Essays in Political Economy

by

Koleman S. Strumpf

B.S., Stanford University (1990)
A.B., Stanford University (1990)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1995

Signature of Author ....
Department of Economics
24 May 1995

Certified by ....
Peter Diamond
Paul A. Samuelson Professor of Economics
Thesis Supervisor

Certified by ....
James Poterba
Professor of Economics
Thesis Supervisor

Accepted by ... .
Richard Eckaus
Chairman, Departmental Committee on Graduate Students

Essays in Political Economy

by

Koleman S. Strumpf

Submitted to the Department of Economics
on 24 May 1995, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Political economy is a useful lens for understanding the forces which shape government decisions. One strand, public choice, recognizes the crucial role of self-interested politicians and bureaucrats. The other main approach, social choice theory, considers how political institutions guide electoral outcomes. This thesis investigates applications of each of these pillars.

The first chapter illustrates the importance of government credibility in the choice of taxes. Local governments can tap a wide variety of tax instruments, and yet they utilize only a few. This work considers the pattern of wage tax take-up among Pennsylvania local governments. All of the hold-outs are adjacent to Philadelphia which is counter to the simplest political economy story; Philadelphia commuters are exempt from home liabilities, so a strict majority in the non-taxing suburbs could benefit from a levy. The solution to this puzzle is that voters do not grant new taxing powers when they doubt property taxes will be equally reduced. After correcting for such credibility concerns, exempt voters are positively associated with wage taxes; the original correlation is due to colinearity between government mistrust and the fraction exempt. This result is further support for the contention that a credibility gap catalyzes taxpayer revolts.

The second chapter shifts focus to social choice theory: how does the sequential nature of American presidential primaries influence the outcome? The prevailing wisdom is that early wins spill-over into later elections, a "system effect" since it involves voter rather than candidate behavior. However, in actual primaries momentum does not always lead to a bandwagon. Equally important, if participating in each election is costly, then a contender may quit the race if his future prospects are poor. This "strategic effect" boosts the victory probability of the candidate whose favorable elections come at the end of the cycle. The intuition is related to the value of commitment: a player with only advantageous elections left can credibly pledge to fight all remaining periods while his opponent cannot. A related explanation is the value of information: candidates only quit after large surprises, so upsets and

not expected victories drive behavior. Informational shocks are most helpful when long-shot elections are front-loaded. The strategic effect is one explanation for why there is so little conflict over the order of elections and why favorite-sons from early primary states rarely gain the nomination.

The final essay is a re-examination of the flypaper effect: local governments spend a larger fraction of lump-sum public income than an equivalent increase in private income. One shortcoming of the related literature is that it presumes all communities have an identical propensity to consume from an intergovernmental grant. This chapter is one attempt to allow for a heterogeneous response. The working conjecture is that government expenditure on administrative overhead is a gauge of voter control over fiscal decisions. High overhead means a low rate of return on tax dollars, and suggests that revenue-maximizing forces dominate the public budget debate. As such, the flypaper effect should be more marked in high overhead communities. The index is applied to the sample of Pennsylvania communities in the Philadelphia metropolitan area considered in the first essay. It is difficult to understand spending propensities out of a tax windfall without the help of the overhead index, supportive evidence for this approach. These results are an incremental step towards the broader project of using Census data on a national sample of state and local governments. The techniques of this paper will predict variations in response to fiscal decentralization, a major policy goal of the current Congress.

Thesis Supervisor: Peter Diamond
Title: Paul A. Samuelson Professor of Economics

Thesis Supervisor: James Poterba
Title: Professor of Economics

For my grandfather who taught me never to stop asking questions.

Acknowledgments

Writing a credits page is almost as difficult as completing the thesis itself. It is impossible to distill down to a few pages all of my interactions with students and teachers over the last five years. Just as daunting, this is the one section that is most likely to be read, so it is important not to leave out anyone. Nonetheless, I will try to briefly thank the several people who played an instrumental role in my life at MIT.

It was a privilege to have Peter Diamond and Jim Poterba as my thesis advisors. My intellectual development as an economist is in large part due to their close attention and helpful guidance. Their combination of theoretical and applied expertise served as a virtually endless font of research suggestions and insightful comments (and meant few of my assertions went unchallenged). They also taught me how to transform my rough thoughts into far crisper and developed forms. And their candid assessments of my progress served as a constant source of motivation. I hope to emulate their roles as mentors in future interactions with students.

I am also indebted to other faculty members who assisted me. At MIT, Professors Jon Gruber, Thomas Piketty, Lones Smith and Jim Snyder were always available to talk though the gritty econometric and modeling details. These sessions helped thresh out several of my results and limited the amount of research wheel-spinning. Discussions with Professors Steven Ansolabehere, Robert Inman and Richard Zeckhauser greatly improved my work.

Without the support and friendship of several of my classmates, I doubt I could have finished this thesis. Bob Majure taught me both about economic theory and the importance of maintaining my convictions. I worked with Aaron Yelowitz both in classes and on research. His advice on economist etiquette, while difficult to follow, was always on target. Also, Dave Frankel, Dan Kessler, Peter Klibanoff, David Laibson and Steve Levitt gave me several useful suggestions on my work. I hope I will be able to continue to interact with them as colleagues.

Several others made my time as a graduate student more enjoyable. Bruce Fischer and Don Yeung were my best friends away from the department, and they helped me keep perspective in good times and bad. I found from hanging out with Ludwig Chincarini and Paul Schulstad that acting irreverent makes life a lot more fun. My friend since our days at Radnor, Neil Cherry, helped remind me of the limits and advantages of being a graduate student. I would also like to thank Marcello Estevao, Guillermo Llorente-Alvarez, Alex Pfaff, Jack Porter, Peter Sorensen, Stacey Tevlin and Beth Anne Wilson for both being supportive and dealing with my often grouchy disposition. I value them all as good friends.

Fellowships from the National Science Foundation and the Harvard/MIT Research Training Group in Positive Political Economy supported my studies for four years while the George Shultz Fund generously provided for research travel. Portions of the data from chapters 1 and 3 were provided by the Pennsylvania Department of Community Affairs and the Delaware Valley Regional Planning Commission.

Finally, I would like to thank my family who always encouraged my intellectual curiosity and without whom I surely would not have reached this stage.

Contents

Chapter 1

Credibility and the Structure of Taxes: The Local Earned Income Tax in Pennsylvania

1.1 Introduction

Research on local fiscal behavior has focused almost exclusively on the level of expenditures (Fisher [21]). The increasing number of radical tax base shifts, such as the recent elimination of school property taxes in Michigan, suggest many interesting questions lie on the revenue side. For example, there is no strong foundation for the structure of taxes: while counties and municipalities have access to numerous instruments ranging from sales to property taxes, few are actually used. And despite such flexibility, over the past twenty years there have been an epidemic of taxpayer revolts placing caps on specific taxes or general revenues.

One largely ignored determinant of revenue policy is government credibility. When voters doubt politician promises, they may reject a change in tax policy which benefits a majority. To illustrate with a recent political example, consider the proposed shift from a flat to graduated income tax in Massachusetts (Questions 6 and 7 on the 1994 ballot). Several studies showed that 93.5% of the voters would see their tax bill fall or stay constant under the proposal, yet it was soundly defeated. The logic for the opposition is the "Grad-Tax Trap": the government will use the Questions' expanded taxing authority to raise taxes in the future, Figure 1. With high state taxes fueling government skepticism, the possibility of a future increase outweighed the guaranteed short-term windfall. This is precisely the mechanism suggested here.

This paper illustrates the importance of credibility in practice. Pennsylvania allows local governments to levy a one percent earned income tax (EIT), essentially a tax on wages. While there is near-universal adoption in the commonwealth, several hold-outs remain along the Philadelphia border. This is puzzling because commuters to Philadelphia are exempt from residential liabilities, so a wage tax would strictly benefit a majority in the near-urban communities. The explanation is that voters fear their government will renege on a promise of property tax relief. Indirect evidence links these doubts to high tax burdens or administrative overhead, and that these variables are the primary deterrent to wage levies on the Philadelphia periphery. In other words, non-credible politicians are denied new tax privileges.

The Pennsylvania EIT is a superior crucible for examining tax structure and taxpayer revolts on several counts. First, there is a panel data of 237 communities over 33 years rather than the usual one-time election in a single state (Courant-Gramlich-Rubinfeld [13], Ladd-Wilson [31], and Sears-Citrin [51]). This diminishes the potential for bias from government-specific fixed effects or year effects. Second, since there is generally a promise of revenue-neutrality, adoption need not require a convoluting change in services. The chief matter at stake is the distribution of taxes. Finally, since the wage tax is capped at only one percent, factor market distortions are much smaller than those from a significant statewide property tax reduction. This also means the tax decision is binary, not complicated by a choice of rates.

This paper complements the tax composition literature. Most closely related is Inman's [26] work on revenue choice in a panel of large cities. He finds that distributional (in addition to efficiency) concerns determine the reliance on property taxes or fees. Unlike the current study he does not explicitly model voter behavior or consider credibility issues. Most importantly he estimates separate regressions for each revenue instrument rather than a simultaneous system, so it is difficult to interpret the (reduced-form) parameters. Sjoquist [52] provides a similar analysis but only estimates a static regression of property tax share. A few papers have directly considered tax initiations at the local level. Blackley-DeBoer [5] and Mikesell [37] show tax exportability and pressure groups (respectively) determine the take-up of a local income tax in Indiana counties. Norstrand [38] finds the dependence on income tax among Danish municipalities is based on resident interest groups, tax exportability and the political affiliation of the city council. Unfortunately these studies are cross-sectional, leaving estimates open to fixed effects bias, and consider specifications which ignore crucial explanatory variables. More work has been conducted on tax composition using the state as the unit of analysis. Hettich-Winer [24] demonstrate that the desire to use the lowest political cost instruments explains states varying reliance on income taxes. Using the Tax Reform Act of 1986 as an exogenous shock, Metcalf [36] shows that the level of income (but not sales) tax is sensitive to price.

The outline of the rest of the paper is as follows. The next section details the Pennsylvania Earned Income Tax law with administrative details in Appendix 1.7. A simple model of interest groups and *tax situs* explains the common pattern of tax propagation such as the case of the Pittsburgh suburbs. This framework fails to explain the spatial pattern of hold-outs in the Philadelphia region suggesting a modified approach in which governments may be non-credible (the more complex analysis with strategic politicians is detailed in Appendix 1.8). Section three motivates the choice of credibility proxies- current tax load, recent changes in tax bills and the level of government overhead. High values of these variables increase politician turnover and shrink property tax relief following a new levy, strong support that the proxies shape voter views of government veracity. The model's empirical implications motivate the regression strategy of section 4. Description of the innovative data set follows with the algorithm used to approximate the commuting matrix relegated to Appendix 1.9. Using either cross-sections or a full dynamic specification, the results in Section 5 confirm that distrust of government is the chief barrier to implementing a wage tax. Including the proxies also reverses the puzzling negative effect of exempt groups on taxing propensity, further support for the important role of credibility. The final section summarizes, addresses a potential criticism and discusses future avenues for research.

1.2 The Pennsylvania Earned Income Tax

The local tax privilege in Pennsylvania was initially granted as an emergency measure in 1947. Fearing the right would be rescinded, many governments chose not to levy. But with the further codification in the 1965 Local Tax Enabling Act, municipalities rushed to implement an earned income tax; in 1992, the most recent year with available statistics, 95% of the commonwealth's 2,546 municipalities had the tax. Almost all hold-outs are concentrated in the Philadelphia standard metropolitan statistical area (SMSA).[1]

Why is the Philadelphia area different? Formal analysis requires a detailed understanding of the Earned Income Tax (EIT) laws. This section includes a primer on tax fundamentals: *tax situs*, special exemptions and procedures for implementation. Administrative details may be found in Appendix 1.7. Then we contrast the archetypal and Philadelphia SMSA take-up experience. A simple interest group model which explains the usual tax propagation is unable to explain the band of hold-outs in the Philadelphia suburbs. This motivates the development of a broader framework in which voters need not trust government promises.

[1]The other non-taxers are clustered in the northeast and northwest corners of the state. The former cannot collect revenues from the large number of their citizens who pay a wage tax in New York city while there is no obvious explanation for the behavior of the western communities.

Type	Number	EIT in 1992?	(1992 $\times 10^3$ \$) Mean Taxes	(1990 $\times 10^3$) Mean Population
City	2	2	8091.3	26.45
Borough	88	52	888.3	4.90
1st Class Township	28	11	5477.0	24.08
2nd Class Township	119	81	1314.1	8.27

Table 1.1: Government Types in the Philadelphia SMSA

1.2.1 Provisions and Implementation

It is essential to understand the basic features[2] of the tax: provisions of the law and rules for implementation. This provides the ground-work for isolating which factors govern tax-setting policy.

There are three key provisions to the Pennsylvania EIT. First, any local government except counties has the power to levy. There are three possible political forms:[3] *cities* are densely populated urban regions governed by a mayor and four member council; *boroughs* are similar to cities but have fewer people, a larger council and a less powerful mayor; *townships* have a 5 to 15 member board of commissioners (first-class) or three supervisors (second-class) to govern. Generally, politician powers, the size of government and the probability of having an income tax are greatest in cities and least in townships; statistics for the Philadelphia suburbs are in Table 1.1.[4] There are also independently governed *school districts* which contain one, or more generally, several communities within their boundaries; these are not included in the sample.[5]

Second, no individual must pay more than one percent of his earned income. This means those with only passive income such as the retired pay no tax. Also, since

[2]Surprisingly, the information in this section cannot be found in a single source and had to be gleaned from conversations with several township managers and tax collectors. A useful starting reference is a publication from the Pennsylvania Department of Community Affairs, [42].

[3][41] gives a detailed treatment of Pennsylvania local governments.

[4]Lower tax revenues in boroughs are due to smaller population size.

[5]In the 1972-1992 period, in only two of twenty-two cases did a school district levy an income tax before one its underlying jurisdictions. Since overlapping jurisdictions whose combined rates exceed the cap must split the tax proceeds, school districts serve to propagate rather than initiate income taxes.

rates almost always hit the cap[6] and the tax is almost never repealed,[7] interesting questions will focus simply on the presence or absence of an income tax.

Tax situs gives residence precedence over work which means taxes are first paid to the home community. A commuter is only liable for taxes where he works if his resident rate is below the one percent cap. There are two exceptions to this rule: commuters to Philadelphia or other states may credit any taxes paid there against their home obligations. Since the former has a non-resident rate in excess of four percent[8] and contiguous states impose significant tax burdens,[9] these commuters are exempt from home community wage taxes.

Politicians do not need explicit voter approval to implement the tax. Generally the earned income tax is proposed as part of the annual budget which of course includes other tax rates and provisions. Due to sunshine laws, there must be a period for public discussion which is formally the only time when citizens can influence the tax decision. But meetings centering on the EIT are often contentious and one required police intervention to prevent a riot (*Philadelphia Inquirer*, [2])! Politicians are reluctant to implement a wage tax in the face of such hostility, so in practice voter preferences are quite important to the levying decision.

1.2.2 Pittsburgh Tax Contagion

Tax situs, the priority of home over workplace obligations, and exemptions for com-

[6]In 1992, 95.8% of Philadelphia region municipalities had rates at zero or at least one percent (including coterminous school district levies).

[7]Of the 146 communities that ever taxed in the Philadelphia area, only one ever repealed its tax, and a few years later that tax was back on the books.

[8]As of 1992 the city wage tax was 4.3125% on commuters and 4.96% on residents.

[9]Credit for non-Pennsylvania taxes must be apportiortioned to state *and* local obligations, so exemption from the local income tax is not automatic. However, Pennsylvania did not have a state tax until 1972 and, of the contiguous states, only New Jersey has had either low enough rates or a tax reciprocity agreement so that full credit was not possible. Due to the administrative complexity of collecting taxes from other states (see Appendix 1.7), several township managers and one New Jersey Fortune 500 personnel manager asssured me that inter-state commuters have always been effectivly exempt from the local income tax.

muters will explain much of the take-up history. Consider a simple model[10] where citizens benefit from some publicly funded service in their home community but dislike taxes. All decisions are by majority rule among residents. Individuals are identical in tastes, income and property ownership, differing only in where they work and if they already pay a wage tax. Taxes over income and property are available, and there is no explicit preference between them. Write the utility function as,[11]

$$U(R, p, t, t_c, c) = B(R) - [p + t + \mathcal{I}_c \times (t_c - t)] \tag{1.1}$$

where $B(R)$ $[B' > 0, B'' < 0]$ is the benefit from the home community's public good when total tax revenues are R, while p and t are the home property and wage tax rates respectively. The last term ensures wage taxes are paid at most once; $\mathcal{I}_c$ is an indicator variable with value one for commuters who face a wage tax of t_c from some other community[12] (such as Philadelphia) or another state.

When the total income and capital stock are equal,

$$R = p + t(1 - f) \tag{1.2}$$

where government revenue is denominated in terms of the capital stock and $f \in [0, 1]$, the fraction of Philadelphia and inter-state commuters, is included to correct for the home tax exemption. Notice that the exempt (Philadelphia and inter-state commuters) have similar preferences as suburban commuters facing a wage tax at work [eq. 1.1], but revenues will be collected only from the latter when a home EIT is imposed [eq. 1.2].

When would a community with only property taxes approve a revenue-neutral

---

[10]My model follows the traditional median voter approach to local public goods, see for example Bergstrom-Goodman [4] or Inman [25].

[11]The comon per capita income and property ownership levels have been appropriately normalized so that tax payments are denominated in utils.

[12]For suburban commuters, the home tax actually has priority, but this functional form suffices as $t_c \equiv 1 \equiv t$ holds in general.

wage tax?[13] Since public spending is unaffected, preference is based on the change in tax burden.[14] The new policy has simple redistributive effects: those who already pay at work (commuters to Philadelphia, out-of-state or taxing suburbs) owe no additional income tax and a smaller property tax while everyone else gets a larger tax bill. Formally, when there is no wage tax all agents face an identical rate, p^*, on property.[15] A shift to an earned income tax of rate $t = 1$ under revenue-neutrality requires (from equation 1.2),

$$\widehat{p} = p^* - (1 - f) < p^* \tag{1.3}$$

So from equation 1.1, the effect on a voter's utility from the regime change is,

$$\triangle U^{\mathrm{EIT}} = \begin{cases} -f & \leq 0 \quad \text{if no wage tax at work} \\ 1 - f & \geq 0 \quad \text{if face wage tax at work} \end{cases} \tag{1.4}$$

This simple model has a powerful prediction: ignoring for now exemptions ($f = 0$), when a central city/commuting center enacts a wage tax, surrounding suburbs should instantly pass their own levy. City commuters strictly favor the shift since it imposes no additional burden on them while suburban workers are indifferent, equation 1.4. Intuitively, consider a world of two communities. When one decides to tax, commuters from the other pay taxes at work. This lowers the burden in the second community of a wage levy, since only those who work at home will see their tax bill increase. In other words, the neighbor lowers the implicit cost (tax burden) while not influencing the benefit (revenue collected) of a new tax. With enough working abroad, the appeal

---

[13]The qualitative results here and in the next sub-section would be unaffected if instead the government devotes the new revenues to a particular project which the voters unanimously support.

[14]Empirical support for this approach may be found in Lucier's ([34], [35]) analysis of tax substitution referenda in Oregon and Washington. He accurately fits county vote outcomes simply assuming that individuals seek to minimize their tax burden.

[15]Existence and uniqueness of this rate follow since the indirect utility function, $B(p) - p$, is single-peaked, so from Black's theorem the social decision is based on the preference of the median voter. The result follows since utility maximization over taxes is a concave programming problem.

of a tax becomes irresistible; enacting an EIT allows such communities to collect taxes which many citizens would have to pay anyway.

This is exactly what we see. Consider the Pittsburgh metropolitan region. Prior to 1954, the city and almost all of its suburban communities did not have a wage tax. When Pittsburgh passed a levy in March 1954, nearby governments instantaneously matched the action. By year's end more than three-fourths of the jurisdictions within twenty-five miles had their own tax! Similar tax clustering occurred in other metropolitan areas such as Scranton, Erie, Johnstown and Altoona.

1.2.3 Philadelphia SMSA: A Puzzle

The city of Philadelphia's special taxing privileges mean suburban commuters are exempt from home obligations. The Philadelphia SMSA is also near commuting centers in other states, such as Wilmington and Trenton, creating a second pool of exempt voters. Still, the decision to levy should follow from majority rule and the preferences given in the simple model summarized in equation 1.4: taxing communities should a have a large fraction of citizens working in Philadelphia, out-of-state or in another taxing suburb. This suggests a tax contagion emanating from the city and state border.

Maps 1 through 3 show the take-up experience in 1970, 1980 and 1992 respectively; for reference Figure 2 shows the fraction of taxing communities over time. While the predicted tax clustering does occur, communities near the Philadelphia and state border are *less* likely to levy. As people tend to work near their residence, these tax hold-outs should have a high fraction of citizens exempt from home earned income taxes; Figures 3 and 4 bear this out. This is in opposition to the self-interest model where exempt citizens rally for enactment combined with property tax relief.

Previous attempts to explain this pattern are unconvincing. The conventional wisdom, summarized in Luce-Summers [33], is that communities impose a wage tax only when their base is large enough, and so near-urban suburbs with a large fraction of exempt city commuters do not bother. The implicit model is that levies only occur when total benefits (tax revenues) exceed costs (say administrative fees or political opposition). But tax decisions are not made on efficiency grounds, but rather though a political mechanism. Instead of looking at aggregate values, it is more appropriate to consider the costs and benefits to particular groups[16] as I have done. Their story ignores the fact that while everyone shares equally in the benefits, certain groups (Philadelphia and inter-state commuters) are exempt from the cost and so should favor a wage tax.

To salvage the positive political economic approach, we must weaken some of the model's assumptions.[17] In town-meetings focusing on budget matters, the primary argument used against wage tax proposals played upon government distrust.[18] With this motivation, we no longer black-box the social decision process but explicitly consider the interaction between voters and politicians. While the earned income tax is usually introduced in the guise of revenue-neutrality, the government has no way to commit to this pledge. Even if property taxes are concurrently lowered, there is nothing to prevent an extraordinary tax increase in the next (or later) budget, a case of fiscal illusion. Suppose that if the government reneges, the revenue windfall has no perceived public benefit due to some combination of voter disillusionment, inefficient spending (empire-building or bureaucratic control), increase in deadweight loss, politician skimming, or voter misperceptions.[19] Then all voters would oppose an income tax when property taxes are not reduced appropriately, eq. 1.1.

We will call a government *credible* if it practices responsible taxation. That is,

---

[16] Oakland [39] and Citrin [12] present regression and survey data which show that interest groups such as homeowners were essential to the passage of Proposition 13 in California.

[17] Economists naturally gravitate towards the tax's effect on factor markets, capital and labor flows. But with the low rate cap of one percent, significant business migrations would require unreasonable elasticities. In formal regressions, available from the author, I have found that the EIT does not significantly alter total capital, employment or population levels.

[18] The review was based on all town-meetings from 1989 to 1992 indexed in *The Philadelphia Inquirer* "Neighbors" section.

[19] Matters are exacerbated if spending already exceeds the median preference following Romer-Rosenthal's [50] monopolistic agenda-setter or Pommerehne-Schneider's [48] fiscal illusion.

politicians will implement as promised non-binding policies, i.e. fully reduce property tax rates (from p^* to $\widehat{p}$) when a wage tax is implemented. At the other extreme, a *non-credible* government does not lower property taxes at all (stays at p^*) while maintaining the current level of services following a levy.

If voters do not know their government's type, the introduction of a wage tax is a gamble: with probability q they get a credible politician who provides tax diversification while with $1-q$ chance the new revenues will be siphoned away. In addition assume that levying the EIT independently lowers utility due to deadweight loss (lower future tax revenues as firms migrate) and psychological opposition to tax introduction (the "slippery slope" argument);[20] call this loss term $a > 0$. Generalizing from equation 1.1, an individual's expected utility from a one percent wage tax ($t = 1$):

$$\begin{aligned} EU^{\text{EIT}} &= q[B(R) - \widehat{p} - 1 + \mathcal{I}_c \times (t_c - 1) - a] + (1-q)[B(R) - p^* - 1 + \mathcal{I}_c \times (t_c - 1) - a] \\ &= [B(R) - p^* - 1 + \mathcal{I}_c \times (t_c - 1)] + [q(1-f) - a] \end{aligned} \tag{1.5}$$

substituting for $\widehat{p}$ from equation 1.3. The new terms are in the second set of square brackets in the bottom line. Credibility probability, q, enters interactively with the effective wage tax base, $1-f$, since the amount of property tax relief depends on how much EIT revenue is collected. The loss from the tax introduction enters directly.

In this framework, the change in utility from enacting the wage tax is,

$$\triangle EU^{\text{EIT}} = \begin{cases} -(1-q) - qf - a & < 0 \quad \text{if no wage tax at work} \\ q(1-f) - a & \lesseqgtr 0 \quad \text{if face wage tax at work} \end{cases} \tag{1.6}$$

As with certainly credible governments, equation 1.4, those who do not yet pay the tax oppose the shift. However, now even those who face a workplace EIT may oppose

[20] In addition, if voters were risk-averse than the loss would include the (negative of the) risk-premium of the wage tax gamble.

a home levy if their government suffers from credibility problems. Formally, for those who already pay the wage tax, there exists a unique credibility threshold $\tilde{q}(f)$ so that

$$\triangle EU^{\text{EIT}}_{\text{pay EIT}} > 0 \Leftrightarrow q > \tilde{q}(f) \tag{1.7}$$

It must be that $\partial\tilde{q}/\partial f > 0$: when there is a smaller wage tax base then the required threshold is higher.

These results indicate the conditions required for a home levy:

Proposition 1 *A wage tax in implemented only when:*

(i) *at least half the voters already pay the tax at work;*

(ii) *the government is credible: $q > \tilde{q}(f)$ where $\partial\tilde{q}/\partial f > 0$.*

This result has implications for where we should and should not expect to see new levies. In particular, when exemptions are not important ($f \to 0$), such as near Pittsburgh, then the credibility hurdle is low and a central city levy should catalyze a tax contagion. Alternatively, when the wage tax base diminishes ($f \to 1$), i.e. the periphery of Philadelphia or along the state border, then credibility concerns are amplified. So even though even a majority of individuals would not pay the home wage tax, near-urban communities may choose not to levy.[21]

In the remainder of the paper I will operationalize this proposition. The presumption that drives the empirical methodology will be:

Assumption *Voters use the history of fiscal policy to make their appraisal of government type. The resulting inter-community heterogeneity in perceived credibility (q) will identify the wage tax take-up regressions.*

[21]Some preliminary support for this contention may be found in Figure 3 of the accompanying paper [54]: no increase in wage tax payments for the median voter is a necessary but not sufficient condition for a home levy.

An alternative story would be based instead on variations in the unobserved taste for the wage tax itself, the parameter a, see equation 1.6. Evidence refuting this story will be presented in Section 1.3 and 1.5.

One potential concern with the approach taken here is that politicians may game on the informational asymmetry and strategically set taxes. For example, a non-credible government may try to signal that it is credible to gain the EIT power and only then escalate tax rates. A full extensive-form model which addresses this point is presented in Appendix 1.8.

There are two tasks left to uphold the political economy model. First, we must find proxies for voter perceptions of government credibility. We do this in the next section. Then using a rigorous statistical approach we must show both that these factors deter and that exempt groups encourage new wage tax levies.

1.3 Credibility Proxies

Citizens must infer their politicians' intentions perhaps using past behavior as a guide. Unfortunately, we cannot directly observe what voters think about their government's reliability, at least with the available data.[22] Instead we must rely on some proxies which voters may use to evaluate their home politicians. While no perfect measure exists, it will be difficult to explain the results of this section without returning to the credibility story.

One likely input in voters' evaluations is the tax burden. There is some evidence in the literature that high tax levels (Bradbury [6], Oakland [39] and Levy [32]) or recent rate increases (Stein-Hamm-Freeman [53]) stir animosity towards fiscal expansion. In particular, Peltzman [40] finds that voters punish politicians who oversee spending growth at the federal and state level, regardless of its composition. Intuitively, tax

[22]One method of proof would be in-depth surveys of why voters oppose income tax proposals. However, such a study would require a time series dating back to the first enactments in the 1960s. Such data was never collected to the best of my knowledge.

bill spikes often cast doubt on the fiscal sensibilities of current incumbents while consistently high taxes foster cynicism towards the government or party machine (if one exists). Thus, some instruments will involve level and change in the prior tax burden.

Consider the common measure of fiscal burden, tax effort. This is defined as revenue collection divided by the level of taxes generated if the mean regional rate prevailed on income and property ("fiscal capacity"); high values indicate extensive taxation. It turns out that mean tax effort increases as we get closer to the city, see Figure 5.[23] This pattern is analogous to the pattern of hold-outs in Map 3, suggesting a possible connection. With this motivation, define low tax municipalities as those whose tax effort is below the mean for that year. Figure 6 plots the fraction of new taxers with low tax effort the year prior to adoption. While there is a mixed relationship in the 1960s and 1970s, by the end of the sample we see a strong positive relationship: in the 1980s, it is mainly low tax communities which take-up taxes.[24] Similar results hold using relative property rates among communities lacking an income tax as the measure of tax burden.[25]

The final instrument captures how efficiently governments deliver useful services. Communities which have higher levels of administrative overhead (and so less funding for public goods) will be viewed with skepticism. First, voters may interpret such spending as indicative of irresponsible fiscal behavior. And if the government receives a revenue windfall, voters would be less likely to receive higher public services. Using overhead as a fraction of the total expenditure is an inappropriate measure, since this

[23]These values actually understate the tax burden gradient since the true tax base would exclude the income of Philadelphia commuters. Nonethless, the annual correlation coefficient between distance and effort is always below -0.75.

[24]The apparent outlier in 1989 can be attributed to a single community while the points in 1990 though 1992 represent several locations.

[25]The property tax norm is restricted to non-taxers since it is inappropriate to compare rates among communities with different number of tax instruments. A community with an income tax can have a low absolute property tax rate which is still exorbitant considering its alternative revenue channels.

ignores the advantage of size in providing government services. So we use the residual from,

$$\ln Overhead_{it} = \alpha_t + \beta_t \ln Expenditure_{it}$$

which allow for fixed costs and returns to scale, each of which may vary over time.[26] In Figure 7, we see overhead also follows a positive gradient approaching Philadelphia.

How reasonable are these proxies? First, their deterrence of new levies would be unambiguous support for the credibility mechanism. High tax burden or overhead both indicate an increased need for new revenue, so our prior is that these variables increase tax-levying propensity. If the variables slow tax propagation, it is information queueing rather than fiscal necessity which generates the statistical relationship.

The proxies are also believable inputs into the public's attitude towards government. The measures are relatively visible and most voters have some feeling for their current burden and level of local services. It is also clear that taxes and benefits are very important in shaping opinions towards politicians.

Following this line of argument, we examine how the proxies influence politician turnover. If our story is correct, adopting the earned income tax will inspire voter backlash only when the government is non-credible. In terms of the proxies, it is the combination of high tax burden (levels or first-difference)/administrative costs and a new levy which should increase turnover. For example, when taxes increase in the year prior to a levy, the model predicts voters rebuff the current incumbents whom they do not trust with a new instrument.[27]

Allowing for period and county fixed effects, how do voters react to a new income

[26]The equivalent model is,

$$Overhead = e^{\alpha_t} Expenditure^{\beta_t}$$

so the α_t measure fixed costs and β_t the returns to scale.

[27]This presumes voters only scrutinize government fiscal policy in the wake of a new earned income tax. In other words, a catalyst is required for citizens to become politically active.

tax? OLS regressions are presented in Table 1.7. An introduction has an insignificant effect on turnover [row 2]. However, if relative tax effort [rows 3-4, column (a)] or administrative overhead was high [rows 7-8, column (c)], politicians are voted out in large numbers.[28] Similar results obtain using the first-difference of tax effort [rows 5-6, column (b)].[29] These results mean that after partialing out the exempt groups, the EIT only catalyzes a backlash against governments with credibility problems. Voters only trust new taxes in the hands of politicians they associate with conservative fiscal policy.

A second aspect of the story is that voters predict only credible governments will maintain revenue-neutrality following the income tax. In other words, non-credible politicians enact the tax for revenue enhancement rather than revenue diversification.

The average change in millage[30] over 1960-1992 is plotted in Figure 8. Three groups are considered: all communities, municipalities enacting the EIT in a particular year and low burden new income taxers. There are diverging experiences. While the average change tends to be slightly positive, a new levy tends to accompany a decline in property tax . This reduction is even larger in communities with low initial tax effort suggesting high burden governments do not provide substantial tax relief.

It turns out that only credible governments use the earned income tax to substitute away from property tax. Table 1.8 shows an OLS regression for change in effective property millage, rates per true market value. Property taxes fall in the year an income tax is initiated [row 2]. However, when this dummy is multiplied by the

[28] With the interaction the total effect of a new wage tax on turnover is

$$\frac{\partial Turnover}{\partial First} = \beta_{First} + \beta_{First \times Proxy} Proxy$$

Substituting large proxy values makes the right-hand side significantly positive.

[29] Of the proxies, only change in tax effort significantly increases turnover in years without a tax introduction (rows 3, 5, 7) which is in keeping with the interpretation that recent tax increases are blamed on incumbents.

[30] Property tax rate is usually stated in "millage" or dollars of tax per one thousand dollars of assessed value. While this does not represent the tax burden due to divergence from true market value, changes in gross rates are the important government control variable.

proxy variable [rows 4, 6, 8 in columns (a), (b), (c) respectively], the total effect is positive.[31] In other words, high/increasing tax or high overhead governments do not significantly lower their property rates when they introduce a wage tax. It is rational for voters living in such communities to forecast a sharp rise in their tax bill when a new instrument is enacted.

As a final note, these figures and regressions run counter to the alternative theory that variation in take-up is due to different taste for wage taxes (Section 1.2.3). Under that lens, Figure 5 and 7 mean that the hold-out communities prefer high total taxes and overhead but totally oppose any form of wage tax! Moreover, it is hard to explain the importance of credibility proxies in citizens' reaction to new levies at the voting booth, Table 1.7, using the alternative hypothesis.

1.4 Empirical Specification and Data

We are now in a position to analyze the important factors in tax-setting behavior. All that remains is the choice of statistical technique and included regressors. Sources and summary statistics for all variables are also provided.

1.4.1 Econometrics

The information we have is a panel of binary decisions, so the appropriate tools should utilize the variation in whether and when a community decides to tax. The simplest approach is to examine static cross-sections at various times and see which regressors enhance the probability that a community has the tax. I will use the familiar probit regression which presumes an observed binary variable is the result of a threshold

[31] That is, the combination of direct [row 2] and interaction [rows 4, 6, 8] parameters evaluated at a typical high proxy value is positive, see earlier note.

hurdle for an underlying continuous variable. Formally,

$$\Pr(Tax_i) = \Phi(X_i\beta) \tag{1.8}$$

where i indexes communities, Tax takes on a value of one when an EIT is present and zero otherwise, X is a vector of regressors and $\Phi(\cdot)$ is the standard normal distribution. The parameter vector β will be estimated from the data with positive values indicating the regressor increases the chance of taxing. The advantage of this approach is that we will be able to easily check if parameters are stable over time, and if not, whether it is shifting regressors in existing taxers or the addition of new taxers that drive the change. But there are several limitations such as ignoring a community's prior history, the fact that tax levies are irreversible, and possible error from year-specific effects.

Duration models, a dynamic generalization of binary choice, overcome these problems. We use the popular proportional hazards functional form,

$$\lambda[t, X_i(t)] \equiv \lambda_0(t)\exp[X_i(t)\beta] \tag{1.9}$$

where $\lambda[t, X_i(t)]$ represents the probability that a community will levy a tax for the first time in the next instant given regressors $X_i(t)$ and a vector of unknown parameters β; $\lambda_0(t)$ is the baseline hazard which captures period specific effects. Notice that the regressors are free to vary over time. A convenient property of this form is that

$$\frac{\partial \log \lambda[t, X_i(t)]}{\partial X_i} = \beta$$

so β_v may be interpreted as the proportional effect of variable v on the probability of a new levy in any period. In other words, a negative parameter value means the regressor delays new taxes.

Since I am generally uninterested in the baseline hazard,[32] I employ a Cox partial likelihood which is indifferent as to the functional form of $\lambda_0(t)$ and appropriately treats taxing as an absorbing state. Heuristically, this method only considers the order of communities non-taxing durations including the special case of never enacting at all. Formally, if we observe N communities tax at distinct times $0 < t_1 < t_2 < ... < t_N$ and $N+1, N+2, ...M$ never tax, then from an independence assumption[33] and Bayes law the probability of observing the jth shortest duration is

$$\frac{\lambda[t_j, X_j(t_j)]}{\sum_{i=j}^{N} \lambda[t_j, X_i(t_j)] + \sum_{i=N+1}^{M} \lambda[t_j, X_i(t_j)]} = \frac{\exp[X_j(t_j)\beta]}{\sum_{i=j}^{M} \exp[X_i(t_j)\beta]}$$

where the baseline hazard cancels out. Kalbfleisch and Prentice [27] show approximations to allow for ties in the tax order. Considering the product of all such terms (i.e. making another independence assumption) and taking logs gives the final form of the likelihood function,

$$\log L(\beta) = \sum_{t=t_1}^{t_N} \left\{ \sum_{j:t_j=t} x_j(t)\beta - d_t \log \left[\sum_{i:t_i \geq t} \exp(x_i(t)\beta) \right] \right\} \quad (1.10)$$

where d_t is the number of new taxers at time t. Again, notice that at any time information from all non-taxing communities plays a role in the likelihood.[34]

A final issue involves non-annual data. Some of the regressors are observed only once a decade. Following Diamond and Hausman [20] we use linear interpolation over the intermediate years.

---

[32] Not estimating the baseline also precludes estimating parameters for environmental variables which vary over time but are constant for all communities. One potentially interesting example would be the Philadelphia tax rate which climbs from two to nearly five percent over the sample period.

[33] This still allows tax-decisions to be inter-dependent on the contiguous taxers as in the Pittsburgh tax contagion. So long as behavior is independent of others in the risk set, communities who have not yet levied a tax, the partial likelihood approach is still appropriate.

[34] One deficiency of the Cox approach is that data from any year in which no community enacts a levy is not used. This occurs in nine of the 33 years in my sample.

1.4.2 Data

The remaining step is to specify which regressors to use in the sample of 237 suburban Philadelphia municipalities ("Minor Civil Divisions" or MCDs) over 1960 to 1992. From the model of the last section we know key variables will include groups which do not pay earned income tax at home, various relative tax burden measures, and administrative overhead.

The primary exempt groups are Philadelphia and out of state commuters. These values come from the decennial Census "Journey-to-Work" file which has sample rates of 25%, 15%, 50% and 100% in 1960, 1970, 1980 and 1990. There are two other groups which receive favorable treatment under the EIT. Senior citizens generally have only passive income which cannot be taxed,[35] while home owners perceive a smaller share of income rather than property tax compared to renters.[36] The fraction of elderly and the ratio of owner to renter occupied housing structures are from decennial Census tract data, [9] and [10].[37] Housing tenure is based on full sample as are the level of senior citizens (except 1960 with 25% sample).

Following the idea of the tax contagion, we should include the percent who face a wage tax at work.[38] Presumably such voters would favor a home tax, since it imposes no additional burden on them. The variable is generated from a matrix of MCD-to-MCD commuter flows, i.e. for each community an exhaustive list of destinations and the fraction who work there. The data for 1980 and 1990 are based on a remarkable and under utilized decennial data set from the Department of Transportation, the Urban Transportation Planning Package (UTPP) [11]. The UTPP did not ex-

[35] Unemployed also do not have current earned income but, unlike the retired, should anticipate future wages. I therefore will not investigate the role of unemployment on tax decisions.

[36] Since renters do not directly pay property tax, several empirical studies indicate they underestimate their share of this tax.

[37] Using Census maps I was able to aggregate tracts into governmental units for each of the years.

[38] Sophisticated communities might also enact a wage tax when the there is a large pool of *non-residents* eligible to pay the home tax. But this revenue source declines with time as more governments implement an EIT, and several township managers told me that tax exporting did not play a role in their community's levying debate. Formal estimates in the accompanying paper [54] (Section 3.2) show that the non-resident tax base is an insignificant correlate in the propensity to levy.

Year	*mean*	*max*	*min*	σ
1970	0.42	0.66	0.08	0.16
1980	0.58	0.90	0.02	0.15
1990	0.57	0.89	0.20	0.14

Table 1.2: Fraction of Government Revenue From Wage Taxes
Sample limited to communities with wage tax in the Philadelphia SMSA.

ist for earlier years,[39] so approximate flows are estimated using the MCD-to-county commuting matrix in the Census "Journey-to-Work" tract data; a full discussion of the algorithm is in the Appendix 1.9. The sample level is identical to the earlier commuting data.

We use two measures of relative tax loads. Both are constructed to be unaffected by time trends, so values simply indicate levels relative to the norm for that year. The first variable is tax effort divided by annual mean ("relative tax effort") while the other is effective property millage relative to the mean among communities without an income tax ("relative property tax"). Effective levels are used since county-level assessments of property diverge from true market value.[40]

The relevant data was compiled from *Local Government Financial Statistics* [43] which are archived at the Department of Community Affairs in Harrisburg. This includes total and earned income tax revenue; because of their central role in our analysis, income tax collections were cross-checked against rate data in [44] to uncover occasional discrepancies. Table 1.2 shows that income tax regularly makes up over half of all revenues, and that this fraction has increased since 1970; the *max* column shows values near unity since 13 communities completely abolished property taxes![41] Nominal millage was multiplied by the ratio of assessed to true market value (state assessor estimate) to get effective millage levels.

[39] An extended version of the Journey-to-Work file, the Census Transportation Planning Package (CTPP), was initiated in 1970, but transportation experts at the Delaware Valley Regional Planning Commission informed me that the Philadelphia SMSA data was quite unreliable.

[40] In 1992, not one of the suburban counties had reassessed in over ten years with one's last change over fifty years prior! True tax burden must correct for this.

[41] For at least some portion of 1960-1992.

Variable	*mean*	*max*	*min*	σ
Revenue ($\times 10^6$ \$)	3.423	40.961	0.108	5.295
Tax effort	0.075	0.389	0.009	0.055
Effective millage	1.720	34.392	0.000	2.556
Expenditure ($\times 10^6$ \$)	3.230	38.706	0.070	4.952
% Overhead	16.31	61.86	1.90	10.25
% Deficit	-9.98	38.71	-306.57	29.17

Table 1.3: 1992 Fiscal Data
Overhead and Deficit are percent of expenditure.
Negative numbers in Deficit row indicate a surplus.

The administrative overhead ("government overhead") data also comes from [43]. The variable is based on the expenditure category, "general government," which includes spending on tax collection, administration, maintenance and planning. Notice that this undermeasures the true level of overhead since some waste may be hidden in public goods categories; for example, salaries for administrators in the police force would not be counted in general government spending.

1992 summary statistics for the fiscal variables may be found in Table 1.3.

We must control for other variables which may influence tax decisions. Communities may implement a wage tax due to a taste or economic shock which requires a large revenue surge. While property taxes could be used, there may be resistance due that tax's clear visibility. To control for this we include the previous period government deficit, expenditures minus revenues,[42] as a percentage of total expenditure, with both variables drawn from [43]; representative values are in Table 1.3. In the regression, a positive parameter on this term means that the wage tax is being used to cover new revenue needs.

We have also seen that government structure might influence tax choice, Table 1.1. Buchanan and Brennan ([7], [8]) suggest that more bureaucratic forms insulate politicians from voters and encourage accumulation of tax rents, the Leviathan theory. I include indicator variables for cities and boroughs (township dummies omitted)

[42]Revenues include not only taxes but also user fees (one-third of municipality revenue in 1988) and state or federal grants (an eighth of revenue).

based on Department of Community Affairs records.

A large business presence might preclude the need for any other tax besides property. Firms do not vote, so their interests may be under-represented. The percent residential and percent commercial land for 1970, 1980 and 1990 is based on analysis of detailed aerial photographs, see [16] and [18].[43] Earlier data is unavailable. Similarly, county dummies are included to account for variation in date of last property assessment.

Finally, the relevant demographic variables must be included to control for variation in demand for public goods. Total population, population growth and median household income[44] are included from Census sources. The income variable is based on 25%, 20%, 19% and 100% sample in the four surveys while population is full sample. The number of jobs located in a community comes from two decennial sources. Employment for 1970 through 1990 comes from the Delaware Valley Regional Planning Commission's modification of the Journey-to-Work and Bureau of Economic Analysis figures- [14], [15] and [17]; this series is unavailable for earlier years. 1960 values are based on a state administered business survey, [45], with 1970 figures comparing favorably to the other series.

