tables III-2-A and B, which add to Table III-1 a $1,401 outlay to be made in Period 6, replicate the actual example in Sunley (1984). (Table III-2 is also in the Appendix.) I have, however, explicitly decomposed the cash flows associated with the asset into revenues and costs. Thus, except for the addition of the $1,401 outlay in Period 6 in Column (1), Columns (1)-(4) of Table III-2-A are identical to Table III-1, while Column (5) computes the value of the final cost, discounted to the end of Periods 0-6. The net value of the asset at each point in time is the difference between Columns (4) and (5), reported in Column (6). Economic depreciation now consists of the successive changes in the entries in Column (6), and is reported in Column (7).

The actual computations of taxable income, tax, and after-tax cash flows are reported in Table III-2-B, which replicates Sunley's observation that economic depreciation again produces tax-invariant valuation. But there are some features to Tables III-2-A

\textsuperscript{64} E.g., Chirelstein (1994) § 6.08(d); Sunley (1984).

\textsuperscript{65} For example, in Period 1, depreciation would be $1,900 - (0.1 \times $3,790.79) = $620.92, as actually reported in Column (5) for Period 1.
and B that merit additional consideration. Note, first, that the introduction of the final cost has reduced the present value of the combined stream of revenues and expenses -- and, presumably, the price an investor would be willing to pay -- from $3,791 to $3,000. Nevertheless, the depreciation properly allowable has increased. This is perhaps the most confusing, potentially misleading, and operationally perplexing aspect of the treatment of future costs. A system of depreciation geared to historical acquisition cost will have an unavoidably hard time dealing with ("properly" depreciable) future costs.

Sunley himself explains the depreciation schedule with the observation that the asset's value declines from its $3,000 acquisition to -$1,400, producing $4,400 in allowable depreciation. As with Table III-1, however, the alternative account suggested by expression (4) is that the depreciation in each period consists of the $1,000 receipt, reduced by the product of the internal rate of return and the asset's value at the beginning of that period. In Table III-2-A, however, the net value of the asset in each period (reported in Column (6)) is, because of the final cost, less than the corresponding value reported in Table III-1. So the allowable depreciation increases. In fact, the depreciation schedule in Table III-2-A is simply the sum of (1) the depreciation schedule from Table III-1, and (2) the periodic change in the present value of the final cost itself (reported in Column (5)). Furthermore, as in Table III-1, the taxable income reported in Table III-2-B can be derived directly as the product of the asset values reported in Table III-2-A, Column (6) for the preceding period and the 10 percent internal rate of return, as implied by Expression (5).

---

66 It is, of course, unlikely that a simple reproducible asset in a market economy would sell for both $3,000 and $3,791, depending on whether it was acquired for a project with or without the future cost. An alternative way of looking at this example would be that, without the final cost, an investor requiring a pre-tax rate of return of 10 percent would pay $3,000 for the asset even if it produced annual revenue of only $791; whereas, with the final cost, she would pay $3,000 only if the project produced annual revenue of $1,000.

67 For example, in Period 1 the present value of the liability grows from $790.83 to $869.91, or by $79.08, which, when added to the $620.92 of Period 1 depreciation reported in Table 1, produces total depreciation of $300.
Tables III-1 and 2 suggest that Samuelson's basic insight applies both to outlays made initially to acquire an asset and to those made at the close of its useful life. In either event, deductions for economic depreciation appear to produce invariant asset valuation (although we have not formally established that this will always be so). That evidently has not, however, served to quiet the debate. For one thing, Samuelson's article is expressly about "depreciation." Adherents of cash-equivalent accounting expressly dispute its validity in the case of outlays other than conventional "investments" in durable goods, made in advance of production. The argument goes essentially like this: if allowing a current deduction for a durable investment is equivalent to exempting it from tax, allowing a deduction in advance of an outlay to be made in the future must be better. Hence accrual of future costs cannot be consistent with an income tax.

Furthermore, even observers like Sunley, who subscribe to accrual, leave almost entirely unexplained (and, indeed, unexplored) the means by which to distinguish between (in Sunley's words) "an investment expenditure that must be incurred in the future if income is going to be earned currently," for which depreciation is appropriate to begin with, and an "ordinary and necessary" expenditure to operate an asset, for which some different treatment presumably would be prescribed. In point of fact, the entire distinction between future "current" and future "capital" outlays, here, as elsewhere, has introduced substantial confusion into the debate. Direct marginal costs of current production are perhaps readily to be distinguished from an investment in the machine with which current output is produced. But the issue is rarely that simple.

3. Costs, Explicitly Considered, and a Natural Capitalization Rule

In light of these considerations, it seems natural to ask what happens in Samuelson's original formulation when costs are explicitly introduced. To that end, define

\[ N(t) = R(t) - C(t), \]

\[ \text{See footnote 1.} \]

113
where \( R(t) \geq 0 \) and \( C(t) \geq 0 \) are gross revenues and costs, respectively, both expressed as continuous functions of time, so that (1) becomes as

\[
(1a) \quad \dot{V}(t) = e^{yt} \int_{s=t}^{T} \left( R(s) - C(s) \right) e^{-ys} ds,
\]

the time derivative of which is now

\[
(2a) \quad \dot{V}'(t) = y \dot{V}(t) - R(t) + C(t).
\]

The general formulation of the cost function \( C(t) \) says nothing in particular about whether, in conventional terms, the costs are "current" or "capital." Some care must therefore be taken in deciding how to rewrite (3). To allow the solution to the problem to tell us just what is deductible, regardless of whether it is deductible as a "current" expense or as "depreciation," I define \( d(t, z) \) to be the total amount deductible from the gross revenue stream, and rewrite (3) as

\[
(3a) \quad \dot{V}(t, z) = e^{y(1-z)t} \int_{s=t}^{T} \left( R(s) - C(s) - z[R(s) - d(s, z)] \right) e^{-y(1-z)s} ds.
\]