We have already considered politician turnover, the fraction of officials up for election who do not return for the next term.[45] Data is hand compiled from two sources: 1960 through 1974 comes from [47] while 1978-1992 figures are based on [19]; data for 1976 is unavailable. Observations are bi-annual (local elections held in odd years) and only include 86 MCDs. The subsample is representative as 45

[43]The sum of these variables is generally less than unity due to omitted categories (such as transportation and undeveloped) which are generally untaxed.

[44]The 1960 and 1970 Census consider the unit "families and unrelated individuals" which is quite close to the "household" group used in later years.

[45]The relevant positions are mayor and councilmen for cities or boroughs and commissioners and supervisors for townships. Terms are staggered in time with a quarter to half of the seats up in each election; term length varies with position and government type. Finally, several councils expanded or contracted in size at the beginning of the sample. Various corrections and data truncations do not markedly influence the results.

Variable	*mean*	*max*	*min*	σ
Median Household Income ($\times 10^3$ \$)	43.42	102.99	20.86	12.03
Population ($\times 10^3$)	9.04	81.18	0.44	11.45
% Seniors	12.61	25.02	2.91	4.17
Jobs ($\times 10^3$)	4.77	46.43	0.05	6.98
Jobs/Population	0.51	3.00	0.03	0.43
Market Value Real Estate ($\times 10^6$ \$)	506.20	6119.90	11.90	709.50
Assessed Value Real Estate ($\times 10^6$ \$)	26.10	324.40	0.80	36.80
Owner/Renter Occupied Housing	4.41	22.52	0.55	3.28
% Residential Land (acres)	28.64	88.35	1.00	17.88
% Commercial Land (acres)	30.30	78.48	0.26	17.94
% Philadelphia commuters	9.95	59.15	0.00	10.29
% Out-of-State commuters	9.44	71.29	0.50	11.92
% Work where EIT (exludes Phil.)	42.15	90.46	2.16	26.40

Table 1.4: 1990 Demographic Data

eventually levy the earned income tax.

A set of summary statistics are given in Table 1.4.

1.5 Results

In this section we identify the key determinants of tax-levying propensity. Without controlling for government credibility (the benchmark), a higher fraction of citizens exempt from the wage tax diminishes the probability of implementation, counter to self-interested behavior. In the augmented regression, proxies for government mistrust have a significant negative effect on taxing. Moreover, inclusion of these variables yields the expected positive coefficient on exempt groups; the sign change in the benchmark case stems from colinearity with the omitted credibility proxy. Time-variation of the proxy parameter is related to the Taxpayer Revolt of the late 1970s. All results are replicated using either probits or the more appropriate Cox hazard technique.

To maintain focus I will not discuss parameters on the control variables listed in the Data section. None play significant roles in the regressions.

1.5.1 Cross-Sections of Taxing Behavior

Probits are run at times when there are different number of taxers in 1970 (18.1%), 1980 (43.9%) and 1992 (61.6%). Results are reported in Tables 1.9-1.11.[46] Consider the first panel of Table 1.9 which contains the benchmark case (no proxies). The most important regressors are Philadelphia/out-of-state commuters, senior citizens, owner-to-renter tenure and those facing a wage tax at work. The interest group theory predicts these favored classes should push for passage of the income tax, so we anticipate positive parameters.

While this generally holds for percent already paying wage taxes, the commuter terms grow increasingly negative while the other two terms are generally insignificant. I re-ran the probits in 1980 and 1992 deleting all MCDs which already had a tax in the earlier regression(s) to ensure that the same effects governed tax decisions in all periods.[47] The commuter parameters become even more negative while the seniors and tenure terms do not noticeably change (regressions omitted). Dropping the county dummies also does not change matters.

As the conceptual framework suggests, ignoring voter mistrust is the root of this problem. To show this, include the credibility proxies- lagged tax level, lagged change in tax burden and lagged administrative overhead-[48] one by one as explanatory variables. The results are listed in the second and third panels of Tables 1.9-1.11 with the new parameters in the second and third row. High values of the proxies (which correspond to low credibility) significantly limit new taxes particularly in the last two years.[49] Even more interesting is the effect on parameters of the exempt groups: seniors and commuters are now significantly positive! Notice the commuter terms

[46]Land use variables are not listed in the table since data limitations preclude their use in the hazard model of the next sub-section. These factors, such as percent land devoted to business use, have insignificant parameters when included in probits.

[47]Recall tax levying is irreversible, so regressor variation in communities which already enacted the EIT is non-informative.

[48]Lagged values are used to reflect prevailing conditions when tax decisions are being made.

[49]The last line contains a log-likelihood test that benchmark is the full specification. Using the $\chi^2(1)$ criterion, in all cases but one we can reject the null even at 99% confidence.

have completely flipped their relationship to taxing propensity; the explanation for this change, omitted variable bias, will be discussed in greater detail in the next subsection. The important point is the positive effect of exempt groups on new levies confirms the interest group model.[50]

1.5.2 Timing of Tax Decisions

The probit approach is suspect since it ignores the historical path until a levy and inappropriately considers variation in regressors of communities which already have the tax. The proportional hazards model overcomes these difficulties with a minimum of functional form assumptions.[51] Still the results corroborate those of the probit model.

The first panel of Tables 1.12-1.14 again includes the benchmark case without the credibility proxies. The regression is run on the full sample of 1960 to 1992 and, given the variation in parameter estimates in the probit, two subsamples. I chose 1975 as the break point since this ends a period of widespread tax introduction (Figure 2).[52] Except for those facing a workplace tax, the parameter for the exempt groups are negative particularly in the second subsample, again counter to the simple self-interest story. The second and third panels report results when the credibility proxies- lagged tax burden (Table 1.12), lagged change in taxes (Table 1.13), lagged government overhead (Table 1.14)- are included. The added regressors are increasingly negative (rows one and two), and they leave a positive sign on the exempt groups, the basic prediction of the credibility model.

[50]I do not take the general insignificance of the housing tenure term as a rejection of the model, since this term's presumed effect was predicated on renters under-estimating their share of property taxes.

[51]For robustness I also used the parameterized version for the baseline suggested in Flinn-Heckman [22] which allows for unobserved heterogeneity. The maximum likelihood estimates were not noticeably different from those listed here (results omitted). This is strong evidence against the variation in taste hypothesis suggested in Section 1.2.3 which should wash-out under this fixed-effects approach.

[52]As the likelihood ratio statistic exceeds $\chi^2(16)$ value even at 99.5% significance, we cannot reject the null of no structural change.

Regressor	Relative Tax Effort(-1) 60-92	60-75	76-92	Relative Property Tax(-1) 60-92	60-75	76-92	Government Overhead(-1) 60-92	60-75	76-92
% Philadelphia Commuter	0.41	0.35	0.47	0.47	0.41	0.53	0.35	0.39	0.29
% Out-of-State Commuter	0.27	0.15	0.31	0.24	0.15	0.28	0.22	0.08	0.27
% Senior Citizen	0.34	0.30	0.39	0.28	0.25	0.37	0.24	0.15	0.30
$\frac{\text{Owner}}{\text{Renter}}$ Tenure	-0.21	-0.14	-0.22	-0.23	-0.24	-0.21	0.24	0.23	0.25
% Face EIT at Work	-0.11	-0.05	-0.15	-0.23	-0.31	-0.12	-0.04	-0.08	-0.02

Table 1.5: Correlation of Exempt Group with Credibility Proxies [Level]

Regressor	ΔTax Effort(-1) 60-92	60-75	76-92	ΔProperty Tax(-1) 60-92	60-75	76-92
% Philadelphia Commuter	0.25	0.21	0.33	0.31	0.40	0.19
% Out-of-State Commuter	0.17	0.13	0.23	0.19	0.15	0.20
% Senior Citizen	0.20	0.17	0.22	0.16	0.11	0.20
$\frac{\text{Owner}}{\text{Renter}}$ Tenure	-0.08	0.01	-0.19	0.02	0.00	0.03
% Face EIT at Work	0.01	0.02	-0.06	-0.07	-0.18	0.02

Table 1.6: Correlation of Exempt Group with Credibility Proxies [Change]

We must search a little deeper to understand the sign change from the benchmark case. It turns out that both commuter terms and the fraction of senior citizens are positively correlated to the proxy variables, Tables 1.5-1.6. Intuition suggests that when the proxy is omitted highly correlated variables will incorporate its effect and suffer from negative parameter bias in either the cross-section or panel model.[53] And this makes sense with our results. In the benchmark case the commuter parameters are most significantly negative in the later subsample, precisely when the proxy terms have their greatest deterrent effect. Including the credibility proxies also shifts upward the elderly parameter though not as noticeably.[54]

One loose end is why credibility gaps become a stronger deterrent after 1975 (see

[53]Kiefer and Skoog [29] show that omitted variable bias in probits locally has the same form as the OLS case while Kiefer [28] claims a similar formula for exponential hazards without duration dependence. These mean positive correlates of a credibility proxy will have a negative bias when the latter is inappropriately excluded from a regression.

[54]As further confirmation, notice the only remaining negative sign on an exempt group, the interstate commuter term in the 1960-1975 hazard using the overhead proxy, may be traced to a small correlation coefficient.

regression results or Figure 6). One explanation involves a changing attitude towards local governments. The Taxpayer Revolt, usually associated with property tax limits, grew into full force in the late 1970s (Ladd-Tideman [30]). Several national surveys indicate that citizens became much more skeptical of the local government's ability to provide services at a reasonable cost; in a review of such polls, Citrin [12] documents that the number who distrusted politicians, felt that taxes were beyond the "breaking point" or believed governments were inefficient rapidly escalated in the 1970s. And in our case, the worsening fiscal situation in the city of Philadelphia surely served as a rallying point for those opposing new public programs. The greater resistance to government expansion in the second half of our sample thus has a basis in shifting voter preferences.[55]

1.6 Discussion

At first blush, the spatial pattern of wage tax resistance in Philadelphia suburbs seems to require majority coalitions foregoing a welfare enhancing policy. But this presumes voters take government promises at face value while in reality they are quite skeptical. We found that governments suffering from low credibility are prevented from enacting an income tax. Correcting for such concerns, a greater numbers of exempt individuals such as Philadelphia commuters increase tax-levying propensity, just as the political economy model predicts. That parameter estimates make sense only when proxies are included is strong support that credibility governs wage tax decisions and suggest that high tax burdens or government waste are key components to tax resistance. A policy implication of this analysis is that new fiscal powers will be granted to an unreliable government only if it can assuage voter concerns, perhaps by placing new revenues in a trust fund.

One potential difficulty remains. If the voter suspicion story is true, why are

[55] In terms of the model, more skeptical voters assign a higher probability that the government is non-credible for any given level of the "objective" proxy variables.

there not other pockets of tax resistance in the commonwealth? First, the revenues generated from a wage tax is lower in the Philadelphia area due the exemption for city commuters. So *ceteris paribus* the expected voter benefit from adopting a tax is higher outside the Philadelphia area. In other words, the minimum level of government credibility required for approval of new taxing powers is lower in communities elsewhere in the state. This point was stated formally in Proposition 1.

Second, the city of Philadelphia was embroiled in several financial crises during this period, unmatched by any other city in the commonwealth. With the central city as the focal point, it seems plausible that Philadelphia suburbanites are more skeptical than their counterparts elsewhere. As evidence consider the returns from the May 1989 Local Tax Reform Act which sought to raise the municipal tax cap in return for property tax reductions. While the measure lost by a 3-to-1 margin statewide, the Philadelphia suburbs were far more skeptical voting greater than 7-to-1 against. The bill's opponents, many centered in the Philadelphia area, successfully channeled voter fears that politicians would abuse their new taxing privileges (*Philadelphia Inquirer*, [1]).

Interacting with these facts is a temporal feature. Central city levies were in effect well before the 1970s, so the tax contagion had time to work before the Taxpayer Revolt heightened voter resistance. It is questionable then whether the Pittsburgh tax contagion or others like it would be repeated in the current climate of government distrust.

The case would be strengthened with evidence from other laboratories of local tax-setting. Unfortunately, with near universal tax-adoption, the rest of Pennsylvania is not the place to look. An alternative path would be to examine other states where local income taxes are allowed but not mandated. The leading contenders are Indiana, Ohio and Kentucky, all of which seem to have some variation in tax adoption. For example, of the 612 Ohio school districts, 96 passed an income tax and 144 rejected it between 1989 and 1993 (*New York Times*, [3]). Whether their experience upholds or disproves the conclusions of this paper is an interesting topic for future research.

1.7 Appendix: EIT Administration[56]

Income tax collection is locally administered. Every employer must submit to its community's earned income tax officer a list of all employees and their incomes; self-employed individuals and those working out of state must in practice submit quarterly estimated taxes to the tax agency. It is the officer's job to see that the proper level of taxes are withheld and that revenues are remitted to the (taxing) home of non-residents. Due to the clear returns to scale, most governments contract out to one of a few private accounting firms to collect and enforce the income taxes. These companies keep a few percent of the revenue as their fee, though much of their profits derive from the interest float generated before they remit the taxes.[57]

Collecting revenues from workers who commute out of state is generally impossible, since no other (contiguous) state either maintains detailed records of worker residence or is willing to remit payments across state lines; attempts to collect delinquent taxes require an expensive and time-consuming civil case and so are generally avoided. Tracking down out-of-state workers is even less likely in communities using a private collection agency which generally view extensive enforcement as a costly diversion from simpler profit opportunities such as interest accumulation. Other loopholes, such as the exemption of S-Corporations[58], also preempt collections.

No state agency has authority to mandate a common enforcement policy, and so there is great variation in how the income tax is administered. The only universal principle of tax collection is minimal enforcement of laborious cases (like inter-state commuters).

[56] I thank Ellie Kirby, a past president of the Pennsylvania Earned Income Tax Association and current tax collector for Palisades school district, for explaining many of the nuances of tax collection law and how they are applied in practice.

[57] Currently, the tax agencies can hold onto the monies for 90 days.

[58] An S-Corporation is a special Pennsylvania tax designation applied to certain partnerships.

1.8 Appendix: Strategic Politician Model

In this section I address concerns over politician gaming on the voters' information extraction. I find that if voters can only imperfectly observe the intended fiscal policy than there is no reason for politicians to mask their true taxing intentions. This means the main conjecture of the empirical paper, only credible governments will be granted an earned income tax (EIT), should hold on average.

The information asymmetry which drives the result is realistic. Voters primary source about government revenue policy is their annual property tax bill. But while municipalities set property millage *rates*, residents only see the *total tax* due. Moreover, payments to the municipality and school district (which has a separate rate structure) are grouped together;[59] while based on a common assessed property value, there is no break down of rates and base. So residents are likely to be well informed about the total level of taxes they pay and still have only an imperfect notion of the government control variable, the municipal millage rate.

The formal approach will utilize a reputation model. The key feature of such games[60] is that an agent, privy to some information, sends a signal to an imperfectly informed principal who then tries to infer something about the private information. Finally, the principal gets to take some action (multiple cycles of signal-inference-action are sometimes used). Generally,[61] the signal will either reveal the information ("separating equilibrium") or be uninformative ("pooling equilibrium") though hybrid forms are possible.

In my credibility model, voters must decide at a town-meeting whether to allow politicians to levy the income tax. If the tax accompanies a drop in property tax, voters benefit (say from diminished deadweight loss). However, politicians might

[59] Complicating matters further, citizens also must pay county property taxes.

[60] Theoretical applications have largely been confined to industrial organization (see the survey in Fudenberg-Tirole [23] Chapter 9) and macroeconomics (Persson-Tabellini [46]).

[61] Since these are dynamic games of incomplete information, the equilibrium concept is sequential or perfect Bayesian, thus ensuring that beliefs are updated in a reasonable fashion.

renege on property tax relief and use the new revenues to line their pocket; voters end-up paying more taxes with nothing to show for it. The twist is that politicians alone know whether they are the greedy or altruistic "type." Voters make an inference about this private information based on the history of tax burden/government overhead, so politicians may game on them. This structure is closest to the Kreps-Wilson/Milgrom-Roberts incumbent-entrant game (see [23]); here there is an additional player and another source of imperfect information. Rogoff-Sibert [49] analyze a similar game where politician competence is the source of information asymmetry and voters simply decide which of two parties is put in office (not whether to give the government new powers as in our model).

The main result: only politicians with a track-record of lower taxes[62] are entrusted with the new tax instrument. The intuition is that it is costly for greedy politicians to signal low taxes in the first period, so they will mask their type only if this gains additional payoff in the later period. But since voters imperfectly observe the rate politicians select, they attribute high rates to system noise rather than politician deviation. Thus, even self-interested politicians will not game on their private information but will reveal their intentions with a high property tax rate; in the language of game theory, there is no pooling equilibrium only a separating one. Clearly such a result does not hold in a model where voters' believe all politician promises.

This claim has three empirical implications which are tested in the main paper. First, high tax burdens should deter new the spread of the EIT. And since politicians do not strategically set rates, only "good" types will get new taxing powers; this means that significant property tax relief should tend to accompany a wage tax. Finally, given the imposition of an income tax, politician turnover should only increase only if the initial tax load is heavy.

[62]It is equally valid to interpret the politician choice variable as the change in tax rates or the level of government overhead, the other proxy variables used in the main text.

1.8.1 Structure

Consider a community using a single revenue instrument, property taxes. The government proposes a revenue-neutral package of property tax relief and a new earned income tax which requires voter approval.[63] There is no way to commit to the reduction in property rates. The voters/principals favor the pledged program since it replaces a large distortion with two smaller ones,[64] but would oppose the EIT alone. The politician/agent who implements the government plan is either an altruistic, voter-welfare maximizer ("good" type) or self-interested rent-seeker ("bad" type). Only the good politician will fulfill his pledge to lower property taxes while the bad type will use the new government revenues for his private gain. Politician type is privately known, though voters can make inferences based on a previous tax policy. Voters know that only good politicians will not renege, and so will oppose the package if they think the government is the non-credible type. Their inference about government intentions is based upon the community's fiscal history.

Formally, let $\theta \in \{G, B\}$ be the politician type where G is good and B is bad. Let q be the voters' prior probability of type G; in the case $q = 1$ credibility is not an issue. Politicians know the values of both of these variables. Let the politician/government and the voters each be represented by a single player.[65]

The timings is as follows (**Diagram 1**). In the first period the politician sets the property tax rate (a telescoped history of tax burden), a noisy version of which the voters observe and pay.[66] Government spending provides no benefit to voters.[67]

[63] While the results of town-meeting are not binding, politicians rarely like to be seen opposing public will especially on such an important issues as taxes. An unscientific sample of town meetings reported in *The Philadelphia Inquirer* from 1989-1992 generally supports this view

[64] Consider the Harberger formulation of deadweight loss.

[65] We abstract from different preferences for exempt and non-exempt voters which drive the interest-group story of the main text.

[66] The noise derives from economic shocks.

[67] This assumption is innocuous. There can be a public good, produced with a lag, fully financed at the low tax rate. Or we can assume that the voters do not discern any additional benefit from spending beyond a threshold as in the empirical paper. The point is to preclude the voters from inferring anything about politician type before the meeting.

Next, at a town meeting the government proposes a wage tax. The voter can veto or approve the plan; the politician then either implements his program or the current tax scheme stays in place (he always gets to stay in office).[68] Finally, both players receive their utility payment.[69]

Let the government's first period rate be $\tau_1 \in \{L, H\}$ where L indicates low and H high taxes. Voters pay a noisy version of this value, ξ: $\tau_1 \rightarrow \tilde{\tau_1} \in \{L, H\}$, with $\Pr[\xi(\tau) = \tau] \rightarrow 1$, i.e. the actual rate tends to be the observed rate; notice that the mapping is independent of the politician type. The politician action space is a pair of tax choices $[\tau_1(\theta, q, \xi), \tau_2(\theta, q, \xi)] \in \{L, H\} \times \{L, H\}$; his second period action is a measure of the aggregate tax burden with both a wage and property tax. Voters must decide whether to accept the plan contingent on the observed history $(\tilde{\tau_1})$, politician strategy (τ), the noise map (ξ) and prior (q), $v(\tilde{\tau_1}|\tau, \xi, q) \in \{0, 1\}$ with 1 an approval of the wage tax. To ease exposition, we restrict both players to pure-strategies. Voters update the priors on θ following the first period using Bayes rule; let $\chi(\tilde{\tau_1}|\tau, \xi, q)$ be the posterior of a good type.

The voter has time-separable preferences over taxes and government-induced distortions:

$$U_{Voter}(\tau, v) = V(\tilde{\tau_1}) + [1 - \mathcal{I}(v = 1)]V(\tilde{\tau_1}) + \mathcal{I}(v = 1)D(\tau_2) \tag{1.11}$$

where $V(\tilde{\tau})$ is the utility from the actualized property tax rate $\tilde{\tau}$ and $D(\tau_2)$ is the second period utility from both a wage and property tax at total rate τ_2. $\mathcal{I}(v = 1)$ is an indicator variable equal to one if the voter approves the wage tax.[70] Voters always prefer low taxes and only want the wage tax if the aggregate rate is low,[71]

[68] The latter assumption is innocuous but simplifies the analysis.

[69] It is easiest to interpret the second period payoffs as a discounted flow from the infinite sequence of all future periods. This allows us to ignore the small noise component, see below.

[70] If the plan is rejected, the *actualized* rate, $\tilde{\tau_1}$, is kept in place.

[71] One interpretation is that with a wage tax in place both the maximum tax burden and the potential efficiency gain is larger.

$$D(L) > V(L) > V(H) > D(H)$$

Politician utility depends on type. Good politicians maximize voter welfare,

$$U^{\theta=G}_{Politician}(\tau, v) = U_{Voter}(\tau, v) \tag{1.12}$$

and so prefers to implement the lowest level of taxes.

The bad type of politician gets utility only from collecting rents.

$$U^{\theta=B}_{Politician}(\tau, v) = P(\tilde{\tau}_1) + [1 - \mathcal{I}(v = 1)]P(\tilde{\tau}_1) + \mathcal{I}(v = 1)W(\tau_2) \tag{1.13}$$

where $P(\tau)$ is the surplus from property tax and $W(\tau)$ from wage and property taxes. Property tax surplus increases with total tax revenue,

$$P(H) > P(L)$$

while for the high aggregate rate more rents are collected under the wage tax,

$$W(H) > P(H)$$

Notice that both types of politicians want the EIT, but for different reasons. The social maximizer wants to lower the total distortion following the "rate-squared" rule while the self-interested politician wants an opportunity to skim more money from government coffers.

To close the model, our equilibrium concept is *perfect Bayesian:* beliefs are updated according to Bayes law whenever possible and strategies must be utility maximizing with respect to these beliefs (a Bayesian equilibrium at each time with respect to any possible history). Fudenberg-Tirole [23] state that for the class of games including this one, perfect Bayesian and the more restrictive *sequential equilibrium* concept coincide (Theorem 8.2).

1.8.2 Equilibrium Results

Good politicians always set taxes at the low level, $\tau_t = L\ \forall t$. In the second period rent-seekers will always set the high rate since there is no threat of punishment. However in the first period they must consider a trade-off: while they prefer higher taxes (and more rent), this may signal their type to the voter (a "separating" equilibrium) who will deny them the wage tax instrument, the richest source of surplus. Instead they may try to pretend to be like the honest type and set $\tau_1 = L$ (a "pooling" equilibrium) to enhance the chance of getting a wage tax.[72] We will see that the system noise rules out the pooling equilibria meaning the bad type sets a high rate in the first period.

First consider the equilibrium conditions (second period actions are suppressed):

	Separating Equilibrium	**Pooling Equilibrium**
Politician	$\tau_1(\theta = G) = L$, $\tau_1(\theta = B) = H$	$\tau_1(\theta = G) = \tau_1(\theta = B) = L$
Beliefs	$\chi(\tilde{\tau}_1 = L) \to 1$, $\chi(\tilde{\tau}_1 = L) \to 0$	$\chi(\tilde{\tau}_1 = L) = \chi(\tilde{\tau}_1 = H) = q$
Voter	$v(\tilde{\tau}_1 = L) = 1$, $v(\tilde{\tau}_1 = H) = 0$	$v(\tilde{\tau}_1 = L) = v(\tilde{\tau}_1 = H) \in \{0, 1\}$

In words, in the separating equilibrium only the bad type sets high taxes, the voter updates his beliefs accordingly and so only let low taxers have the income tax. Under the pooling equilibrium, both types initially set low taxes, and any observed high rate is attributed to noise; the signal is completely uninformative so the posterior is the same as the prior (Bayes). Thus, regardless of the observed rate, the voter will approve the EIT only when he was initially confident enough that the politician is honest.[73] No other pure-strategy equilibria is possible since this would require either

[72] Ruling out randomizations precludes hybrid equilibria. There are no qualititative changes if we allow the voter to mix.

[73] Formally,

the good type to set high taxes (violates individual rationality) or the public to vote for high taxes (violates updating rule or utility maximization).[74]

We now turn to our first result:

Lemma 2 *There cannot be a pooling equilibrium.*

PROOF: Focus on the bad politician. In this equilibrium the voter ignores the observed rate and approves or denies the new tax based simply on his prior information (see footnote above). Thus the politician's first period action simply influence his return from that period. Since the actual rate tends to be the observed rate ($\Pr[\xi(\tau) = \tau] \to 1$), the politician does strictly better ($P(H) > P(L)$) by setting a high rate in the first period. This profitable deviation precludes a pooling equilibrium.

□

This result is quite intuitive. The only reason a rent-seeker would set a low rate is if this was the only way to obtain the wage tax. In the pooling equilibrium, the first period rate is irrelevant to the voter decision, so the politician may as well go ahead and try to extract a large rent.

There is no such problem with the separating equilibrium. Clearly the good politician will never set high taxes (lowers social welfare) while the voter cannot

---

$$V(\tilde{\tau}_1|\tau^{pool}, \xi, q) = \begin{cases} 1 & \text{if } q \geq \overline{q} \\ 0 & \text{otherwise} \end{cases}$$

where $\overline{q}$ is the minimal *ex ante* probability of a good type required to approve of the EIT. Notice that the voting strategy does not depend on either the observed action or the functional form of the noise.

[74]Under mixed strategies, voters could randomize over low taxes (and reject high taxers) or accept low taxers (and mix over high taxers). This could be compatible with separating, pooling or hybrid equilibria, see footnote below.

profitably deviate so long as the noise component is relatively small.[75] As for the bad politician, he will abide by the equilibrium strategy so long as

$$\begin{array}{rccc} & U^{\theta=B}_{Politician}(\tau_1 = H, v^*) & \geq & U^{\theta=B}_{Politician}(\tau_1 = L, v^*) \\ \Rightarrow & 2P(H) & \geq & P(L) + W(H) \end{array}$$

where $v^*(\tilde{\tau}_1)$ is the voter strategy. In words, the rent-seeker will not deviate so long as the surplus under both taxes is not too much larger than under the property tax alone. There is nothing which precludes this condition.[76]

Lemma 3 *There may be a separating equilibrium.*

Thus only separating equilibria are possible under pure strategies meaning politicians are not gaming on their superior information. This has two main implications:

Proposition 4 *The wage tax will only be granted when the observed property tax rate is low. Moreover, rent-seeking politicians will set high rates, so in general only good types ("credible") will have the EIT power.*

Both of these points are empirically verifiable. We can test the first claim by seeing if high rates deter taxing propensity; this is verified in the hazard results. As to which type gets the new taxing authority, we can see how property tax rates respond to an EIT. This is presented in the test of the credibility proxies.

Finally, a simple reinterpretation of the model gives another testable implication.

Proposition 5 *Let the government unilaterally implement the wage tax, but now a no vote at the town-meeting ousts the politician from office. Then only politicians in high tax communities should be voted out of office in greater than usual numbers following a wage tax.*

---

[75]This means the voter is very confident that low taxes indicate a good type and high rates mean a bad type.

[76]When the condition is violated, the only equilbrium involves randomized strategies: the voter sets either (i) $v(\tilde{\tau}_1 = L) = 1$, $v(\tilde{\tau}_1 = H)$ mixed; or (ii) $v(\tilde{\tau}_1 = L)$ mixed, $v(\tilde{\tau}_1 = H) = 0$ while $\theta = B$ mixes in the first period. This does not violate the qualitative results presented next.

A check on this is to see how politician turnover varies with expanded tax powers and the prior tax record. Favorable results are given in the proxies section of the empirical paper.

1.9 Appendix: Commuting Data

A critical variable for this analysis is the percent of workers who face an income tax at their workplace. Those commuters whose home do not levy an EIT are obligated to pay at work, so being surrounded by taxers lowers the additional burden of a wage tax.

To compute this variable an annual matrix of taxing and commuter flow data is needed. The second must include the fraction of citizens working at all possible destinations and so is quite large. For 1990 and 1980, when commuting data is available in the Urban Transportation Planning Package, this matrix regularly exceeds 10,000 rows (on average more than 50 destinations per community). For 1960 and 1970 such detailed information is unavailable. However, MCD-to-*county* flows are given in the Census "Journey-to-Work" file. Recall that there are five Pennsylvania counties in the Philadelphia SMSA one of which is the city itself. This information will be used to estimate the commuting matrix using a simplistic costly travel model.

Imagine that transportation costs are uniform for any location within a given radius, but more jobs are available in some places. Abstract away from all issues of job quality and wages, and assume employment location decisions are optimal but stochastic. So citizens prefer to work at home, but then are indifferent between all jobs up to five miles from the MCD centroid and also between companies from five to ten miles away. Workers distribute to locations within each of the bands based on the relative levels of employment.

Under these assumptions it is easy to use the information in the county flows to estimate the employment matrix for 1960 and 1970.[77] Define y_{ict} as the fraction of citizens in MCD i who work in county c in year t. Let d_{cj} be an indicator variable if community j is in county c, and e_{jt} the employment level for community j in period t.

[77] What drives the estimation procedure is variation across communities in the number of contiguous counties. Border municipalities have more neighbors in other counties than do places near the county center.

Then with $r5_{ij}$ and $r10_{ij}$ dummies for communities within 5 and 10 miles respectively of community i and $r0_{ij}$ for the origin,

$$\widetilde{x0_{itc}} \equiv \sum_{j=1}^{N} d_{cj} r0_{ij} e_{jt} \quad \widetilde{x5_{itc}} \equiv \sum_{j=1}^{N} d_{cj} r5_{ij} e_{jt} \quad \widetilde{x10_{itc}} \equiv \sum_{j=1}^{N} d_{cj} r10_{ij} e_{jt} \tag{1.14}$$

with N is the number of possible destinations (remember that some people work outside the Pennsylvania portion of the SMSA so N is strictly larger than the number of suburbs). These values are county specific totals of employment for MCDs within the given range from origin i.

Finally, sum all variables related to a given community to get the number of jobs located within a ten mile radius, and divide into each vector to get the employment fraction in each neighboring county. Call these vectors $x0_{itc}$, $x5_{itc}$ and $x10_{itc}$ respectively.

The regression is

$$y_{ijt} = \alpha_t + \beta_{0t} x0_{ijt} + \beta_{5t} x5_{ijt} + \beta_{10t} x10_{ijt} + \epsilon_{ijt} \tag{1.15}$$

where ϵ_{ijt} is the error term. Again remember that y_{ijt} is observed while each of the regressors is synthesized from known employment data. The parameters are to be estimated with OLS though presumably $\beta_0 > \beta_5 > \beta_{10}$ for each year from the costly transportation supposition. The actual estimated parameters for 1960 and 1970 are listed in the table below with β_0 normalized to one to ease comparison.

parameter	1960	1970
$\widehat{\beta_0}$	1.00	1.00
$\widehat{\beta_5}$	0.62	0.73
$\widehat{\beta_{10}}$	0.33	0.24

The final step is to assign workers to MCD destinations. One simply uses the estimated parameters and relative employment value to assign the relevant weight; renormalization to ensure the values sum to one is necessary. For example, the fraction of citizens living in origin i and working in destination j within ten miles is $k_{it}\widehat{\beta_{10t}}e_{jt}$ where k_{it} is the renormalization factor and $\widehat{\beta_{10t}}$ the estimated parameter. To test the algorithm, the same matrix was estimated using 1980 county flow data. Over 85% of the predicted destinations were correct.

This matrix then can be multiplied by the known set of EIT dummies to obtain an estimate of the fraction of workers who face a wage tax at work. I chose to break out both Philadelphia and out-of-state workers due their special status under the tax law. Hence neither of these are counted in the fraction taxed. For non-SMSA Pennsylvania locations, however, the appropriate tax dummy based on [43] was used.

Finally some data notes. First, for non-SMSA destinations a modification of the Census estimate of employment was used. In addition, the Philadelphia job figure was scaled down since the city is so large (many jobs are inaccessible from any given suburban location). Also, since exact data is available for Philadelphia commuters, I used this rather than the estimated value; again this requires renormalizing the numbers for each MCD origin. Finally, for years up through 1980 a category "place of work unreported" was included. Here these commuters were assumed to be a representative sample, and so were simply removed from the normalization factors. For 1990 the Census itself used an incompletely documented algorithm to assign such workers to destinations.

Bibliography

[1] Reassessment: Poison or Panacea? The Philadelphia Inquirer, 11 March 1990. p. H4.

[2] Wage Tax Approval Sets off Uproar. The Philadelphia Inquirer, 21 January 1990. p. 4H-26H.

[3] Two Suburbs, A School Tax and Two Attitudes About It. The New York Times, 25 August 1993. p. A13.

[4] BERGSTROM, THEODORE C. AND ROBERT P. GOODMAN. Private Demands for Public Goods. *The American Economic Review*, 63(3):280–296, 1973.

[5] BLACKLEY, PAUL R. AND LARRY DEBOER. Tax Base Choice by Local Governments. *Land Economics*, 63(3):227–236, 1987.

[6] BRADBURY, KATHERINE L. Can Local Governments Give Citizens What They Want? Referendum Outcomes in Massachusetts. *New England Economic Review*, pages 3–22, May/June 1991.

[7] BRENNAN, GEOFFREY AND JAMES BUCHANAN. Towards a Tax Constitution for Leviathan. *Journal of Public Economics*, 8:255–273, 1977.

[8] BRENNAN, GEOFFREY AND JAMES BUCHANAN. *The Power to Tax: Analytical Foundations of a Fiscal Constitution.* Cambridge University Press, Cambridge, 1980.

[9] BUREAU OF THE CENSUS, Washington, DC. *Census of Population and Housing: Summary Characteristics for Pennsylvania [Census Tracts]*, 1960-1990.

[10] BUREAU OF THE CENSUS, Washington, DC. *Census of Population: Characteristics of the Population for Pennsylvania*, 1960-1990.

[11] BUREAU OF TRANSPORTATION STATISTICS (DEPT. TRANSPORTATION), Washington D.C. *Census Transportation Planning Package*, 1980 and 1990.

[12] CITRIN, JACK. Do People Want Something for Nothing: Public Opinion on Taxes and Government Spending. *National Tax Journal*, 32(2):113–129, 1979.

[13] COURANT, PAUL N., EDWARD M. GRAMLICH AND DANIEL L. RUBINFELD. Why Voters Support Tax Limitation Amendments: The Michigan Case. *National Tax Journal*, 33(1):1–20, 1980.

[14] DELAWARE VALLEY REGIONAL PLANNING COMMISSION, Philadelphia. *Demographic and Employment Data for Transportation Planning: Summaries by Minor Civil Divisions*, 1979.

[15] DELAWARE VALLEY REGIONAL PLANNING COMMISSION, Philadelphia. *1980 Data Bank for Transportation Planning and Travel Simulation*, 1984.

[16] DELAWARE VALLEY REGIONAL PLANNING COMMISSION, Philadelphia. *Land Use in the Delaware Valley: 1970 and 1980 Municipal Data in 12 Categories*, 1984.

[17] DELAWARE VALLEY REGIONAL PLANNING COMMISSION, Philadelphia. *Year 2020 County and Municipal Interim Population and Employment Forecasts*, 1993.

[18] DELAWARE VALLEY REGIONAL PLANNING COMMISSION, Philadelphia. *Delaware Valley Data: 1990 Land Use Information*, 1994.

[19] DEPARTMENT OF GENERAL SERVICES, Harrisburg. *The Pennsylvania Manual*, 1978-1993.

[20] DIAMOND, PETER A. AND JERRY A. HAUSMAN. The Retirement and Unemployment Behavior of Older Men. In H. J. Aaron and G. Burtless, editors, *Retirement and Economic Behavior*, chapter 4, pages 97–134. The Brookings Institution, Washington D.C., 1984.

[21] FISHER, RONALD C. *State and Local Public Finance.* Scott, Foresman Company, Glenview, IL, 1988.

[22] FLINN, C. J. AND J. J. HECKMAN. Models for the Analysis of Labor Force Dynamics. *Advances in Econometrics*, 1(1):35–95, 1982.

[23] FUDENBERG, DREW AND JEAN TIROLE. *Game Theory.* The MIT Press, Cambridge MA, 1992.

[24] HETTICH, WALTER AND STANLEY WINER. A Positive Model of Tax Structure. *Journal of Public Economics*, 24:67–87, 1984.

[25] INMAN, ROBERT P. Testing Political Economy's 'as if' Proposition: is the Median Income Voter Really Decisive? *Public Choice*, 33(4):45–65, 1978.

[26] INMAN, ROBERT P. The Local Decision to Tax: Evidence from Large U.S. Cities. *Regional Science and Urban Economics*, 19:455–491, 1989.

[27] KALBFLEISCH, JOHN D. AND ROSS L. PRENTICE. *The Statistical Analysis of Failure Time Data.* John Wiley and Sons, New York, 1980.

[28] KIEFER, NICHOLAS M. Econometric Duration Data and Hazard Functions. *Journal of Economic Literature*, 26(2):646–679, 1988.

[29] KIEFER, NICHOLAS M. AND GARY SKOOG. Local Asymptotic Specification Error Analysis. *Econometrica*, 52(4):873–885, 1984.

[30] LADD, HELEN F. AND T. NICOLAS TIDEMAN. *Tax and Expenditure Limitations.* The Urban Institute Press, Washington D.C., 1981.

[31] LADD, HELEN F. AND JULIE B. WILSON. Why Voters Support Tax Limitations: Evidence from Massachusetts' Proposition $2\frac{1}{2}$. *National Tax Journal*, 35(2):121–148, 1982.

[32] LEVY, MICKEY D. A Comparison of Voting Behavior on Tax Initiatives with Different Perceived Distribution Consequences. *National Tax Journal*, 32(4):551–554, 1979.

[33] LUCE, THOMAS F. AND ANITA A. SUMMERS. *Local Fiscal Issues in the Philadelphia Metropolitan Area.* University of Pennsylvania Press, Philadelphia, 1987.

[34] LUCIER, RICHARD L. The Oregon Tax Subsititution Referendum: The Prediction of Voting Behavior. *National Tax Journal*, 24(1):87–90, 1971.

[35] LUCIER, RICHARD L. The Prediction of Public Choice Behavior in the Washington Tax Substitution Referendum. *National Tax Journal*, 26(4):625–630, 1973.

[36] METCALF, GILBERT E. Tax Exporting, Federal Deductibiliy, and State Tax Structure. *Journal of Policy Analysis and Management*, 12(1):109–126, 1993.

[37] MIKESELL, JOHN L. Local Legislative Behavior in Tax Substitution. *Quarterly Review of Economics and Business*, 14:83–91, 1974.

[38] NORSTRAND, ROLF. The Choice Between Income Tax and Land Tax in Danish Municipalities. *Public Finance*, 35(3):412–424, 1980.

[39] OAKLAND, WILLIAM H. Proposition 13: Genesis and Consequences. *National Tax Journal*, 32(2):387–409, 1979.

[40] PELTZMAN, SAM. Voters As Fiscal Conservatives. *The Quarterly Journal of Economics*, 107(2):327–361, 1992.

[41] PENNSYLVANIA DEPARTMENT OF COMMUNITY AFFAIRS, Harrisburg. *Citizen's Guide to Pennsylvania Local Government*, 1991.

[42] PENNSYLVANIA DEPARTMENT OF COMMUNITY AFFAIRS, Harrisburg. *Taxation Manual*, 1993.

[43] PENNSYLVANIA DEPARTMENT OF COMMUNITY AFFAIRS, Harrisburg. *Local Government Financial Statistics*, various years.

[44] PENNSYLVANIA DEPARTMENT OF COMMUNITY AFFAIRS, Harrisburg. *Register for Earned Income Taxes*, various years.

[45] PENNSYLVANIA DEPARTMENT OF INTERNAL AFFAIRS, Harrisburg. *County Industry Report [Census of Manufacturing Industries]*, 1961.

[46] PERSSON, TORSTEN AND GUIDO TABELLINI. *Monetary and Fiscal Policy. Volume 1: Credibility.* The MIT Press, Cambridge MA, 1994.

[47] THE PHILADELPHIA BULLETIN ALMANAC, Philadelphia. *The Evening and Sunday Bulletin*, 1960-1975.

[48] POMMEREHNE, WERNER W. AND FRIEDRICH SCHNEIDER. Fiscal Illusion, Political Institutions, and Local Public Spending. *Kyklos*, 31(3):381–408, 1978.

[49] ROGOFF, KENNETH AND ANNE SIBERT. Election and Macroeconomic Policy Cycles. *Review of Economic Studies*, 55(1):1–16, 1988.

[50] ROMER, THOMAS AND HOWARD ROSENTHAL. Bureaucrats versus Voters: On the Political Economy of Resource Allocation by Direct Democracy. *Quarterly Journal of Economics*, 93:563–587, 1979.

[51] SEARS, DAVID O. AND JACK CITRIN. *Tax Revolt: Something for Nothing in California*. Harvard University Press, Cambridge, 1982.

[52] SJOQUIST, DAVID L. A Median Voter Analysis of Variations in the Use of Property Taxes Among Local Governments. *Public Choice*, 36:273–285, 1981.

[53] STEIN, ROBERT M., KEITH E. HAMM AND PATRICIA K. FREEMAN. An Analysis of Support for Tax Limitation Referenda. *Public Choice*, 40:187–194, 1983.

[54] STRUMPF, KOLEMAN S. A Predictive Index for the Flypaper Effect. Xerox, MIT Economics Department, 1995.

Regressor	(a)	(b)	(c)
constant	-39.712	-47.591	-39.001
	(-1.41)	(-1.50)	(-1.38)
$\mathcal{I}$(EIT First Year)	0.359	0.292	0.523
	(0.22)	(0.16)	(0.31)
Relative Tax Effort(-1)	1.976	—	—
	(1.45)		
$\mathcal{I}$(EIT First) × Relative Tax Effort(-1)	4.526	—	—
	(1.94)		
ΔTax Effort(-1)	—	61.232	—
		(2.28)	
$\mathcal{I}$(EIT First) × ΔTax Effort(-1)	—	230.184	—
		(2.46)	
Government Overhead(-1)	—	—	0.415
			(0.40)
$\mathcal{I}$(EIT First) × Government Overhead(-1)	—	—	9.395
			(2.45)
% Philadelphia Commuter	-0.121	-0.247	-0.199
	(-0.22)	(-0.42)	(-0.37)
% Out-State Commuter	-0.466	-0.736	-0.494
	(-0.92)	(-1.36)	(-0.97)
% Seniors	-1.932	-2.898	-1.965
	(-1.73)	(-2.43)	(-1.75)
$\frac{\text{Owner}}{\text{Renter}}$ Tenure	0.355	0.134	0.356
	(0.44)	(0.15)	(0.44)
% Face EIT at Work	-0.210	-0.342	-0.263
	(-0.53)	(-0.81)	(-0.66)
% Deficit(-1)	0.099	0.121	0.115
	(0.33)	(0.57)	(0.43)
Median HH Income [$\times 10^3$ $]	2.929	3.807	2.879
	(1.00)	(1.14)	(0.98)
Population [$\times 10^3$]	1.954	1.309	1.510
	(1.08)	(0.69)	(0.87)
Population Growth	-0.235	-0.378	-0.221
	(-1.10)	(-1.67)	(-1.03)
Employment [$\times 10^3$]	0.602	0.563	0.616
	(0.81)	(0.70)	(0.83)
Market Value [$\times 10^3$ $]	-0.997	-0.333	-0.651
	(-0.50)	(-0.15)	(-0.33)
City	5.364	5.475	5.081
	(2.04)	(2.08)	(2.03)
Borough	6.967	6.848	6.543
	(4.37)	(4.62)	(4.51)
N	1009	929	1009
R^2	0.101	0.107	0.099

Table 1.7: Politician Turnover: A Test for Credibility Proxies

(a): Relative tax effort [level]. (b): Δtax effort [first-difference]. (c): Overhead expenditure (residual). Sample: biannual 1962-1992 with 86 MCDs. All terms in logs except dummies, population growth, change in tax effort, government overhead. All regressions include county and period dummies.

(t-statistics)

Regressor	(a)	(b)	(c)
constant	-1.532	-1.434	-1.543
	(-4.16)	(-3.90)	(-4.19)
$\mathcal{I}$(EIT First Year)	-0.311	-0.290	-0.299
	(-4.89)	(-4.28)	(-4.86)
Relative Tax Effort(-1)	0.049	—	—
	(2.08)		
$\mathcal{I}$(EIT First) × Relative Tax Effort(-1)	0.223	—	—
	(2.02)		
ΔTax Effort(-1)	—	3.980	—
		(7.14)	
$\mathcal{I}$(EIT First) × ΔTax Effort(-1)	—	9.629	—
		(3.21)	
Government Overhead(-1)	—	—	0.036
			(1.61)
$\mathcal{I}$(EIT First) × Government Overhead(-1)	—	—	0.518
			(3.72)
% Philadelphia Commuter	0.022	0.021	0.019
	(1.68)	(1.61)	(1.48)
% Out-State Commuter	0.023	0.023	0.022
	(2.32)	(2.40)	(2.29)
% Seniors	0.029	0.024	0.030
	(0.99)	(0.85)	(1.02)
$\frac{\text{Owner}}{\text{Renter}}$ Tenure	-0.011	-0.008	-0.012
	(-0.55)	(-0.44)	(-0.62)
% Face EIT at Work	0.008	0.007	0.006
	(0.80)	(0.77)	(0.64)
% Deficit(-1)	0.043	0.037	0.041
	(1.12)	(1.01)	(1.13)
Median HH Income [$\times 10^3$ \$]	0.219	0.210	0.224
	(4.49)	(4.38)	(4.65)
Population [$\times 10^3$]	0.243	0.232	0.242
	(9.44)	(9.11)	(9.50)
Population Growth	0.001	0.002	0.001
	(0.26)	(0.52)	(0.22)
Employment [$\times 10^3$]	-0.047	-0.061	-0.055
	(-1.83)	(-2.43)	(-1.97)
Market Value [$\times 10^3$ \$]	-0.242	-0.233	-0.244
	(-9.86)	(-9.50)	(-9.95)
City	0.178	0.165	0.154
	(1.78)	(1.68)	(1.56)
Borough	-0.037	-0.043	-0.042
	(-1.19)	(-1.62)	(-1.57)
N	7579	7579	7579
R^2	0.136	0.142	0.116

Table 1.8: Change in Property Tax Rates: A Test for Credibility Proxies

(a): Relative tax effort [level]. (b): Δtax effort [first-difference]. (c): Overhead expenditure (residual). Sample: annual 1960-1992 with 237 MCDs. All terms in logs except dummies, population growth, change in tax effort, government overhead. All regressions include county and period dummies.

(t-statistics)

Regressor	Benchmark 1970	Benchmark 1980	Benchmark 1992	Tax Effort(-1) 1970	Tax Effort(-1) 1980	Tax Effort(-1) 1992	Property Tax(-1) 1970	Property Tax(-1) 1980	Property Tax(-1) 1992
constant	1.085	-2.533	-8.538	-1.354	-3.855	-11.646	1.456	-1.699	-8.065
	(0.88)	(-1.82)	(-3.67)	(-0.78)	(-2.54)	(-3.59)	(1.17)	(-0.96)	(-3.30)
Relative Tax	—	—	—	-1.416	-1.782	-2.040	—	—	—
Effort(-1)				(-1.29)	(-2.75)	(-2.68)			
Relative Property	—	—	—	—	—	—	-0.875	-3.060	-2.431
Tax(-1)							(-1.83)	(-3.85)	(-3.11)
% Philadelphia	-0.041	-0.068	-0.065	0.202	0.046	0.106	0.141	0.065	0.091
Commuter	(-0.63)	(-1.93)	(-1.94)	(2.40)	(1.40)	(2.63)	(2.29)	(1.71)	(2.40)
% Out-of-State	-0.034	-0.033	-0.48	0.109	0.036	0.037	0.082	0.054	0.029
Commuter	(-1.42)	(-1.95)	(-2.02)	(2.22)	(2.22)	(1.39)	(2.70)	(2.59)	(1.57)
% Seniors	0.016	-0.033	-0.036	0.033	0.054	0.088	0.021	0.034	0.065
	(0.30)	(-0.62)	(-0.61)	(0.48)	(1.23)	(1.84)	(0.40)	(0.90)	(1.43)
$\frac{\text{Owner}}{\text{Renter}}$ Tenure	-0.092	-0.030	0.134	-0.076	-0.063	0.148	-0.109	-0.012	0.135
	(-0.85)	(-0.44)	(1.39)	(-0.48)	(-0.75)	(1.23)	(-0.99)	(-0.16)	(1.40)
% Face EIT at Work	0.005	0.086	0.184	-0.008	0.083	0.205	0.002	0.094	0.181
	(0.66)	(6.19)	(5.48)	(-0.74)	(5.71)	(4.69)	(0.26)	(5.19)	(5.30)
% Deficit(-1)	0.002	0.005	0.001	0.002	0.004	0.001	0.001	0.003	0.000
	(0.23)	(0.42)	(0.09)	(0.21)	(0.35)	(0.08)	(0.16)	(0.29)	(0.03)
Median HH Income	-0.051	-0.036	-0.017	-0.050	-0.008	-0.002	-0.022	-0.077	-0.020
[$\times 10^3$ $]	(-0.51)	(-0.95)	(-0.77)	(-0.35)	(-0.20)	(-0.09)	(-0.21)	(-1.69)	(-0.89)
Population [$\times 10^3$]	0.132	-0.012	-0.004	0.019	-0.049	-0.015	0.181	0.100	0.000
	(1.99)	(-0.17)	(-0.06)	(0.24)	(-0.67)	(-0.17)	(2.50)	(1.30)	(0.00)
Population Growth	0.029	0.132	-0.079	0.003	0.142	-0.071	0.015	0.052	-0.089
	(0.45)	(1.58)	(-0.68)	(0.03)	(1.61)	(-0.48)	(0.22)	(0.52)	(-0.75)
Employment [$\times 10^3$]	-0.001	-0.053	0.011	-0.378	-0.052	-0.004	0.030	-0.064	0.007
	(-0.10)	(-0.70)	(0.17)	(-2.72)	(-0.69)	(-0.05)	(0.30)	(-0.70)	(0.10)
City	-0.190	8.127	0.965	-0.747	5.621	0.279	0.436	17.796	2.977
	(-0.13)	(0.00)	(0.00)	(-0.50)	(0.00)	(0.00)	(0.28)	(0.00)	(0.00)
Borough	-0.023	0.750	1.850	-2.178	-0.549	1.976	0.445	2.350	1.926
	(-0.05)	(1.29)	(2.43)	(-3.09)	(-0.69)	(1.97)	(0.90)	(2.93)	(2.50)
County Dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	237	237	236	237	237	236	237	237	236
$\log L$	-63.60	-47.15	-29.71	-41.81	-42.50	-24.01	-51.87	-36.23	-25.50
LR omitted variable	—	—	—	43.58	9.30	11.40	23.46	21.84	8.42

Table 1.9: Probit Model of Income Tax Propensity

Dependent variable: EIT dummy. Years: 1970, 1980, 1992. Sample: 237 MCDs in Philadelphia SMSA.
Panels: Benchmark (1), Credibility Proxies [Levels] (2,3).
Last row: LR test of omitted variable for credibility proxy.

(t-statistics)

Regressor	Benchmark 1970	1980	1992	ΔTax Effort(-1) 1970	1980	1992	ΔProperty Tax(-1) 1970	1980	1992
constant	1.085	-2.533	-8.538	0.955	-2.47	-8.850	1.170	-2.534	-7.895
	(0.88)	(-1.82)	(-3.67)	(0.74)	(-1.76)	(-3.66)	(0.93)	(-1.82)	(-3.52)
ΔTax Effort(-1)	—	—	—	-30.086	-38.870	-41.102	—	—	—
				(-1.56)	(-2.70)	(-2.37)			
ΔProperty Tax(-1)	—	—	—	—	—	—	-0.086	-0.791	-3.921
							(-0.35)	(-1.93)	(-2.05)
% Philadelphia	-0.041	-0.068	-0.065	0.173	0.048	0.070	0.103	0.035	0.065
Commuter	(-0.63)	(-1.93)	(-1.94)	(2.72)	(1.97)	(1.99)	(1.75)	(1.46)	(1.78)
% Out-of-State	-0.034	-0.033	-0.48	0.082	0.033	0.040	0.054	0.033	0.034
Commuter	(-1.42)	(-1.95)	(-2.02)	(2.75)	(1.96)	(1.59)	(1.83)	(1.95)	(1.50)
% Seniors	0.016	-0.033	-0.036	0.016	0.036	0.063	0.015	0.033	0.072
	(0.30)	(-0.62)	(-0.61)	(0.31)	(0.66)	(1.09)	(0.28)	(0.62)	(1.35)
$\frac{\text{Owner}}{\text{Renter}}$ Tenure	-0.092	-0.030	0.134	-0.088	-0.030	0.127	-0.092	-0.030	0.122
	(-0.85)	(-0.44)	(1.39)	(-0.80)	(-0.44)	(1.30)	(-0.85)	(-0.44)	(1.22)
% Face EIT at Work	0.005	0.086	0.184	0.005	0.086	0.191	0.005	0.086	0.180
	(0.66)	(6.19)	(5.48)	(0.66)	(6.17)	(5.35)	(0.66)	(6.16)	(5.39)
% Deficit(-1)	0.002	0.005	0.001	0.003	0.005	0.002	0.002	0.005	0.001
	(0.23)	(0.42)	(0.09)	(0.23)	(0.42)	(0.13)	(0.22)	(0.40)	(0.10)
Median HH Income	-0.051	-0.036	-0.017	-0.026	-0.037	-0.016	-0.056	-0.036	-0.014
[$\times 10^3$ $]	(-0.51)	(-0.95)	(-0.77)	(-0.24)	(-0.97)	(-0.73)	(-0.55)	(-0.94)	(-0.63)
Population [$\times 10^3$]	0.132	-0.012	-0.004	0.137	0.013	-0.004	0.131	0.012	0.007
	(1.99)	(-0.17)	(-0.06)	(2.01)	(0.18)	(-0.05)	(1.95)	(0.17)	(0.10)
Population Growth	0.029	0.132	-0.079	0.038	0.132	-0.087	0.031	0.132	-0.095
	(0.45)	(1.58)	(-0.68)	(0.58)	(1.58)	(-0.73)	(0.48)	(1.57)	(-0.78)
Employment [$\times 10^3$]	-0.001	-0.053	0.011	-0.080	-0.056	0.009	-0.010	-0.053	0.019
	(-0.10)	(-0.70)	(0.17)	(-0.78)	(-0.73)	(0.13)	(-0.11)	(-0.69)	(0.27)
City	-0.190	8.127	0.965	0.739	8.374	1.195	-0.246	7.744	1.355
	(-0.13)	(0.00)	(0.00)	(0.52)	(0.00)	(0.00)	(-0.17)	(0.00)	(0.00)
Borough	-0.023	0.750	1.850	-0.086	0.752	2.050	-0.038	0.747	1.668
	(-0.05)	(1.29)	(2.43)	(-0.20)	(1.29)	(2.68)	(-0.09)	(1.24)	(2.22)
County Dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	237	237	236	237	237	235	237	237	235
$\log L$	-63.60	-47.15	-29.71	-55.34	-43.11	-25.40	-56.54	-41.15	-25.40
LR omitted variable	—	—	—	16.52	8.08	8.62	14.12	12.00	8.62

Table 1.10: Probit Model of Income Tax Propensity

Dependent variable: EIT dummy. Years: 1970, 1980, 1992. Sample: 237 MCDs in Philadelphia SMSA.
Panels: Benchmark (1), Credibility Proxies [First-Difference] (2,3).
Last row: LR test of omitted variable for credibility proxy.

(t-statistics)

Regressor	Benchmark 1970	1980	1992	Government Overhead(-1) 1970	1980	1992
constant	1.085	-2.533	-8.538	1.131	-2.919	-8.719
	(0.88)	(-1.82)	(-3.67)	(0.91)	(-1.91)	(-3.65)
Government Overhead(-1)	—	—	—	-0.201	-0.782	-0.802
				(-0.62)	(-1.84)	(-1.74)
% Philadelphia Commuter	-0.041	-0.068	-0.065	0.158	0.042	0.065
	(-0.63)	(-1.93)	(-1.94)	(2.58)	(1.60)	(1.94)
% Out-of-State Commuter	-0.034	-0.033	-0.48	0.085	0.031	0.023
	(-1.42)	(-1.95)	(-2.02)	(2.86)	(1.81)	(0.96)
% Seniors	0.016	-0.033	-0.036	0.018	0.033	0.060
	(0.30)	(-0.62)	(-0.61)	(0.35)	(0.59)	(1.01)
$\frac{\text{Owner}}{\text{Renter}}$ Tenure	-0.092	-0.030	0.134	0.100	0.055	0.198
	(-0.85)	(-0.44)	(1.39)	(0.91)	(0.81)	(2.05)
% Face EIT at Work	0.005	0.086	0.184	0.005	0.090	0.185
	(0.66)	(6.19)	(5.48)	(0.64)	(5.76)	(5.44)
% Deficit(-1)	0.002	0.005	0.001	0.002	0.006	0.000
	(0.23)	(0.42)	(0.09)	(0.25)	(0.49)	(0.07)
Median HH Income [$\times 10^3$ $]	-0.051	-0.036	-0.017	-0.052	-0.045	-0.017
	(-0.51)	(-0.95)	(-0.77)	(-0.52)	(-1.14)	(-0.80)
Population [$\times 10^3$]	0.132	-0.012	-0.004	0.132	0.013	-0.004
	(1.99)	(-0.17)	(-0.06)	(1.97)	(0.18)	(-0.06)
Population Growth	0.029	0.132	-0.079	0.022	0.160	-0.071
	(0.45)	(1.58)	(-0.68)	(0.33)	(1.79)	(-0.60)
Employment [$\times 10^3$]	-0.001	-0.053	0.011	-0.014	-0.042	0.010
	(-0.10)	(-0.70)	(0.17)	(-0.14)	(-0.53)	(0.16)
City	-0.190	8.127	0.965	-0.338	7.426	0.984
	(-0.13)	(0.00)	(0.00)	(-0.23)	(0.00)	(0.00)
Borough	-0.023	0.750	1.850	0.003	0.872	1.989
	(-0.05)	(1.29)	(2.43)	(0.01)	(1.41)	(2.43)
County Dummies?	Yes	Yes	Yes	Yes	Yes	Yes
N	237	237	236	237	237	236
$\log L$	-63.60	-47.15	-29.71	-58.41	-43.26	-26.59
LR omitted variable	—	—	—	10.38	7.78	6.24

Table 1.11: Probit Model of Income Tax Propensity

Dependent variable: EIT dummy. Years: 1970, 1980, 1992. Sample: 237 MCDs in Philadelphia SMSA.
Panels: Benchmark (1), Credibility Proxies [Levels] (2).
Last row: LR test of omitted variable for credibility proxy.

(t-statistics)

Regressor	Benchmark 1960-92	Benchmark 1960-75	Benchmark 1976-92	Tax Effort(-1) 1960-92	Tax Effort(-1) 1960-75	Tax Effort(-1) 1976-92	Property Tax(-1) 1960-92	Property Tax(-1) 1960-75	Property Tax(-1) 1976-92
Relative Tax	—	—	—	-0.955	-0.842	-1.698	—	—	—
Effort(-1)				(-4.05)	(-2.54)	(-3.98)			
Relative Property	—	—	—	—	—	—	-0.652	-0.843	-0.789
Tax(-1)							(-2.12)	(-2.02)	(-3.41)
% Philadelphia	-0.046	-0.015	-0.055	0.037	0.043	0.084	0.031	0.037	0.069
Commuter	(-2.40)	(-0.39)	(-2.21)	(1.82)	(1.50)	(2.20)	(1.71)	(1.43)	(1.81)
% Out-of-State	-0.021	-0.023	-0.020	0.019	0.024	0.036	0.019	0.008	0.049
Commuter	(-1.62)	(-0.84)	(-1.55)	(1.65)	(1.99)	(1.56)	(1.45)	(0.57)	(1.71)
% Seniors	-0.055	-0.084	-0.057	0.059	0.042	0.093	0.073	0.050	0.101
	(-1.65)	(-1.63)	(-1.25)	(1.69)	(0.78)	(2.08)	(1.95)	(1.05)	(2.00)
$\frac{\text{Owner}}{\text{Renter}}$ Tenure	-0.024	-0.036	-0.035	-0.018	-0.028	0.003	-0.028	-0.025	-0.033
	(-0.42)	(-0.43)	(-0.39)	(-0.31)	(-0.33)	(0.03)	(-0.50)	(-0.31)	(-0.37)
% Face EIT at	0.051	0.048	0.091	0.047	0.044	0.087	0.051	0.048	0.091
Work	(11.27)	(9.12)	(6.91)	(10.11)	(8.21)	(6.32)	(11.27)	(9.16)	(6.92)
% Deficit(-1)	0.005	0.007	0.002	0.004	0.005	0.002	0.003	0.005	0.000
	(0.67)	(0.88)	(0.19)	(0.57)	(0.65)	(0.21)	(0.47)	(0.61)	(0.11)
Median HH Income	-0.004	-0.023	0.003	0.003	0.002	0.026	-0.006	-0.036	0.007
[$\times 10^3$ \$]	(-0.24)	(-0.35)	(0.12)	(0.15)	(0.02)	(0.97)	(-0.29)	(-0.56)	(0.25)
Population [$\times 10^3$]	0.025	0.044	0.004	0.012	0.027	0.001	0.032	0.082	0.001
	(1.47)	(0.99)	(0.11)	(0.65)	(0.63)	(0.02)	(1.74)	(1.80)	(0.03)
Population Growth	0.030	0.070	-0.050	0.042	0.065	-0.005	0.020	0.038	-0.046
	(0.63)	(1.28)	(-0.49)	(0.87)	(1.19)	(-0.04)	(0.40)	(0.69)	(-0.45)
Employment [$\times 10^3$]	0.004	0.075	-0.002	-0.003	0.031	0.006	0.001	0.072	0.003
	(0.11)	(0.97)	(-0.04)	(-0.09)	(0.40)	(0.13)	(0.04)	(1.03)	(0.06)
City	1.319	1.061	—	0.221	0.323	—	1.707	2.211	—
	(1.61)	(1.20)		(0.26)	(0.35)		(1.90)	(2.18)	
Borough	0.815	0.647	0.541	0.122	0.118	0.182	0.998	1.265	0.520
	(3.11)	(1.91)	(0.93)	(0.40)	(0.30)	(0.28)	(3.22)	(3.09)	(0.89)
County Dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	7607	3582	4025	7607	3582	4025	7607	3582	4025
$\log L$	-564.52	-384.72	-158.88	-550.22	-379.91	-151.67	-555.93	-381.70	-154.62
LR omitted variable	—	—	—	28.60	9.62	14.42	17.18	6.04	8.52

Table 1.12: Hazard Model of Income Tax Propensity

Cox Partial Likelihood of Proportional Hazards with Time-Varying Regressors.

Dependent variable: EIT dummy. Years: 1960-1992, 1960-1975, 1976-1992. Sample: 237 MCDs in Philadelphia SMSA.

Panels: Benchmark (1), Credibility Proxies [Levels] (2,3).

1976-1992: *Cities* omitted since both are left-censored. Last row: LR test of omitted variable for credibility proxy.

(t-statistics)

Regressor	Benchmark 1960-92	Benchmark 1960-75	Benchmark 1976-92	ΔTax Effort(-1) 1960-92	ΔTax Effort(-1) 1960-75	ΔTax Effort(-1) 1976-92	ΔProperty Tax(-1) 1960-92	ΔProperty Tax(-1) 1960-75	ΔProperty Tax(-1) 1976-92
ΔTax Effort(-1)	—	—	—	-23.711	-13.666	-38.647	—	—	—
				(-6.55)	(-2.94)	(-3.82)			
ΔProperty Tax(-1)	—	—	—	—	—	—	-0.720	-0.540	-0.900
							(-5.46)	(-2.92)	(-4.85)
% Philadelphia	-0.046	-0.015	-0.055	0.041	0.046	0.068	0.055	0.046	0.056
Commuter	(-2.40)	(-0.39)	(-2.21)	(2.16)	(1.59)	(1.99)	(2.66)	(1.58)	(1.66)
% Out-of-State	-0.021	-0.023	-0.020	0.025	0.021	0.040	0.024	0.008	0.045
Commuter	(-1.62)	(-0.84)	(-1.55)	(1.82)	(1.43)	(1.32)	(1.79)	(0.58)	(1.54)
% Seniors	-0.055	-0.084	-0.057	0.071	0.073	0.074	0.056	0.096	0.083
	(-1.65)	(-1.63)	(-1.25)	(1.80)	(1.41)	(1.68)	(1.62)	(1.82)	(1.83)
$\frac{\text{Owner}}{\text{Renter}}$ Tenure	-0.024	-0.036	-0.035	-0.031	-0.037	-0.022	-0.030	-0.042	-0.050
	(-0.42)	(-0.43)	(-0.39)	(-0.54)	(-0.44)	(-0.25)	(-0.53)	(-0.51)	(-0.56)
% Face EIT at	0.051	0.048	0.091	0.047	0.043	0.075	0.047	0.045	0.085
Work	(11.27)	(9.12)	(6.91)	(10.15)	(8.04)	(5.38)	(10.26)	(8.49)	(6.45)
% Deficit(-1)	0.005	0.007	0.002	0.005	0.006	0.002	0.005	0.007	0.002
	(0.67)	(0.88)	(0.19)	(0.65)	(0.79)	(0.20)	(0.66)	(0.85)	(0.22)
Median HH Income	-0.004	-0.023	0.003	-0.005	0.007	0.002	0.001	-0.013	0.013
[$\times 10^3$ \$]	(-0.24)	(-0.35)	(0.12)	(-0.24)	(0.11)	(0.06)	(0.07)	(-0.20)	(0.50)
Population [$\times 10^3$]	0.025	0.044	0.003	0.030	0.053	0.016	0.021	0.030	0.005
	(1.47)	(0.99)	(0.10)	(1.68)	(1.15)	(0.49)	(1.22)	(0.67)	(0.16)
Population Growth	0.030	0.070	-0.050	0.049	0.064	0.009	0.016	0.050	-0.053
	(0.63)	(1.28)	(-0.48)	(1.04)	(1.19)	(0.08)	(0.32)	(0.89)	(-0.51)
Employment [$\times 10^3$]	0.004	0.075	-0.002	-0.005	0.081	-0.044	0.010	0.091	0.005
	(0.11)	(0.97)	(-0.04)	(-0.15)	(1.01)	(-0.90)	(0.29)	(1.14)	(0.11)
City	1.319	1.061	—	0.953	0.688	—	1.182	0.964	—
	(1.61)	(1.20)		(1.15)	(0.77)		(1.37)	(1.03)	
Borough	0.815	0.647	0.540	0.564	0.449	0.292	0.665	0.519	0.571
	(3.11)	(1.91)	(0.93)	(2.09)	(1.28)	(0.49)	(2.53)	(1.55)	(0.97)
County Dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	7390	3366	4024	7390	3366	4024	7390	3366	4024
$\log L$	-564.52	-384.72	-158.87	-548.88	-377.44	-153.15	-551.13	-377.94	-151.26
LR omitted variable	—	—	—	31.28	14.56	11.44	26.78	13.56	15.22

Table 1.13: Hazard Model of Income Tax Propensity

Cox Partial Likelihood of Proportional Hazards with Time-Varying Regressors.
Dependent variable: EIT dummy. Years: 1960-1992, 1960-1975, 1976-1992. Sample: 237 MCDs in Philadelphia SMSA.
Panels: Benchmark (1), Credibility Proxies [First-Difference] (2,3).
1976-1992: *Cities* omitted since both are left-censored. Last row: LR test of omitted variable for credibility proxy.

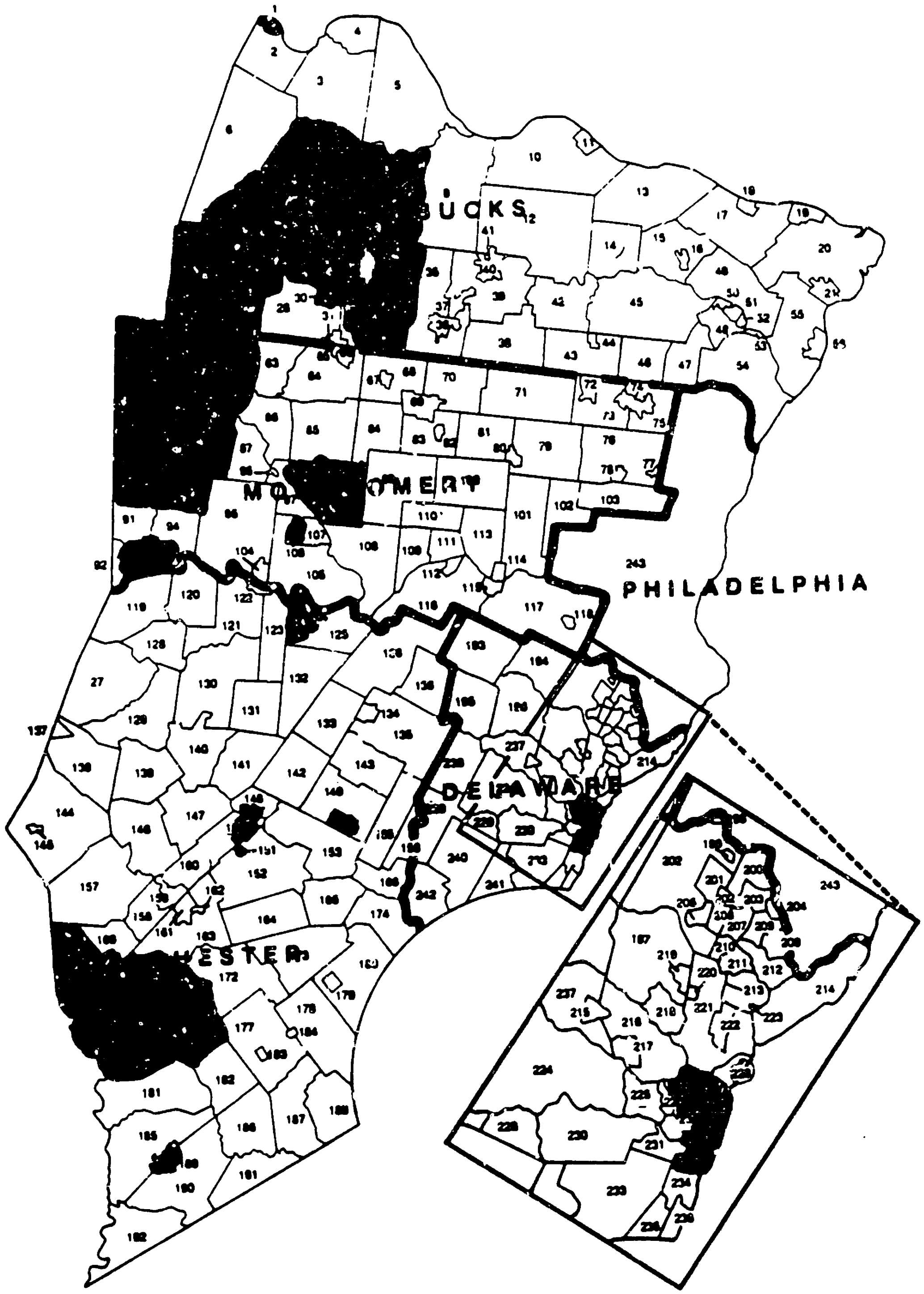
(t-statistics)

Regressor	Benchmark 1960-92	Benchmark 1960-75	Benchmark 1976-92	Government Overhead(-1) 1960-92	Government Overhead(-1) 1960-75	Government Overhead(-1) 1976-92
Government Overhead(-1)	—	—	—	-0.417	-0.204	-0.790
				(-1.86)	(-0.72)	(-1.71)
% Philadelphia	-0.046	-0.015	-0.055	0.037	0.024	0.062
Commuter	(-2.40)	(-0.39)	(-2.21)	(1.92)	(0.84)	(1.79)
% Out-of-State	-0.021	-0.023	-0.020	0.022	-0.007	0.054
Commuter	(-1.62)	(-0.84)	(-1.55)	(1.68)	(-0.51)	(1.79)
% Seniors	-0.055	-0.084	-0.057	0.051	0.043	0.069
	(-1.65)	(-1.63)	(-1.25)	(1.51)	(0.83)	(1.43)
$\frac{\text{Owner}}{\text{Renter}}$ Tenure	-0.024	-0.036	-0.035	0.028	-0.041	0.056
	(-0.42)	(-0.43)	(-0.39)	(0.51)	(-0.49)	(0.62)
% Face EIT at	0.051	0.048	0.091	0.051	0.048	0.091
Work	(11.27)	(9.12)	(6.91)	(11.29)	(9.13)	(6.90)
% Deficit(-1)	0.005	0.007	0.002	0.002	0.005	0.000
	(0.67)	(0.88)	(0.19)	(0.36)	(0.60)	(0.05)
Median HH Income	-0.004	-0.023	0.003	-0.009	-0.036	0.004
[$\times 10^3$ \$]	(-0.24)	(-0.35)	(0.12)	(-0.49)	(-0.54)	(0.15)
Population [$\times 10^3$]	0.025	0.044	0.004	0.026	0.042	0.007
	(1.47)	(0.99)	(0.11)	(1.50)	(0.93)	(0.22)
Population Growth	0.030	0.070	-0.050	0.029	0.063	-0.042
	(0.63)	(1.28)	(-0.49)	(0.61)	(1.15)	(-0.41)
Employment [$\times 10^3$]	0.004	0.075	-0.002	0.006	0.090	-0.001
	(0.11)	(0.97)	(-0.04)	(0.17)	(1.12)	(-0.01)
City	1.319	1.061	—	1.092	0.743	—
	(1.61)	(1.20)		(1.31)	(0.81)	
Borough	0.815	0.647	0.541	0.872	0.628	0.758
	(3.11)	(1.91)	(0.93)	(3.33)	(1.85)	(1.19)
County Dummies?	Yes	Yes	Yes	Yes	Yes	Yes
N	7607	3582	4025	7607	3582	4025
$\log L$	-564.52	-384.72	-158.88	-558.75	-382.27	-155.51
LR omitted variable	—	—	—	11.54	4.90	6.74

Table 1.14: Hazard Model of Income Tax Propensity

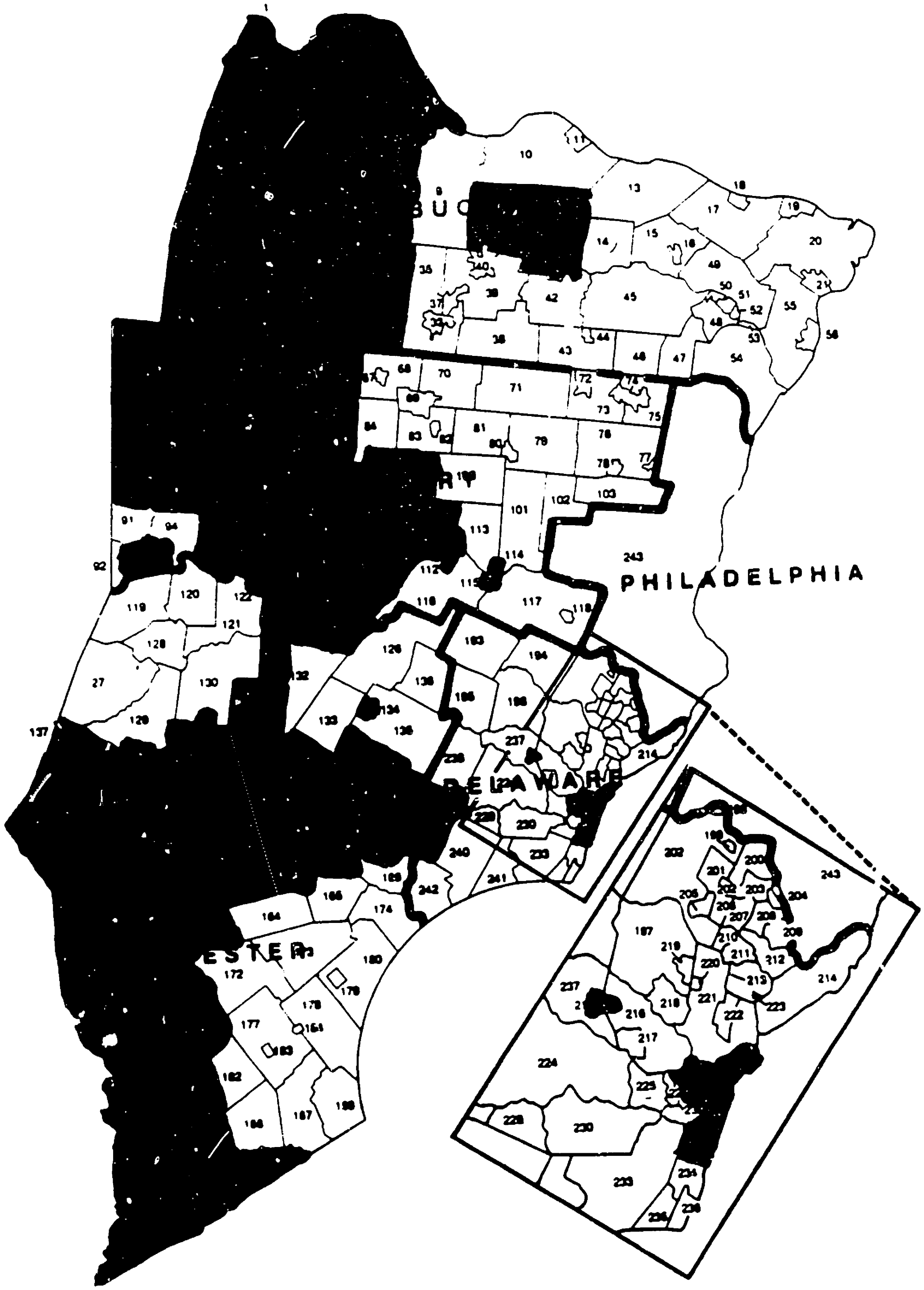
Cox Partial Likelihood of Proportional Hazards with Time-Varying Regressors.
Dependent variable: EIT dummy. Years: 1960-1992, 1960-1975, 1976-1992. Sample: 237 MCDs in Philadelphia SMSA.
Panels: Benchmark (1), Credibility Proxy [Levels] (2).
1976-1992: *Cities* omitted since both are left-censored. Last row: LR test of omitted variable for credibility proxy.

(t-statistics)



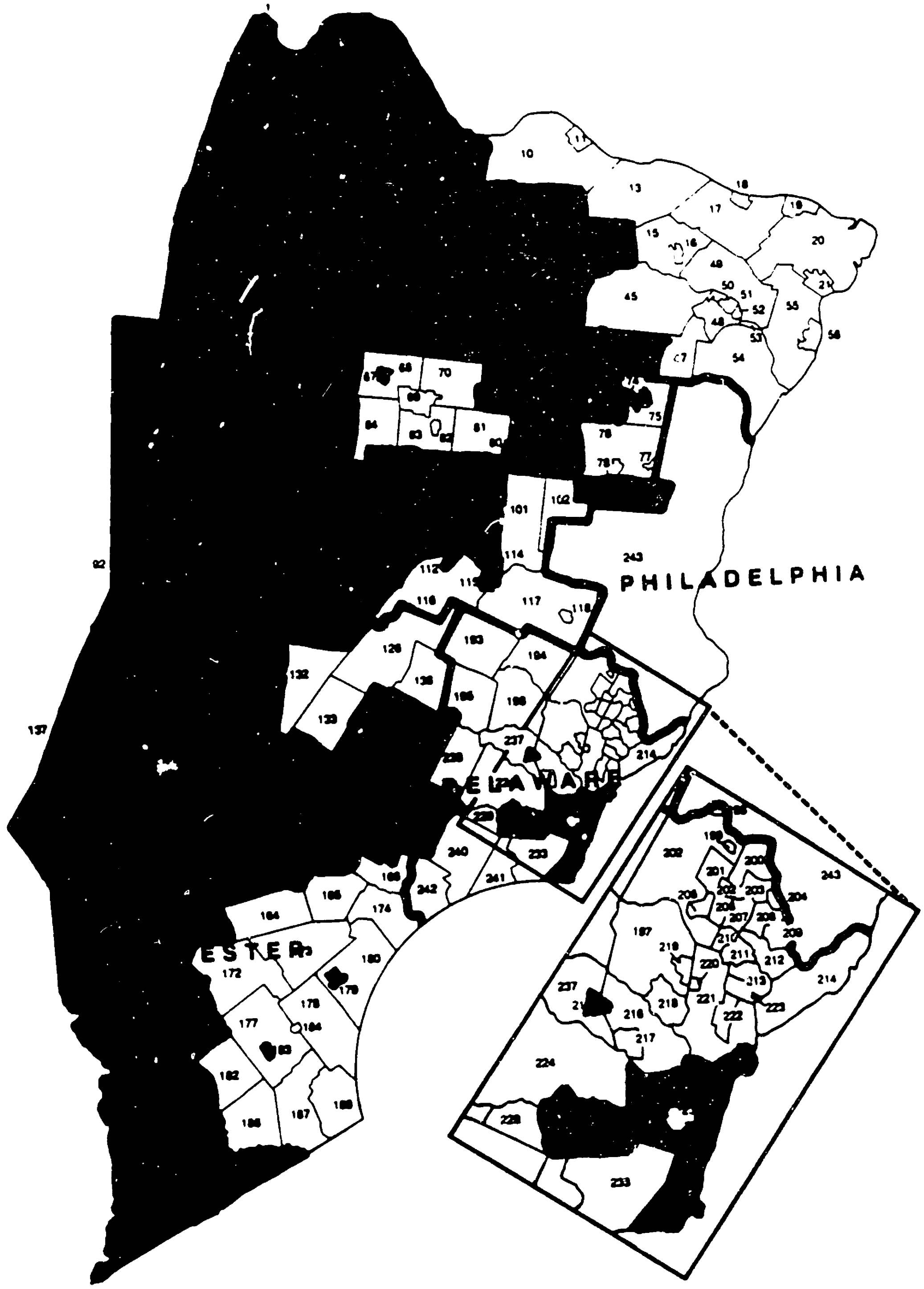
Map 1: Earned Income Tax in 1970

Shaded communities have tax



Map 2: Earned Income Tax in 1980

Shaded communities have tax



Map 3: Earned Income Tax in 1992
Shaded communities have tax

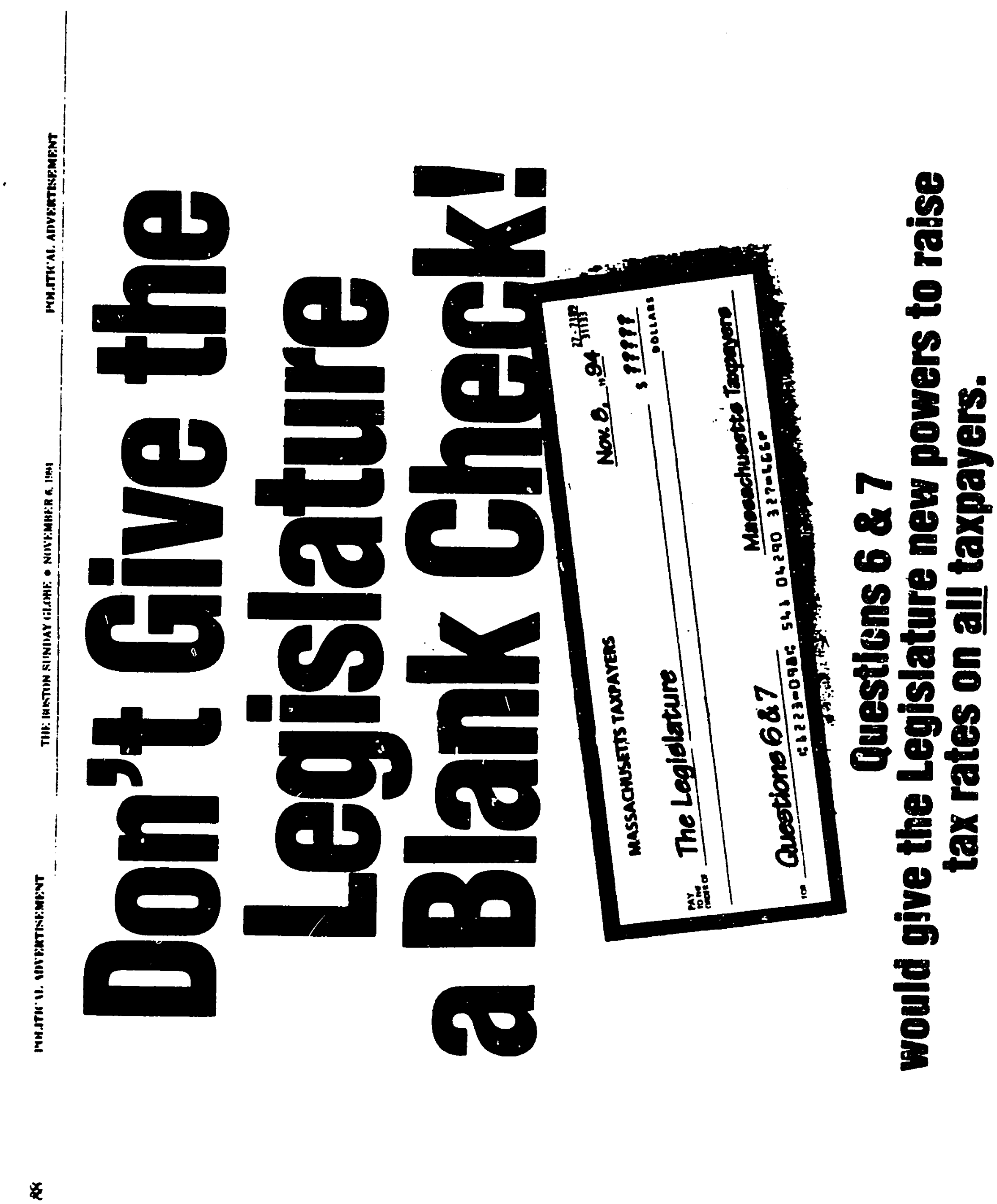

POLITICAL ADVERTISEMENT
POLITICAL ADVERTISEMENT
Don't Give the
Legislature
a Blank Check!
MASSACHUSETTS TAXPAYERS
Nov. 8, 94
$?????
DOLLARS
PAY TO THE ORDER OF
The Legislature
Massachusetts Taxpayers
FOR
Questions 6 & 7
Questions 6 & 7
would give the Legislature new powers to raise
tax rates on all taxpayers.