The derivative of (3a) with respect to time is

\[
(7) \quad \frac{\partial \dot{V}(t, z)}{\partial t} = y(1 - z) \dot{V}(t, z) - R(t) + C(t) + z R(t) - z d(t, z).
\]

As in Samuelson, if (3a) is to be independent of \( z \), that is, if we are to have

\[
\dot{V}(t, z) = \dot{V}(t, 0) = \dot{V}(t),
\]

then we can equate (2a) and (7) to find that

\[
d(t, z) = d(t)
\]

\[
= R(t) - y \dot{V}(t)
\]

\[
= - \dot{V}'(t) + C(t).
\]
The tax-invariant depreciation schedule in the reformulated problem appears to differ in an important respect from expression (4). Bearing in mind that, by definition, \( V(t) = \hat{V}(t) \), the last line of (4a) suggests that the revised schedule exceeds the original schedule by \( C(t) \), indicating that the costs at time \( t \), in addition to economic depreciation, are properly allowable as deductions. In fact, however, the appearance of gross revenues in (2D) limits the extent to which the costs may be deducted. To be completely explicit (if somewhat redundant) about (4a), the amount deductible at \( t \) will be:

\[
d(t) = \begin{cases} 
N + C - y\hat{V}, & R > C, \\
R - y\hat{V}, & 0 < R < C, \\
- y\hat{V}, & R = 0.
\end{cases}
\]

Thus, if revenues exceed costs, the costs are allowable as deductions. If, however, revenues are less than costs, the amount allowable is limited to revenues (reduced by the instantaneous accrual of yield). In short, costs in excess of current gross revenues are not deductible. They are, instead, to be accounted for through depreciation.

At first this may seem strange. But the intuition again is relatively simple. In Samuelson's formulation, depreciation is a function of the change over time in the value of the net revenue stream \( N(t) \). To the extent that current expenses do not exceed current gross revenues, they can be applied to reduce gross revenue and subsumed into the depreciation schedule itself. To the extent, however, that current costs exceed current revenue, they will -- irrespective of whether they are conventionally capital or current -- influence the change in the asset's value over a longer horizon than the current period, and therefore should properly be accounted for through "depreciation".

In a sense, this is a kind of "natural" capitalization rule. It offers some insight into otherwise puzzling facts. Recall that the intuition underlying Samuelson's depreciation schedule is that a receipt, formally includible in gross income, is exactly offset by a decline in the remaining net income from the asset. Exactly the same is true of current outlays. The outlay itself, although reducing the holder's net worth, is exactly
offset by a reduction in the holder’s remaining payment obligation, so that, once again, the net consequence is a wash.\textsuperscript{69}

The insight that conventionally "current" expenditures have no impact on an investor’s net worth has exactly the same flavor as the insight underlying the capitalization requirement itself: the exchange of, \textit{e.g.}, cash, for a durable income-producing asset, does not affect the purchaser’s net worth, and it is for that very reason that the outlay may not be currently deducted. The similarity between the underlying insights suggests, in turn, that despite our instinctive beliefs that certain kinds of outlays are inherently "current" while others are naturally "capital," the analysis exemplified by Sunley (1984) is properly to be applied to any current period expenditure in excess of current revenue, in \textit{any} "current" period, regardless of whether the expenditure conventionally seems to be "ordinary and necessary" or not. It thus has the virtue (at least in theory) of providing a guide to determining when an income-producing expenditure is properly to be accounted for through depreciation, regardless of whether it occurs near the beginning, in the middle or at the end of an asset’s productive life.

In fact, this generalized "depreciation" schedule can easily accommodate the initial (capitalized) investment in the asset itself, which typically is treated instead as a separate event, distinct from the matter of depreciation. To see this, assume that (1) $R(0) = 0$, (2) $C(t) = 0$ and $R(t) = N(t)$ for $0 < t \leq T$, and (3) set $C_0 = -\dot{V}(0)$ (determined without regard for $C_0$). "Cost" now consists solely of a Time 0 outlay equal to the present value of the entire net revenue stream, so that overall this is a zero-net present value investment. Now, in accordance with (4a), the amount deductible in connection with the initial outlay is

$$-y\dot{V}(0) = 0,$$

\textsuperscript{69} It should be kept in mind at this juncture that, by hypothesis, both $R(t)$ and $C(t)$ are assumed to be known in advance. As discussed in the next Section, somewhat different considerations arise with respect to unanticipated costs.
since $R(0) = \dot{V}(0) = 0$. Thus, the natural capitalization rule replicates conventional capitalization of an investment made in advance.

In practice, however, the implications of this analysis will not be quite as radical as at first they might appear. Both the Samuelson paper and the modest elaboration offered here assume that revenues and expenses are known with certainty in advance. Outlays that are directly related to current production (and conventionally regarded as deductible) in excess of current revenues will typically be contemplated in advance only by a producer who plans on pricing below marginal cost, or, given the absence of uncertainty, otherwise engaging in suboptimal conduct like unnecessarily stockpiling raw materials in advance. With no uncertainty, it is therefore reasonable to expect that the relationship between revenues and costs that determines whether, in the analysis above, an outlay is to be "depreciated," will conform to conventional understandings of whether it is "capital" or "current." If, however, some originally unanticipated outlay becomes necessary, as when a producer discovers only ex post that its activities have been causing environmental degradation, modifications to the basic analysis may be required. In the balance of the paper, then, I go on to illustrate the application of the analysis to anticipated costs, after which I turn to those not anticipated in advance.