Figure 1

Figure 2: Percent of Communities Levying Income Tax

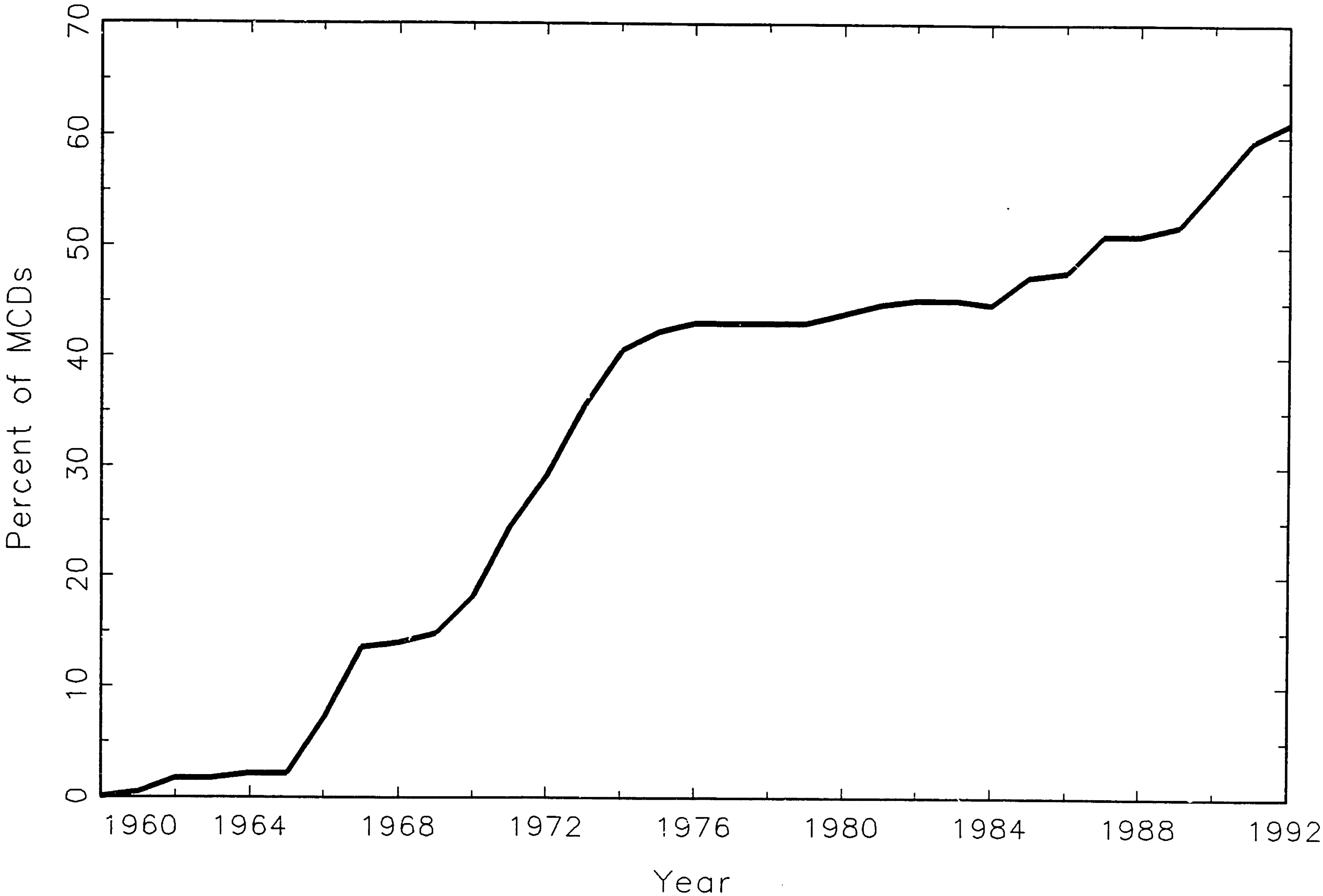


Figure 3: Fraction of Communities Levying Income Tax

as Function of Philadelphia Commuters: 1992

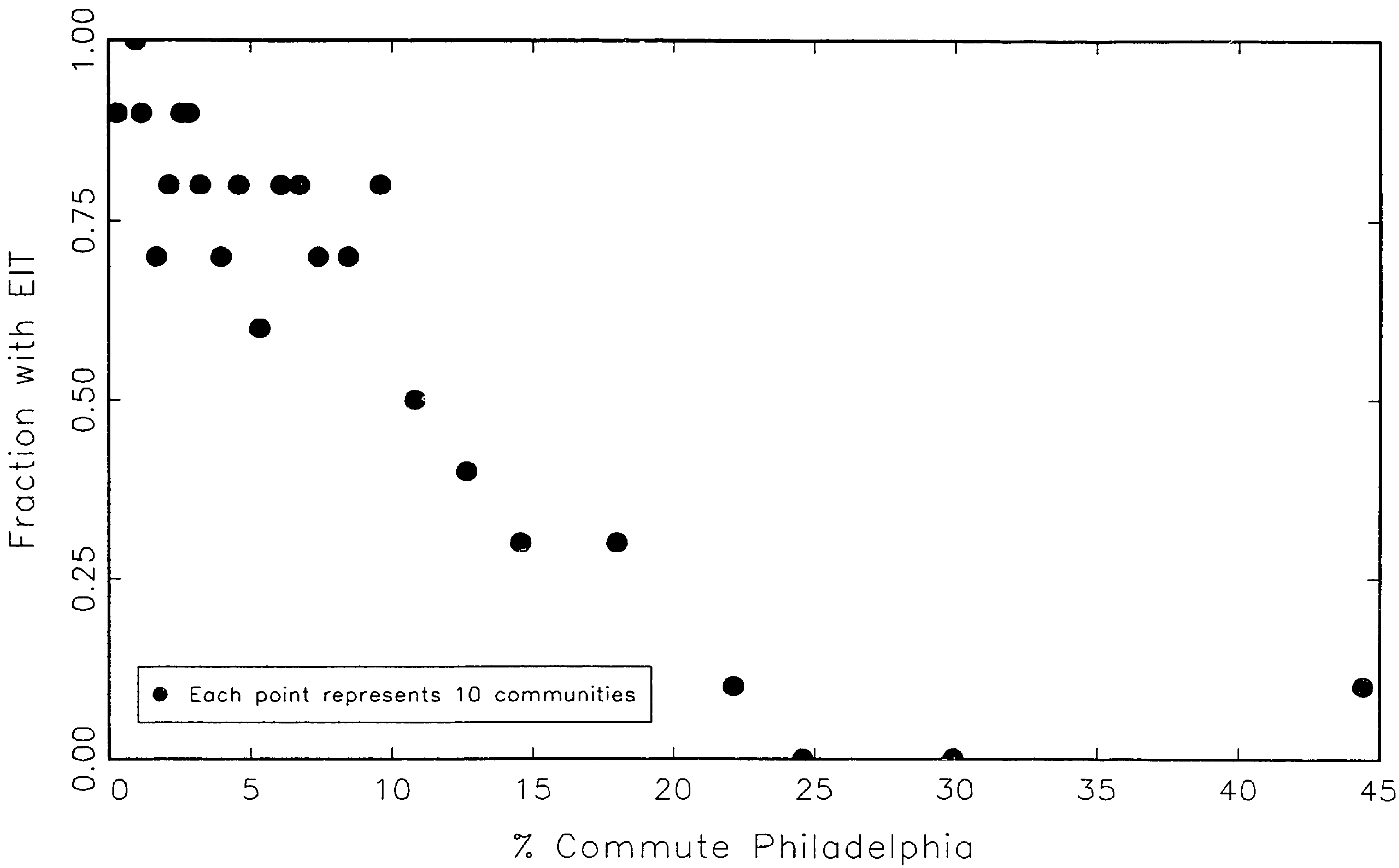


Figure 4: Fraction of Communities Levying Income Tax

as Function of Out-of-State Commuters: 1992

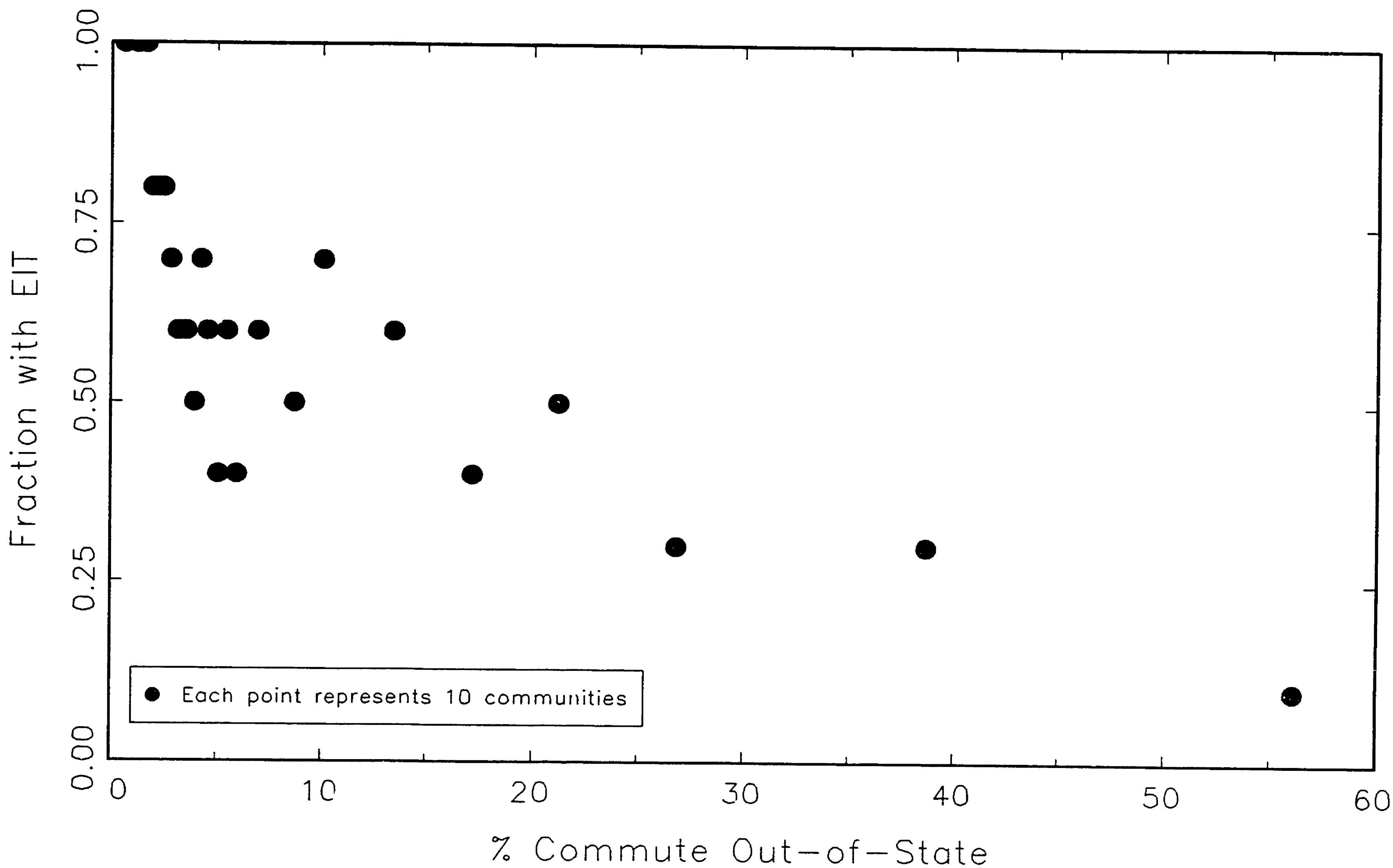


Figure 5: Tax Effort against Distance from Philadelphia

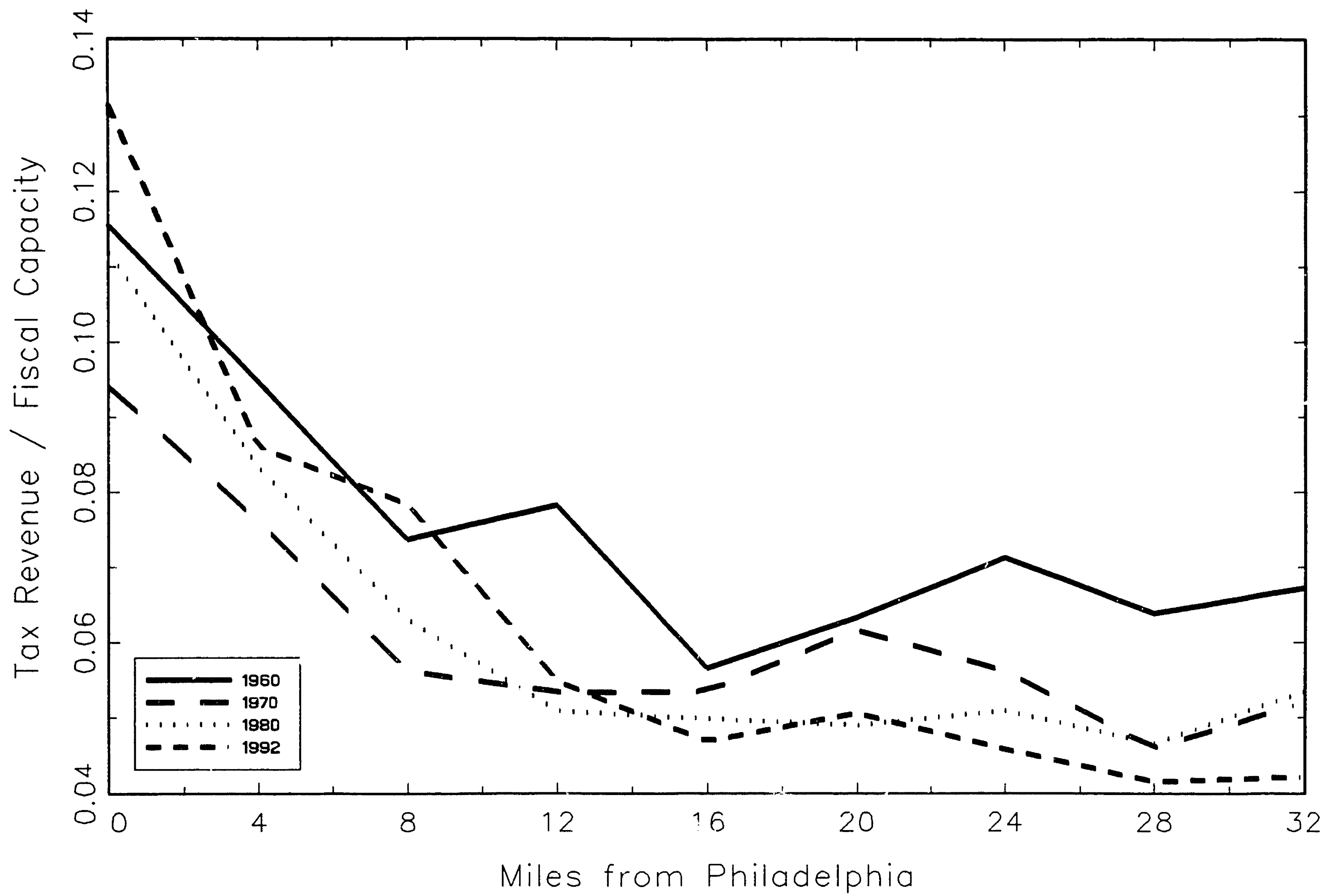


Figure 6: Fraction of First-Time Taxers with LOW Tax Effort

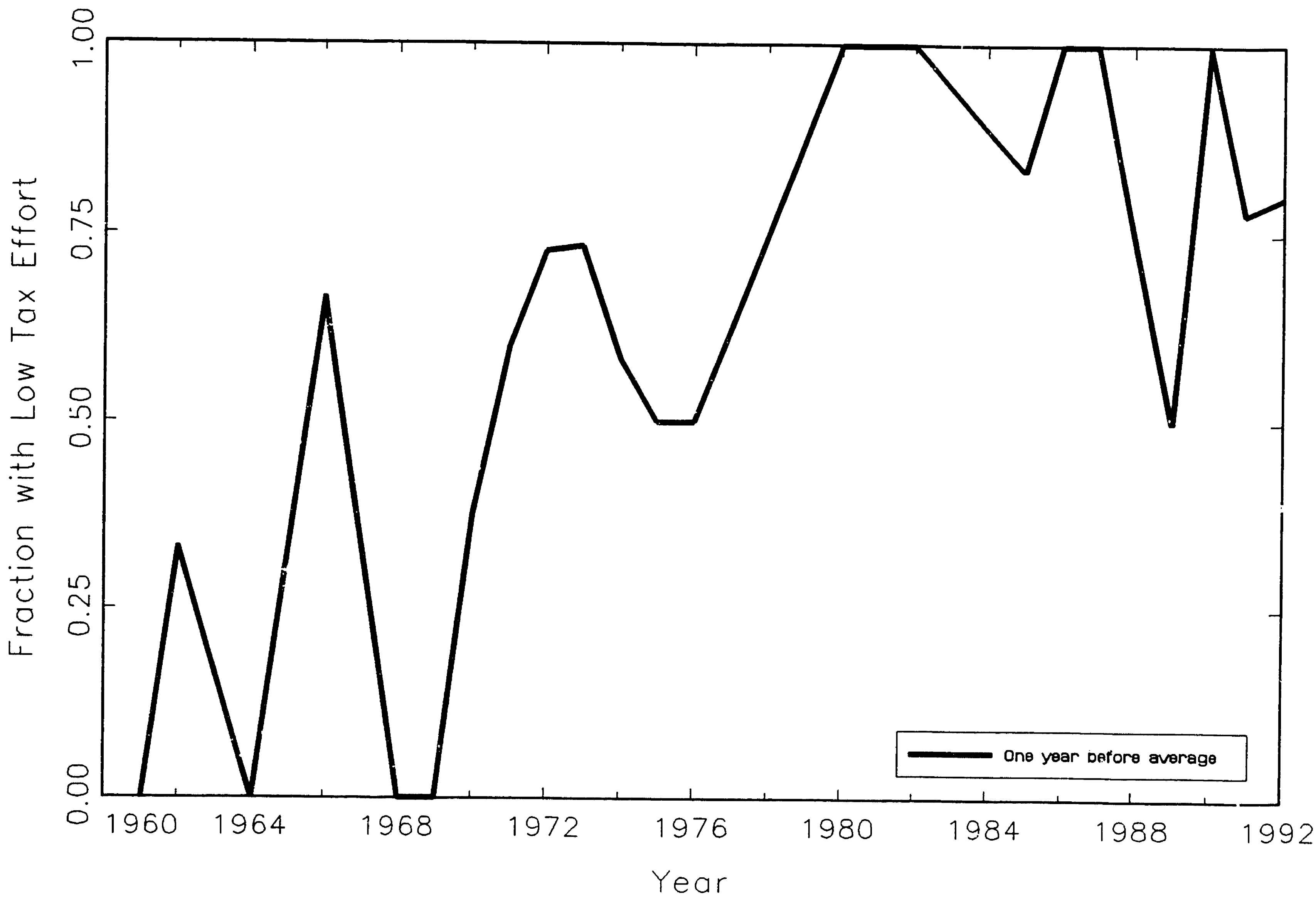


Figure 7: Government Overhead vs Distance from Philadelphia

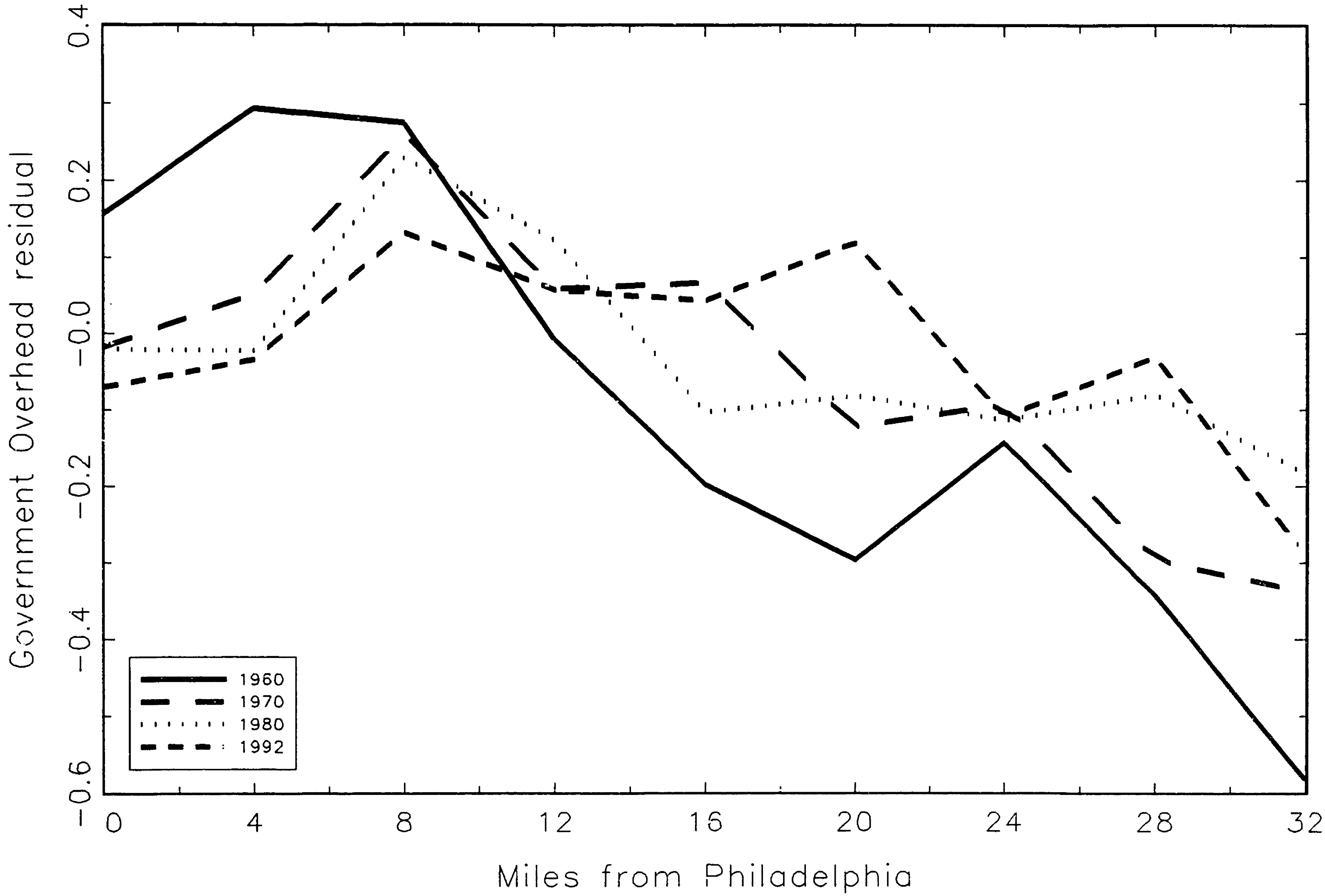


Figure 8: Average Change in Millage

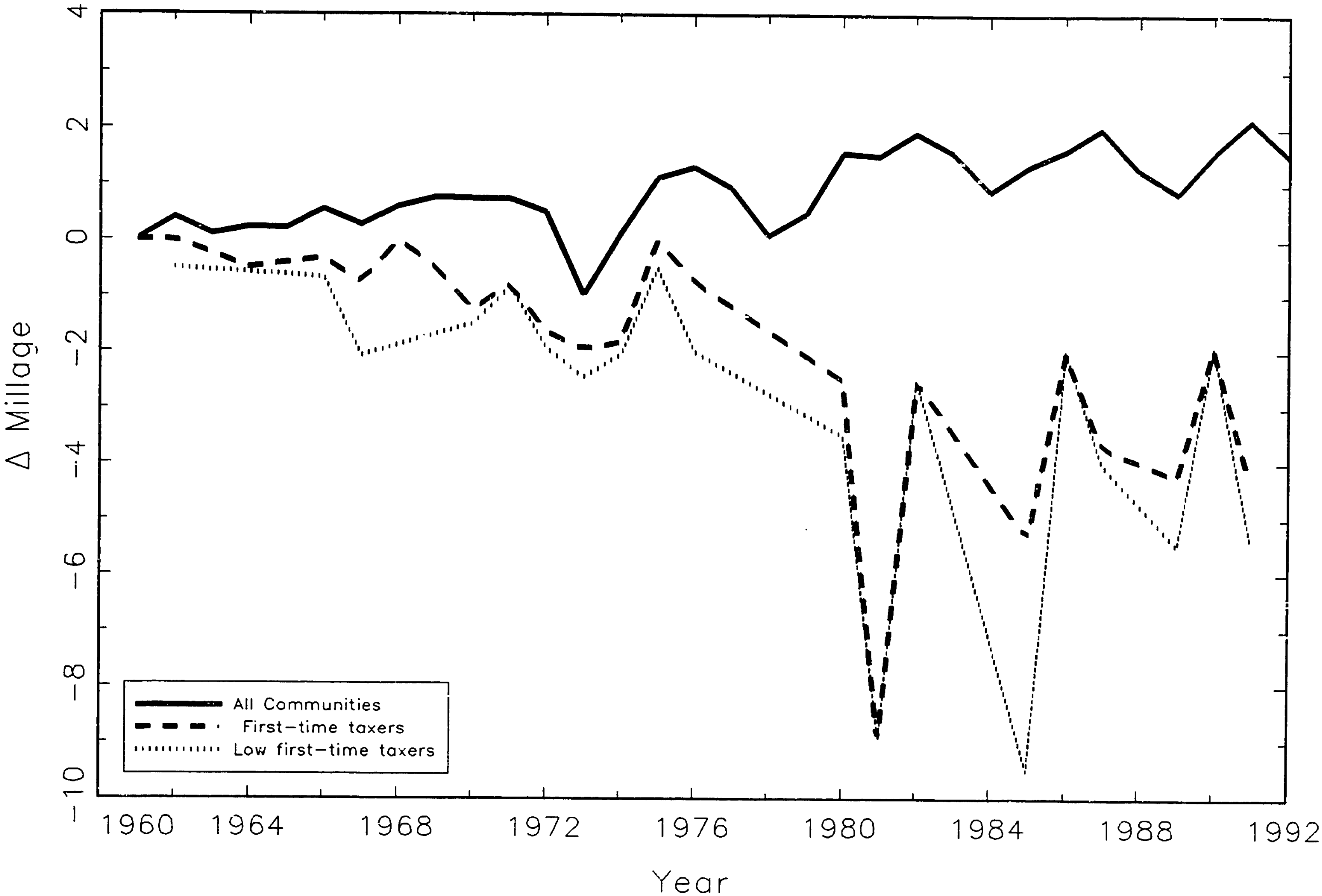
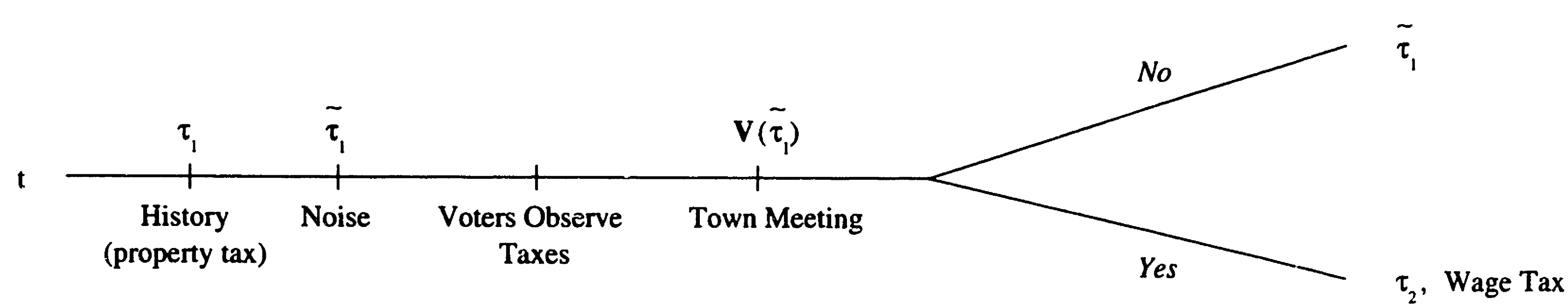


Diagram 1: Timing of the Game



Chapter 2

Sequential Election Contests: Strategic Effects Oppose Momentum in Presidential Primaries

2.1 Introduction

The most fundamental characteristic of an electoral system is how it translates voter preferences into election outcomes. As Myerson [17] notes in his recent survey of democratic institutions, formal analysis of this question is one of the central pillars of social choice research. This paper contributes to the literature by exploring sequential elections with particular focus on the American presidential primary.

The two major political parties select their nominee though a state-by-state ballot. Each state has a separate primary and awards delegates to candidates based on their vote percentage. Roughly, the contender who wins a majority of the available delegates from all elections becomes the party nominee. A peculiar feature is that the state elections are staggered in time. While this allows candidates to devote their resources (time and money) to a small group of voters for at least a short time, there may also be spill-overs between elections. In the conventional story, winning an early election increases vote total in subsequent states, *momentum*. I call this a "system" characteristic since the mechanism involves voter learning rather than candidate behavior.[1]

Momentum favors the victor of early elections. This means that the appropriate sequencing could yield a winner who would have lost in a simultaneous election, see Aldrich [3] for a simple model. But do such bandwagons occur in practice?

Consider the 1976 Republican race between Gerald Ford and Ronald Reagan. Before the primary season began, Ford led both in campaign resources (Federal Election Commission disclosure form [10]) and national polling (Aldrich [4] Figure 5.3). As shown in Figure 1, many of the first primary states favored Ford. The conventional wisdom predicts that Ford would jump out to an early lead, and the resulting momentum would propel him to an easy victory; since participating in elections is costly in time and money, Reagan should have simply dropped-out of the contest. Yet this is not what happened: while Ford did win the first four elections, Reagan managed to stave off a bandwagon, garner key wins in North Carolina and Texas, and keep the contest competitive all the way to the last election in his home state of California. The main lesson is that momentum effects need not be important, particularly at the end of the election cycle when voters are more informed about the candidates (Bartels [6] chapters 6, 7, 9). Still, even dissipated momentum suggests an easy Ford victory.

This suggests an opposing effect is at work. Reagan could credibly commit to stay in the race even following his early string of losses, since he anticipated key victories in later states. Alternatively, had Ford been upset in an early election, he would have to quit since he had few favorable states left and faced a determined opponent. In another words, a candidate may benefit when his best elections come at the end of the cycle. The mechanism is a "strategic" effect since it explicitly involves interacted

[1]There is debate over which information cue is most important with the leading contenders policy position (McKelvey-Ordeshook [15] and [16]), candidate valence (Cukierman [9] and Banerjee [5]) and the identity of the front-runner (Bartels [6] and Bikhchandani-Hirshleifer-Welch [7]).
A second aspect of momentum is the boost in campaign contributions and media attention which increase candidate exposure in later elections (and hence voter learning).

candidate choice (whereas momentum does not).

I formalize[2] this intuition in a war of attrition: two forward-looking candidates compete in a series of costly battles (the elections) for a single prize (the nomination). I argue that anticipation can generate a "force" which often works in opposition to inter-election momentum. A candidate who only has unfavorable states left may choose to drop-out rather than incur participation costs. Conversely, his opponent can *credibly threaten* to stay in every future election. A second explanation for the strategic effect is the *value of information*: when a candidate gets a large dose of bad news, like being upset in an election, he often quits the race. Back-loaded election sequences best allow a candidate to capitalize on these surprises.

This theory is one explanation for why Reagan choose to stay in the 1976 election and why successful nominees have not been based in early primary states (like New Hampshire and Iowa). And since in reality momentum plays some role, the optimal ordering of a given set of elections is ambiguous. The tension between strategic effects and the bandwagon might explain the relative stability of the order of states in the primary schedule (see Figure 2), and why there is no reform movement for a one-shot election.

My formal approach generalizes the war of attrition by allowing non-stationary environments and including an evolving state variable (a success gauge for one of the players). This technique can be applied to other dynamic games with costly participation. For example, in which house of Congress should a bill be initiated given their different ideological leanings? My theory suggests that supporters might be more effective in dousing out active opposition if they first propose it in the less favorable venue. In another example, the non-stationarity can be used to enrich

[2]The only theoretical paper I am aware of which explicitly considers dynamic strategies in presidential primaries is Brams-Davis [8] who impose symmetric candidate resources and strong functional forms. Their main contribution is showing the existence of an equal spending in every period equilibrium, a result which gets us little closer to understanding the role of time in sequential political contests. Aldrich [2] examines a discrete choice model somewhat similar to mine, but he does not formalize any dynamics.

existing war of attrition models.[3] One prominent case is the patent race game where two firms engage in research competition (Fudenberg-Gilbert-Stiglitz-Tirole [12]). If research progress is publicly observed, is it better to be a "quick" or "slow" researcher? This paper suggests that if a firm expecting quick results has difficulties getting out of the blocks, it will likely abandon the R&D process. The slower firm is willing to continue and may have an advantage in equilibrium.

The outline of the rest of the paper is as follows. The next section presents the formal model with the relevant characteristics of the system and candidate objectives. Some useful tools are also presented. The third section contains the solution to a simple three period model which illustrates both the general advantage of back-loaded sequences and the rich array of equilibrium outcomes. This motivates examination of general horizon problems under a simplifying assumption about participation costs. A candidate's probability of clinching the nomination is almost always greater than half in a sequence where a string of unfavorable elections precedes an equal number of favorably biased ones. The model is generalized in Section 2.5 to allow for endogenous fighting costs, and the conditions for favorable back-loading are derived. The last section examines the empirical implications of the model and discusses directions for future research.

2.2 Model

In presidential primaries candidates ("players") wage a costly battle over an indivisible prize, the party nomination. I model this contest as a finite-horizon *war of attrition*: two players fight a sequence of costly battles until one quits leaving the prize for his opponent. Each period of combat represents a contested election; when one player concedes, his opponent gets the nomination. There is one important exception: if neither player drops-out, the prize goes to the candidate who wins the majority of

[3]See Fudenberg-Tirole [11] for general results and a survey of economic applications.

elections. The outcome of an election is revealed only after candidates fight there, but *ex ante* each stage has a commonly known bias towards one contestant.

Below I formalize this story in a simple model. I describe the nature of the dynamic system and then player's objectives and strategies. Finally I present some useful tools and motivate the restrictions I use in this paper.

2.2.1 System

There is a sequence of states/elections,[4] $1, 2, ..., T$, T odd. Each state, t, awards a single delegate to the candidate who receives a majority of votes in its election. The *ex ante* probability that player one wins election t when both players compete is $p_t \in [0, 1]$.[5] These commonly known parameters stem from electorate preferences. So if player one is perceived as having a policy position near the median or having strong valence attributes (charisma, political ability), then p_t is large.[6] Due to imperfect voter polling and incompletely informed citizens, the election outcome is uncertain, e.g. the parameter can be interior to zero and one.

The cycle is called *front-loaded* for player i if his best-elections come in the beginning of the sequence,

$$\sum_{s<t_m} p_s > \sum_{s>t_m} p_s$$

where t_m is the middle election, $\frac{T+1}{2}$. With two contenders, a front-loaded sequence for a player is *back-loaded* for his opponent.

Let $\omega(t)$ be one if player one wins election t and zero otherwise. This variable is observable at the beginning of period $t+1$. Define:

$$W(t) = \sum_{s=1}^{t-1} \omega(s) \tag{2.1}$$

[4]Elections are staggered in time; at any time, t, results from $s \leq t-1$ are known.

[5]The value is exogenous: neither players nor the prior path influence its value. In particular there is no momentum effect.

[6]Some empirical evidence on the importance of policy positions is presented in Gopoian [13].

player one's cumulative win count after t elections. This is the state variable. Notice:

$$\begin{array}{lcl} W(t) > \frac{T-1}{2} & \Rightarrow & \text{2 eliminated} \\ W(t) < t - \frac{T+1}{2} & \Rightarrow & \text{1 eliminated} \end{array}$$

where "eliminated" means that the player cannot win a majority of elections. A state-time pair, $[t, W(t)]$, will be referred to as a *node*.

Two other bits of jargon will help ease the exposition later. A candidate is said to *clinch the nomination* either due to his opponent's mathematical elimination or quitting the race (below). A *decisive election* is a fight in which the winner clinches the nomination.

I now turn to the candidates whose decisions are simply whether to quit at a give node.

2.2.2 Players

There are two players labeled $i \in \{1, 2\}$. Remaining in the game has benefits and costs: a candidate pays cost c_t if he participates in election t, whereas continuing means he is still in contention for the nomination which has unit value. Costs stem from some combination of financial (advertising expenditure), political (political capital expenditure) and personal (psychological duress) factors; notice both players pay the same value for any given stage. There are two ways for candidate one to win the contest: (a) his opponent quits first, or (b) both contest all T elections and $W(T+1) > 0.5T$, he has a larger cumulative delegate count.

So if the game ends at the start of period t:

$$U_i(t) = \mathcal{I}_{i,t}(win) - \sum_{s=1}^{t-1} c_t \tag{2.2}$$

where $\mathcal{I}_{i,t}(\cdot)$ is the indicator function for player one at time t. Notice if candidates quit simultaneously than neither wins the nomination. Also, if the one quits the winner incurs no fighting costs for the terminal period.

A strategy is a contingent decision whether to stop for all possible nodes:

$$\mathbf{a_i} \equiv \{a_i[t, W(t)] \in [0,1] \; \forall t\}$$

A value of one means the player surely stops in period t given the cumulative delegate count and no one has stopped yet; players cannot return to the contest after they drop-out. Mixed strategies are permitted.

Utility maximization guides the players' strategy choice. A subgame-perfect equilibrium is a strategy pair from which neither player wants to deviate even at nodes reached with probability zero on the equilibrium path. A player quits with probability one only when his future costs along the equilibrium path exceed the benefit and the other player chooses to stay.[7] In other words, current actions are based upon *continuation value*:

$$V_i[W(t), \mathbf{a_1}, \mathbf{a_2}] = \sum_{s=t}^{T+1} Pr(s) \left[Pr(i,s) - \mathcal{I}(s \neq t) \sum_{r=t}^{s-1} c_r \right] \tag{2.3}$$

where $Pr(s)$ is the probability either player quits for the first time in period s and $Pr(i,s)$ is the chance that i wins at s, defined with respect to the equilibrium path. Players ignore sunk (fighting) costs and evaluate their future prospects. Hence, this variable makes sense only with respect to a given equilibrium strategy pair and current node. To conserve space I will henceforth suppress the strategy arguments.

When one's opponent continues, it will be more useful to use the inductive form:

$$V_i[W_n(t)] = \max\langle p_t V_i[W_{n+1}(t+1)] + (1-p_t)V_i[W_n(t+1)] - c_t, 0\rangle \tag{2.4}$$

where in abuse of notation $W_n(t)$ indicates that player one has won n elections before period t. The first term is the expected value if the player continues: he pays the

[7]No equilibrium has both players drop-out with probability one at any node.

fighting cost and weights his value at the possible nodes next period with the current win probability. The other term is the payoff from quitting, zero. So this says if $V_i[W(t)]$ is positive, then player i's optimal strategy is to continue from this sub-game; if $V_i[W(t)] < 0$ and $V_j[W(t)] > 0$ than player i should quit. Mixed strategies are invoked when both players have negative continuation payoffs. An equilibrium strategy pair yields behavior which satisfies these conditions at every possible history.

This completes the discussion of the game and its players.

2.2.3 Useful Tools

Before turning to substantive analysis, there are two lemmata which simplify the calculations. The first result will allow us to conserve space and omit one player's continuation value.

Lemma 1 *At any node the following properties always hold:*

(i) *the future costs of both players are equal;*

(ii) *the sum of benefits equals one (only under pure strategies);*

The first point holds since there is a common equilibrium path and the second since probabilities sum to one (under mixed strategies both players may quit at a given node so total benefits may be less than one).

The next lemma describes some common situations which will or will not trigger a candidate to quit the race.

Lemma 2 *In equilibrium:*

(i) *mathematically eliminated candidates exit immediately;*

(ii) *if player i quits at some $\overline{W}(t)$, than he also quits at any earlier node in which he cannot attain a more favorable state at t;*

(iii) *a player never drops-out following a win.*

PROOF: See Appendix 2.7.1.

The broad message is that with uncertain elections, any victory contains a positive informational shock. Alternatively, the same outcome is bad news to the loser. It is this shifting in relative positions which induce quits. This insight will prove useful later.

2.2.4 Restrictions

The problem so far is quite general. In particular the set of possible p_t sequences is a continuum. I limit the range of election biases.

Assumption 1. *A state may favor one or the other candidate by a fixed amount or it is unbiased:*

$$p_t \in \{p, 0.5, 1-p\}, \quad p \leq 0.5$$

While in practice there are more than three flavors of election bias, the model still captures the broad notion that states may favor one candidate. I will refer to the unbiased state as the *toss-up* while a loss in a state with a favorable bias will be termed an *upset.*

In addition, the family of sequences which I examine impose no bias on average.

Assumption 2. *The election cycle belongs to the unbiased class $\mathcal{F}(T)$:*

$$\mathcal{F}(T) \equiv \left\{ (p_t)_{t=1}^{T} : \frac{1}{T} \sum_{t=1}^{T} p_t = 0.5 \right\}$$

In words, this is the set of T period election sequences where there are equal number of states favoring each candidate. It is not hard to show that any sequence in $\mathcal{F}(T)$ yields a fifty percent probability of player one winning when candidates commit never to quit. With strategic behavior, however, there will be a variety of significant variations from this outcome.

Finally, assume that fighting costs are exogenous and constant across players and periods.

Assumption 3. *Constant fighting costs:*

$$c_t \equiv c \in (0, T^{-1}) \quad \forall t$$

The upper bound means that if a player knows he will win the contest, then he should be willing to fight every election. In Section 2.5 I relax the constant cost assumption and endogenize c_t.

We can now examine some equilibrium outcomes.

2.3 Worked Example: Three Period Back-Loading

In finite multi-stage games with observed actions subgame-perfection and backwards induction are equivalent: players recognize that actions in earlier periods influence what happens later. So to find equilibria, first solve for the optimal action pair at each possible node in the final period. Then do likewise in the penultimate period taking into account the solution next period (note the critical role for continuation values). Repeating this process yields the (unique) equilibrium pair of best actions for every possible node under any particular (c, p) combination.

In this section I solve the game in the class $\mathcal{F}(3)$ which is back-loaded for player one: $\{p, 0.5, 1-p\}$. The main objective is to solve for the winning probability of player one and compare it to the outcome when there are no strategic quits. The results are summarized in Figure 3: the equilibrium depends in a highly non-linear way on both the relative and absolute values of cost and election bias. Still in most of the parameter space player one wins with probability exceeding one half.[8] Favorable

[8]Recall that any member of $\mathcal{F}(T)$ would have either player winning with equal probability if there were no strategic quits.

elections at the end rather than beginning allow commitment to never quit, and so forces the opponent to concede if upset. Since only big information shocks influence behavior, players prefer early surprises to be favorable, e.g. be the underdog in the beginning.

2.3.1 Solution

The problem will be solved using backwards-induction. To conserve space I suppress nodes in which a player has been eliminated [Lemma 2 (**i**)]. It will be easiest to follow the presentation using Diagram 1. Time flows from left to right and the cumulative win count (prior to that election) is on the vertical axis. Each line represents a node with a surrounding circle indicating that a player has mathematically clinched.

Let $p_1 = p$, $p_2 = 0.5$, $p_3 = 1 - p$ and consider $t = 3$. At $W(3) = 2$ one has clinched while $W(3) = 0$ means he has been eliminated. Both are still in contention at $W(3) = 1$. Using eq. 2.4, find the continuation value presuming both continue:[9]

$$\begin{aligned} \tilde{V}_1[W(3) = 1] &= p \times 0 + (1 - p) \times 1 - c \\ &\lesseqgtr 0 \leftrightarrow 1 - p \lesseqgtr c \end{aligned} \tag{2.5}$$

$$\begin{aligned} \tilde{V}_2[W(3) = 1] &= p \times 1 + (1 - p) \times 0 - c \\ &\lesseqgtr 0 \leftrightarrow p \lesseqgtr c \end{aligned} \tag{2.6}$$

In this terminal period, the winner of the election gets the prize and the loser gets nothing (it is decisive). While the candidates have equal standing coming into the election, player one has the advantage since his is expected to win the election with probability $1 - p > 0.5$.

Presume first that both players continue at this point and roll-back to the unbiased section election. To conserve space, following Lemma 1[10] I suppress the continuation

[9] The notation $\tilde{V}_i[W(t)]$ is used to indicate continuation value when the player is not permitted to quit at the current stage.

[10] In this part of the problem costs are never so high that players resort to mixed strategies, so

value of the leading candidate. From eq. 2.4,

(1): $c \leq p$

$$\begin{aligned}
\widetilde{V}_2[W(2)=1] &= 0.5 \times V_2[W(3)=1] + 0.5 \times V_2[W(3)=2] - c \\
&= 0.5 \times (p-c) + 0.5 \times 0 - c \\
&= 0.5p - 1.5c \\
&\lesseqgtr 0 \;\; \leftrightarrow \tfrac{p}{3} \lesseqgtr c
\end{aligned} \tag{2.7}$$

$$\begin{aligned}
\widetilde{V}_1[W(2)=0] &= 0.5 \times V_1[W(3)=0] + 0.5 \times V_1[W(3)=1] - c \\
&= 0.5 \times 0 + 0.5 \times (1-p-c) - c \\
&= 0.5(1-p) - 1.5c \\
&\lesseqgtr 0 \;\; \leftrightarrow \tfrac{1-p}{3} \lesseqgtr c
\end{aligned} \tag{2.8}$$

These are the values for the losers of the first round election; one more loss and they are surely out. Relatively speaking, player one is better off (higher continuation value) in such a state of the world since (i) he has a favorable election to come and (ii) he was the underdog in the first election anyway.

Continue under the assumption that $c \leq p$. Presume both continue from the second election.

(1a): $c \leq \frac{p}{3}$

$$\begin{aligned}
\widetilde{V}_1[W(1)=0] &= (1-p) \times V_1[W(2)=0] + p \times V_1[W(2)=1] - c \\
&= (1-p) \times [0.5(1-p) - 1.5c] + p \times [(1-0.5p) - 1.5c] - c \\
&= 0.5 - 2.5c \\
\widetilde{V}_2[W(1)=0] &=
\end{aligned}$$

---

this is appropriate.

This term is always positive. Hence, no one quits unless mathematically eliminated in the last round. The equilibrium outcome for the whole election sequence:

CONCLUSION **(1a)**: *When* $c \leq \frac{p}{3}$, $\Pr(1 \text{ win}) = 0.5$

Now allow costs to be just high enough that two quits after he loses the first election but a similarly situated player one will stay, compare eq. 2.7 and 2.8.

(1b): $c \in (\frac{p}{3}, \frac{1-p}{3}]$ AND $c \leq p$

$$\begin{aligned}
\tilde{V}_1[W(1)=0] &= (1-p) \times V_1[W(2)=0] + p \times V_1[W(2)=1] - c \\
&= (1-p) \times [0.5(1-p) - 1.5c] + p \times 1 - c \\
&= 0.5(1+p^2) - (2.5 - 1.5p)c \\
&> 0
\end{aligned}$$

so from Lemma 1: $\tilde{V}_2[W(1) = 0] = 0.5(1 + p^2) - (2.5 - 1.5p)c \lesseqgtr 0$. In this case player two may not even participate in the nomination contest since he knows a loss in the first round will force a quit later. Even if not, player one wins with probability greater than half (the non-strategic outcome), since the only strategic quit would be his opponent quitting after losing the first election.

CONCLUSION **(1b)**: *When* $c \in (\frac{p}{3}, \frac{1-p}{3}]$ *and* $c \leq p$,

$$\Pr(1 \text{ win}) = \begin{cases} 0.5(1+p^2) & c \leq \frac{1-p^2}{5-3p} \\ 1 & c \in (\frac{1-p^2}{5-3p}, \min\left\langle p, \frac{1-p}{3}\right\rangle] \end{cases}$$

Next allow large enough costs so that both players quit after a first round loss, but not so high that player two exits before $W(3) = 1$.

(1c): $c \in (\frac{1-p}{3}, p]$

$$\begin{aligned}\tilde{V}_1[W(1)=0] &= (1-p)\times V_1[W(2)=0] + p\times V_1[W(2)=1] - c \\ &= (1-p)\times 0 + p\times 1 - c \\ &= p - c \\ &\geq 0\end{aligned}$$

so $\tilde{V}_2[W(1)=0] = 1-p-c$. As the loser of the first round exits, it is clearly best to be the leader in that election.

CONCLUSION **(1c)**: *When* $c \in (\frac{1-p}{3}, p]$, $\Pr(1 \text{ win}) = p$

The only remaining case is when costs exceed the probability of the underdog winning an election. This is an important scenario since it is the minimum cost which the leader in the final round needs to force out his opponent, see Section 2.4.

(2): $c > p$

Since $c < T^{-1} \equiv \frac{1}{3} < 1-p$, from eq. 2.5 and 2.6 $\tilde{V}_1[W(3)=1] > 0 > \tilde{V}_1[W(3)=1]$, so two does not contest the third election. An application of Lemma 2 **(ii)** implies that he will also drop out at $W(2) = 1$, since he is effectively eliminated there. Hence, $V_1[W(2)=1] = V_1[W(3)=1] \equiv 1$ and $V_2[W(2)=1] = V_2[W(3)=1] \equiv 0$. For the other node before the second election,

$$\begin{aligned}\tilde{V}_1[W(2)=0] &= 0.5\times V_1[W(3)=0] + 0.5\times V_1[W(3)=1] - c \\ &= 0.5\times 0 + 0.5\times 1 - c \\ &= 0.5 - c \\ \tilde{V}_2[W(2)=0] &= \end{aligned}$$

The loser of this decisive election quits, so with an unbiased electorate there are equal (positive) continuation values. Finally we solve for the first period:

$$\begin{aligned}\tilde{V}_1[W(1)=0] &= (1-p)\times V_1[W(2)=0] + p\times V_1[W(2)=1] - c \\ &= (1-p)\times(0.5-c) + p\times 1 - c \\ &= (0.5+p) - (2-p)c \\ &\lesseqgtr 0\end{aligned}$$

so $\tilde{V}_2[W(1)=0] = (0.5-p)-(2-p)c \lesseqgtr 0$. In general one wins with probability greater than half. Unfortunately, for high enough costs ($c > \frac{0.5+p}{2-p}$, say $c = \frac{1}{3}$ and $p = 0.1$) then *both* players have a negative continuation value if they do not drop and so must resort to mixed strategies; each must win with equal probability in equilibrium.[11]

CONCLUSION **(2)**: *When $c > p$,*

$$\Pr(1 \text{ win}) = \begin{cases} 0.5(1+p) & c \leq \frac{1-p}{2(2-p)} \\ 1 & c \in \left(\frac{1-p}{2(2-p)}, \frac{1+p}{2(2-p)}\right] \\ \frac{(2-p)c}{[0.5(1+p)+(2-p)c][0.5(1-p)+(2-p)c]} & c > \frac{1+p}{2(2-p)} \end{cases}$$

Player two always wins with probability less than a half; for high enough costs, he will quit. For even higher costs, mixed strategies are invoked and with some probability no one enters the contest! Clearly the latter is a degenerate case.

---

[11]Mixed strategies must leave each player indifferent between staying and quitting, i.e. $EU_i = 0$. It is not hard to show that,

$$\begin{aligned}\Pr(1 \text{ win}) &\equiv \theta_1 = \frac{-0.5(1-p)+(2-p)c}{0.5(1+p)+(2-p)c} \\ \Pr(2 \text{ win}) &\equiv \theta_2 = \frac{-0.5(1+p)+(2-p)c}{0.5(1-p)+(2-p)c}\end{aligned}$$

Then it follows that,

$$\begin{aligned}\Pr(1 \text{ win}) &= (1-\theta_1)\left\{(1-\theta_2)\times\tfrac{1+p}{2} + \theta_2\times 1\right\} \\ &= \tfrac{(2-p)c}{[0.5(1+p)+(2-p)c][0.5(1-p)+(2-p)c]} \\ &= (1-\theta_2)\left\{(1-\theta_1)\times\tfrac{1-p}{2} + \theta_1\times 1\right\} \\ &= \Pr(2 \text{ win})\end{aligned}$$

The first term in the curly bracket is when both contest the election while the second is for when the player is unopposed.

This concludes the solution.

2.3.2 Interpretation

The equilibrium win probability for player one is graphed in (c, p) space in Figure 3. Except for one region, under back-loading player one always does as least as well his opponent. Notice there are no simple comparative statics with respect to changes in either parameter. Still higher fighting costs tend to increase one's chance of winning the nomination (except for very high costs) as does greater election bias, $p \to 0$ (again exempting high cost values). The intuition is that either of these factors increase the chance of two quitting in the final period.

The general advantage of back-loading may be framed using the *value of commitment.*[12] Even when competition is expensive, an early loss does not cause one to exit since he still has two reasonably favorable elections; he can credibly pledge to fight the remainder of the sequence. Alternatively, if two is upset in the beginning he faces difficult elections and an unyielding opponent. Thus two quits if he loses the first election [**(1b)**] or is tied coming into the last election [**(2)**]; in fact for high costs and low bias he will not even contest the first election due to his poor final period prospects.[13]

A related explanation for **(1b)** is the *value of information.* Under the monotone increasing sequence, player two winning the first election does not induce one to quit since this was largely anticipated. A win for one, however, is an informational shock and may disparage two enough that he quits. Only surprising revelations induce quits, and so it is best for a player to first have elections in which big shocks are also favorable shocks.

[12] Of course credible threats is the underlying motivation for subgame perfect refinements of Nash equilibria. I thank David Frankel for suggesting this conceptual structure.

[13] This concept can also explain the one region where one does poorly, [**(1c)**]: costs are just high enough there that he also cannot commit to remain after a first period loss. In this case, being first out of the blocks is best.

Overall, this three period model shows the general advantages of back-loading for a candidate. It also illustrates the inherent complexity of this problem: solving any dynamic, non-stationary game is bound to be troublesome. Still, with the general intuition from this section we will be able generate some results for the arbitrary horizon problem.

2.4 General Horizon Contests

It is difficult to solve the election game as the number of periods grow large due to the multiplicity of contingencies.[14] However, an assumption on costs greatly simplifies the analysis. In this case the general advantage of back-loading generalizes to the arbitrary horizon problem: with enough periods and a monotone sequence of elections, player two does not participate at all and so one wins with certainty.

The cost assumption:

Assumption 4. *Fighting costs exceed the probability of winning an unfavorable state:*

$$c = p + \epsilon, \quad \epsilon \to 0^+$$

An underdog will not contest a decisive election, a fight in which the winner clinches the nomination either due to his opponent's mathematical elimination or quitting the race. Using the intuition from Section 2.3.2, the value of information is paramount: the bad news component of any upset, even in a relatively unbiased election ($p_t \to 0.5^-$), will be enough to induce exit. Notice that back-loading elevates win probability under this restriction in the three period sequence $\{p, 0.5, 1-p\}$, see Figure 3. On a more mundane level, this cost structure greatly reduces the number of contingencies since both players will remain only in a toss-up election.

[14]It is also unclear what is the right generalization of the three period model. For example, how many elections with $p_t = 0.5$ should be included?

I defer justification of Assumption 4 until Section 2.5 where fighting costs become endogenous.

Now the first main result of the paper. Consider the multi-period analogue [in the class of unbiased elections $\mathcal{F}(T)$] of the example from Section 2.3: $\{p, ..., p, 0.5, 1 - p, ..., 1 - p\}$ where there $\frac{T-1}{2}$ elections which favor each candidate. Back-loading increases winning probability:

Proposition 3 *Assume elections are unbiased on average, there is exactly one competitive state, and costs follow Assumption 4. When the states favoring player one come at the end and the even state is in the middle of the cycle, then player one wins with probability greater than half. As either the number of periods grows unbounded ($T \to \infty$) or the bias grows smaller ($p \to 0.5$), player two quits immediately and one is the uncontested winner:*

$$\Pr(\mathit{1\ win}) = \begin{cases} 1 - 0.5(1-p)^{0.5(T-1)} & T \leq T^*(p) \\ 1 & T > T^*(p) \end{cases} \tag{2.9}$$

where $T^*(p) \equiv -2\frac{\ln(1.5-p)}{\ln(1-p)} + 1$ *so* $\frac{\partial T^*}{\partial p} < 0$.

PROOF: See Appendix 2.7.2.

The intuition for the result follows from backwards induction. In the second half of the cycle, elections are biased towards player one, who thus need only quit if mathematically eliminated; in turn, this allows him to fight in all the earlier elections even if he happens to lose every one. However, an upset in the beginning is devastating for two since the remainder of the schedule is unfavorable and his opponent commits to not quitting. Losses induce immediate exit. Thus two gains the nomination only if he garners a majority of the delegates: he must win *every* favorable election as well as the toss-up, i.e. the first $0.5(T + 1)$ elections, which occurs with probability $0.5(1 - p)^{0.5(T-1)}$ and has expected cost approximately $\left[1 - (1 - p)^{0.5(T+1)}\right]$. As the benefit is decreasing and the cost increasing in T, for a long enough horizon two will not even enter the first election. This threshold, $T^*(p)$,[15] is decreasing in p (see Figure 4) meaning smaller bias increases the minimum horizon length needed for an uncontested nomination. The explanation is that the cost term increases in p faster than the winning probability, so with less biased elections it becomes relatively more expensive to remain in the race.[16]

Notice even if two does enter the contest, a longer horizon benefits player one, since this increases the number of elections two must win to mathematically clinch the nomination. Formally, the top line of eq. 2.9 is increasing in T.

More generally Proposition 3 means that the strategic advantage of back-loading strengthens as the schedule lengthens whereas it may be argued that the momentum is likely to dwindle in importance for long contests.[17] As recent primaries have encompassed thirty-five states (see Figure 2), the strategic effect may be the dominating mechanism in practice.

It is important to test the robustness of this result even within the very special class of elections $\mathcal{F}(T)$. For example, what happens as we permute the placement of the toss-up election while keeping player one's favored elections after player two's?

Corollary 4 *Consider the class of T period election games where $0.5(T-1)$ contests favoring player two precede an equal number favoring one. Then if a single toss-up state is inserted at or before the median position in the sequence, player one wins with probability at least half. However, later placement can leave two with an uncontested nomination. There are three possible cases:*

---

[15]That is:

$$\text{two does not contest nomination} \Leftrightarrow T > T^*(p)$$

[16]This contrasts with the intuition in the three period case (Section 2.3.2), since there cost and bias are decoupled.

[17]If momentum stems from voter cues and all voters grow more informed about candidates with time, then the momentum effect should only be important in early elections.

(i) $\tilde{t}(p,T) > T$*:*

$$\Pr(\mathit{1\ win}) = \begin{cases} 1 - 0.5(1-p)^{0.5(t_{0.5}-1)} & t_{0.5} \leq \frac{T+1}{2} \\ 1 - [1 - 0.5(1-p)^{t_{0.5}-0.5(T+1)}](1-p)^{0.5(T-1)} & \frac{T+1}{2} < t_{0.5} \end{cases} \quad (2.10)$$

(ii) $\tilde{t}(p,T) \leq T$ AND $t^*(p) > \frac{T+1}{2}$*:*

$$\Pr(\mathit{1\ win}) = \begin{cases} 1 - 0.5(1-p)^{0.5(t_{0.5}-1)} & t_{0.5} \leq \frac{T+1}{2} \\ 1 - [1 - 0.5(1-p)^{t_{0.5}-0.5(T+1)}](1-p)^{0.5(T-1)} & \frac{T+1}{2} < t_{0.5} \leq \tilde{t}(p,T) \\ 1 & \tilde{t}(p,T) < t_{0.5} \end{cases} \quad (2.11)$$

(iii) $t^*(p) \leq \frac{T+1}{2}$*:*

$$\Pr(\mathit{1\ win}) = \begin{cases} 1 - 0.5(1-p)^{0.5(t_{0.5}-1)} & t_{0.5} \leq t^*(p) \\ 1 & t^*(p) < t_{0.5} \leq t^*(p) + \frac{T-1}{2} \\ 0 & t^*(p) + \frac{T-1}{2} < t_{0.5} \end{cases} \quad (2.12)$$

where $t_{0.5}$ *is the period of the toss-up election,* $t^*(p) \equiv 1 - \frac{\ln(1.5-p)}{\ln(1-p)}$*, and* $\tilde{t}(p,T) \equiv \frac{T+1}{2} - \ln(1-p)^{-1}[\ln[0.5(T-1) - \ln[-\ln(1-p)] + \ln(0.5-p)]$*. Notice that the victory probability exceeds half in every case but the one region in* **(iii)**.

PROOF: See Appendix 2.7.3.

While this is messy looking, in practice it is quite simple; the three possible cases are presented in Figures 5a-5c. Case **(i)** with full participation in the first election requires very short and biased (and so low fighting cost) elections. The second possibility holds under a very narrow range of parameter values. The most likely case is **(iii)** which involves one player not participating in the contest at all (generally player two, see Figure 5c).[18]

A few general points always apply. Pushing the toss-up election closer to the

[18]Notice that Proposition 3 is a special case of **(iii)** where $t^*(p) = 0.5[T^*(p) + 1]$.

median position, $\frac{T+1}{2}$, from the left strictly increases candidate one's winning percentage. The intuition is that two quits if he loses *any* election which precedes the toss-up, so pushing that contest later increases the chance that he exits. In fact if the even election is placed late enough and costs are sufficiently high, then two will not contest even the first election. Approaching from the other side of the median gives the opposite conclusion: one will not enter the race when the toss-up comes at the end of the schedule. The explanation is that since two has a not unfavorable election at the end of the cycle, late losses need not induce exit; one then faces a long series of costly battles and a committed opponent. There is an exception: if costs are moderate, one can commit to remaining even if the toss-up is the very last election and candidate two may decide not to enter the contest at all since clinching the nomination will require too many periods of combat.

In addition, less biased elections ($p \to 0.5$) tend to benefit player one. When no one quits his victory probability is increasing while the cutoff values, $t^*(p)$ and $\tilde{t}(p, T)$, are decreasing in p; the latter means higher p may preclude player two from entering the contest. The intuition is that as p increases fighting costs rise and this induces more exit. This also explains why less biased elections may hurt player one when the toss-up election falls at the end of the cycle, since this case will require him to fight for more periods.

This corollary shows that the general advantage of back-loading is robust to shifts in the toss-up election. We can be confident that strategic effects tend to advantage a candidate who has his best elections at the end of the cycle.

In the next section players are allowed to select the value of fighting costs at every node. The level will in general correspond to the form supposed in Assumption 4.

2.5 Endogenous Costs

In this section we drop Assumptions 3 and 4 and allow endogenous costs. In each election a subgame is added in which players "bid" a cost. Only the high bidder is allowed to continue so candidates will bid up to their continuation value at that node. This means either one player quits ($p_t \neq 0.5$) or both players have the same continuation value ($p_t = 0.5$) in every election, and so results in less cumbersome proofs than with exogenous costs. More important, the general spirit of the results from the previous section hold true in this more general framework.

The formal structure:[19]

Cost-Bidding Subgame: At every node $W(t)$ both players simultaneously announce costs c_i. The low bidder ($c_i < c_{-i}$) is allowed to match or quit. If he matches, both pay the larger bid and the game continues as before. Otherwise, neither have to pay current fighting costs and the high bidder captures the nomination.

This process naturally captures the intended role of fighting costs. If c is considered political advertising, this says that a candidate who is unwilling to match his opponent's expenses must drop from the election; the smoothed cost choice captures the spirit of the static campaign spending literature (Snyder [19]) and the broader question of effort choice in tournament competition (Rosen [18]). Alternatively, interpreting costs as political capital, then when one candidate raises the stakes in a particular state (through political spin-doctoring), then the election becomes equally crucial to the other contender.

Since players are willing to bid up to their continuation value, the leader can force his opponent to exit with a high enough cost. Such a "knock-out" expenditure strategy is the often employed by front-runners, e.g. see Wilcox [20].

[19]This also means that strategies are no longer simply exit decisions but instead are a first and second stage bid for all possible contingencies. It is not hard to modify the relevant notation in Section 2.2.2.

With this intuition we can rapidly analyze the general horizon game from Proposition 3:

Proposition 5 *Assume elections are unbiased on average, there is exactly one competitive state, and now costs are determined through the cost-bidding subgame (rather than Assumptions 3 and 4). When the states favoring player one come at the end and the even state is in the middle of the cycle, then player one is the unopposed winner of the party nomination:*

$$\Pr(\textit{1 win}) = 1 \tag{2.13}$$

PROOF: See Appendix 2.7.4.

Player one can set costs high enough in later periods, where he is favored, to force two's departure. This means that the median positioned toss-up election is decisive if two has won all prior elections. Thus candidate two is disadvantaged in the immediately preceding election since an upset loss there results in his exit while even a win will only leave him on equal ground with his opponent. Hence one can bid a high enough cost to drive two out here and in any earlier period even if two has won all elections up to that point. In particular, two will exit from the very first election and so hand the first candidate an uncontested nomination.

This result shows the advantage of back-loading is enhanced under endogenous costs. The key difference from before is that player one can set costs high enough in the election immediately preceding the toss-up to drive out his opponent (even if two has won all prior elections). In the exogenous costs case player one could not "go for the kill" in this election and so could have allowed two to remain in the game.

The final step is to check the outcome of endogenous cost games when the position of the toss-up election is rotated.

Corollary 6 *Consider the class of T period election games where $0.5(T-1)$ contests favoring player two precede an equal number favoring one. Period fighting costs are* *determined through the cost-bidding subgame. Then if a single toss-up state is inserted after the first election and no later than the median election, player one gains the nomination with probability one. Later placement results in a certain win for two while a first-period toss-up is a decisive election:*

$$\Pr(\mathit{1\ win}) = \begin{cases} 1 & 1 < t_{0.5} \leq \frac{T+1}{2} \\ 0.5 & t_{0.5} = 1 \\ 0 & t_{0.5} > \frac{T+1}{2} \end{cases} \tag{2.14}$$

where $t_{0.5}$ is the position of the unbiased election.

PROOF: See Appendix 2.7.5.

The intuition follows immediately from Proposition 5 and Corollary 4. Since players compete away rents during the toss-up election, in the prior period the player who clinches the nomination with a win has a strictly higher continuation value. Hence he can set costs high enough to force the other to exit, which, through a (backwards) induction argument, can extend all the way back to the first election. When the toss-up comes after the median, nodes in which candidate one is a loss away from being mathematically eliminated means two is advantaged with the converse holding when the toss-up comes in the beginning (focus there on nodes in which two is a loss away from quitting next period). This also shows why $t_{0.5} = 1$ has a special status: there is no earlier election where the leader can exploit his position and set high costs.[20]

While this corollary is not entirely supportive of the back-loading advantage, one variant of this concept does explain the pattern of wins. In cases where one loses for sure, the toss-up election comes at the end of the cycle. This is in one sense back-loading for candidate two, since he now has more to look forward to at the end of the

[20]In the previous section when $c = p^+$, one also won the tournament with probability one-half when the toss-up election was first, see Corollary 4.

cycle, and front-loading for one, since his earliest favored election is pushed forward to the median period.

2.6 Conclusion

The sequential nature of primaries helps shape the outcome. I have tried to explain drop-out behavior in a model of forward-looking rational expectations. Behavior in later periods influences candidate behavior in the beginning of the cycle. Generally, back-loaded elections increase victory percentage because it allows credible commitment to stay the course even in the face of consecutive losses; at the same time, the opponent who is favored in the initial elections must exit if he is upset even once.

Unfortunately the model has few testable implications beyond the somewhat prosaic "candidates only drop-out after an upset."[21] Still, the general intuition allows us to understand the strategic decisions of candidates in a two-man race. Return to the 1976 Republican primary discussed in the introduction. If momentum was the dominating factor, then Reagan should have quit after he lost the first four largely competitive elections. But recognizing his best assets came later he persevered. Momentum did not bandwagon for Ford. In this aspect the model is helpful. However, testable empirical implications involve one player conceding; as both Reagan and Ford contested all elections, it is unclear how important are strategic quits.[22]

The model also sheds light on why there is so little dispute over the order of primaries.[23] If momentum is the sole dynamic force, then every state should seek to be first in the primary schedule to magnify their importance in the nomination process.[24] Alternatively, we should see many "favorite-sons" from the early states of

[21] There is another problem: there have been only two examples of two candidate contests since primaries became binding in 1972.

[22] The other leading two-man race, the 1980 Democratic contest between Carter and Kennedy, also had both contesting all elections though the momentum effect was stronger than in 1976.

[23] Thomas Piketty suggested this point.

[24] This presumes states value their role in the nomination process. State party officials, who effectively control the timing of the election, may prefer a late primary so as to not cede power to

New Hampshire and Iowa running. Neither of these is true. From 1972 (the first contest where state votes were binding on convention delegates) to 1992 the primary schedule was relatively stable. Figure 2 shows that the only significant innovations during this period were an increase in number of elections and the creation of the "Super-Tuesday" primary date in early March. As for the other point, no recent nominee and only a few actual candidates were from an early primary state. These facts together suggest that momentum is not too powerful. One explanation is that there is an opposing force such as the strategic effect of this paper.

The strategic theory also has implications for changes in the 1996 schedule. Several states, notably California and New York, have shifted their primaries to an earlier date (though still not to the first position). The conventional wisdom is that this will boost the fortunes of candidates from the leap-frogging states, *The New York Times* [1]. This paper suggests a muted or even opposite effect.

There are a variety of extensions which would undoubtedly make the model more realistic. For example, it is not just the ordinal ordering but also the time between elections that plays a role. A longer period allows for momentum to dissipate and so may lead the strategic effect to become relatively more important. One approach would be a continuous time formulation with elections coming at discrete points. This renders the problem as a game-theoretic version of the Black-Scholes option pricing model which is in general intractable (see for example Huang-Li [14]), so this point serves mainly as a caveat to the policy implications.

Other issues should be possible to incorporate into the model. Foremost the game must allow multiple players. The early dynamics of most primaries are characterized by a "winnowing out" of a plethora of weaker contenders. Such N-person games in general have multiple equilibria and will likely require using cooperative theory. Similarly, this model takes the presence of candidates as exogenous whereas the entry decision is based on an implicit coordination process. It seems possible to add a "pre-

national leaders and to maintain focus on local issues.

game" period where potential candidates from some pool decide whether or not to enter the contest. The final extension would allow candidates to refine their estimate of the election bias as the primary unfolds.[25] This should lead to a strategic variant of the momentum effect as early losses lead a candidate to reassess downwards his future prospects, possibly inducing him to exit from the race.

[25] I thank Dan Kessler for this suggestion.

2.7 Appendix: Proofs

2.7.1 Proof of Lemma 2

The first point is self-evident: a mathematically eliminated candidate cannot win a primary in which no one quits and his opponent can credibly commit to remain the rest of the way. Thus to minimize fighting costs he should exit immediately.

Now for **(ii)** let $W_{\overline{w}}(t)$ indicate that one has won $\overline{w}$ prior elections and presume that two quits at this node (the proof for one is symmetric); than, $V_1[W_{\overline{w}}(t)] = 1$ and $V_2[W_{\overline{w}}(t)] = 0$. The claim is that two will also quit at all $W_n(u)$ where $n \geq \overline{w}$ and $u \leq t$. Now fixing on some period, s, increasing the number of one's prior wins leaves him no worse off and his opponent no better:

$$\begin{aligned} V_1[W_m(s)] &\geq V_1[W_n(s)] \\ V_2[W_m(s)] &\leq V_2[W_n(s)] \end{aligned} \tag{2.15}$$

for $\forall m > n$. The proof follows from induction on the time index. The result trivially holds for $s = T$. Assume that the result holds for period $s+1$ and show it for period s. It will be sufficient to prove the claim for two adjacent nodes, $W_n(s)$ and $W_{n+1}(s)$:

$$\begin{aligned} V_i[W_n(s)] &= \max\langle p_s V_i[W_{n+1}(s+1)] + (1-p_s)V_i[W_n(s+1)] - c_s, 0\rangle \\ V_i[W_{n+1}(s)] &= \max\langle p_s V_i[W_{n+2}(s+1)] + (1-p_s)V_i[W_{n+1}(s+1)] - c_s, 0\rangle \end{aligned}$$

where the max is used to allow for the option of quitting. Then eq. 2.15 follows from the inductive assumption and the fact that a linear combination of no smaller (no greater) terms is itself no smaller (no greater). Continual application of eq. 2.15 shows point **(ii)**. For example, compare our original term $W_{\overline{w}}(t)$ to the node in the previous period with the same state variable, $W_{\overline{w}}(t-1)$:

$$V_i[W_{\overline{w}}(t-1)] = \max\langle p_{t-1}V_i[W_{\overline{w}+1}(t)] + (1-p_{t-1})V_i[W_{\overline{w}}(t)] - c_{t-1}, 0\rangle$$

From the inductive result, $V_2[W_{\overline{w}}(t-1)] \leq V_2[W_{\overline{w}}(t)] = 0$ and $V_1[W_{\overline{w}}(t-1)] \geq 0$ [since $V_1[W_{\overline{w}}(t)] = V_1[W_{\overline{w}+1}(t)] \equiv 1$ and $c_{t-1} < 1$]. Hence two quits and one stays.

The intuition for **(iii)** is straightforward: if false than the player's continuation value must have been negative the period before, so he should have dropped out then and saved the fighting cost, contradiction. The result will be proved for player one. Presume at node $W_n(t)$ he continues, wins and then quits. So,

$$V_1[W_n(t)] = \max\langle p_t V_1[W_{n+1}(t+1)] + (1-p_t)V_1[W_n(t+1)] - c_t, 0\rangle$$

Since he quits following a win, $V_1[W_{n+1}(t+1)] < 0$; by **(ii)**, $V_1[W_n(t+1)] < V_1[W_{n+1}(t+1)]$ so the first term in the brackets is negative. It is optimal to quit now.

□

2.7.2 Proof of Proposition 3

It will be easiest to follow the proof using the graphical depiction in Diagram 2a. The triangular regions on the right are nodes in which one player has been mathematically eliminated. Again we will proceed using backwards induction.

Consider the ordered sequence $\{p, ..., p, 0.5, 1-p, ..., 1-p\}$ with equal number of p and $1-p$ terms. At stage T, only at $W(T) = \frac{T-1}{2}$ will neither have mathematically clinched; in this decisive election, two quits since $\tilde{V}_2 = p - c < 0$ and $\tilde{V}_1 = (1-p) - c > 0$ [$c = p^+ < 1-p$]. In election $T-1$, there are two nodes where neither has cliched. From Lemma 2 **(ii)** two quits at $W(T-1) = \frac{T-1}{2}$. He also exits at $W(T-1) = \frac{T-3}{2}$ since it is decisive (the loser exits) and $\tilde{V}_2 < 0 < \tilde{V}_1$ just as with $W(T) = \frac{T-1}{2}$. This can be repeated inductively for all $t \in (\frac{T+3}{2}, T]$: initially there are $T-t+1$ undecided nodes and from the solution at the $T-t$ later states along with application of Lemma 2 **(ii)**, two quits at the $T-t$ nodes with highest values of $W(t)$. In the one remaining case, $W(t) = \frac{2t-T-1}{2}$, two quits since $\widetilde{V}_2 = p - c < 0 < (1-p) - c = \widetilde{V}_1$ (NB: this is the node on the diagonal right above the bottom triangle in Diagram 2a). The earliest election favoring player one, $t = \frac{T+3}{2}$, has two exit at all states except $W(\frac{T+3}{2}) = 0$ where he has mathematically clinched (one quits there).

The toss-up election with $p_t = 0.5$, $t = \frac{T+1}{2}$, is next. While no one has mathematically clinched at any node, from the last paragraph and Lemma 2 **(ii)** player two must quit at $W(\frac{T+1}{2}) = 1, 2, ..., \frac{T+1}{2}$. Both players contest the remaining node which is a decisive election:

$$\begin{aligned}\widetilde{V}_1[W(\tfrac{T+1}{2}) = 0] &= 0.5 - c > 0 \\ \widetilde{V}_2[W(\tfrac{T+1}{2}) = 0] &= \end{aligned}$$

The remaining stages $t = 1, ..., \frac{T-1}{2}$ all favor player two, $p_t = p$. Consider $t = \frac{T-1}{2}$. From our work so far and Lemma 2 only at $W(\frac{T-1}{2}) = 0$ will two remain, and there,

$$\begin{aligned}\widetilde{V}_1[W(\tfrac{T-1}{2}) = 0] &= (1-p)\widetilde{V}_1[W(\tfrac{T+1}{2}) = 0] + p\widetilde{V}_1[W(\tfrac{T+1}{2}) = 1] - c \\ &= (1-p)(0.5-p) + p(1) - c \\ &= 0.5(1+p) - (2-p)c \geq 0 \\ \widetilde{V}_2[W(\tfrac{T-1}{2}) = 0] &= (1-p)(0.5-p) + p(0) - c \\ &= 0.5(1-p) - (2-p)c \lesseqgtr 0\end{aligned}$$

Either player two quits, in which case repeated application of Lemma 2 shows he never enters, or he remains. Continuing with the latter case, consider any $t = 1, ..., \frac{T-3}{2}$ and presume two will not quit at $W(u) = 0$ for $u > t$. We know he will exit at any $W(t) > 0$, so the only way that two gains the nomination is if he wins every one of his remaining favored elections plus the toss-up. It is not hard to show that,

$$\begin{aligned}
\tilde{V}_2[W(t)=0] &= 0.5(1-p)^n - \left\{p \sum_{s=1}^{n}(1-p)^s s + (1-p)^{n+1} n + 1\right\} c \\
&= 0.5(1-p)^n - \left\{p(1-p) \sum_{s=1}^{n} \tfrac{\partial}{\partial(1-p)}(1-p)^s + (1-p)^{n+1} n + 1\right\} c \\
&= 0.5(1-p)^n - \left[1-(1-p)^{n+1}\right] \tfrac{c}{p} \\
&\approx 0.5(1-p)^n - \left[1-(1-p)^{n+1}\right]
\end{aligned}$$

where $n \equiv \frac{T+1}{2} - t$, the number of elections until the toss-up. Using Lemma 1,

$$\begin{aligned}
\tilde{V}_1[W(t)=0] &\approx \left[1-0.5(1-p)^n\right] - \left[1-(1-p)^{n+1}\right] \\
&= (0.5-p)(1-p)^n
\end{aligned}$$

which is strictly positive, i.e. one can always commit to remain. For either player, the first term on the top line is the benefit of remaining, the probability of winning, while the second is the expected cost of remaining. For player two, the benefit is falling and the cost rising in n, so if there are enough periods (T large and consider $t = 1$) than he exits. In general two quits n periods before the toss-up if $n > n^*(p)$ where,

$$n^*(p) \equiv \frac{-\ln(1.5-p)}{\ln(1-p)} > 0$$

Notice that,

$$\frac{\partial n^*}{\partial p} < 0$$

This means a longer horizon is required to induce two's exit as the elections grow more biased.

Hence, what we have shown is that for the entire game:

$$\Pr(1 \text{ win}) = \begin{cases} 1 - 0.5(1-p)^n & T \leq T^*(p) \\ 1 & T > T^*(p) \end{cases}$$

where $T^*(p) \equiv 2n^*(p) + 1$ so $\frac{\partial T^*}{\partial p} < 0$.

Diagram 2b shows the exit and continuation decisions at all nodes for the case where $T > T^*(p)$.

□

2.7.3 Proof of Corollary 4

The proof proceeds by showing the existence and solving for the three participation cutoff values. The final step is to show the relationship of the exit conditions in the intervals between these cutoffs.

Denote the period of the toss-up election as $t_{0.5}$. First consider the set of games where $t_{0.5} < \frac{T+1}{2}$, the fair election is before the median position in the sequence: $\{p, ..., p, 0.5, p, ..., p, 1-p, ..., 1-p\}$. As the last half of the cycle is identical to the case from Proposition 3, from that proof we know that in periods $t \in [\frac{T+3}{2}, T]$ that two will quit at every node except those in which he has mathematically clinched; in particular, for period $\frac{T+3}{2}$ he remains only for $W(\frac{T+3}{2}) = 0$. From Lemma 2 **(ii)** two quits at $W(t) > 0 \; \forall t \leq \frac{T+3}{2}$, i.e. losing any of the early elections forces him to retire from the contest. However for elections which favor candidate two following the toss-up, one exits if he has not won an state yet, $W(t=0) \; \forall t \in (t_{0.5}, \frac{T+1}{2}]$; to see this, observe $\tilde{V}_1[W(\frac{T+1}{2}) = 0] = p = c < 0 < 1 - p - c = \tilde{V}_2[W(\frac{T+1}{2}) = 0]$ and roll back the game tree period-by-period. At $t = t_{0.5}$ both players get $\tilde{V}_i[W(t_{0.5}) = 0] = 0.5 - c > 0$ so they continue. For the earlier nodes, $t < t_{0.5}$:

$$\begin{aligned}
\tilde{V}_2[W(t) = 0] &= [0.5(1-p)^n] - \left[p \sum_{s=1}^{n} (1-p)^s s + (1-p)^n n(1-p) + 1\right] c \\
&= [0.5(1-p)^n] - \left\{\tfrac{1-p}{p}[1 - (1+np)(1-p)^n] + (1-p)^n n(1-p) + 1\right\} c \\
&\approx -1 + (1.5 - p)(1-p)^n \\
&\lesseqgtr 0
\end{aligned}$$

where $n \equiv t_{0.5} - t$. From Lemma 1,

$$\begin{aligned}\widetilde{V}_1[W(t) = 0] &\approx [1 - 0.5(1-p)^n] - [1 - (1-p)^{n+1}] \\ &= (0.5 - p)(1-p)^n \\ &> 0\end{aligned}$$

Notice that as $n \uparrow$ player two's benefits (first term on top line) falls and costs (second term) rise; as $n \to \infty$ (i.e. $t_{0.5}$ approaches the median position $\frac{T+1}{2}$ and T large) then he surely quits. Let $t^*(p)$ be the cutoff for $t_{0.5} < \frac{T+1}{2}$ so that two quits the first election (he does not enter the contest) if $t_{0.5} > t^*(p)$:

$$\begin{aligned}t^*(p) &\equiv \arg\min_{t_{0.5}} \Big\{ \widetilde{V}_2[W(1) = 0] |_{t_{0.5}} < 0 \Big\} \\ &= 1 - \tfrac{\ln(1.5-p)}{\ln(1-p)}\end{aligned} \tag{2.16}$$

It is not hard to show that

$$\frac{\partial t^*(p)}{\partial p} < 0$$

Finally,

$$\Pr(1 \text{ win}) = \begin{cases} 1 - 0.5(1-p)^{t_{0.5}-1} \geq 0.5 & t_{0.5} \leq \min\Big\langle t^*(p), \frac{T+1}{2} \Big\rangle \\ 1 & t_{0.5} \in (t^*(p), \frac{T+1}{2}] \end{cases} \tag{2.17}$$

Player one does unambiguously better under less biased elections ($p \to 0.5$) which either induce player two to quit ($\frac{\partial t^*(p)}{\partial p} < 0$) or increase one's victory probability if he stays (top line of eq. 2.17). Notice also that an extra case, $t_{0.5} = \frac{T+1}{2}$, solved in Proposition 3, has been included in the formula.

Turn to the remaining cases, $t_{0.5} > \frac{T+1}{2}$: $\{p, ..., p, 1-p, ..., 1-p, 0.5, 1-p, ..., 1-p\}$. Following the previous paragraph two quits at all nodes without a mathematical winner for $\forall t > t_{0.5}$. At $t_{0.5}$ and $W(t_{0.5}) = 0.5(2t_{0.5} - T - 1)$, the node right above the two clinch range, both players have equal continuation value: $\widetilde{V}_i[W(t_{0.5}) = 0.5(2t_{0.5} - T - 1)] = 0.5 - c > 0$. As we move back to earlier periods on this diagonal $[W(t) =$

$0.5(2t - T - 1)]$, one's benefit of remaining diminishes since losing any single election means he is mathematically eliminated. In the extreme case, $\frac{T+1}{2}$, presuming that one has not quit along the diagonal,

$$\begin{aligned} \widetilde{V}_1[W(\tfrac{T+1}{2}) = 0] &= 0.5(1-p)^n - \left\{p \sum_{s=1}^{n} (1-p)^s s + (1-p)^n n(1-p) + 1\right\} c \\ &\approx 0.5(1-p)^n - [1 - (1-p)^{n+1}] \\ &= -1 + (1.5 - p)(1-p)^n \end{aligned} \tag{2.18}$$

with $n \equiv t_{0.5} - \frac{T+1}{2}$. It is not hard to show $\widetilde{V}_2[W(\frac{T+1}{2}) = 0] > 0$, so one quits when $n > n^{**}(p) \equiv \frac{-\ln(1.5-p)}{\ln(1-p)}$ or $t_{0.5} > t^{**}(p)$ where

$$t^{**}(p) \equiv \frac{T+1}{2} - \frac{\ln(1.5-p)}{\ln(1-p)} \tag{2.19}$$

Now examine each case, $t_{0.5} > t^{**}(p)$ and $t_{0.5} \in (\frac{T+1}{2}, t^{**}(p)]$, in turn.

When the toss-up election is placed at the end of the cycle, $t_{0.5} > t^{**}(p)$, by definition one drops out at $W(\frac{T+1}{2}) = 0$. But then he must also quit at $W(t) = 0$ for all earlier elections, $t < \frac{T+1}{2}$, which are biased towards candidate two. For example,

$$\begin{aligned} \widetilde{V}_1[W(\tfrac{T-1}{2}) = 0] &= pV_1[W(\tfrac{T+1}{2}) = 1] + (1-p) \times 0 - c \\ &< 0 \end{aligned}$$

as you can show $V_1[W(\frac{T+1}{2}) = 1] = 1$ and $p < c$ while

$$\begin{aligned} \widetilde{V}_2[W(\tfrac{T-1}{2}) = 0] &= pV_2[W(\tfrac{T+1}{2}) = 1] + (1-p) \times 1 - c \\ &> 0 \end{aligned}$$

This equation can be iterated all the way back to $W(1) = 0$ and so one chooses not to participate in the election at all:

$$\Pr(1 \text{ win}) = 0 \quad \text{if } t_{0.5} > t^{**}(p) \tag{2.20}$$

Finally consider when the toss-up is shortly after the median, $t_{0.5} \in (\frac{T+1}{2}, t^{**}(p)]$. This means one remains at $W(\frac{T+1}{2}) = 0$. Notice that two will exit along the diagonal $W(t) = 0.5(2t - T + 1)$ for all $t \geq \frac{T+1}{2}$. To see this first consider the node at which a loss for player one results in the decisive election at $t_{0.5}$:

$$\begin{aligned} \widetilde{V}_1[W(t_{0.5} - 1) = 0.5(2t_{0.5} - T - 1)] &= p(0.5 - c) + (1 - p) \times 1 - c \\ &> 0 \end{aligned}$$

and

$$\begin{aligned} \widetilde{V}_2[W(t_{0.5} - 1) = 0.5(2t_{0.5} - T - 1)] &= p(0.5 - c) + (1 - p) \times 0 - c \\ &< 0 \end{aligned}$$

so two exits here. For the remaining terms on the diagonal:

$$\begin{aligned} \widetilde{V}_1[W(t) = 0.5(2t - T + 1)] &= pV_1[W(t + 1) = 0.5(2t - T + 1)] + (1 - p) \times 1 - c \\ &> 0 \end{aligned}$$

while

$$\begin{aligned} \widetilde{V}_2[W(t) = 0.5(2t - T + 1)] &= pV_2[W(t + 1) = 0.5(2t - T + 1)] + (1 - p) \times 0 - c \\ &< 0 \end{aligned}$$

Repeated application of Lemma 2 **(ii)** shows that two quits every election for $t \in [1, \frac{T+1}{2}]$ in which $W(t) > 0$. Consider the remaining nodes, $W(t) = 0$ and $t \leq \frac{T+1}{2}$, those where two has neither lost an election nor clinched. As $t \to 1$, the expected probability of one winning increases assuming (for now) that no one quits at $W(u) = 0 \; \forall u \in (t, \frac{T+1}{2}]$. To show this contention first solve for $W(\frac{T-1}{2}) = 0$, the last election biased towards player two:

$$\begin{aligned}\widetilde{V}_1[W(\tfrac{T-1}{2})=0] &= p\times 1+(1-p)\times V_1[W(\tfrac{T+1}{2})=0]-c\\ &= [p+0.5(1-p)^{n+1}]-[1-(1-p)^{n+2}]\tfrac{c}{p}\\ &\approx [-1+(1.5-p)(1-p)^n](1-p)\\ &> 0\end{aligned}$$

substituting $V_1[W(\frac{T+1}{2})=0]$ from eq. 2.18 and $n\equiv t_{0.5}-\frac{T+1}{2}$; the inequality follows from the presumption that $V_1[W(\frac{T+1}{2})=0]>0$, i.e. $t_{0.5}\leq t^{**}(p)$, see eq. 2.18. For his opponent,

$$\begin{aligned}\widetilde{V}_2[W(\tfrac{T-1}{2})=0] &= [1-p-0.5(1-p)^{n+1}]-[1-(1-p)^{n+2}]\tfrac{c}{p}\\ &\approx (0.5-p)(1-p)^{n+1}-p\end{aligned}$$

which is in general positive. Using this fact, we can derive the continuation value at $W(t)=0$ for any $t<\frac{T-1}{2}$:

$$\begin{aligned}\widetilde{V}_1[W(t)=0] &= \left[p\sum_{s=0}^{m-1}(1-p)^s+0.5(1-p)^{n+m}\right]-\\ &\quad \left\{1+p\sum_{s=1}^{m-1}(1-p)^s s+(1-p)^m[(m-1)+p^{-1}(1-(1-p)^{n+1})]\right\}c\\ &= [1-(1-p)^m+0.5(1-p)^{n+m}]-[1-(1-p)^{m+n+1}]\tfrac{c}{p}\\ &\approx [(1.5-p)(1-p)^n-1](1-p)^m\\ &> 0\end{aligned}$$

where $m\equiv\frac{T+1}{2}-t$. The inequality again follows from $V_1[W(\frac{T+1}{2})=0]>0$ and eq. 2.18. Then for player two:

$$\begin{aligned}\widetilde{V}_2[W(t)=0] &= [(1-p)^m-0.5(1-p)^{n+m}]-[1-(1-p)^{m+n+1}]\tfrac{c}{p}\\ &\approx -1+(1-p)^m+(0.5-p)(1-p)^{m+n}\\ &\lesssim 0\end{aligned}$$

Notice that $\frac{\partial \widetilde{V}_2[W(t)=0]}{\partial m} < 0$ so the value falls as $t \to 1$. In particular for the very first election,

$$\begin{aligned}\widetilde{V}_2[W(1)=0] &\approx -1 + (1-p)^{0.5(T-1)} + (0.5-p)(1-p)^{t_{0.5}-1} \\ &= -1 + [1 + (0.5-p)(1-p)^{t_{0.5}-0.5(T+1)}](1-p)^{0.5(T-1)}\end{aligned}$$

which is decreasing in $t_{0.5}$, T, and p. The intuition for the $t_{0.5}$ comparative static is that when the toss-up election is further in the future it becomes less valuable to player two when he is on the diagonal $W(t) = 0.5(2t-T-1)$; if $t_{0.5}$ is one period later, the cost of the extra fight (c) offsets the new opportunity of mathematically clinching (occurs with probability p), so the dominating effect is the reduction in probability of ever reaching the toss-up election. Let $\tilde{t}(p,T)$ be the cutoff for $t_{0.5} > \frac{T+1}{2}$ so that two quits the first election (he does not enter the contest) if $t_{0.5} > \tilde{t}(p,T)$ and $t_{0.5} \leq \min\langle t^{**}(p), T\rangle$:

$$\begin{aligned}\tilde{t}(p,T) &\equiv \arg\min_{t_{0.5}}\left\{\widetilde{V}_2[W(1)=0]|_{t_{0.5}} < 0\right\} \\ &\approx \tfrac{T+1}{2} - \ln(1-p)^{-1}\left\{-\ln[0.5(T-1)] - \ln[-\ln(1-p)] + \ln[0.5-p]\right\}\end{aligned} \quad (2.21)$$

where the second line follows from a first-order Taylor series approximation on the $\widetilde{V}_2[W(1)=0]$ equation. It is not hard to show,

$$\frac{\partial \tilde{t}(p,T)}{\partial p} < 0$$

Hence we have,

$$\Pr(1 \text{ win}) = \begin{cases} 1 - \left[1 - 0.5(1-p)^{t_{0.5}-0.5(T+1)}\right](1-p)^{0.5(T-1)} & t_{0.5} \in (\frac{T+1}{2}, \min\left\langle\tilde{t}(p,T), T\right\rangle] \\ 1 & t_{0.5} \in (\tilde{t}(p,T), \min\langle t^{**}(p), T\rangle] \end{cases} \quad (2.22)$$

The top line is always greater than half and decreasing in $t_{0.5}$ (a later toss-up means there is a greater chance of losing a favored election before reaching $t_{0.5}$); also it is increasing in p.