4. Future Environmental Costs

Advocates of accrual accounting for future costs have frequently appealed to tax-rate invariance as the standard by which different methodologies are to be judged. As a casual matter, it is hard to imagine that the Samuelson property could be judged desirable with respect to investments in conventionally capitalized depreciable assets but not with respect to future costs. One virtue of the analysis above is that, as a formal matter, it justifies those appeals.

As an illustration, we can justify formally the treatment of the example in Tables III-2-A and B. To do so, we begin by assuming first that the asset can be described by (1), with pre-tax value $V(t)$, and then add the following assumptions:
\[ R(t) = \begin{cases} N(t), & 0 \leq t < T, \\ R(t) \geq 0, & t = T, \end{cases} \]

\[ C(t) = \begin{cases} 0, & 0 \leq t < T, \\ C_T > R(T) \geq 0, & t = T. \end{cases} \]

Thus, gross revenues are simply equal to net revenues, except for a single expenditure (for remediation) made at time \( T \). \( C_T > R(T) \geq 0 \) implies that the (pre-tax) value of the asset, taking account of the remediation expense, is

\[ (1b) \quad \tilde{V}(t) = e^{yt} \left\{ \int_{s-t}^{T} N(s) e^{-ys} ds - e^{-yT} C_T \right\} < V(t). \]

The tax-invariant depreciation schedule for (1b), the negative of its time derivative, or

\[ (4b) \quad N(t) - y\tilde{V}(t), \]

has the same form as Samuelson’s depreciation schedule (4). But \( \tilde{V}(t) < V(t) \) implies that the amount allowable as depreciation at time \( t \) exceeds the depreciation that would have been allowable in the absence of the final payment \( C_T \) by

\[ ye^{y(t-T)} C_T, \]

the instantaneous increase at time \( t \) in the present value of \( C_T. \)

Each conclusion conforms to what we observed in Table III-2. The pre-tax present value of the asset declined; total depreciation increased; and the amount by which

\[ ^{70} \text{Aggregate additional depreciation just equals the change in the present value of } C_T \text{ between } 0 \text{ and } T. \text{ It can be obtained by integrating the expression in the text from } 0 \text{ to } T, \text{ and is given by} \]

\[ (1 - e^{-yT}) C_T. \]
depreciation in each period increased equalled the change in the present value of the final, $1,401 payment ($C_T$).

5. Economic Depreciation and Cash-Equivalent Accounting

A second virtue of the analysis is that it can be used to pinpoint more precisely the difference between economic depreciation and "cash equivalent" accounting for future costs. For, despite all the legislation enacted in 1984, there remain serious differences of opinion about what is theoretically correct. We can, however, use the analysis above to shed light on just why it is that economic depreciation, elsewhere so widely regarded as theoretically appropriate, has been so vehemently resisted in the case of future costs. In fact, Samuelson's analysis can in principle be applied to future costs. But the tax-invariant depreciation schedule turns out to require (at least apparently) a zero rate of tax, a conclusion that is hard to reconcile with an income tax.

To pursue this, I assume that the investor in the asset described by (1b) operates in an economic environment such that she can increase her prices so as to raise additional revenues having an aggregate pre-tax present value equal to the required final payment $C_T$. Specifically, I assume that the additional revenues are described by some continuous function $\tilde{f}(t) \geq 0$, defined on $0 \leq t \leq T$, satisfying

\begin{equation}
\int_{s=0}^{T} \tilde{f}(s) e^{-\gamma s} ds = e^{-\gamma T} C_T.
\end{equation}

With hesitation, I suggest thinking of $\tilde{f}(t)$ as generating a "fund" dedicated to satisfying $C_T$.\footnote{Alternatively, one could imagine the investor as dedicating a portion of the original revenue stream $R(t)$ to satisfying that liability, but the result would be equivalent and the effects somewhat more difficult to disentangle.} As so considered, it is clear that the value of the original asset, freed of the burden of satisfying the final payment, would once again be given by $V(t)$ (rather than $\tilde{V}(t) < V(t)$), and depreciation of the asset itself could, once again, be geared to its historical cost.
The important question is how to account for "the fund." Advocates of cash-equivalent accounting, such as Fieksowsky (1984), would claim that I have already misspecified the problem. They assert that the additional revenues collected to meet the final payment should be defined by the assumption that, when reduced to present value using an after-tax discount rate, they equal the present value of \( C_T \). Formally, the required revenues should be given by some other function \( g(t) \) satisfying

\[
(8a) \quad \int_{s=0}^{T} g(s) e^{-\gamma (1-z)s} \, ds = C_T e^{-\gamma (1-z)T}.
\]

With those assumptions, and the additional assumption that the final payment itself would be deductible when made, it follows automatically that consistency is achieved only if (1) the revenues "contributed" to "the fund" are deductible when contributed, but (2) the "fund" itself is taxed. That is because, by definition

\[
e^{\gamma (1-z)T} \int_{s=0}^{T} (1-z) g(s) e^{-\gamma (1-z)s} \, ds = e^{\gamma (1-z)T} (1-z) C_T e^{-\gamma (1-z)T}
= (1-z) C_T.
\]

That treatment, however, sacrifices the Samuelson property. Specifically, the additional revenue required to fund the final liability will increase with a producer's marginal rate of tax, as can easily be seen by differentiating either side of (8a) with respect to \( z \). So if, for example, the ability to raise prices to provide for the final liability were to be constrained by competition, those taxed at lower marginal rates would be better situated to be able to do so.