This exhausts the set of possible games. All that remains to do is to compare the various cut-off values: equations 2.16, 2.19 and 2.21. First notice that $t^{**}(p) = t^{*}(p) + \frac{T-1}{2}$, see eq. 2.16 and 2.19, so player one not contesting the nomination when $t_{0.5} = \bar{t} > \frac{T+1}{2}$ is equivalent to the condition that two never enter when $t_{0.5} = \bar{t} - \frac{T-1}{2} \leq \frac{T+1}{2}$ [Proof: Take the case of $\bar{t} = T$. One not contesting any election $\Rightarrow T > t^{**}(p) \Leftrightarrow \frac{T-1}{2} > -\frac{\ln(1.5-p)}{\ln(1-p)}$ from eq. 2.19. This is identical to the condition for two to quit at $t_{0.5} = \frac{T+1}{2}$, see eq. 2.16, and hence not enter at all]. Due to this result the variable $t^{**}(p)$ is not used in the statement of the corollary. Second, if two does not enter the contest when $t^{*}(p) < t_{0.5} \leq \frac{T+1}{2}$, then he will not seek the nomination for any $\frac{T+1}{2} < t_{0.5} \leq t^{**}(p)$. In other words, we can ignore the $\tilde{t}(p,T)$ threshold, see eq. 2.22, when parameters values induce two to exit for some $t_{0.5} \leq \frac{T+1}{2}$ [Proof: This follows from comparing the functional forms. Some simple algebra shows that $\tilde{t}(p,T) - \frac{T+1}{2} < t^{*}(p) - 1 \Leftrightarrow \frac{0.5-p}{[0.5(T-1)][-\ln(1-p)]} < 1.5 - p$ which must hold for $T \geq 3$ and $p \leq 0.5$. Intuitively, if $\frac{T+1}{2} < t_{0.5} < t^{**}(p)$ then by definition one will not exit along the $W(t) = 0.5(2t - T - 1)$ diagonal. So from the perspective of the very first period, two is worse off than even $t_{0.5} = \frac{T+1}{2}$ (where he quits) since he has to incur more fighting costs until the toss-up]. This means that in sequences where $t^{*}(p) < t_{0.5}$ candidate two will not enter when $t_{0.5} \in (t^{*}(p), t^{**}(p)]$. Alternatively, it is possible for two to quit only when the toss-up is after the median position [when $\frac{T+1}{2} < t^{*}(p)$ and $\tilde{t}(p,T) < T$]. The claims from this paragraph show that there are only the three possible cases listed in the corollary.

□

2.7.4 Proof of Proposition 5

Candidate two quits during $t \in [\frac{T+3}{2}, T]$, the elections favoring player one, at all nodes in which no one has mathematically clinched the nomination, $W(t) \in [\frac{2t-T-1}{2}, \frac{T-1}{2}]$.

To see this first consider $W(T) = \frac{T-1}{2}$ which is a decisive election. Since the winner gains the nomination either player is willing to bid up to his continuation value, $c_1 = 1 - p > p = c_2$. So one's optimal bid is any $c \in (p, 1-p]$ which forces two's exit here and, from Lemma 2 **(ii)**, at $W(t) = \frac{T-1}{2}$ for all earlier periods. Using this approach we can inductively work down the diagonal $W(t) = 0.5(2t - T - 1)$: each node is decisive as a win for two clinches the nomination while a win for one means that two quits in the next period (from the inductive assumption); since all elections favor candidate one, he can outbid two and force his exit.

For the earlier periods, $t \in [1, \frac{T+1}{2}]$, the only nodes left are $W(t) = 0$. At the toss-up election, $t = \frac{T+1}{2}$, both players bid the same cost (at a value up to $p_t = 0.5$) and have the same continuation value, $V_1[W(\frac{T+1}{2}) = 0] = V_2[W(\frac{T+1}{2}) = 0]$. In the preceding period only player one has a positive continuation value since a win here forces two to exit (from the last paragraph):

$$\begin{aligned}
\widetilde{V}_1[W(\tfrac{T-1}{2}) = 0] &= (1-p) \times V_1[W(\tfrac{T+1}{2}) = 0] + p \times V_1[W(\tfrac{T+1}{2}) = 1] - \widetilde{c} \\
&= (1-p) \times 1 + p \times V_1[W(\tfrac{T+1}{2}) = 1] - \widetilde{c} \\
&> (1-p) \times 0 + p \times V_2[W(\tfrac{T+1}{2}) = 1] - \widetilde{c} \\
&= \widetilde{V}_2[W(\tfrac{T-1}{2}) = 0]
\end{aligned}$$

where $\widetilde{c}$ is the equilibrium fighting cost if both players continue. The players are willing to set $c_1 \geq 1 - p > c_2$, so two will be forced to exit. Lemma 2 **(ii)** again shows that two must quit at $W(1) = 0$ and so will not even contest the nomination.

□

2.7.5 Proof of Corollary 6

The proof parallels that of Corollary 4. For sequences in which $t_{0.5} < \frac{T+1}{2}$, $\{p, ..., p, 0.5, p, ..., p, 1-p, ..., 1-p\}$, the analysis for periods $t_{0.5} > \frac{T+1}{2}$ is identical to Proposition 5. We saw there that candidate two quits at every node above the diagonal $W(t) = 0.5(2t-T-1)$

and from Lemma 2 **(ii)** he also quits at $W(t) > 0 \; \forall t \leq \frac{T+3}{2}$. The only remaining nodes are $W(t) = 0$ for $t \leq \frac{T+1}{2}$. Now $W(\frac{T+1}{2}) = 0$ is a decisive election: if two wins he mathematically clinches while if he loses he must exit. Since this election favors two, he sets $c_2 = p^+$ and forces one to exit. Applying this same technique inductively to $W(t) = 0$ for $t \in (t_{0.5}, \frac{T+1}{2}]$ we find that two prices one out of the election with $c_2 = p^+$. $W(t_{0.5}) = 0$ is also decisive: both win with equal probability, bid some fighting cost $c \in [0, 0.5]$, and have equal continuation value, $V_1[W(\frac{T+1}{2}) = 0] = V_2[W(\frac{T+1}{2}) = 0]$. In the preceding period only player one surely has a positive continuation value (since a win here forces two to exit):

$$\begin{aligned}
\tilde{V}_1[W(t_{0.5} - 1) = 0] &= (1-p) \times V_1[W(t_{0.5}) = 0] + p \times V_1[W(t_{0.5}) = 1] - \tilde{c} \\
&= (1-p) \times 1 + p \times V_1[W(t_{0.5}) = 1] - \tilde{c} \\
&> (1-p) \times 0 + p \times V_2[W(t_{0.5}) = 1] - \tilde{c} \\
&= \tilde{V}_2[W(t_{0.5} - 1) = 0]
\end{aligned}$$

where $\tilde{c}$ is the equilibrium fighting cost if both players continue. The players are willing to set $c_1 \geq 1 - p > p \geq c_2$, so two will be forced to exit. A final application of Lemma 2 **(ii)** shows that two must quit at $W(1) = 0$ and so will not even contest the nomination. This means:

$$\Pr(1 \text{ win}) = \begin{cases} 0.5 & t_{0.5} = 1 \\ 1 & 1 < t_{0.5} \leq \frac{T+1}{2} \end{cases} \tag{2.23}$$

The reason $t_{0.5} = 1$ is different is that one has no opportunity prior to the toss-up election to take advantage of his favorable position and drive up costs (and two from the contest). Notice also that $t_{0.5} = \frac{T+1}{2}$ is included since it follows the intuition of this paragraph.

Turn to the remaining cases, $t_{0.5} > \frac{T+1}{2}$: $\{p, ..., p, 1-p, ..., 1-p, 0.5, 1-p, ..., 1-p\}$. Following the previous paragraph two quits at all nodes without a mathematical winner for $\forall t > t_{0.5}$. At $t_{0.5}$ and $W(t_{0.5}) = 0.5(2t_{0.5} - T - 1)$, the node right above the two clinch range, both players have equal continuation value: $\widetilde{V}_i[W(t_{0.5}) = 0.5(2t_{0.5} - T - 1)] = 0.5 - c \geq 0$ for some $c \in [0, 0.5]$. Now player one will quit at all earlier nodes along this diagonal, $t \in [\frac{T+1}{2}, t_{0.5})$ and $W(t) = 0.5(2t - T - 1)$. To show this look at $t_{0.5} - 1$,

$$\begin{aligned}\widetilde{V}_1[W(t_{0.5}-1) = 0.5(2t_{0.5} - T - 3)] &= p \times 0 + (1-p) \times V_1[W(t_{0.5}) = 0.5(2t_{0.5} - T - 1)] - \tilde{c} \\ &< p \times 1 + (1-p) \times V_2[W(t_{0.5}) = 0.5(2t_{0.5} - T - 1)] - \tilde{c} \\ &= \widetilde{V}_2[W(t_{0.5}-1) = 0.5(2t_{0.5} - T - 3)]\end{aligned}$$

since a win for two gives him the mathematical clinch while a loss means both face the same continuation value next period. So player two will set costs just high enough to force one's exit. Applying Lemma 2 **(ii)** shows that one will also quit on the other nodes, in particular $W(\frac{T+1}{2}) = 0$. At $W(\frac{T-1}{2}) = 0$ if two sets $c_2 = p^+$ he can force one's ouster since this election is biased in his favor and we just found one exits if he loses this stage. This technique can be applied inductively to each $W(t) = 0$ for $t \leq \frac{T-1}{2}$. In particular one exits at $W(1) = 0$ so:

$$\Pr(1 \text{ win}) = 0 \quad \text{if } t_{0.5} > \frac{T+1}{2} \tag{2.24}$$

□

Bibliography

[1] California Race May Be the Key to White House. The New York Times, 21 March 1994. p. A1.

[2] ALDRICH, JOHN H. A Model of the U.S. Presidential Primary Campaign. In S. J. Brams, A. Schotter and G. Schwodiauer, editor, *Applied Game Theory.* Physica-Verlag, Wurzburg, 1979.

[3] ALDRICH, JOHN H. A Dynamic Model of Presidential Nomination Campaigns. *American Political Science Review*, 74:651–669, 1980.

[4] ALDRICH, JOHN H. *Before The Convention: Strategies and Choices in Presidential Nomination Campaigns.* The University of Chicago Press, Chicago, 1980.

[5] BANERJEE, ABHIJIT. A Simple Model of Herd Behavior. *Quarterly Journal of Economics*, 107:797–817, 1992.

[6] BARTELS, LARRY M. *Presidential Primaries and the Dynamics of Public Choice.* Princeton University Press, Princeton, 1988.

[7] BIKHCHANDANI, SUSHIL, DAVID HIRSHLEIFER AND IVO WELCH. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy*, 100(5):992–1026, 1992.

[8] BRAMS, STEVEN J. AND MORTON D. DAVIS. Optimal Resource Allocation in Presidential Primaries. *Mathematical Social Sciences*, 3:373–388, 1982.

[9] CUKIERMAN, ALEX. Asymmetric Information and the Electoral Momentum of Public Opinion Polls. *Public Choice*, 70:181–213, 1991.

[10] FEDERAL ELECTION COMMISSION, Washington, DC. *1976 Presidential Campaign Receipts and Expenditures*, no. 7 edition, 1977. FEC Disclosure Series.

[11] FUDENBERG, DREW AND JEAN TIROLE. *Game Theory.* The MIT Press, Cambridge, MA, 1992.

[12] FUDENBERG, DREW, R. GILBERT, JOSEPH STIGLITZ AND JEAN TIROLE. Preemption, Leapfrogging, and Competition in Patent Races. *European Economic Review*, 22:3–31, 1983.

[13] GOPOIAN, J. DAVID. Issue Preferences and Candidate Choice in Presidential Primaries. *American Journal of Political Science*, 26(3):523–546, 1982.

[14] HUANG, CHI-FU AND LODE LI. Continuous Time Stopping Games with Monotone Reward Structures. *Mathematics of Operations Research*, 15(3):496–507, 1990.

[15] MCKELVEY, RICHARD D. AND PETER C. ORDESHOOK. Elections with Limited Information: A Fufilled Expecations Model Using Contemporaneous Poll and Endorsement Data as Information Sources. *Journal of Economic Theory*, 36(1):55–85, 1985.

[16] MCKELVEY, RICHARD D. AND PETER C. ORDESHOOK. Sequential Elections with Limited Information. *American Journal of Political Science*, 29:480–512, 1985.

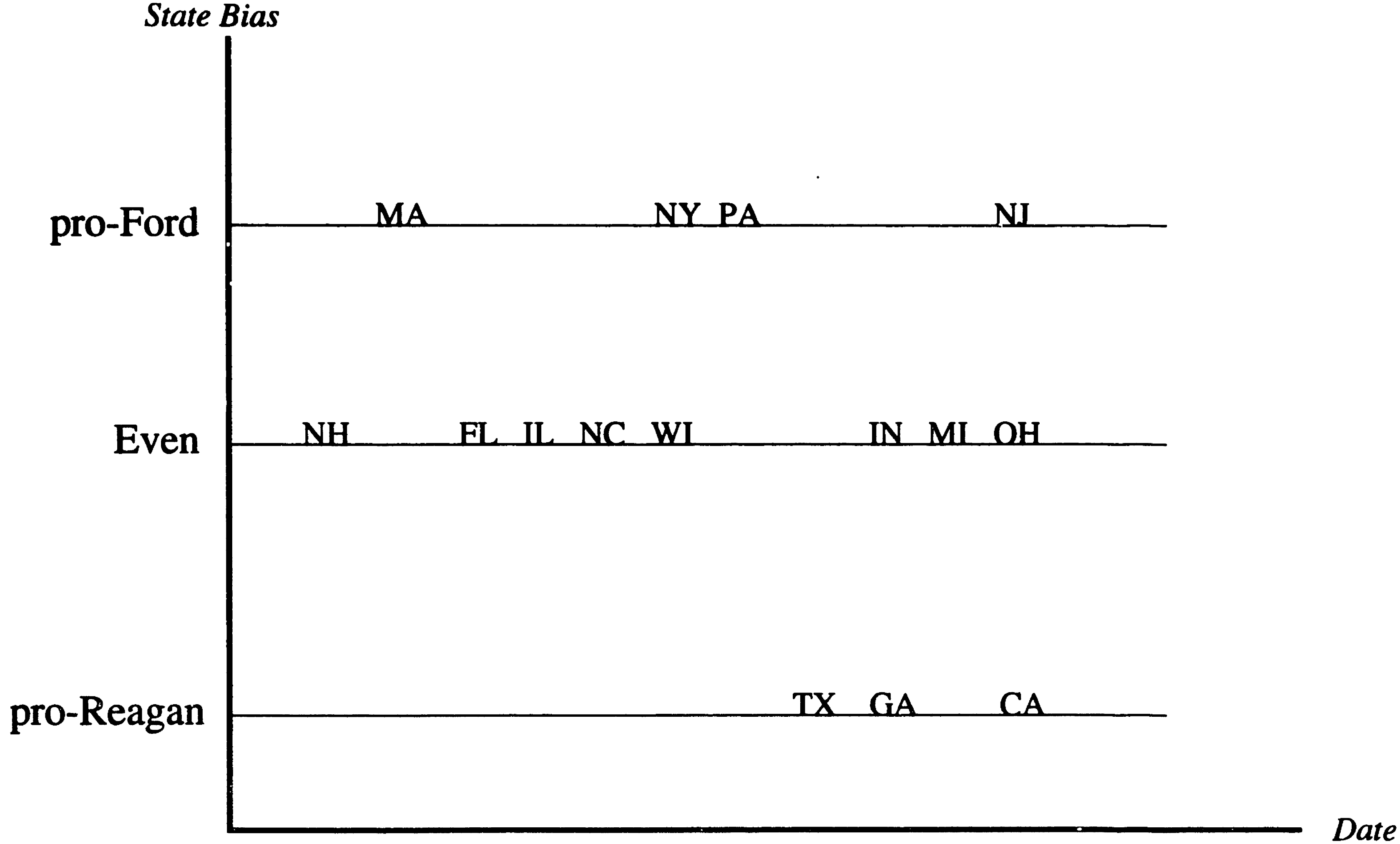
[17] MYERSON, ROGER B. Analysis of Democratic Institutions: Structure, Conduct and Performance. *Journal of Economic Perspectives*, 9(1):77–89, 1995.

[18] ROSEN, SHERWIN. Prizes and Incentives in Elimination Tournaments. *The American Economic Review*, 76(4):701–715, 1986.

[19] SNYDER, JAMES M. Election Goals and the Allocation of Campaign Resources. *Econometrica*, 57(3):637–660, 1989.

[20] WILCOX, CLYDE. Financing the 1988 Prenomination Campaigns. In E. H. Buell Jr. and L. Sigelman, editor, *Nominating the President*. The University of Tennessee Press, Knoxville, 1991.

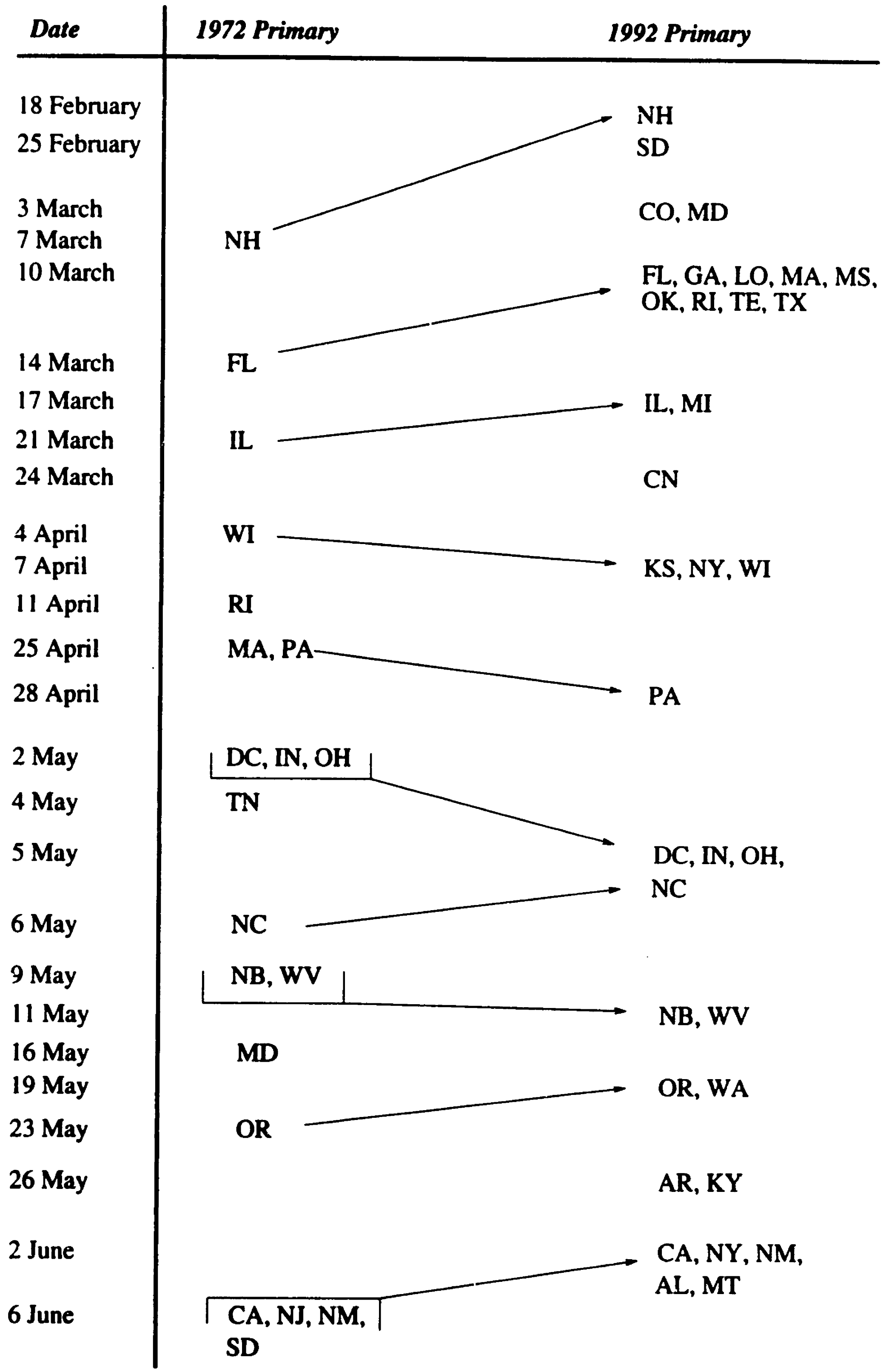
Figure 1: The 1976 Republican Primary



Sources: Aldrich Table 6.4 and Bartels Table 9.3

Note: Does not include small, late states

Figure 2: 1972 vs. 1992 Primary Schedule



Source: *Congressional Quarterly*, 22 January 1972 and 7 September 1991

Note: Does not include 1972 dates for Arkansas (no primary held) and New York or Alabama (candidate names not listed on ballot)

Figure 3: Pr(1 Win) For Election Sequence {p, 0.5, 1-p}

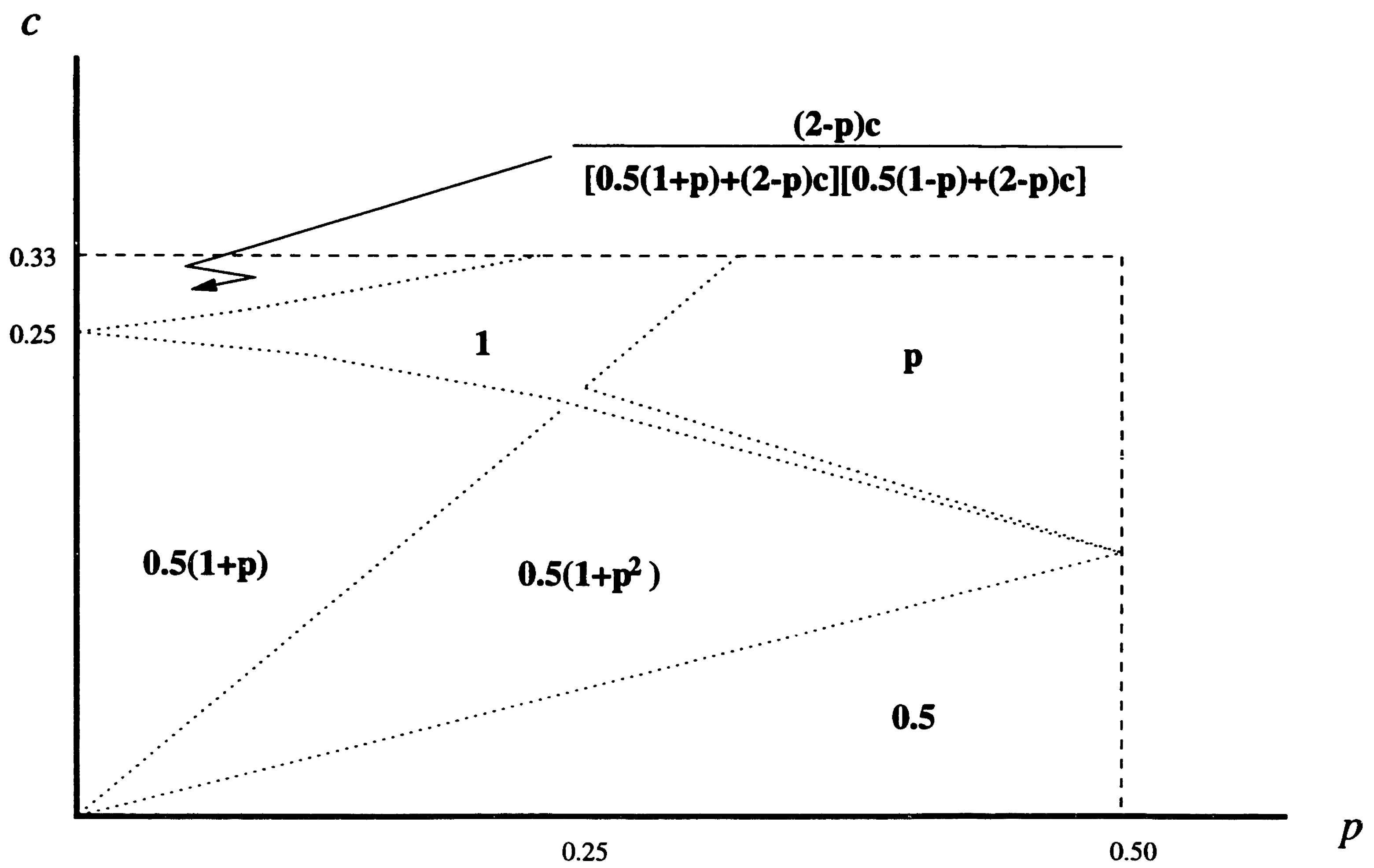


Figure 4: Horizon Cutoff for Proposition 3

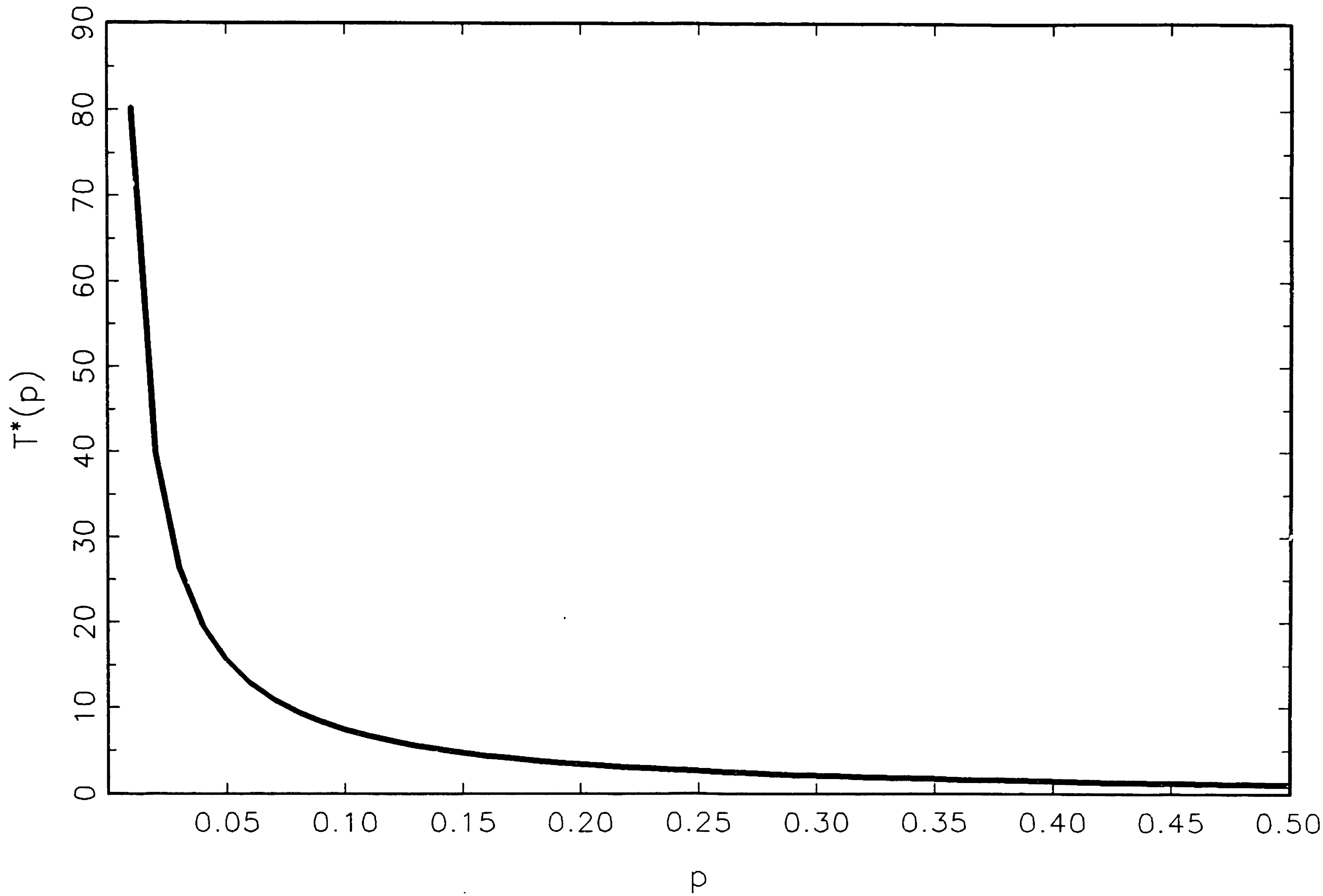


Figure 5a: Pr(1 Win) in Corollary 4 (i)
T=25, p=0.025

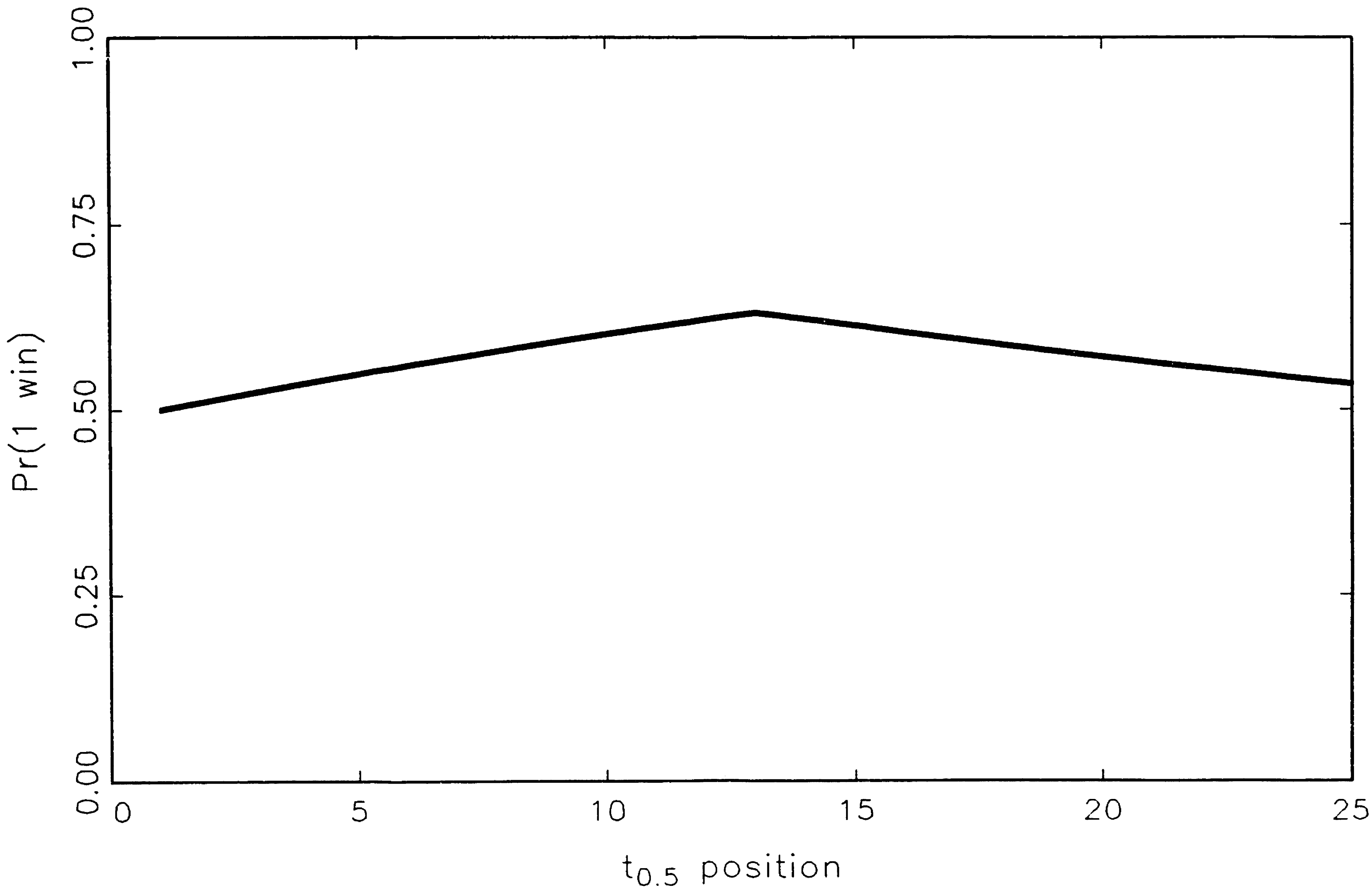


Figure 5b: Pr(1 Win) in Corollary 4 (ii)
T=25, p=0.031

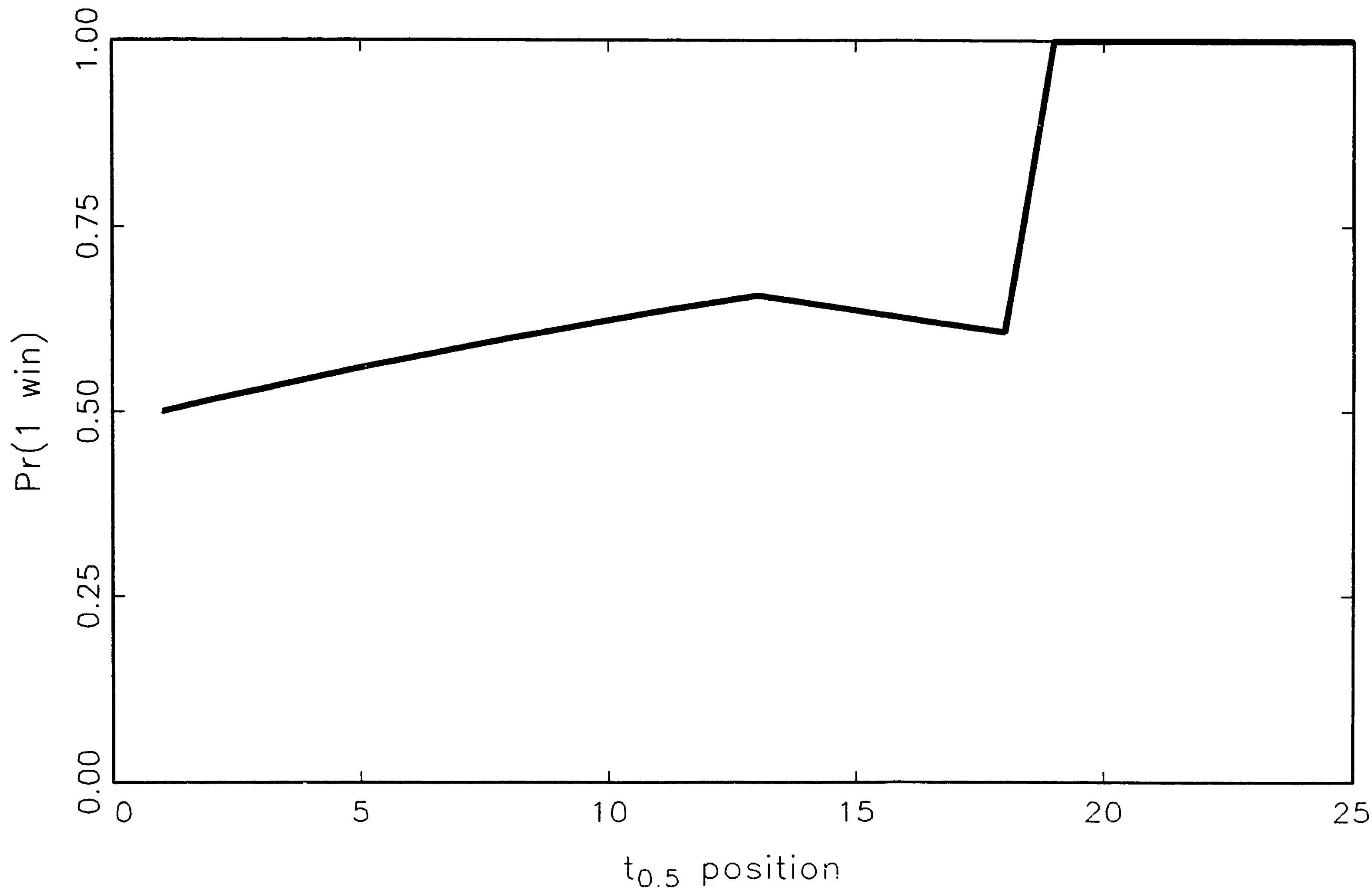


Figure 5c: Pr[1 Win] in Corollary 4 (iii)
T=25, p=0.05

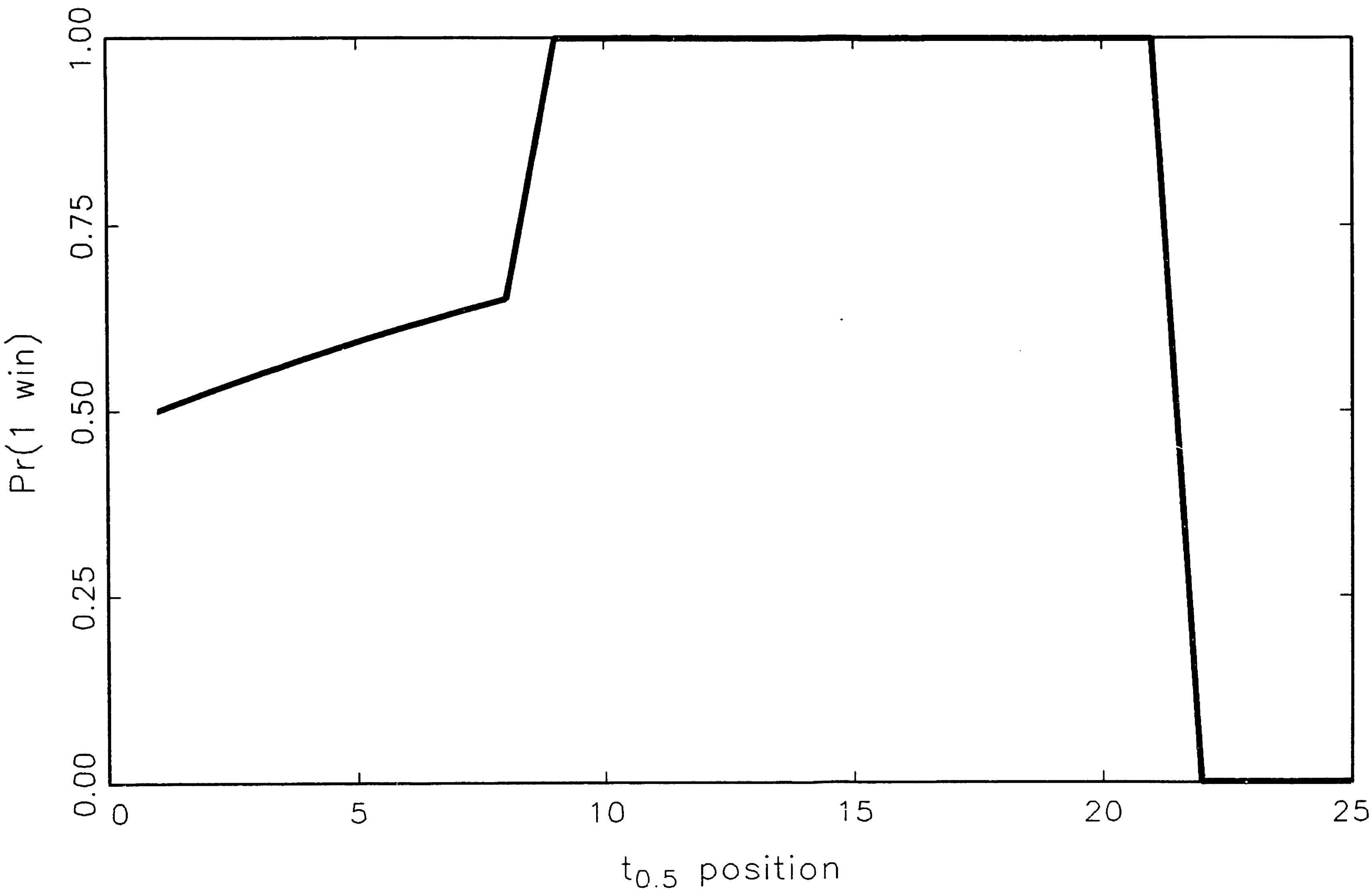


Diagram 1: Contest {p, 0.5, 1-p}

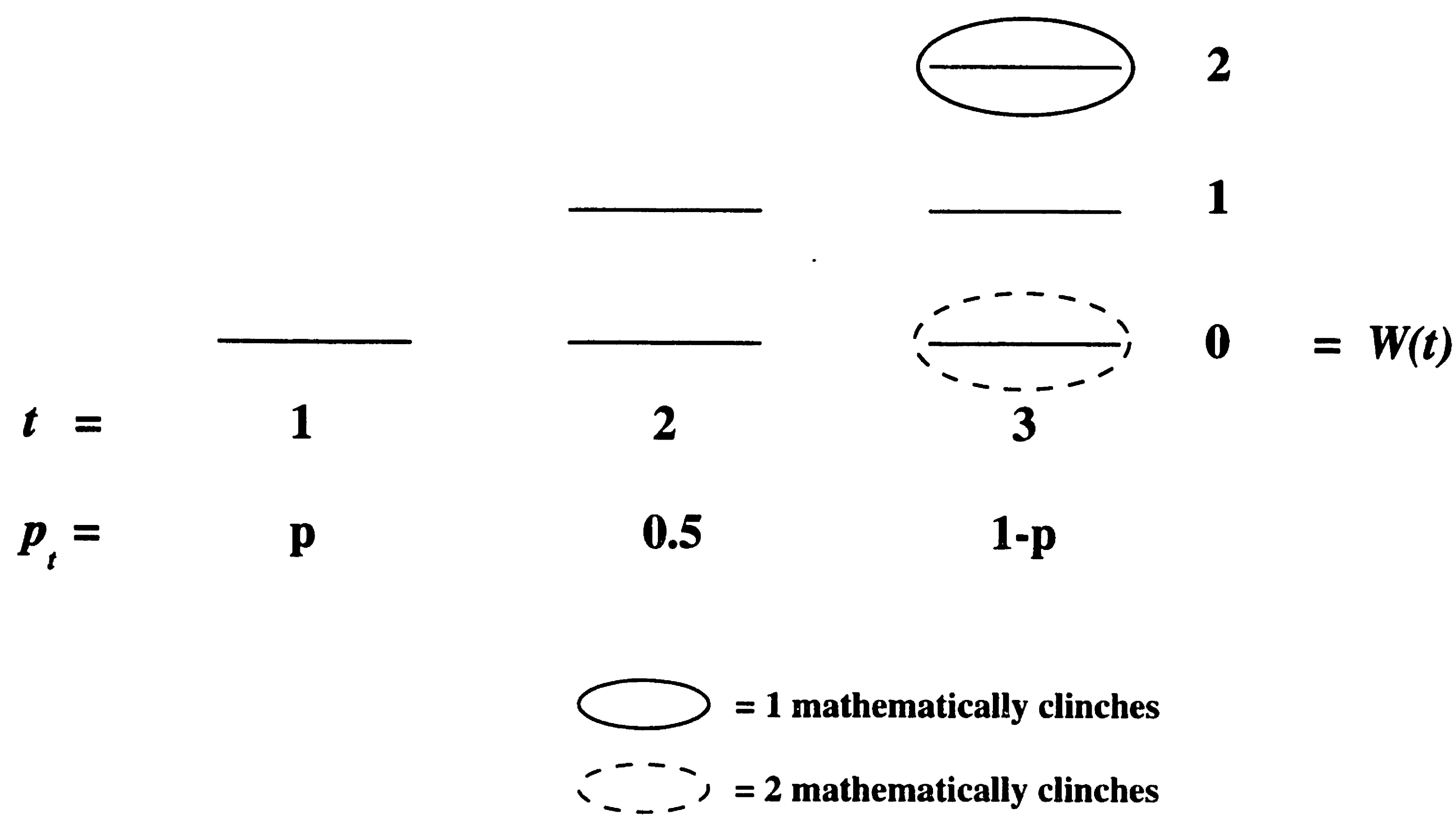


Diagram 2a: *T*-Period Game

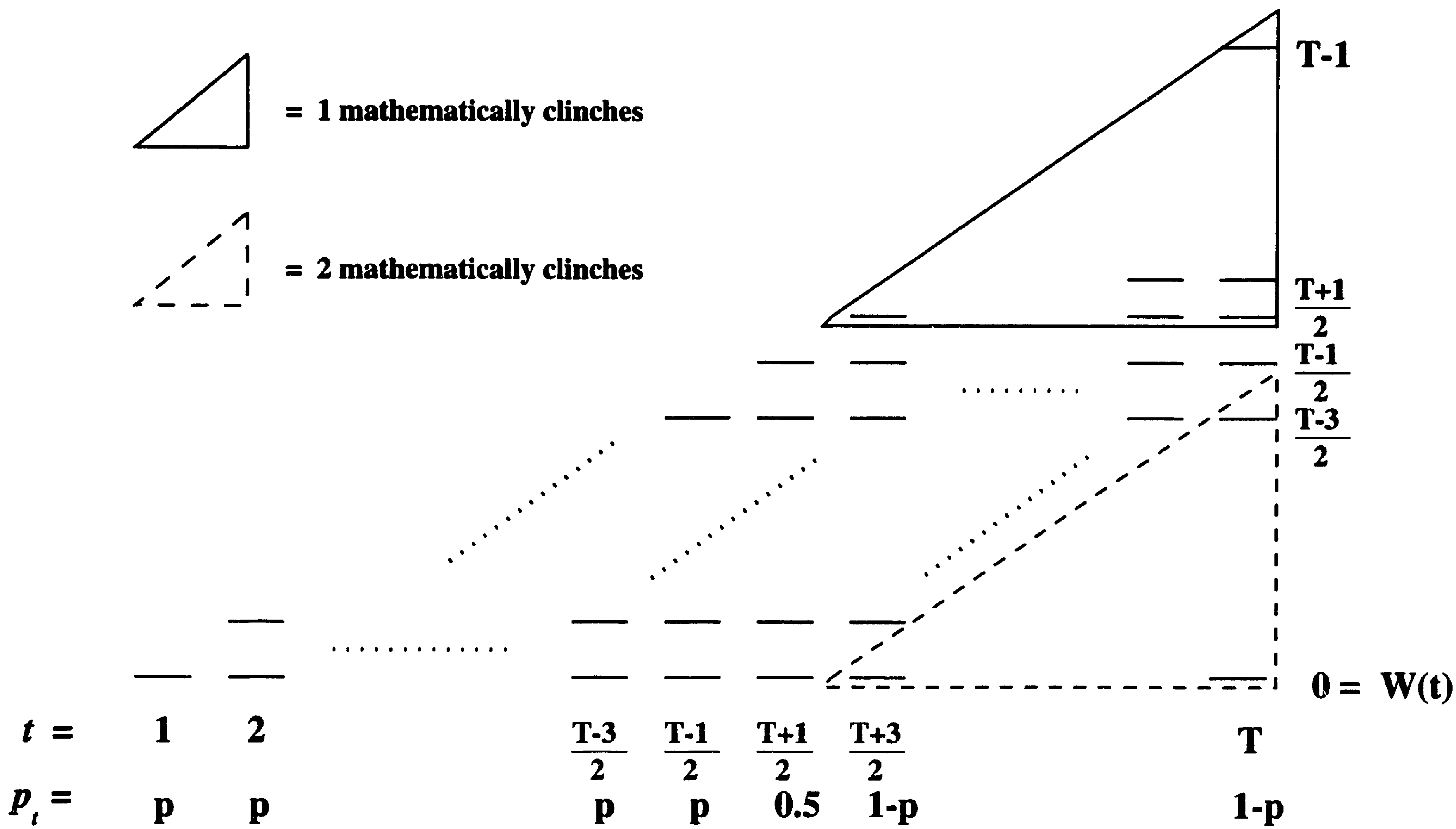
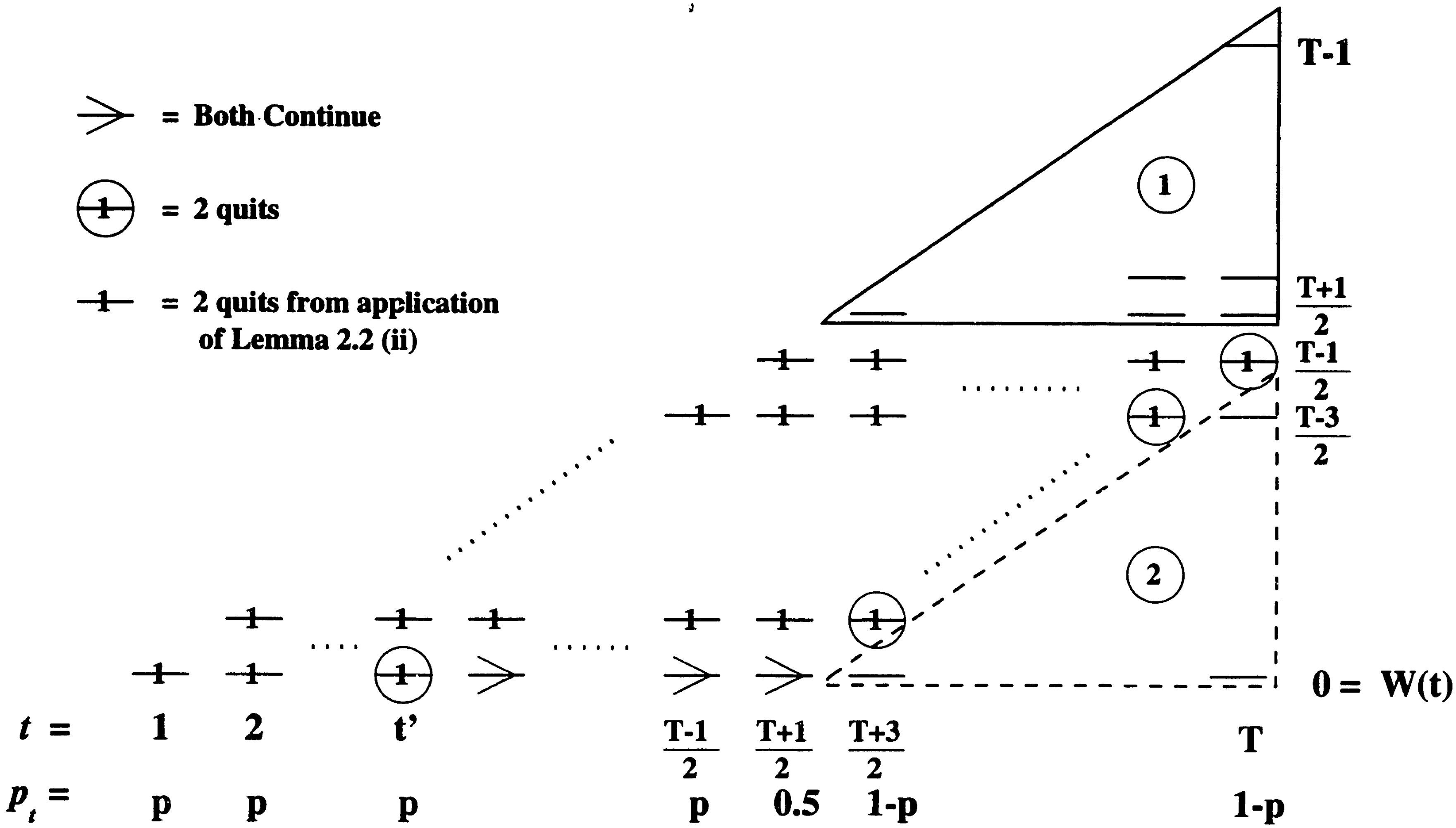


Diagram 2b: Solved T-Period Game where two exits at W(t')=0



Chapter 3

A Predictive Index for the Flypaper Effect

3.1 Introduction

The fiscal decentralization plans of the 104th Congress has stirred interest in the effects of block grants on state and local spending. One of the more consistent findings in the empirical literature is that lump-sum intergovernmental aid boosts spending more than an equivalent increase in private income, referred to as the "flypaper effect" (because grant money sticks where it hits). Determining the theoretical roots and the extent of the flypaper are essential to understanding the nature of local decision-making and predicting the effect of shifting responsibilities away from the central government.

In the neoclassical model of local government, a fully informed electorate and political competition result in the implementation of the decisive voter's policy preference. In this framework, exogenous private income (wages) and public income (grants) are perfect substitutes, so there is no way to explain the flypaper effect. Alternatively, a budget-maximizing politician can take advantage of imperfect voter information (about the social budget) and/or lax discipline at the polls to spend a disproportionate share of public windfalls.

These are extreme views with reality falling somewhere in between. We would expect the flypaper to be more serious when voter activism is limited. Presumably, an index which measures how closely the electorate follows their government would also predict the seriousness of the flypaper effect.

In this paper I use the level of government spending devoted to administrative overhead (correcting for possible returns to scale) as such a measure. The level of overhead is determined by a variety of factors: the degree of politician rent-seeking, the strength of public sector unions, or the importance of bureaucracy. But no matter the source, consistently high levels of overhead mean that voters are getting low return on their tax dollars, an indicator they have limited control over fiscal decisions.[1] Since the other institutions all favor government expansion, this suggests a more marked flypaper effect for high overhead communities.

The wide availability of administrative expense data means this gauge can be applied to national samples of various levels of government. For example, the Census annually compiles overhead spending at the state level ([6]) and on a decennial basis for counties ([4]) and municipalities ([5]). Thus policy-makers can use this approach to predict not just the aggregate spending change resulting from intergovernmental aid but in which communities there will be the greatest stimulative effect. In an era where both revenues and responsibilities are being transferred from the federal to the state and local level such a tool should find manifold uses.

This paper is a first step towards that broader project. The index is used to predict the fiscal response of Philadelphia suburbs to a windfall which accompanies levying a home earned income tax.[2] High overhead governments devote a disproportionate amount of the new money towards spending rather than reduced property taxes. At the same time low overhead municipalities consume about the same fraction of the revenue spike as they would from an equivalent increase in private income. Since most

[1]The appropriate concept is marginal return on taxes, but the data only provides an average. Correcting for returns to scale eliminates the chief sources of a wedge, start-up costs and scarce resource bottle-necks.

[2]Enacting the tax yields a windfall from non-residents and captures taxes which residents previously paid at their workplace.

of the communities that levy the wage tax fall in the second category, on average there is little extraordinary spending. Omitting the overhead index precludes identification of the flypaper effect.

The road-map for the remainder of the paper: the following section briefly surveys the existing flypaper literature. Next, I elaborate on the institutions of the earned income tax and address a potential concern about endogenous levying decisions. A simple model of community decision-making under partial voter control and politician rent-seeking motivates the estimation strategy; proof of the key proposition is relegated to an appendix. Section 3.5 comments on the data with particular focus on how government overhead is used to index the extent of voter control. The empirical estimates show that the overhead index successfully predicts which communities devote a larger portion of the tax windfall to new expenditure. The results are robust to removing outliers and to different forms of the index. Section 3.7 concludes and discusses avenues for future research.

3.2 Literature Review

This section surveys the leading papers on the flypaper effect. Empirical studies which document its existence are followed by the leading challenges of its importance (these attacks are either irrelevant or muted when applied to the current study). Finally, the various mechanisms which could explain the flypaper are presented and critiqued.

Virtually every empirical paper to study the effect of grants has documented the enhanced propensity to consume from lump-sum public aid over a similar jump in private income. The leading surveys are Gramlich [15], Fisher [13] and Wycoff [32] all of whom find significant flypaper effects in local spending. There is also some evidence that federal lump-sum aid is a stronger stimulant to local expenditure than state transfers (Grossman [16]). Ladd [20] shows that states whose taxable income definition tracks the federal base retained a significant portion of the windfall resulting from the Tax Reform Act of 1986's broadening of the federal tax base.

Nonetheless, several authors have challenged the existence of the flypaper, attributing the measured effect to econometric problems or misanalysis. The actual transfers which are studied, such as the General Revenue Sharing program or state equalization grants, are based on complicated formulae with both matching and lump-sum elements (see for example, Rothstein [27]). A simple explanation for the flypaper is that matching grants, as opposed to lump-sums, have a price as well as an income effect. The problem is compounded in the more realistic case where the grant has a closed-end matching component: the community's budget constraint is kinked and OLS estimates are inconsistent (as with labor supply estimates under progressive taxation); spending beyond the cap results in both a higher price and an increase in (virtual) income, so observational errors will be positively correlated with these regressors. Moffitt [22] and Megdal [21] claim the resulting positive bias explains most of the observed flypaper.

Similarly, the very nature of the grant can induce a statistical flypaper. Chernick [8] argues that even conditional[3] closed lump-sums are often targeted to communities willing to spend the largest amount of own funds, so these are effectively matching funds; a similar problem occurs with grants that are allocated based on prior tax effort ([12]). Zampelli [33] points out that only the fungible portion of conditional aid could possibly be rebated to citizens, so the income equivalent of such grants is actually much smaller than the total transfer. Accounting for voter illusion over the cost of other communities' subsidies also diminishes the flypaper effect (Holsey [19]).

The main lesson from this literature is to be skeptical of papers which do not investigate the nature of the public grant formula. Fortunately, my study is immune to this critique since the revenues are explicitly lump-sum with no matching component (see next section).

The other main challenge to the flypaper is omitted variable bias. B. Hamilton

[3]Conditional monies must be devoted to a particular type of project.

[17] argues that less spending for particular public services such as schooling or police protection is necessary if the community has certain unobserved socioeconomic characteristics. As median income is likely to be positively correlated with such terms, this imparts a negative bias on the propensity to spend private income. Wycoff [32] disputes the empirical importance of this point. For the Pennsylvania communities which I study, schooling is handled by a separate authority so the severity of the bias should be minimal.

Granting the existence of a flypaper effect, there are several schools of thought on its origin. In the canonical papers Courant-Gramlich-Rubinfeld [9] and Oates [23] assume voters do not know their community receives a grant and that they infer the marginal price of public goods from the ratio of services to tax payments, i.e. the average price. With lump-sum aid, then even with a constant returns production technology the average cost will be below marginal cost. This is price illusion, so the demand for government services will exceed an equivalent increase in private income (consider the Slutsky decomposition). One problem is the simultaneity of spending and information extraction; [32] computes 2SLS estimates and finds little support for the voter error hypothesis. More fundamental, in the first year of the grant voters perceive that public spending has become substantially cheaper but do not then (or later) wonder how this happened.

Along similar lines, Filimon-Romer-Rosenthal [11] show when voters are aware of the grant but underestimate its size (grant illusion), a budget-maximizing official can set spending beyond the ideal level of the median voter. The informational inconsistency is even more extreme here: misperception persists over time even though public spending is outside the voters' perceived feasible set.

A final use of fiscal illusion involves the federal government taxing residents and then remitting the funds to the local government. While this is potentially revenue-neutral, if the median voter has a smaller federal than local tax share (Fisher [12]) or perceives a net subsidy from the payments of other regions (Winer [30]) then spending will increase. While more appealing since it relies less on information shortfalls, in practice this effect is likely to explain only a small portion of the flypaper effect ([13]) and does not apply to most of the revenue-sharing programs which have been studied.

The Romer-Rosenthal [25] spending reversion argument has also been applied. If politicians seek to maximize expenditure, have agenda power and if rejection of their proposal results in an exogenous budget (the reversion), then they can implement the highest level of spending which leaves the median voter just indifferent to the reversion. When consumer preferences are single-peaked in the public goods, then spending is strictly higher then the median's ideal. However, this does not explain the differential effect of private and public monies. In a later paper [26] the authors show that depending on initial conditions the model may be consistent with the flypaper or its converse.

A potpourri of other theoretical explanations have been proposed which do not rely on voter illusion: the deadweight loss of taxation (J. Hamilton [18]), the lobbying of interest groups (Dougan and Kenyon [10]), the asymmetry in voter bargaining power from private versus public grants (Wycoff [31]), or uncertainty about the size of the tax base or intergovernmental grants (Turnbull [29]). The importance of these theories in explaining the flypaper has not yet been evaluated, and none get us closer to predicting when the effect should be strongest.

To date, only Wycoff [32] has conducted a serious horse-race among even a subset of these theories. While documenting the existence of the flypaper, he finds little support for either the fiscal illusion arguments or the various econometric and mis-specification critiques. Unfortunately, several of the school districts he studies have negative matching rates and hence non-convex budget sets, so the demand curves should be discontinuous. Wycoff seems to ignore jumps when he estimates the government spending equation, resulting in biased parameters.

3.3 The Pennsylvania Earned Income Tax

With some understanding of the state of play, I turn briefly to the institution studied in this paper. The revenue windfall associated with taxing non-residents under the Pennsylvania local wage tax provides an excellent crucible for examining the flypaper effect: there is no complicated grant formula and the monies are unconditional lump-sums. As such the results are immune from the econometric critiques of previous work relating to matching elements, project targeting or non-fungibility of specific grants, see last section. In addition, the flat tax means that the Bergstrom-Goodman [1] condition for the median income to belong to the median preference voter is likely satisfied (see discussion in [14], chapter 14). The concluding portion of the section addresses a potential concern showing that *tax situs* makes the levying decision effectively exogenous.

3.3.1 Institutions and the Wage Tax Windfall

Municipalities in Pennsylvania have the option of levying an earned income tax (EIT) at a rate no more than one percent.[4] There is no double taxation and residence has priority: citizens from a taxing community pay only to their home government.[5] Alternatively, individuals whose home does not levy are liable for any wage taxes imposed at their workplace. It is collections from in-commuting non-residents which is the basis for the EIT "windfall."

EIT revenues comprise over half of the tax collections for the levying communities. Figure 1 shows the mean wage tax contribution as a function of years from the initial levy: after rising sharply between the first and second year, undoubtedly due to a

[4]In practice, the tax is always set at the full 1% cap. A detailed description of the Pennsylvania wage tax is presented in my earlier paper, [28].

[5]Philadelphia alone has a special taxing privilege: its wage tax takes precedence over home levies. In addition, those who commute to another state or have no earned income (the retired) will be exempt from any home levies.

divergence between the fiscal and calendar year[6] and a learning curve for collection procedures, the fraction slowly tapers off.[7] The explanation is that as more neighboring communities decide to levy, fewer in-commuters must pay taxes at work. Figure 2 plots the proportion of wage tax revenues collected from non-residents against years from the enactment.[8] There is a linear decline, starting off at over one quarter of wage taxes and dropping to near zero after thirty years.[9] The fall in wage tax revenues in the earlier graph maps neatly into the decline of non-resident collections.

Importantly for our purposes, since collections from residents will not decline, this variation will allow us to separately measure the effect on spending of taxes from these two groups. Also a comparison in the two figures shows that the windfall collection is quite significant, contributing over an eighth of total tax collections during the five years immediately following the levy.

3.3.2 Exogenous Levies

In principle, the decision to levy is completely internal, based only on the tastes of community members. In this case the timing of the EIT may reflect budgetary or economic circumstances, so it is invalid to make the assumption (needed to test for the flypaper effect) that the windfall is an exogenous lump-sum.

However, in practice communities tend not to actively seek out a wage tax but rather have it foisted upon them by tax hemorrhages to neighboring jurisdictions. Prior to a levy, residents are liable for any earnings taxes at their workplace. A home enactment is costless to those already paying "abroad"[10] since they simply transfer their taxes to the home government. In the spirit of the political economy

[6]Some communities also enacted the wage tax in the middle of a calendar year, so first year collection figures do not cover a full twelve months.

[7]Values more than 25 years after the levy should be interpreted with some caution since they represent only a few communities, the earliest taxers. See Figure 2 in [28].

[8]Construction of this variable is discussed in Section 3.5.2.

[9]The annual standard error remains steady in the interval between 0.11 and and 0.15 for all years except for a spike in the initial year (0.41).

[10]I will use this term to refer to any local government besides the home government.

approach, it is natural to think that communities implement the EIT only when it does not increase the tax burden of the median individual (read voter); that is, for the majority of citizens the wage tax is a free source of funds.[11]

Motivation may be found in Maps 1-3, repeated from my earlier paper, [28]. Communities tend not to tax in geographic isolation but rather in clumps. It seems a necessary (though not sufficient) condition to levy is that ones neighbors already have a tax. As people tend to work near where they live, this is anecdotal support for the median voter approach.

Figure 3 graphs the density among leviers of the population fraction which already pays wage taxes or would be exempted from home taxes[12] in the year *prior* to a home enactment. In only 2 of the 146 initiations would over fifty percent of the population see their tax burden rise. Formally, the null "a necessary condition for a home levy is that the median voter[13] will not face additional liability" cannot be rejected at even 95% confidence.[14] Alternatively, we can reject the hypothesis that the taxing decision is completely unrelated to the median's status even with 99.5% confidence.[15]

This evidence suggests that EIT implementation is an exogenous decision to the

[11]There is an auxillary assumption that the wage tax has no inherent attraction over other revenue instruments such as the property tax. This is reasonable in light of the univeral opposition to *new* taxes as opposed to gradual increases in pre-existing rates.

[12]The groups meeting this criterion are those who work in a levying municipality (including Philadelphia) or another state, and those without current or future earned income (senior citizens).

[13]It is crucial to be explicit about who is counted as a voter in determining the median. I restrict the electorate to resident workers and senior citizens (there is no double counting since presumably the elderly do not work).

[14]Assume the probability that the median resident in a community will see a tax increase from a home levy is p, i.e. each incidence is an i.i.d. Bernoulli variable. To test the null H_0: $p = \widetilde{p}$ given an empirical frequency $\widehat{p}$ among N communities simply calculate the test statistic $Z = (\widehat{p} - \widetilde{p})/\sigma_p$ where σ_p is the standard deviation under the null, $\sqrt{\widetilde{p}(1 - \widetilde{p})/N}$. The null is rejected when Z exceeds the relevant t distribution value (at N degrees of freedom and some level of confidence).

In evaluating the null $p = \widetilde{p} \equiv 0$ the standard deviation from the empirical rather than assumed probability was used to avoid a zero variance. Here, the empirical frequency is $\widehat{p} = 2/146 \equiv 0.014$ and $N = 146$, so the calculated test statistic is 1.424 which is within the expected bounds at 95% confidence.

[15]Following the previous footnote, we test $H_0 : \widetilde{p} = 0.5$. Again using $\widehat{p} = 2/146$ and $N = 146$, the calculated test statistic is $Z = 50.552$, so the null can be rejected at even 99.5% confidence.

extent that neighboring governments determine whether one even considers the levy.[16]

Still, conditional on being in a "taxing clump," it might be that communities are predisposed to levy if they have a revenue shortfall (corner solution). This would bias in favor of finding a flypaper since such governments will spend a disproportionate amount of the new levy (i.e. not provide one-for-one tax relief in the pre-existing instruments). There are two responses to this critique. First, in the context of a hazard model for propensity to levy, neither revenue need (as measured by deficit spending) nor the potential size of the wage tax payments from non-residents have a significant influence while the percent paying the EIT abroad is the main positive factor,

$$\begin{aligned} \Pr(\text{levy}_{it}) = \quad & 0.007 \times \%\ \text{Deficit}_{it} \quad - 0.019 \times \text{Windfall}_{it} \quad + 0.047 \times \%\ \text{Pay EIT}_{it} \\ & (0.87) \qquad\qquad\qquad (-0.95) \qquad\qquad\qquad (10.10) \\ N = 7607, \quad & \log L = -555.49 \end{aligned}$$

where t-statistics are in parentheses and $\Pr(\text{levy}_{it})$ is the probability that community i levies for the first time in period t, % Deficit is the government budget shortfall as a percentage of expenditure, Windfall is the expected first year wage tax collections from non-residents divided by government expenditure, % Pay EIT is the fraction already paying the wage tax at work, and there is a suppressed matrix of control variables (see the accompanying paper [28] for details). Second, if this point is correct, then all revenue-starved communities should spend excessive amounts of the EIT revenue. We will see that lagged deficits have little predictive power over how much of the wage tax windfall communities decide to spend.

[16]This point also presumes commuting decisions are unaffected by wage taxes. With the exception of the Philadelphia levy, this seems true (regressions omitted).

3.4 Theoretical Framework

This section models the choice of government spending given a private and community budget constraint. We consider social decisions under various assumptions about the degree of electorate information. When politicians have no informational advantage, the optimal level of spending for the median voter is implemented. Here a government windfall has the same effect as an equal increase in private income. However, when voters are ill-informed about the public budget then rent-seeking politicians are not subject to strict discipline and will rebate less of the windfall than desired by the median voter, the government spending flypaper. While the actual formulation is new, it is in the flavor of previous approaches such as [11].

The main use of the developed machinery is to determine the fraction of the wage tax collections which are recycled as increased government spending rather than property tax reductions. Following the discussion in the previous section, I assume the levying decision is exogenous and instead focus only on its expected effect upon public expenditure.

Consider a community where individuals are identical in tastes, income and property ownership but exogenously work at home or abroad; non-residents own no capital but may work here. The government provides a single local public good[17] which it funds with a non-distorting[18] variable rate property and (possibly) a fixed rate earned income tax.[19] Recall that home income tax has priority, so when the community implements an EIT, it gains a windfall from all in-commuters whose home does not yet tax; alternatively, prior to the levy, home citizens are liable for any taxes in their

[17] A local public good is a non-rival, non-exclusive output which benefits only those members of a given community (no spillovers).

[18] That is, wages and property ownership are exogenous, not responding to the tax environment. In formal regressions (omitted) neither the earned income tax or property millage had a significant effect on labor or capital flows. Considering the low rates involved this is hardly surprising.

[19] We can allow heterogeneity over income and property ownership, but then the wage tax might shift the identity of the median voter. This needlessly complicates the analysis. Again I defer to the Bergstrom-Goodman [1] conditions under which the median income voter is decisive.

workplace abroad.[20]

The relative bargaining strength and information asymmetries between the government and the electorate drive the results. Politicians are assigned agenda control[21] over all fiscal decisions, in particular the level of public spending, subject to maintaining some level of voter satisfaction (see below). The constraint arises due to the possibility that the incumbent politician will not be re-elected if he selects policies too far removed from the preferences of the electorate. *Ceteris paribus* this condition will be more binding if there is extensive competition for political office.

Presume that voters are imperfectly informed about *shocks* to public income. That is, voters know their own income, tax burden and level of government services, but only have partial information about the size of any public windfall, in this case wage tax collections.[22] One interpretation of this assumption is that it is costless for a citizen to know the value of any variable which directly impacts him, but it is costly to monitor the community finance restraint.[23] Completely informed politicians prefer higher levels of government spending, and the electorate's imperfect information following a levy provides the slack needed to for rent-seeking.

The government's problem at any time is:

$$\begin{aligned} \max_{G} \quad & G \\ \text{st} \quad & X + \tau_P P = y_D \\ & \sum_{i=1}^{N} \tau_P P + \mathcal{I}(EIT)\overline{\tau_Y} \sum_{i=1}^{N+M} y_i = p_G G \\ & U(X, G) \geq \overline{U}(c) \end{aligned} \tag{3.1}$$

where G is public good consumption, X private consumption (whose price serves as

[20]We ignore the special exemptions discussed in Section 3.3.1 in this stylized model.

[21]In Pennsylvania, politicians write the municipal budgets which are not subject to voter approval such as the referenda required in several states.

[22]It might seem more reasonable that voters are fully informed about collections from residents but not in-commuters. In the econometric implementation I will allow for this possibility.

[23]Alternatively, we can presume that it is costly to join the political process. This will also provide an opportunity for politician rent-seeking.

numeraire), P the assessed value of the individual's property, τ_P the property tax rate, y the level of wages, $\overline{\tau_Y}$ the (fixed rate) wage tax, N the number of home citizens, M the number of in-commuters who do not face a home wage tax, and p_G the price of the public good. $U(\cdot)$ is an increasing, strictly quasiconcave utility function which represents the preferences of a home resident.

The politician is a revenue maximizer subject to the private and public budget as well as a voter individual rationality condition. The first restraint is an individual's spending frontier: private consumption plus property tax must equal to disposable income (y_D), wages minus any local income tax paid at home or abroad.[24] The second constraint states that government spending equals revenues, property tax plus the income tax (if levied- $\mathcal{I}$(EIT) is an indicator variable). Notice that the government gets a windfall from the M non-residents if a wage tax is in place.

These constraints can be combined:

$$X = y_D + \mathcal{I}(\text{EIT})(\overline{\tau_Y} y + \frac{W}{N}) - \frac{p_G G}{N} \tag{3.2}$$

where W is the tax revenue collected from non-residents ($= \overline{\tau_Y} \sum_{i=1}^{M} y_i$). Three comments: first, property taxes have disappeared, set implicitly by the level of public spending. Also, from the individual's perspective personal and lump-sum public income are perfectly substitutable,[25] a manifestation of Bradford-Oates' [2] equivalence theorem. Finally, the home EIT enriches all residents since taxes are no longer "wasted" abroad and a windfall is generated from taxing in-commuters.[26]

[24] $y_D = (1 - \overline{\tau_Y})y$ when a wage tax is paid and $y_D = y$ otherwise.

[25] This can be seen in the combined budget constraint when a home wage tax is in place [$\mathcal{I}$(EIT) = 1],

$$X = y + \frac{W}{N} - \frac{p_G G}{N}$$

Notice in this case all individuals face an identical budget.

[26] This should be obvious for those not yet paying the wage tax while those with a workplace tax pay no higher taxes [$y_D = (1 - \overline{\tau_Y})y$ before and after] but get a share of the collections from non-residents.

The final equation in (3.1) captures both the voter's imperfect information about the public budget and the re-election constraint. Since a voter may not know the level of wage tax collections, his *perceived* public budget is

$$X = y_D + \mathcal{I}(\text{EIT})(1-c)(\overline{\tau_Y} y + \frac{W}{N}) - \frac{p_G G}{N} \tag{3.3}$$

where $c \in [0,1]$ indicates the degree of fiscal illusion or the politician's relative bargaining strength. $c \to 1$ means voters are completely uninformed (or it is expensive to discipline politicians) while $c \to 0$ indicates full information (or voters have large bargaining power). Then the final constraint in the politician's program means that fiscal policies must ensure the representative voter at least his first-best utility under the perceived public budget, $\overline{U}(c)$.[27]

It is easiest to understand this condition graphically, see Diagram 1. Initially, there is no wage tax and so no room for politician rent-seeking; the community meets the demand of the decisive voter, G_0^*.[28] A positive windfall moves out the effective budget of the decisive voter. If the voter is completely informed and it is costless to discipline the government ($c = 0$), he will insist on a bundle, G_1^*, which maximizes his utility under the true budget. However, under complete fiscal illusion ($c = 1$) he will not notice any of the new windfall, and resists only utility diminishing bundles; a revenue-maximizing agenda setter can set $G_1^{Rent} \gg G_1^*$. Notice that the government spends more than one-for-one out of the new revenue (it actually raises property taxes) since voters are benefiting from additional spending.[29] It should be clear that an intermediate level of leverage results in spending bound between these two values.

This intuition motivates the main result of this section. The level of government spending following a levy is determined by the degree of voter knowledge/bargaining

[27] Formally, $\overline{U}(c)$ is the indirect utility function derived from maximizing individual preferences subject to (3.3).

[28] There is fiscal illusion only in response to shocks to the government budget.

[29] Nonetheless, we presume that voters cannot use the newly observed level of government spending to infer the size of the windfall, the same information inconistency pinned on [9], [11] and [23]. It is more appropriate to interpret this model as short-run, before the economy re-equilibrates.

power.

Proposition 1 *When the electorate is incompletely informed and/or has only partial control over fiscal decisions, there is a rent-seeking agenda setter, and private demand for public services is normal, the marginal propensity to spend out of government windfalls exceeds that from private windfalls. The flypaper effect strengthens as the information asymmetry grows. Under complete fiscal illusion, public spending increases by more than the size of the windfall while perfect information results in no flypaper.*

PROOF: See Appendix 3.8.

The implicit solution to the politician's problem (3.1) and the proposition motivate a log-linearized demand for government services,

$$\begin{aligned}\widetilde{G}_{it} &= \beta_0 + \beta_1 y_{it} + \beta_2 \widetilde{p_G}_{it} + \gamma X_{it} \\ &\quad + \mathcal{I}(\text{levy}) \left[\left(\theta_1 + \theta_2 \widetilde{W}_{it} + \theta_3 \widetilde{R}_{it} \right) c_{it} + \theta_4 \widetilde{W}_{it} + \theta_5 \widetilde{R}_{it} \right] + \epsilon_{it}\end{aligned} \tag{3.4}$$

where all non-indicators are in natural logarithms and the tilde indicates per-capita terms. This formula says that government spending ($\widetilde{G}$) in community i at time t is determined by the median household income (y), price of public spending ($\widetilde{p_G}$), and a matrix of supplemental factors (X). In addition, if a wage tax[30] is in place, some of the collections from home residents ($\widetilde{R}$) and the windfall from commuters ($\widetilde{W}$) will go to public spending, with the level potentially depending on the degree of voter discipline over the government (c).[31] The innovation is the interaction terms

[30] The wage tax rate $\overline{\tau_Y}$ does not appear since it is by assumption constant.

[31] Revenues from residents and commuters are separated for two reasons. First, while the expected value of collections from residents is widely publicized prior to a home levy, only highly informed voters would have some estimate of non-resident revenues. Also, those voters who do not pay the wage tax prior to a home levy are less likely to view the government collections from residents as "net wealth." Both points suggest a prior of larger spending effect from commuter-based revenues.

Voter Control (H1)	$\theta_1 = \theta_2 = \theta_3 = 0$	$\theta_2 \bar{c} + \theta_4 = \beta_1$
Intermediate Control (H2)	$\theta_1, \theta_2, \theta_3 > 0$	$\theta_2 \bar{c} + \theta_4 > \beta_1$
Politician Control (H3)	$\theta_1 = \theta_2 = \theta_3 = 0$	$\theta_2 \bar{c} + \theta_4 > \beta_1$

Table 3.1: Leading Models and Their Restrictions on Spending Regression (3.4)

involving the windfall and the degree of public information. Such an effect has not been included in previous estimates of the spending function.

This equation embeds the leading views of government reaction to revenue windfalls. The general case we have examined so far allows partial politician control (H2), and suggests flypaper consumption only occurs when voter discipline is lax, i.e. high c communities. There are also two leading special cases. Under voter control of social decisions (H1) there is never an information asymmetry, politicians always implement the median citizen's desired spending and from the proposition there should be no flypaper (this is the neoclassical model). A wage tax windfall is equivalent to an increase in private income while the level of overhead is irrelevant. At the other extreme, complete fiscal illusion/politician control (H3), voters are always perfectly ignorant of the public windfall (this is the basic assumption in [9] and [23]). Here the level of voter discipline is irrelevant and a substantial flypaper effect should follow a windfall in any community.

For reference, each theory's predicted restrictions on equation 3.4 are summarized in Table 3.1. The first column involves the relative importance of the index regressors while the second considers whether there should be a flypaper when there is significant fiscal illusion, $\bar{c}$.

3.5 Empirical Specification and Data

Before turning to the results, we need to describe the index of voter activism and the variables of the empirical model.

3.5.1 Index of Voter Activism: Overhead

Crucial to both the framework of the last section and the empirical implementation to come is a gauge for voter activism. The measure I use is reported government spending on administrative overhead.[32] The idea is that such expenses provide no service to voters of all ideological stripes while they are the most fungible kind for a rent-seeking politician; as such, higher overhead should be indicative of limited voter control.

The raw data must be modified to account for returns to scale. Figure 4 shows that overhead is closely related to total expenditures. A formal regression which allows for fixed costs and scale effects,

$$\begin{aligned} \ln Overhead &= -0.475 + \;\; 0.870 \ln Expenditure \\ &\quad\; (-11.30) \qquad (265.96) \\ N = 7817, &\quad R^2 = 0.90 \end{aligned} \tag{3.5}$$

with t-statistics in parentheses. The index will be the residual from this equation. That is, the excess of administrative expenses over the average for a government of that size is dubbed "wasteful overhead." As a robustness check, other versions of the overhead variable will be used for the index, see Section 3.5.2.

The basic index has several desirable features besides its intuitive appeal. First, due to the log-linear regression, the residuals are scale independent and so easily comparable across communities and time. Also, at least for this sample, the index is quite stable over time for a given community, so it makes sense to label a particular community as "high overhead" or "low overhead."[33] Finally, if the values help explain

[32]In Pennsylvania, local government overhead is self-reported to state authorities. It is defined to include spending on tax collection, legal staffs, personnel administration, maintenance and planning. This likely undermeasures the true level of overhead since some waste may be hidden in public goods categories.

[33]Communities which only infrequently shift from one side of the aggregate mean to the other are said to exhibit persistent overhead status (i.e. consistently "high" or "low"). Formally, for community i,

the distribution of marginal propensities to spend out of the windfall, it will be hard to attribute this to non-informational stories.

Figure 5 shows the measure's density for the set of all 237 Philadelphia suburbs from 1960 through 1992. Generally values are clustered near a mean of zero, but there is significant dispersion, i.e. the returns to scale regression does not capture all the variation. Notice that unlike in the model, the index is free to take on any value, but higher numbers are still indicative of minimal voter control. Figure 6 looks at the density for communities in the year before they levy the EIT. There is a noticeable shift to the left from the previous distribution meaning that enacting municipalities tend to have low levels of overhead; the accompanying paper [28] uses a government credibility interpretation to explain this point. The importance to the current study is that if overhead is related to the propensity to spend the wage tax windfall, then the flypaper effect should only be important for high overhead communities. Due to the paucity of such wasteful spenders in the sample of taxers, the *average* propensity to consume from the windfall will be little higher than an equivalent income boost. Public expenditure regressions which omit the index will find little excess spending while the full specification should document a flypaper which is more extreme for high overhead communities.

3.5.2 Data

The remaining step before turning to the estimates is to detail the variables in equation 3.4. The data is annual observations from 1960-1992 for the 237 Pennsylvania minor civil divisions (MCDs) in the Philadelphia SMSA. A full discussion of sources

---

$$\text{Persistence}_i = 1 - \left(\frac{\text{Number Switches}}{\text{Number Years} - 1}\right)_i$$

The statistic may take on any value from 0 to 1 with higher values indicating greater persistence. The average value over all communities is 0.988 indicating a high degree of stability in overhead status.

may be found in the accompanying paper [28].

The dependent variable, per capita[34] public spending ($\tilde{G}$), and earned income tax status [$\mathcal{I}$(levy)] are based on data from *Local Government Financial Statistics* [24] which is archived at the Department of Community Affairs in Harrisburg. School districts, which are governed separately from municipalities, can also levy a wage tax, in which case they split revenue with the underlying (taxing) municipalities; calculation of EIT contributions (see below) were appropriately adjusted in such cases.

The commuting matrix is based on decennial data from the Census [3] and Department of Transportation [7].[35] While the destinations of SMSA residents is quite completely tabulated, in-flows to suburban MCDs only include workers who themselves reside in the Philadelphia area. That is, workers who commute from outside the metropolitan area into the Philadelphia suburbs are not included in the data. This means I underestimate the size of the windfall from a home levy (below), and this imparts an upwards bias on the associated parameters (θ_2 and θ_4) in the spending regression (3.4).[36]

We use this information, along with the Census' median household income (y), to partition wage tax revenue into collections from residents (R) and from in-commuters, the tax windfall (W). Formally, for all observations when an EIT is in place, I calculate the median income[37] times eligible numbers[38] for both residents and non-

[34]Following the discussion in Section 3.3.2, I will use the *voting* population, defined as the total number of workers plus senior citizens, as the norm for per capita calculations.

[35]The commuting data for 1971 through 1992 is interpolated based on decennial MCD-to-MCD flows; earlier observations are estimated from MCD-to-county flows. A full description of the approach is presented in the Appendix in [28].

[36]In reality the under-estimate of the windfall from in-commuters is small. We can ignore interstate commuters who are themselves exempt from local Pennsylvania taxes. This only leaves Pennsylvanians from outside the SMSA who commute to a Philadelphia SMSA *and* do not face a home levy. As most communities on the cusp of the metropolitan area had their own levy before 1970, there should be few such individuals. In addition, the tabulated in-flows compare quite closely to the (known) employment level within each suburb.

[37]There are two potential problems here. One is using the median (rather than the more appropriate mean) income to proxy for the actual income of a given individual. Also, the tax is strictly on wages, so I have implicitly assumed that earned to unearned income is a constant for all levels of income.

[38]Recall, those residents who work in Philadelphia or another state are exempt while commuters

residents; the ratio of these terms represents the relative contributions of the two groups to wage tax collections.[39]

The crucial overhead index (c) follows the construction from Section 3.5.1. Rather than fixing on the value in the year prior to the levy, the year-to-year values are used. This does not cause a difficulty as the index is quite persistent (Section 3.5.1) and does not vary in the face of a wage tax levy (regression omitted). Modified versions of the overhead data are also tried as the index: the ratio of log overhead to log expenditure (c') and also log overhead (c''). In light of equation 3.5 only the former should approximate the degree of politician rent-seeking/voter apathy.

For the price of per capita public expenditure ($\widetilde{p_G}$) I follow the literature and use the median voter's tax share. When there is no home levy, this is the ratio of median to mean property value; the former comes from the Census [3] and the average is based on *Local Government Financial Statistics* [24]. Under a wage tax, the share is the average[40] of the median's relative property ownership and relative taxable income, his wages relative to the mean. Unfortunately, mean income is not available for all years, so I assumed no dispersion (wage share equal to unity).[41]

Finally, we must specify the elements of the control matrix, X. As discussed in Section 3.3.2, it may be that current cash flows determine how much of a public windfall will be spent. To check this, I interact the previous period government deficit ($deficit$), as a proportion of total expenditure, with the two wage tax revenue terms. Next, include the fraction of voters who pay local wage taxes ($payEIT$). As taxes lower disposable income, this variable should have a negative effect on public

whose home governments have a tax are not liable for a workplace EIT.

[39] To check this procedure, I generated an expected wage tax collection based on the income and commuting data, an assumption that three-quarters of income is from wages, and the EIT rate (usually one percent). The constructed revenue explains over 80% of the actual level of collections.