The revenues raised to provide for the final payment \( C_T \) can, alternatively, be taxed so as to make them tax-invariant. To derive the solution, let the present value at time \( t \) of the remaining stream of receipts, net of the present value of the final payment \( C_T \), be given by
\begin{equation}
F(t) = e^{yt} \left\{ \int_{s=0}^{T} f(s) e^{-ys} \, ds - e^{-yT} C_T \right\}.
\end{equation}

Note, first, that by reason of (8), \( F(0) = 0 \), and, second, that, as time passes, the value of the remaining stream of receipts declines, while the present value of the final payment grows, so that \( F(t) < 0 \) for \( t > 0 \). Finally, since (9) has the same form as (1b), we can expect to find a stream of deductions that will render it tax-invariant.

But we can be more specific. Using the properties of the integral and (8), (9) can be rewritten as

\[
e^{yt} \left\{ \int_{s=0}^{T} f(s) e^{-ys} \, ds - \int_{s=0}^{t} f(s) e^{-ys} \, ds \right\} - e^{-yT} C_T
\]

\begin{equation}
(9a)
= -e^{yt} \int_{s=0}^{t} f(s) e^{-ys} \, ds,
\end{equation}

What (9a) says is that the amounts already received \( f(t) \) from time 0 to any time \( t \), together with accumulated interest, will be just sufficient to offset the amount by which the present value of the final liability exceeds the present value of the revenues yet to be received. In the abstract, this really is a representation of "a fund," held to satisfy the final payment \( C_T \). It then follows from Samuelson's original derivation that the amount deductible from the revenues \( f(t) \) to insure invariant valuation of the fund is given by

\begin{equation}
(10)
-F'(t) = f(t) - yF(t) > f(t) > 0
\end{equation}

since \( F(t) < 0 \).

Expression (10) says simply that invariant valuation of a fund, held to satisfy a future liability, can be achieved in the presence of an income tax, but if (and only if) the investor is permitted to deduct the sum of (1) the revenues received to fund the lia-
bility, plus (2) the interest accruing on the accumulated fund. Suppose, then, we were to imagine funds being "set aside" to satisfy a future cost, and suppose also that we were to assume that both contributions to the fund and earnings from investing those contributions were to be taxed, subject only to the allowance of a deduction sufficient to preserve tax-invariance. According to (10), the required deduction would equal the fund’s entire gross income, producing (at least apparently) a zero rate of tax. That is, a zero rate of tax, and only a zero rate of tax, can insure invariant valuation of a "fund" held to satisfy a future cost.

This finding serves to explain why those who have searched for "neutral" means by which to tax future costs have so frequently produced solutions that seem to entail a zero rate of tax. Expression (10) suggests that this feature will be exhibited by any tax-rate invariant solution to the problem of future costs. It also helps to explain

\[ -yF(t) = ye^{y(t-T)}CT. \]

the accrual of yield at time \( t \) on the amount initially received.

This is the case, studied in detail by Aidinoff & Lopata (1980), Halperin (1986), and more recently by Halperin & Klein (1988), in which the single payment received to fund the future liability is analogized to a loan, repayment of which is evidenced by a pure discount bond. By appeal to that analogy, those observers have argued that (1) the "loan proceeds" should be excluded from gross income when received, (2) the "interest" on the loan should (as with an OID obligation subject to section 1272(a)) be deductible as it accrues, and (3) "repayment" of the "loan" should not be allowable as a deduction.

I invoke this analogy with even greater reservation than I invoked the analogy to a "fund." It leads to the same conclusions as Samuelson’s analysis. But the latter is derived from the sole objective of achieving invariant taxation of the investor. In contrast, the loan analogy,
why this problem has so sharply divided those who have studied this problem, many of them committed to an income tax. To those who advocate cash-equivalent accounting, it simply must be that earnings on assets held to satisfy a future cost are taxed, and that revenues must be "set aside" to satisfy the liability on that assumption. But it generally is not the case that producers taxed at different rates will be free to charge different prices for their products. And for a given amount of revenue, cash-equivalent solutions generally will not be tax-invariant. For advocates of economic depreciation (or "neutrality," by any name), on the other hand, tax-invariant solutions are available, but they seem to entail taxation of producers that is not easily reconciled with an "income" tax. Expression (10) suggests that this will always be the case. Thus, the analysis developed here more clearly outlines the implications of the two basic alternatives identified at the start.

D. Haig-Simons Taxation and Financial Distress

Section C suggests that, if the right answer is relevant to congressional outcomes, Congress did not very well understand the problem of future costs. Where debt is concerned it has done little better, and very possibly worse. In a few important respects the taxation of debt has improved in the past dozen years. In most, however, it has gotten significantly worse.

Take the simple but significant case of a pure discount bond. Congress did not until 1982 enact legislation that taxes such instruments in accordance with the exponential pattern in which interest accrues. Even now, Congress and the Internal Revenue

while often quite informative, can, I think, also be misleading. In particular, it leads to the search for someone else -- specifically for a "lender" -- to whom the "interest" on the "loan," not taxed to the investor, properly "should be" taxed. There is much to be said for such a search in the case of consensual financial transactions -- deferred compensation, in particular, and things like "Mooney bonds." It strikes me as quite possibly less compelling in the case of future costs associated with the production of goods. At this juncture it is not obvious to me in just which situations it would be right to insist that someone else be taxed on the income otherwise untaxed by reason of (10).

The strongest claim of this sort I have ever heard I attribute to Daniel I. Halperin, who once suggested that the right answer was not "completely irrelevant" to the outcome.

74
Service have not recognized in full generality that all debt instruments -- including those that provide for periodic payments (of "interest") -- in theory are characterized by exponential growth. So they have failed to grasp that the implication of the invariance theorem is that accrued net changes in value -- not simply the payment of interest -- properly should govern the timing of the taxation of additions to the wealth of holders of debt. Equally to the point, it appears to have been lost on those responsible for formulating income tax policy that, from the standpoint of the issuer of debt, interest should be deductible as it accrues, even if the issuer should happen in general to be on the "cash receipts and disbursements" method (rather than the "accrual" method) of tax accounting.  

1. The Treatment of Interest in Arrears and Debt in Default

These failures have important implications for the structure of the tax rules that govern the treatment of financial distress. Two fundamental questions are addressed by those rules. The first is simply the deductibility of interest in default. In general, an accrual method debtor may continue to deduct the unpaid interest whereas a cash method debtor may not. The propriety of this distinction does not appear to be a matter of real dispute. But it has generated litigation as the result of efforts by cash method debtors to circumvent the impact of the limitation. In what seems like a triumph of formalism, the rule has evidently emerged that, if a cash method debtor pays interest on an existing loan by taking out a new loan from a different lender, the interest that is paid is deductible; whereas, if the second loan is taken out from the primary lender, or if the primary

---

75 Under I.R.C. §§ 446 and 461, a "cash method" taxpayer is allowed to deduct items (including interest) only when they have been actually "paid." Where debt is concerned the strictures of the "cash method" are overridden only for instruments issued with "original issue discount," for which (discrete) exponential accrual is now explicitly prescribed by I.R.C. § 1272(a). There are, of course, a number of instances in which interest is not deductible at all, as for example is the case under I.R.C. § 163(h) with respect to so-called "personal interest," but I confine my attention here to interest that is deductible at some time.
loan simply goes into arrears (even if the arrearages bear interest themselves), a deduction isn't allowed.\textsuperscript{76}

The second tax question implicated by the onset of financial distress, generally regarded both as more important and more controversial, is the treatment of a loan that irremediably goes into default. The resulting "forgiveness" or "cancellation of indebtedness," usually referred to as "COD," is includible in the debtor's gross income at that time. The premise is that, since the original loan proceeds were not includible in income when they were received -- they were regarded as in some sense "offset" by the resulting obligation to repay, and as therefore not altering the debtor's net wealth -- the debtor should be taxed when the "loan" turns out ex post not "really" to have been a loan.

Existing analysis suggests, however, that the COD rule is deficient. Since COD is invariably taken into income later than when the loan was first taken out, the present value of the resulting taxes will be less than if the repayment obligation had been disregarded ex ante and the proceeds been taxed when initially received. In particular, it has been suggested that the COD rule restores the debtor ex post to the same ex ante financial position only when the loan contract calls for periodic payments of interest, and, even then, only if those payment obligations are currently serviced up to the point of default. In effect the claim is that, if interest accrues as a deduction but isn't actually paid (as is the case with a pure discount bond, and may be the case with an accrual method debtor), or if interest isn't paid because it goes into default (whether or not it continues to accrue as a deduction for tax purposes), the COD rule will under-tax debtors ex post. Coven (1986, 1991).

In the balance of this Section I establish two propositions that, in light of the actual state of the law, turn out to be somewhat ironic. I show first that the current (and generally undisputed) treatment of interest in default by a cash method taxpayer is theoretically wrong. By simply restating Samuelson's invariance result in terms suitable to debt I note first that interest should be deductible as it accrues, regardless of when it is

\textsuperscript{76} E.g., Battlestein v. Comm'r, 631 F.2d 1182 (1981).
actually paid; and, by a simple application of that result to interest that is in "arrears" (in a sense to be precisely defined, one that is consistent with the treatment typical of actual arrearages), I show also that, by allowing interest in arrears to be deducted, even by a cash-method debtor, Samuelson invariance is preserved. That is Part B. In Part C I then show that, conditional on having a system of Samuelson-invariant deductibility of interest (even by cash method debtors), the existing COD rule is -- in contrast to prevailing belief -- theoretically correct. By specializing an earlier result, also due to Samuelson (1937), I show that, under a system in which interest is always deductible as it accrues, COD will always ex post restore a defaulting debtor to the same position -- adjusted for the intervening effects of having made payments (if any) on the loan and having deducted the interest -- as he would have been in had the loan proceeds been taxed at the time of receipt.

In sum, what I show is that the COD rule does not now work properly, but not through any fault of its own. Were we to take the simple step of prescribing accrual as the touchstone to the timing of all interest deductions (as theoretically justified in Section B), we would simultaneously achieve Haig-Simons consistent taxation of COD, thereby getting two fixes for the price of just one.

2. Interest in Default and Accrual Invariance

A debt instrument with maturity T that pays principal and interest of \( p(t) \) for \( t \in [0, T) \), and (possibly) provides for repayment at maturity of any then unrepaid principal \( p_T \), can be represented by:

\[
V(0) = \int_{s=0}^{T} p(s) e^{-\gamma s} ds + e^{-\gamma T} p_T.
\]

(11)

(where for convenience I assume \( p(t) \geq 0 \)). Setting \( p(t) = 0 \) for \( t < T \) is a pure discount bond; \( p(t) = yp_T \) is a simple coupon bond with no amortization of principal; \( p(t) = k \) and \( p_T = 0 \) is a level-payment fully self-amortizing loan (such as a conventional home mortgage) for \( T < \infty \), and a consol for \( T = \infty \). Given \( V(0) \) and \( p(s) \), the yield-to-maturity of a debt instrument solves (11).
At any \( t \in (0, T) \), the value of the remaining obligation of the issuer of (11) has the same form as (1), or

\[
D(t, 0) = - V(t, 0) = - \left\{ e^{\gamma t} \int_{s=t}^{T} p(s) e^{-\gamma s} \, ds + e^{\gamma(t-T)} p_T \right\},
\]

and its tax-invariant form (with the final payment suppressed for convenience) is

\[
D(t) = - V(t)
\]

\[
= - e^{\gamma(1-z) t} \int_{s=t}^{T} e^{-\gamma(1-z)s} \left\{ p(s) - z y V(s) \right\} \, ds.
\]

Here the only net effect on the debtor's wealth, properly deductible for tax purposes, is the accrual of interest on the value of the unpaid balance, given by \( y V(t) \). When the debtor is allowed the tax benefit of such deductions (\( z y V(t) \)) the value of the obligation is independent of \( z \). If not, the instrument will be (read \textit{has been}) vulnerable to manipulation by debtors and creditors taxed at different rates.\(^{77}\) By itself, then, Samuelson's theorem implies that, under an accretion-type tax, interest should be deducted (taxed) as it \textit{accrues}, irrespective of whether the debtor (creditor) is otherwise on the accrual method of accounting. That treatment is generally prescribed, however, \textit{only} for original issue discount debt.\(^{78}\)

I first show that, if the original terms of a debt obligation are defined by a stream of payments \( p(t) \), but if, contrary to the terms of the contract, the payments of interest should go into arrears, the characteristics of the instrument will be unaffected provided \textit{only} that the arrearages themselves are required to bear interest at the instrument's underlying yield-to-maturity. That is:

\(^{77}\) Until 1982, obligations originally issued at a discount were taxed by \textit{pro-rating} the aggregate discount over the life of the loan, producing interest accruals that were faster than exponential growth, leading to the manipulative issuance of discount debt by taxable debtors to tax-exempt creditors, as noted in Sims (1992).

\(^{78}\) Under I.R.C. §1286, a coupon bond that is disassembled into its constituent pieces ("stripped") is taxed under the original issue discount rules.
PROPOSITION I (INTEREST ACCRUAL INVARIANCE): If (1) payments on a debt instrument defined by (5a) go into arrears, and (2) all outstanding arrearages accrue interest at the instrument's yield-to-maturity (y) until they are repaid, tax-invariance will be preserved by permitting the debtor to deduct the interest that accrues on the unpaid balance of the instrument, including the accrued arrearages.

PROOF. We can define an \( \text{arrearage} \, p^a(t) \) on the domain of \( p(t) \) by requiring that it satisfy:

\[
p^a(t) = \begin{cases} 
    p(t), & t \in [t_1, t_2], \\
    \int_{s=t_1}^{t} [p(s) - p^a(s)] \, ds \geq 0, & t \in [t_1, t_2], \\
    \int_{s=t_1}^{t_2} [p(s) - p^a(s)] \, ds = 0.
\end{cases}
\]

Thus, \( p^a(t) \) is identical to \( p(t) \), except that on some interval \([t_1, t_2] \subset [0, T]\) it is at first less than \( p(t) \) (the arrearage) but that over the entire interval the arrearage is eventually repaid. Now define an \emph{arrearage function} \((a(t))\) on \([t_1, t_2]\) by:

\[
(12) \quad a(t) = p(t) - p^a(t).
\]

The relationships among \( p(t), p^a(t) \) and \( a(t) \) are exemplified in Figure 1.

Figure III-1
An Arrearage Function

[Diagram of arrearage function]
The arrearages to time \( t \in [t_1, t_2] \), together with accrued interest, can be written

\[
A(t) = \int_{s=t_1}^{t} a(s)e^{y(t-s)}ds = e^{yt} \int_{s=t_1}^{t} a(s)e^{-ys}ds.
\]

Now the value of the debtor's obligation at \( t \) will be given by \( D(t) \) minus the arrearages \( A(t) \), or (using 12, and with \( p_T \) again suppressed for convenience)

\[
- V^a(t) = - V(t) - A(t)
\]

\[
= - e^{yt} \left\{ \int_{s=t_1}^{T} p(s)e^{-ys}ds + \int_{s=t_1}^{t} a(s)e^{-ys}ds \right\}
\]

Thus, the value of the obligation remaining at time \( t \) is the value of the original obligation at time \( t_1 \), plus accrued interest, reduced by the payments actually made under the payment function in arrears (\( p^a(t) \)). By Samuelson's theorem, a taxable counterpart of (14) will be invariant to the tax rate only if

\[
d(z, t) = d(t) = - \dot{V}^a(t)
\]

\[
= \frac{d}{dt} \left[ - e^{yt} \left\{ \int_{s=t_1}^{T} p(s)e^{-ys}ds - \int_{s=t_1}^{t} p^a(s)e^{-ys}ds \right\} \right]
\]

\[
= - ye^{yt} \left\{ \int_{s=t_1}^{T} p(s)e^{-ys}ds - \int_{s=t_1}^{t} p^a(s)e^{-ys}ds \right\} + p^a(t)
\]

(15a)

Using (15a), the tax-invariant counterpart to (14) is now

\[
- e^{y(1-z)t} \left\{ \int_{s=t_1}^{T} p(s)e^{-y(1-z)s}ds \right. \\
\left. - \int_{s=t_1}^{t} e^{-y(1-z)s} \left[ p^a(s) - zy V^a(s) \right] ds \right\}.
\]

(14a)
By (14a) and (15), tax-invariance is preserved by allowing the debtor to deduct the interest that accrues on the sum of the unpaid balance on the original loan, plus the total arrearage (with interest) to time $t$. Since $t$ was selected arbitrarily from $[t_1, t_2]$, the conclusion follows.

3. Haig-Simons Neutrality of COD

I now confine my attention to instruments that are taxed in accordance with the conditions of Proposition I. That is, I consider debt instruments the interest on which is deductible for income tax purposes at it accrues, regardless of when the interest is actually paid and regardless of the obligor's "method of accounting" (cash or accrual) for tax purposes. In particular, I assume that any interest that is unpaid (whether in accordance with the terms of the contract or by reason of an arrearage) itself accrues (deductible) interest at the instrument's underlying yield-to-maturity. For such instruments I consider a tax rule under which, in the event of default, the entire unpaid balance -- principal plus accrued and deducted but unpaid interest -- is included in the debtor's gross income at that time. I show that the effect of this rule is to put the obligor in essentially the same position as if no repayment obligation had arisen when the proceeds of the loan were received (at time 0) and those proceeds had instead been included in gross income at that time. Formally,

**Proposition II (COD Equivalence).** The unrepaid balance at any time $t$ after issue of a debt instrument whose economic yield is deductible as accrued at rate $y$, together with the cost of all payments on the instrument, and the value of all tax savings from the accrual of yield, prior to time $t$, equals the value at time $t$ of the instrument's issue price.

**Proof.** The proof is a simple specialization of an earlier result by Samuelson (1937) to a taxable environment. For a debt instrument described by (5a) we can write its value at time $0$ as:

$$-V(0, z) = -V(0)$$

$$= - \int_{s=0}^{T} e^{-y(T-t)s} \{ p(s) - z \ y \ V(s) \} \ ds$$

130
\[- \left\{ \int_{s=0}^{t} e^{-y(1-z)s} \{ p(s) - zyV(s) \} \, ds + \int_{s=t}^{T} e^{-y(1-z)s} \{ p(s) - zyV(s) \} \, ds \right\}.

Multiplying through by \( e^{y(1-z)t} \) and rearranging yields:

\[- V(0) e^{y(1-z)t} + \int_{s=0}^{t} e^{y(1-z)(t-s)} \{ p(s) - zyV(s) \} \, ds = - e^{y(1-z)t} \int_{s=t}^{T} e^{-y(1-z)s} \{ p(s) - zyV(s) \} \, ds = - V(t, z) = - V(t) \]  

(16)

(where the last step again follows from the invariance theorem).

Note first that since \( V(t) \) is independent of \( z \), so is the left-hand side of (16). The first term on the left side of (16) is simply the original issue price together with interest continuously compounded at the after-tax yield-to-maturity \( y(I - z) \). The integral is the aggregate of all payments (with continuously compounded interest) made up to time \( t \), reduced by the tax savings (also with interest) from having deducted \( yV(t) \). What (10) therefore says is that, at any time after issue, the value of the initial payment obligation, reduced by the value of the payments already made and increased by the value of the taxes already saved, exactly equals the then outstanding balance on the loan. This result holds, moreover, independently of \( z \).

**Corollary.** Taxing the issuer of a debt instrument at rate \( z \) on the unpaid balance at time \( t \) is equivalent to taxing the issuer on the sum of the issue price plus the present value of the tax savings from deducting the yield on the instrument as it accrued, minus the present value of the payments made under the contract prior to \( t \).

**Proof.** This follows immediately from Proposition II.
The intuition behind this result is straightforward. A debtor in fact is *not* in the same position as someone who received proceeds without any ex ante obligation to repay. Under the debt contract before the time of default, he will be worse off by having made payments on the contract, and better off by the tax savings from having accrued the interest as deductions, than someone who received unrestricted funds. The import of the corollary is that taxing the debtor at time $t$ on the unpaid balance on the debt leads to the inclusion in income of an amount *exactly equal* to the original issue price plus the net value of all costs and all tax savings attributable to the debt contract up to the point of default. Hence, as far as the original loan proceeds *themselves* are concerned, the debtor is no better off ex post. (The result can easily be specialized to the case of a pure discount bond, which produces only tax savings prior to default.)

4. Conclusion

Taken together, Propositions I and II establish that in principle the COD rule operates to achieve proper ex post Haig-Simons taxation as long as interest is deductible as it accrues, even if the debtor is on the cash method of accounting in general. (By a parity of reasoning, it says also that the income of a creditor whose debtor has defaulted will be properly measured in a Haig-Simons sense only if, in addition to being allowed a loss deduction at the time of default, the creditor has taken the "interest" on the loan into income as it accrues). They also establish that, at least as far as interest (including interest in default) is concerned, accrual is the only proper method of income tax accounting. With respect to financial distress itself, it is worth noting that, as a practical matter, much income from cancellation of debt is not actually taxed at the point of default, either because of explicit exceptions in the Internal Revenue Code, or because the debtor is already in a net loss position. and those losses may not be carried back indefinitely (and used immediately). Even so, many debtors successfully emerge from reorg-

---

79. As will typically be the case under I.R.C. § 166.

80. Under I.R.C. § 108, COD is technically excludible from the gross income of insolvent or bankrupt debtors, although the resulting exclusion must be applied to reduce other tax-
anization, and there is merit in having measured accurately the amount of COD that was included in gross income when financial distress occurred. More importantly, a great many defaults occur on a project basis, as for example with syndicated investments in real estate, with the resulting income being taxed currently to otherwise solvent debtors. In such instances achieving a proper measure of income on default will be of immediate practical as well as theoretical importance. Even more generally, it has been suggested (for example by Coven (1991)) that the shortcomings in the COD rule make it vital to know accurately in advance whether a transaction structured as loan is "really" a loan. But the two propositions established in this final Essay show that, in principle at least, we can always get matters right ex post.

attributes of the debtor, including their "net operating loss carryovers" and the basis of depreciable assets. In some measure these exclusions have roughly the same effect as including the COD in income, which by itself would reduce the amount of the debtor's "net operating loss" under I.R.C. § 172. Under the latter provision, an NOL may be carried forward nearly indefinitely, but may (with some exceptions) be carried back only to the three years preceding the year in which the loss arose.

As others have noted, moreover, there is little excuse for treating the cancellation of indebtedness from a project default as though it were part of the "gain" from a disposition of the project, possibly to be taxed as long-term capital gain, rather than as (ordinary) income from cancellation of the debt itself. See Barnett (1983).
<table>
<thead>
<tr>
<th>Period</th>
<th>Outlay</th>
<th>Revenue</th>
<th>PV</th>
<th>Σ PV at Period</th>
<th>Taxable Income</th>
<th>AT Cash Flow</th>
<th>PV (7%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3790.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1000.00</td>
<td>999.09</td>
<td>3169.87</td>
<td>620.92</td>
<td>379.08</td>
<td>828.30</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1000.00</td>
<td>999.09</td>
<td>2668.65</td>
<td>620.92</td>
<td>379.08</td>
<td>828.30</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1000.00</td>
<td>999.09</td>
<td>2668.65</td>
<td>620.92</td>
<td>379.08</td>
<td>828.30</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>1000.00</td>
<td>999.09</td>
<td>2668.65</td>
<td>620.92</td>
<td>379.08</td>
<td>828.30</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>1000.00</td>
<td>999.09</td>
<td>2668.65</td>
<td>620.92</td>
<td>379.08</td>
<td>828.30</td>
</tr>
</tbody>
</table>

**Table III-1**  

Economic Depreciation  

<table>
<thead>
<tr>
<th>Column (2)</th>
<th>Column (3)</th>
<th>Column (4)</th>
<th>Column (5)</th>
<th>Column (6)</th>
<th>Column (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
</tr>
</tbody>
</table>

The payment in each period in column (2), discounted for n periods at 10 percent.

The present values of all remaining payments as of the end of each period n = 0, 1, 2, 3, 4, consisting of the sum of the first 5 - n entries in Column (3).  

The change from period n - 1 to period n in the present value of all remaining payments, as given in Column (4).

Column (2) minus Column (5).

Column (6) times .30.

Column (2) minus Column (7).

The amount in each period n in column (8), discounted for n periods at 7 percent.
Table III-2-A
Computation of Economic Depreciation
With Final Cost

<table>
<thead>
<tr>
<th>Period</th>
<th>(1) Expense [Initial Outlay]</th>
<th>(2) Revenue</th>
<th>(3)(^{90}) PV (Revenue) @10%</th>
<th>(4)(^{90}) Σ PV (Revenue) at Period (n) @10%</th>
<th>(5)(^{91}) PV (Cost) at Period (n) @10%</th>
<th>(6)(^{92}) Net Value at Period (n)</th>
<th>(7)(^{93}) Depreciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>[3000.00]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2999.96</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>1000.00</td>
<td>909.09</td>
<td>3169.87</td>
<td>869.91</td>
<td>2299.95</td>
<td>700.00</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>1000.00</td>
<td>826.45</td>
<td>2486.85</td>
<td>956.90</td>
<td>1529.95</td>
<td>770.00</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>1000.00</td>
<td>751.31</td>
<td>1735.54</td>
<td>1052.59</td>
<td>682.95</td>
<td>847.00</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>1000.00</td>
<td>683.01</td>
<td>908.09</td>
<td>1157.85</td>
<td>-248.76</td>
<td>931.71</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>1000.00</td>
<td>620.92</td>
<td>0.00</td>
<td>1273.64</td>
<td>-1273.64</td>
<td>1024.88</td>
</tr>
<tr>
<td>6</td>
<td>1401.00</td>
<td>-</td>
<td></td>
<td>1401.00</td>
<td>-1401.00</td>
<td>127.36</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4400.96</td>
</tr>
</tbody>
</table>

\(^{90}\) The payment in each period \(n\) in column (2), discounted for \(n\) periods at 10 percent.

\(^{91}\) The present values of all remaining payments as of the end of each period \(n = 0, 1, 2, 3, 4\), consisting of the sum of the first \(5 - n\) entries in Column (3).

\(^{92}\) The present value of the outlay in Period 6 in Column (1), discounted at 10 percent to the end of period \(n = 0, \ldots, 6\).

\(^{93}\) Column (4) - Column (5).
Table III-2-B
Computation of Taxable Income
With Final Cost
Using Economic Depreciation

<table>
<thead>
<tr>
<th></th>
<th>(1) Revenue</th>
<th>(2) Expense</th>
<th>(3) Depreciation</th>
<th>(4)*6</th>
<th>(5)*5</th>
<th>(6)*6</th>
<th>(7)*7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1000.00</td>
<td>-</td>
<td>700.00</td>
<td>300.00</td>
<td>90.00</td>
<td>910.00</td>
<td>850.47</td>
</tr>
<tr>
<td>2</td>
<td>1000.00</td>
<td>-</td>
<td>770.00</td>
<td>230.00</td>
<td>69.00</td>
<td>931.00</td>
<td>813.17</td>
</tr>
<tr>
<td>3</td>
<td>1000.00</td>
<td>-</td>
<td>847.00</td>
<td>153.00</td>
<td>45.90</td>
<td>954.10</td>
<td>778.83</td>
</tr>
<tr>
<td>4</td>
<td>1000.00</td>
<td>-</td>
<td>931.71</td>
<td>68.29</td>
<td>20.49</td>
<td>979.51</td>
<td>747.26</td>
</tr>
<tr>
<td>5</td>
<td>1000.00</td>
<td>-</td>
<td>1024.88</td>
<td>-24.88</td>
<td>-7.46</td>
<td>1007.46</td>
<td>718.31</td>
</tr>
<tr>
<td>6</td>
<td>0.00</td>
<td>1401.00</td>
<td>127.36</td>
<td>-127.36</td>
<td>-38.21</td>
<td>-1362.79</td>
<td>-908.09</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2999.96</td>
</tr>
</tbody>
</table>

*6 Column (1) minus Column (3).
*5 Column (4) times .30.
*6 Column (1) minus Column (2) minus Column (5).
*7 The amount in each period n in column (6), discounted for n periods at 7 percent.
REFERENCES

Essays 1-2


Essay 3