[40] In fact property and wages are differentially taxed. Still, since we have seen that the EIT generally contributes half of total tax collections (and property tax the remainder), this is a rough approximation

[41] For available years, 1970-1990, a community's average income rarely exceeds the median by more than ten percent.

spending.[42] We must also add the (per capita) size of the property tax base ($\widetilde{M}$), the main alternative revenue source as well as an additional source of wealth for residents. Market rather than assessed values are used to ensure comparability. Along similar lines, the fraction of land devoted to residential (*residential*) and commercial (*biz*) use is included for 1970-1992 to allow for varying tastes for public services between these groups.[43] Finally, when not using fixed effects, we include dummies for counties,[44] to control for non-municipality provision of local services and heterogeneity in assessment timing,[45] and government form.[46]

Several other variables are included to make the results comparable to estimates in the literature. Presumably, higher population growth (*popg*) should require new capital projects and hence increase the level of spending; this variable is considered in annual percentage terms. Bedroom communities, characterized by a small ratio of home jobs to population ($\widetilde{jobs}$), should have fewer needs for daytime services. It may be that homeowners have different tastes as well as tax burdens for public services relative to renters; I include the home ownership tenure ratio (*owner*), the ratio of owner to renter occupied dwellings. Finally, not only do senior citizens (*senior*) have special tax status under the EIT, they may require special supplemental services or, alternatively, may be more fiscally conservative and push for low taxes. This variable is considered as a percentage of total population.

A set of summary statistics for all included variables is presented in Table 3.2.

[42]In light of the theoretical framework, it might be more appropriate to focus on the tax status of the median rather than the fraction paying. Nonetheless, if there is some uncaptured variation- measurement error, stochastic voter participation, etc.- than the identity of the decisive voter is uncertain and it makes sense to substitute a "smoothed" version of the median, the fraction paying the wage tax.

[43]Two data notes: earlier observations are unavailable while the omitted category is undeveloped or transportation-related property.

[44]There are four suburban counties in the SMSA: Bucks, Chester, Delaware and Montgomery. The omitted group is Montgomery county.

[45]In Pennsylvania, all property assessment is centralized at the county-level. There is no maximum period between assessments, so counties vary in the timing of each reassessment.

[46]There are three possible political structures: cities, boroughs and townships. In general, politician power and the size of bureacracy are greatest in cities and least in townships. The omitted group will be townships.

Variables		Mean	σ	Max	Min
PC gov. spending	$\ln(\tilde{G})$	4.421	0.796	7.210	1.317
Median household income	$\ln(y)$	9.718	0.678	11.652	7.736
PC price gov. spending	$\ln(\widetilde{p_G})$	−0.713	0.377	0.231	−2.172
PC property tax base	$\ln(\widetilde{M})$	8.651	0.631	10.729	6.007
Gov. deficit (% expend.)	$deficit$	−0.113	0.201	0.433	−4.769
Fraction pay wage tax	$payEIT$	0.433	0.296	0.959	0.000
Fraction resident. land	$residential$	0.263	0.178	0.923	0.006
Fraction comm. land	biz	0.310	0.192	0.812	0.000
Population growth	$popg$	1.435	2.222	13.889	−6.422
Jobs / population	$\widetilde{jobs}$	0.811	0.567	8.909	0.000
Owner / renter homes	$owner$	3.970	2.717	25.360	0.380
Seniors / population	$senior$	10.213	3.965	32.100	1.800
Wage tax revenue: residents	$\ln(\tilde{R})$	4.426	0.561	5.713	2.319
Wage tax revenue: non-res.	$\ln(\widetilde{W})$	1.819	1.989	6.042	−3.597
Overhead index when levy	c_{EIT}	-1.104	0.286	1.299	-2.371
Overhead index when no levy	$c_{No\ EIT}$	0.067	0.454	1.699	-2.564
$\frac{\ln(\text{overhead})}{\ln(\text{expenditure})}$ when levy	c'_{EIT}	0.672	0.139	0.861	0.606
$\frac{\ln(\text{overhead})}{\ln(\text{expenditure})}$ when no levy	$c'_{No\ EIT}$	0.876	0.208	0.969	0.629
ln(overhead) when levy	c''_{EIT}	10.980	1.598	14.561	6.265
ln(overhead) when no levy	$c''_{No\ EIT}$	10.447	1.248	14.943	6.148

Table 3.2: Descriptive Statistics ("PC" = per capita)

3.6 Results

We consider the effect of the wage tax windfall on public spending through the log-linear form of (3.4). There is little support for either version H2 or H3 of the flypaper effect in the benchmark regression which omits the index of voter control. Inclusion of this term improves the fit under either OLS or fixed effects and shows that the propensity to consume out of a public windfall varies markedly depending on the prior level of administrative expenses: high overhead governments spend significantly more. The reason these results differ from the benchmark is the preponderance of low overhead values among the taxing communities, Section 3.5.1; the equation suppressing the index picks up the average propensity to consume, only slightly elevated due to the self-selection. In the remainder of this section, to maintain focus I will not discuss the parameters on the control variables.

Table 3.3 contains the estimates.[47] The first and third columns contain the estimates for the benchmark case (no index terms) using OLS and fixed effects. In line with previous studies of local spending (see [14]), government spending is a necessity with income elasticity of 0.3-0.4 (though a correlated term, per capita property value, has an elasticity of 0.75). The tax price has the expected negative effect with an elasticity in the range commonly found. And the potential worry that cash-starved communities would spend a significantly higher portion of windfalls proves to be unfounded.

More important to this study are the parameters on the wage tax revenue, the regressors in bold. A one percent increase in public monies increase spending by roughly one-quarter percent in the case of resident collections and two-fifths of a percent for non-residents. While we anticipated that there would be a larger propensity to spend out of the windfall (after all, this is "free money"), the parameter is only slightly greater than that on private income, favorable support for the voter control

[47] Land use variables are not listed in the regression results since they are only available from 1970. These factors have insignificant parameters when included in a truncated sample regression.

hypothesis (H1) from Section 3.4.

But this result disappears when we include the overhead variable and the interaction terms, columns two and four. Focus first on collections from non-residents: the large parameter on the interaction with administrative overhead means the total impact on spending can be significantly larger than the income effect. In particular, for high overhead communities ($c > 1$ see Table 3.2 or Figure 5), the elasticity of consumption from the non-resident windfall exceeds one![48] At the same time, low overhead communities ($c < -1$) spend no more than 30% of the windfall, less then what we would expect from a spike to private income.

Notice that administrative expenses have a much more muted effect on the propensity to consume out of collections from residents. This makes sense: nearly all individuals are informed that a home wage tax has been put in place and are able to make a guess as to the level of collections. As such, we would expect the revenues to be spent in roughly the same manner that the median would decide, i.e. the income elasticity.[49]

As a final piece of evidence, compare the overhead parameters for years with and without the EIT. High overhead has only a slight negative influence on spending when there is no home levy (insignificant under fixed effects), and it provides a positive boost in the presence of a wage tax.[50] This affirms the assumption from Section 3.4

[48]With the interaction term, the percent increase in government spending from a one percent rise in wage tax collections from in-commuters is,

$$\frac{\partial \ln(\widetilde{G})}{\partial \ln(\widetilde{W})} = \theta_{\ln(\widetilde{W})} + \theta_{c \times \ln(\widetilde{W})} c + \beta_{deficit \times \ln(\widetilde{W})} deficit$$

The term for residents is analogous. A particular community's elasticity value depends on its overhead index value (and also the deficit level but its low parameter value makes this term inconsequential).

Notice that the population denominator on the spending and collections term cancel, so it is equally valid to interpret the parameters in terms of total dollars, $\partial \ln(G)/\partial \ln(W) = \partial \ln(\widetilde{G})/\partial \ln(\widetilde{W})$.

[49]Recall that in all (but two) communities the median voter did not see his wage tax bill rise when the home levy was enacted. So to the decisive voter even his own tax payment can be considered a windfall since he simply transfers what he was already paying to the home coffers.

[50]Again the full effect would take into account the interactions with the wage tax terms.

that voter illusion is important only in the face of unexpected and significant revenue surges such as the wage tax.

It is important to check whether large outliers drive the results. Visual inspection of the residual plot from the spending regression when the overhead index is included (Figure 7) or not (Figure 8) do not reveal any extreme observations. As a formal robustness check, deleting observations with a residual larger than 0.75 (103 of 7607 observations) in the regression including the index did not qualitatively change the parameter estimates (regression omitted).[51] Results were also unaffected in the non-index regression following truncation of residuals greater than 1.5 in absolute value.

Other forms of the overhead index are applied. The first and third columns of Table 3.4 use the ratio of log overhead to log expenditure. Based on the estimated equation 3.5, this variable will be similar to the original index except the variation will be damped for larger governments.[52] Thus, the results should not be very different which is exactly what I find (the drop in significance of the index regressors is due to the variable compression problem). The other index used is simply log of overhead, columns two and four of Table 3.4. Most of the variation in this variable reflects returns to scale (see Figure 4) rather than the desired concept of wasteful/rent-seeking spending. It should come as no surprise that this variable does little to predict the extent to which wage tax collections are recycled as government spending, see interaction terms.[53]

[51]Only positive values are excluded since the alternative null is that there is a negative (or no) relationship between the index and spending, and the positive parameter estimate is due only to a few extreme values.

[52]Presume the true model is

$$\ln(\textit{overhead}) = \alpha + \beta \ln(\textit{expenditure}) + \epsilon$$

Only ϵ represents the "excessive" part of the spending, but the measure under consideration compresses the variation for high levels of spending,

$$\frac{\ln(\textit{overhead})}{\ln(\textit{expenditure})} = \frac{\alpha}{\ln(\textit{expenditure})} + \beta + \frac{\epsilon}{\ln(\textit{expenditure})}$$

see the last term. Also, since $\beta > 0$ there is positive shift relative to the index.

[53]Using instead ln(*overhead per capita*) gives similiar results.

These results taken together support the intermediate voter control hypothesis (H2) and reject the other two. The spending elasticity in the benchmark case is the average of direct and interactive terms in the full specification which, due to the preponderance of low overhead communities among taxers, is not significantly higher than the income elasticity. As only supposition H2 predicts, overhead independently raises the consumed proportion of the windfall. While the evidence is indirect, any competing model must be able to explain why low overhead places are fiscally conservative in response to windfall revenues. Such an alternative theory is difficult to envision.

3.7 Conclusion

The overhead index explains why Philadelphia suburbs seem to have a low spending propensity from a tax windfall. High overhead is indicative of an inattentive electorate, so only in such communities will there be politician rent-seeking and extraordinary consumption of the windfall, Hypothesis H2. The small number of such cases means on average the marginal propensity to consume is not elevated; since a regression omitting the index only captures the mean effect, such an estimate will not register evidence of the flypaper effect. While other settings may not exhibit such extreme selection issues, in this application at least the overhead index is instrumental in understanding the government reaction function.

As discussed in the introduction, this paper should be viewed as an incremental step towards the bigger project involving a national sample of state or local governments. Since the Census tracks overhead spending across these units ([4], [5] and [6]), it should be relatively straightforward to apply the techniques developed here to a variety of questions stemming from the unconditional lump-sum grants planned by the Republican-controlled Congress. For example, it would be interesting to apply the overhead index to states and predict how they might react to a transfer intended to finance welfare. At the very least, this research program should encourage local public economists to go beyond simply documenting the existence of the flypaper and towards understanding how it varies between governments.

3.8 Appendix: Proof of Proposition 1

We can rewrite the voter's perceived budget constraint (3.3) when a wage tax is in place as,

$$X + p_G \frac{G}{N} = (1 - c\overline{\tau_Y})y + (1 - c)\frac{W}{N}$$

It is clear that greater information (lower c) is like an increase to income. Assuming that X and G are normal goods, then optimal demands, and hence the level of reservation utility $\overline{U}(c)$, are strictly decreasing in c.

Now consider utility values which are feasible given the actual budget constraint ($c = 0$). Convexity in (G, X) space ensures that the associated indifference curves intersect the budget frontier twice (except for the one case of tangency); label each point by the level of government services, $G^{\min}(c)$ and $G^{\max}(c)$. Assume that an increase in c does not raise the maximal level of public spending, $G^{\max}(\overline{c}) \leq G^{\max}(\underline{c})$ for $\overline{c} \geq \underline{c}$. We can immediately dismiss $G^{\max}(\underline{c}) = G^{\max}(\overline{c})$ since this would require the indifference curves to cross. Consider the portion of the budget frontier between $G^{\min}(\underline{c})$ to $G^{\max}(\underline{c})$. By definition of strict quasiconcavity, any point on this segment gives higher utility than $\overline{U}(\underline{c})$. But we have assumed $G^{\max}(\overline{c})$ is along this segment $[G^{\max}(\overline{c}) < G^{\max}(\underline{c})]$, which contradicts reservation utility monotone decreasing in c. Hence, a rent-seeking politician is able to attain higher spending when voters are less informed. This in turn means the *increase* in spending is larger since the economy was presumed to initially be in equilibrium.

When $c = 0$, the demands are by definition equal to the optimum at the true budget constraint. In this case there will be an identical response in government spending to a public windfall and a private income spike since they are perfectly substitutable, see (3.3). The result from the previous paragraph shows there will be a flypaper effect when $c > 0$. Alternatively, if $c = 1$ the perceived budget is unchanged and if the government devotes the entire windfall to G, the voter's utility would increase (X is unchanged). So clearly the government could increase G even more.

$\square$

Bibliography

[1] BERGSTROM, THEODORE C. AND ROBERT P. GOODMAN. Private Demands for Public Goods. *The American Economic Review*, 63(3):280–296, 1973.

[2] BRADFORD, DAVID F. AND WALLACE E. OATES. The Analysis of Revenue Sharing in a New Approach to Fiscal Decisions. *Quarterly Journal of Economics*, 35:416–439, 1971.

[3] BUREAU OF THE CENSUS, Washington, D.C. *Census of Population and Housing: Summary Characteristics for Pennsylvania [Census Tracts]*, 1960-1990.

[4] BUREAU OF THE CENSUS, Washington, D.C. *Government Finances: Finances of County Governments*, various years.

[5] BUREAU OF THE CENSUS, Washington, D.C. *Government Finances: Finances of Municipal and Township Governments*, various years.

[6] BUREAU OF THE CENSUS, Washington, D.C. *State Government Finances*, various years.

[7] BUREAU OF TRANSPORTATION STATISTICS (DEPT. TRANSPORTATION), Washington D.C. *Census Transportation Planning Package*, 1980 and 1990.

[8] CHERNICK, HOWARD A. An Economic Model of the Distribution of Project Grants. In P. Mieszkowski and W. Oakland, editors, *Fiscal Federalism and Grants-in-Aid*. The Urban Institute, Washington D.C., 1979.

[9] COURANT, PAUL N., EDWARD M. GRAMLICH AND DANIEL RUBINFELD. The Stimulative Effects of Intergovernmental Grants: Or Why Money Sticks Where It Hits. In P. Mieszkowski and W. Oakland, editors, *Fiscal Federalism and Grants-in-Aid.* The Urban Institute, Washington D.C., 1979.

[10] DOUGAN, WILLIAM R. AND DAPHNE A. KENYON. Pressure Groups and Public Expenditures: The Flypaper Effect Revisited. *Economic Inquiry*, 26:159–170, 1988.

[11] FILIMON, RADU, THOMAS ROMER AND HOWARD ROSENTHAL. Asymmetric Information and Agenda Control: The Bases of Monopoly Power in Public Spending. *Journal of Public Economics*, 17:51–70, 1982.

[12] FISHER, RONALD C. A Theoretical View of Revenue Sharing Programs. *National Tax Journal*, 32:173–184, 1979.

[13] FISHER, RONALD C. Income and Grant Effects on Local Expenditure: The Flypaper Effect and Other Difficulties. *Journal of Urban Economics*, 12:324–345, 1982.

[14] FISHER, RONALD C. *State and Local Public Finance.* Scott, Foresman Company, Glenview, IL, 1988.

[15] GRAMLICH, EDWARD M. Intergovernmental Grants: A Review of the Empirical Literature. In W. Oates, editor, *Political Economy of Fiscal Federalism.* D.C. Heath, Lexington, MA, 1977.

[16] GROSSMAN, PHILIP J. The Impact of Federal and State Grants on Local Government Spending: A Test of the Fiscal Illusion Hypothesis. *Public Finance Quaterly*, 18(3):313–327, 1990.

[17] HAMILTON, BRUCE W. The Flypaper Effect and Other Anomalies. *Journal of Public Economics*, 22:347–361, 1983.

[18] HAMILTON, JONATHAN H. The Flypaper Effect and the Deadweight Loss from Taxation. *Journal of Urban Economics*, 19:148–155, 1986.

[19] HOLSEY, CHERYL M. Price and Income Distortions Under Separate Spending and Taxing Decisions. *Journal of Public Economics*, 50:93–114, 1993.

[20] LADD, HELEN F. State Responses to the TRA85 Revenue Windfalls: A New Test of the Flypaper Effect. *Journal of Policy Analysis and Management*, 12(1):82–103, 1993.

[21] MEGDAL, SHARON BERNSTEIN. The Flypaper Effect Revisited: An Econometric Explanation. *The Review of Economics and Statistics*, 69:347–351, 1987.

[22] MOFFITT, ROBERT A. The Effects of Grants-in-Aid on State and Local Expenditures: The Case of AFDC. *Journal of Public Economics*, 23:279–305, 1984.

[23] OATES, WALLACE E. Lump-Sum Intergovernmental Grants Have Price Effects. In P. Mieszkowski and W. Oakland, editors, *Fiscal Federalism and Grants-in-Aid.* The Urban Institute, Washington D.C., 1979.

[24] PENNSYLVANIA DEPARTMENT OF COMMUNITY AFFAIRS, Harrisburg. *Local Government Financial Statistics*, various years.

[25] ROMER, THOMAS AND HOWARD ROSENTHAL. Bureaucrats versus Voters: On the Political Economy of Resource Allocation by Direct Democracy. *Quarterly Journal of Economics*, 93:563–587, 1979.

[26] ROMER, THOMAS AND HOWARD ROSENTHAL. An Institutional Theory of the Effects of Intergovernmental Grants. *National Tax Journal*, 33:451–458, 1980.

[27] ROTHSTEIN, PAUL. The Demand for Education with 'Power Equalizing' Aid: Estimation and Simulation. *Journal of Public Economics*, 49:135–162, 1992.

[28] STRUMPF, KOLEMAN S. Credibility and the Structure of Taxes: The Local Earned Income Tax in Pennsylvania. Xerox, MIT Economics Department, 1995.

[29] TURNBULL, GEOFFREY K. Fiscal Illusion, Uncertainty, and the Flypaper Effect. *Journal of Public Economics*, 48:207–223, 1992.

[30] WINER, STANLEY L. Some Evidence on the Effect of the Separation of Spending and Taxing Decisions. *Journal of Political Economy*, 91(1):126–140, 1983.

[31] WYCOFF, PAUL GARY. A Bureacratic Theory of Flypaper Effects. *Journal of Urban Economics*, 23:115–129, 1988.

[32] WYCOFF, PAUL GARY. The Elusive Flypaper Effect. *Journal of Urban Economics*, 30:310–328, 1991.

[33] ZAMPELLI, ERNEST M. Resource Fungibility, the Flypaper Effect, and the Expenditure Impact of Grants-in-Aid. *The Review of Economics and Statistics*, 68:33–40, 1986.

Regressors	Dependent variable: $\ln(\widetilde{G})$ (OLS) No Index	(OLS) Index	(FE) No Index	(FE) Index
constant	-5.444	-5.826	—	—
	(-68.05)	(-70.01)		
$\ln(y)$	0.305	0.341	0.385	0.390
	(28.65)	(31.62)	(40.90)	(40.70)
$\ln(\widetilde{M})$	0.783	0.784	0.728	0.730
	(74.33)	(75.16)	(78.82)	(78.92)
$\ln(\widetilde{p_G})$	-0.382	-0.283	-0.368	-0.265
	(-15.54)	(-11.12)	(-17.98)	(-10.85)
$\mathbf{ln(\tilde{R})}$	0.268	0.256	0.228	0.239
	(31.44)	(20.82)	(36.58)	(27.00)
$\mathbf{ln(\widetilde{W})}$	0.410	0.682	0.391	0.651
	(48.72)	(87.30)	(53.11)	(77.21)
$\mathbf{c} \times \mathbf{ln(\tilde{R})}$	—	0.131	—	0.121
		(5.96)		(7.26)
$\mathbf{c} \times \mathbf{ln(\widetilde{W})}$	—	0.345	—	0.300
		(49.85)		(51.89)
$\mathbf{c_{EIT}}$	—	0.058	—	0.065
		(5.88)		(6.74)
$\mathbf{c_{No\ EIT}}$	—	-0.039	—	-0.022
		(-3.29)		(-1.33)
$deficit \times \ln(\tilde{R})$	0.099	0.033	0.082	0.056
	(0.73)	(0.87)	(1.11)	(0.99)
$deficit \times \ln(\widetilde{W})$	0.013	0.019	0.010	0.022
	(0.42)	(0.57)	(0.51)	(0.73)
payEIT	-0.245	-0.281	-0.245	-0.218
	(-6.55)	(-7.33)	(-8.43)	(7.10)
popg	-0.054	-0.051	-0.011	-0.011
	(-23.82)	(-22.36)	(-5.49)	(-5.67)
$\widetilde{jobs}$	0.021	0.027	0.120	0.120
	(2.35)	(3.06)	(13.17)	(13.11)
owner	-0.019	-0.016	-0.032	-0.031
	(-9.66)	(-8.08)	(-15.69)	(-14.75)
senior	0.008	0.006	0.005	0.005
	(5.75)	(4.88)	(2.92)	(2.82)
$\mathcal{I}(city)$	0.943	0.973	—	—
	(19.16)	(19.96)		
$\mathcal{I}(borough)$	0.327	0.314	—	—
	(26.89)	(26.01)		
$\mathcal{I}(county)$?	Yes	Yes	No	No
N	7607	7607	7607	7607
R^2	0.724	0.878	0.827	0.938

Table 3.3: Government Expenditure Estimate: Equation 3.4

Sample: 1960-1992, 237 Philadelphia SMSA MCDs.

FE = Fixed effects regression. Key regressors are in bold.

$\ln(\tilde{R})$ and $\ln(\widetilde{W})$ only for observations with levy in place.

(t-statistics)

Regressors	Dependent variable: $\ln(\widetilde{G})$			
	(OLS) **Index'**	(OLS) **Index"**	(FE) **Index'**	(FE) **Index"**
constant	-6.156	-5.748	—	—
	(-47.06)	(-73.04)		
$\ln(y)$	0.393	0.144	0.476	0.237
	(31.68)	(15.01)	(46.31)	(15.97)
$\ln(\widetilde{M})$	0.780	0.837	0.742	0.721
	(74.64)	(93.02)	(82.64)	(80.79)
$\ln(\widetilde{p_G})$	-0.302	-0.236	-0.355	-0.259
	(-10.11)	(-8.29)	(-11.65)	(-9.31)
$\mathbf{\ln(\widetilde{R})}$	0.175	0.290	0.161	0.261
	(8.23)	(11.61)	(7.70)	(12.70)
$\mathbf{\ln(\widetilde{W})}$	0.209	0.456	0.265	0.331
	(11.04)	(18.88)	(8.36)	(10.46)
$\mathbf{c \times \ln(\widetilde{R})}$	0.298	-0.046	0.265	-0.035
	(9.84)	(-19.03)	(6.57)	(-16.39)
$\mathbf{c \times \ln(\widetilde{W})}$	0.835	0.029	0.751	0.019
	(19.31)	(8.33)	(13.93)	(6.54)
$\mathbf{c_{EIT}}$	0.276	0.132	0.256	0.151
	(1.54)	(16.57)	(1.59)	(14.89)
$\mathbf{c_{No\ EIT}}$	-1.274	0.110	-0.279	0.180
	(-7.76)	(25.64)	(-2.15)	(21.40)
$deficit \times \ln(\widetilde{R})$	0.111	0.477	0.096	0.378
	(1.00)	(0.88)	(0.94)	(0.69)
$deficit \times \ln(\widetilde{W})$	0.037	0.016	0.051	0.017
	(0.33)	(0.08)	(0.83)	(0.13)
payEIT	-0.080	-0.089	-0.134	-0.273
	(-1.89)	(-2.48)	(-3.92)	(-8.21)
popg	-0.052	-0.039	-0.010	-0.009
	(-22.58)	(-19.84)	(-5.33)	(-4.82)
$\widetilde{jobs}$	0.021	0.009	0.099	0.112
	(2.45)	(1.23)	(11.08)	(12.81)
owner	-0.016	-0.008	-0.023	-0.022
	(-8.05)	(-4.99)	(-11.34)	(-11.53)
senior	0.005	0.001	-0.001	-0.000
	(3.62)	(1.05)	(-0.55)	(-0.08)
$\mathcal{I}(city)$	0.908	0.441	—	—
	(18.45)	(9.87)		
$\mathcal{I}(borough)$	0.319	0.451	—	—
	(26.39)	(42.69)		
$\mathcal{I}(county)$?	Yes	Yes	No	No
N	7607	7607	7607	7607
R^2	0.819	0.778	0.894	0.810

Table 3.4: Government Expenditure Estimates Under Modified Overhead Index

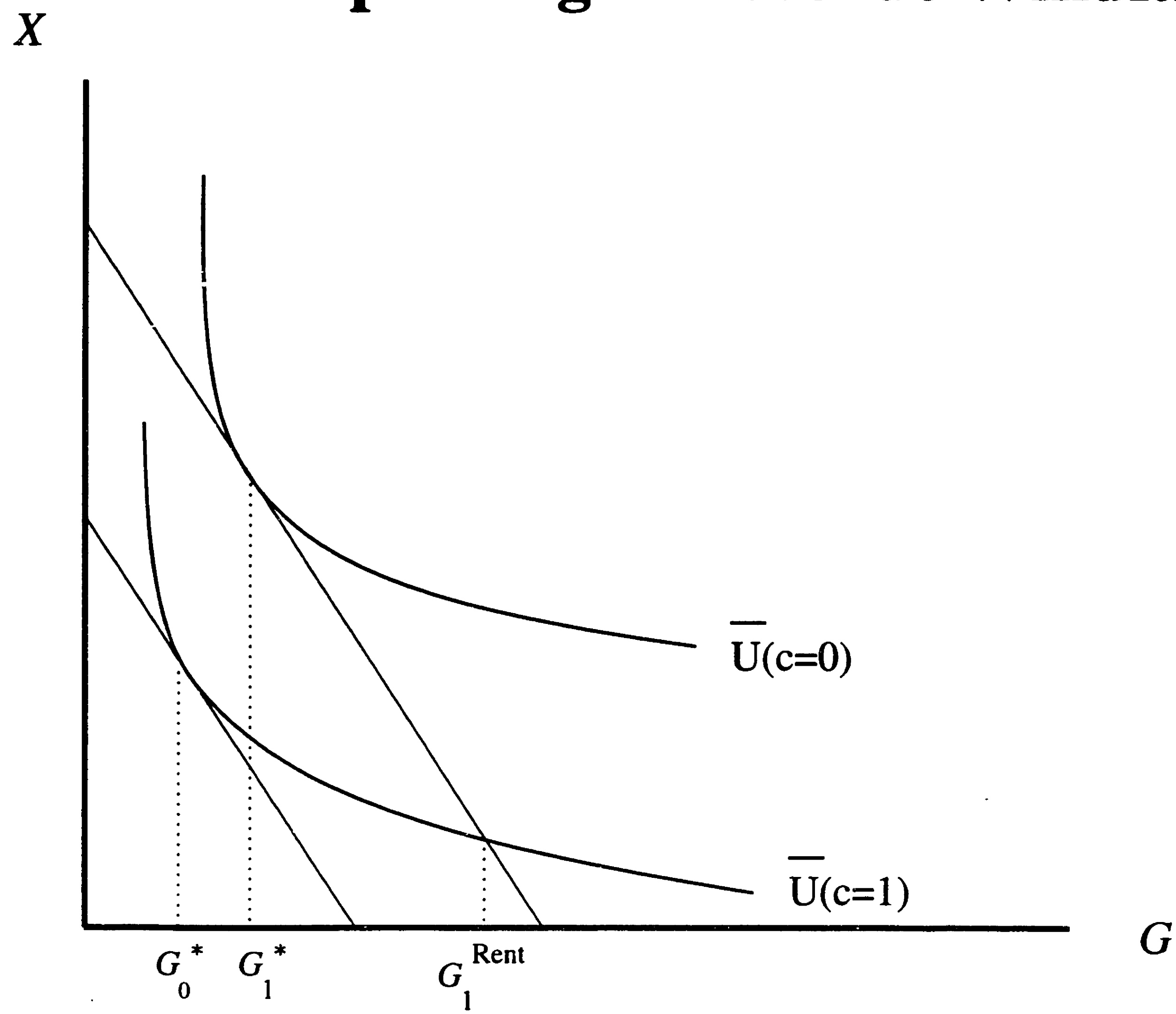
FE = Fixed effects regression. Key regressors are in bold.

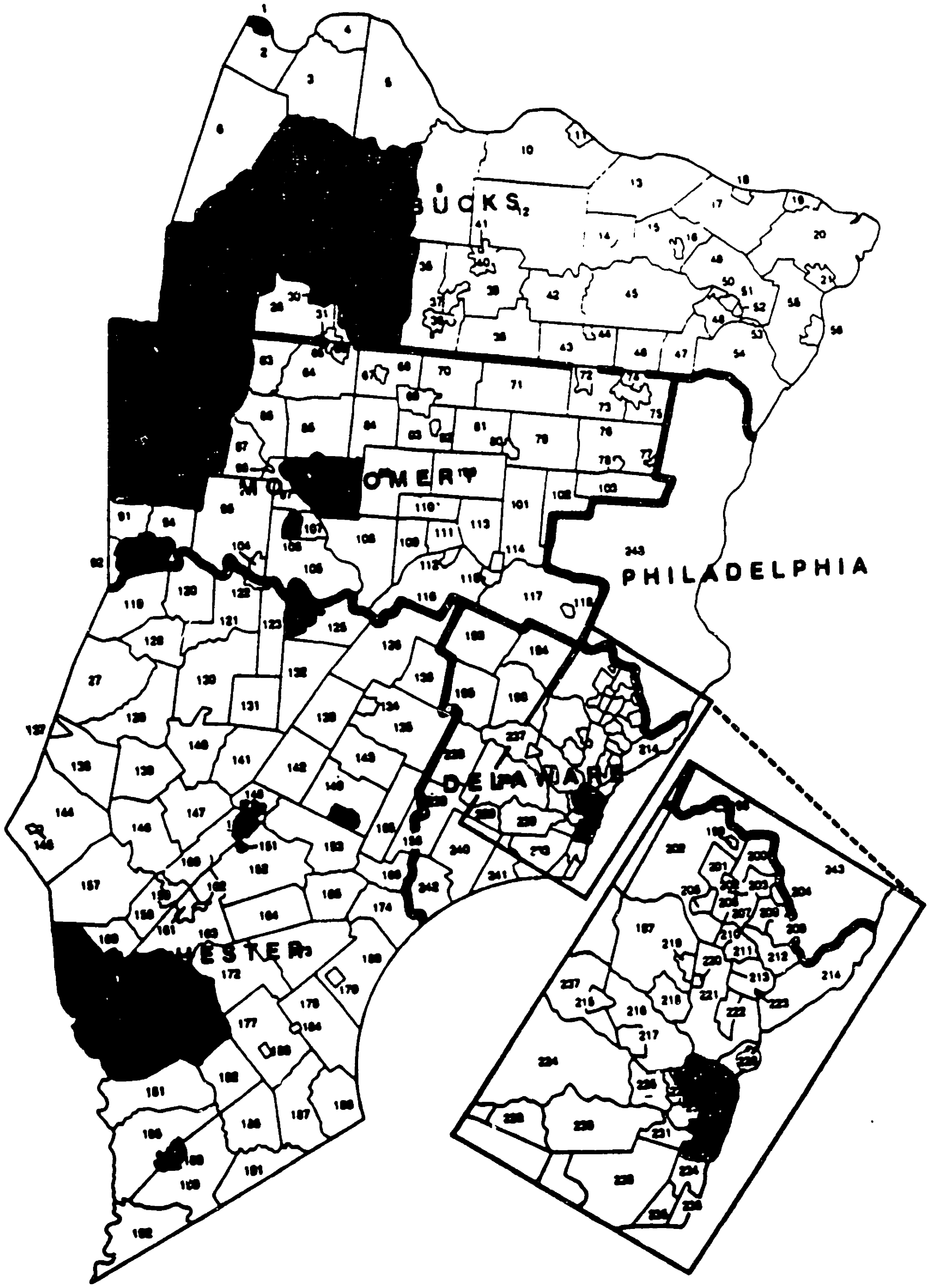
Index' means $c = \ln(\text{overhead})/\ln(\text{expenditure})$ and Index" has $c = \ln(\text{overhead})$.

$\ln(\widetilde{R})$ and $\ln(\widetilde{W})$ only for observations with levy in place.

(t-statistics)

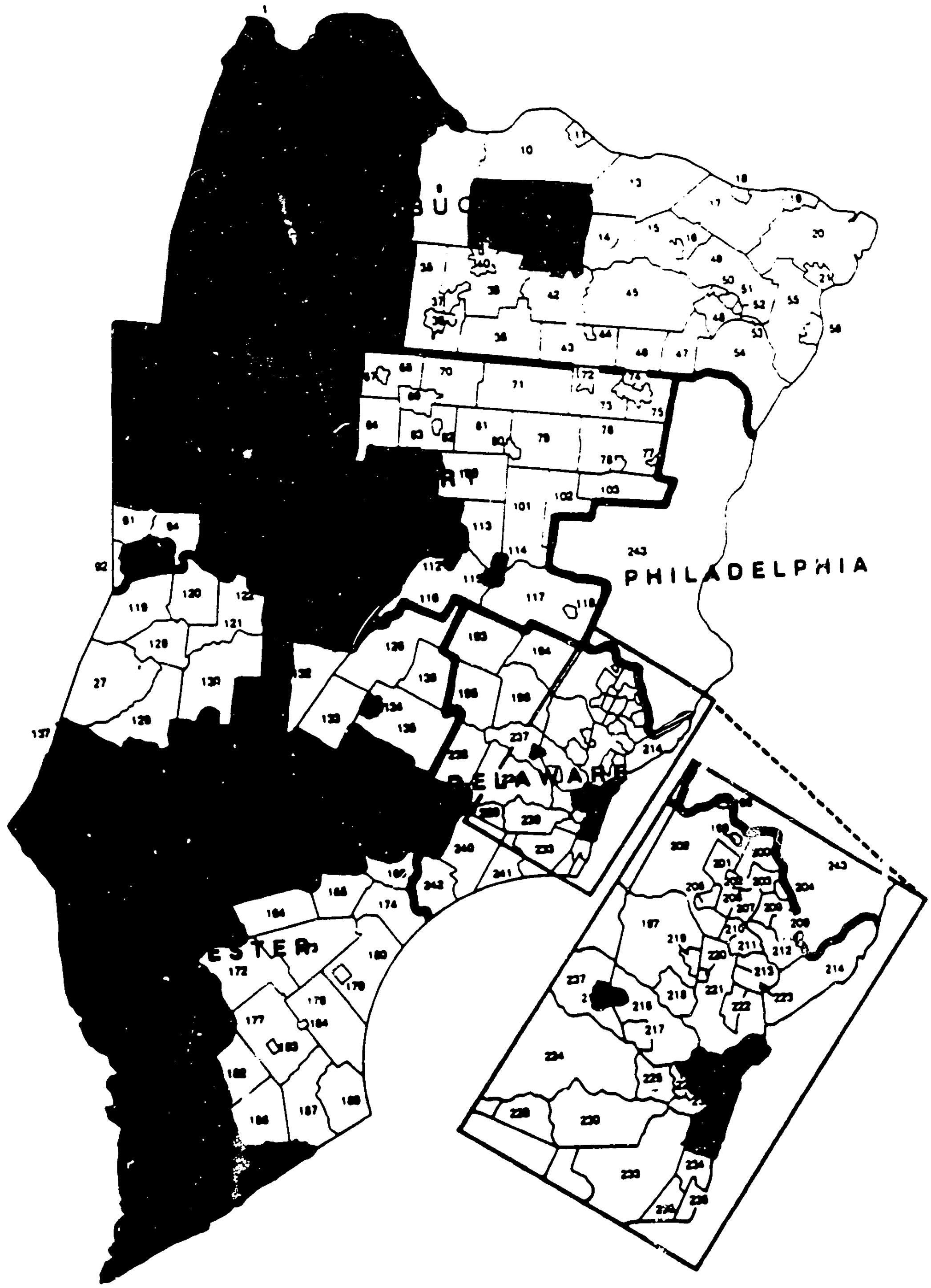
Diagram 1: Effect on Public Spending of Revenue Windfall





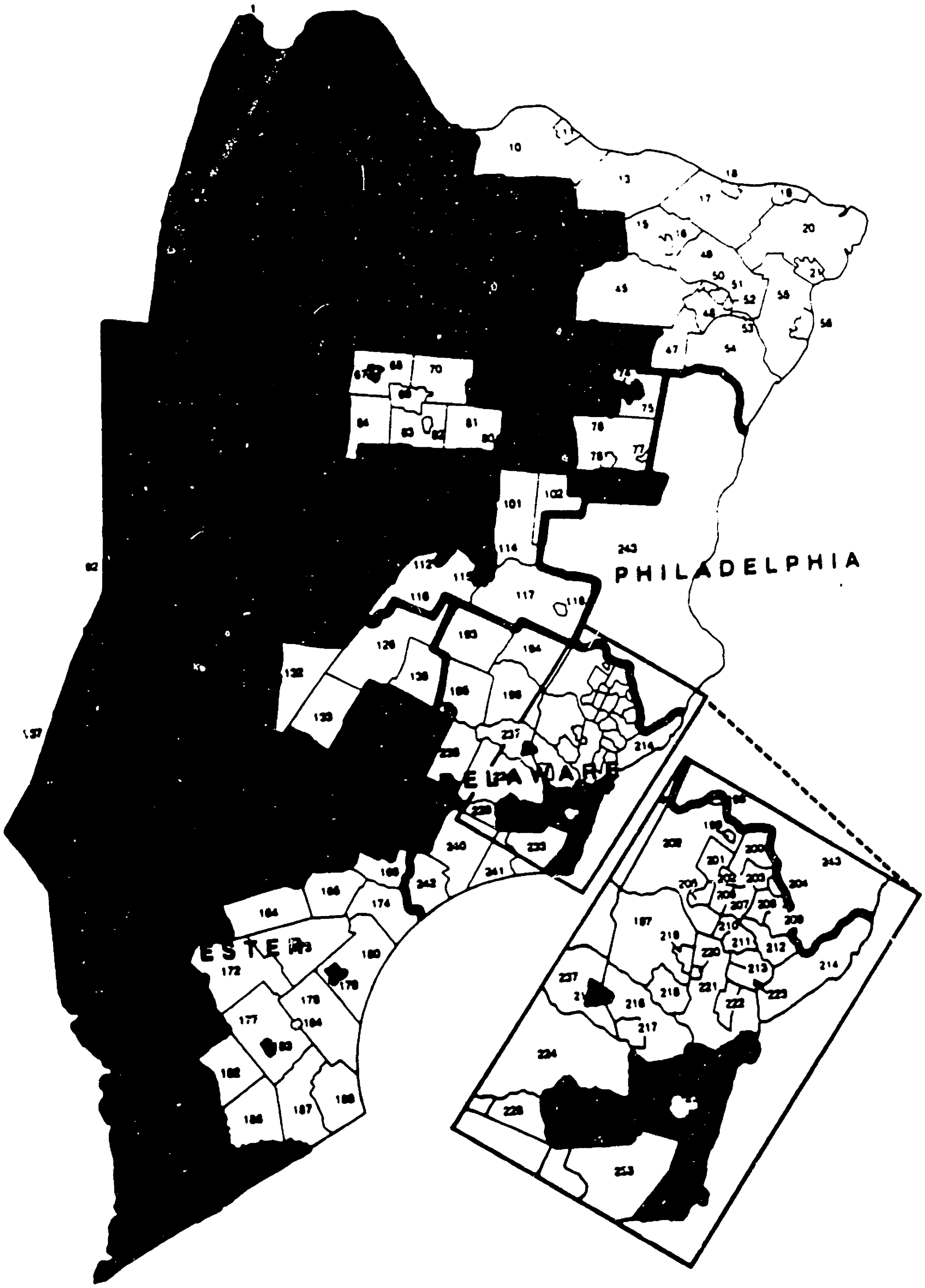
Map 1: Earned Income Tax in 1970

Shaded communities have tax



Map 2: Earned Income Tax in 1980

Shaded communities have tax



Map 3: Earned Income Tax in 1992

Shaded communities have tax

Figure 1: Fraction of Total Tax Revenue

from EIT by Years after First Levy

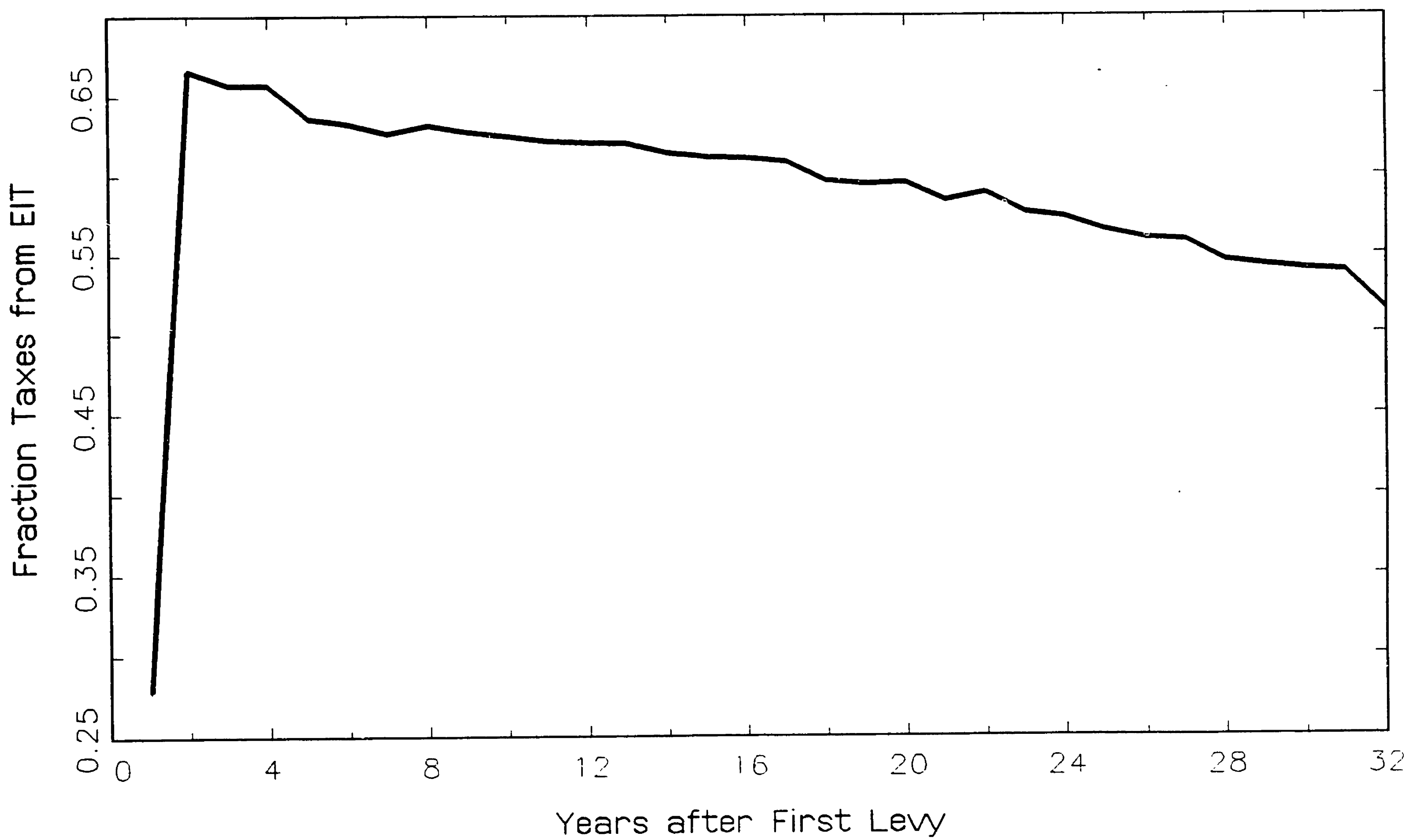


Figure 2: Fraction of EIT Revenue from Non-

Residents by Years after First Levy

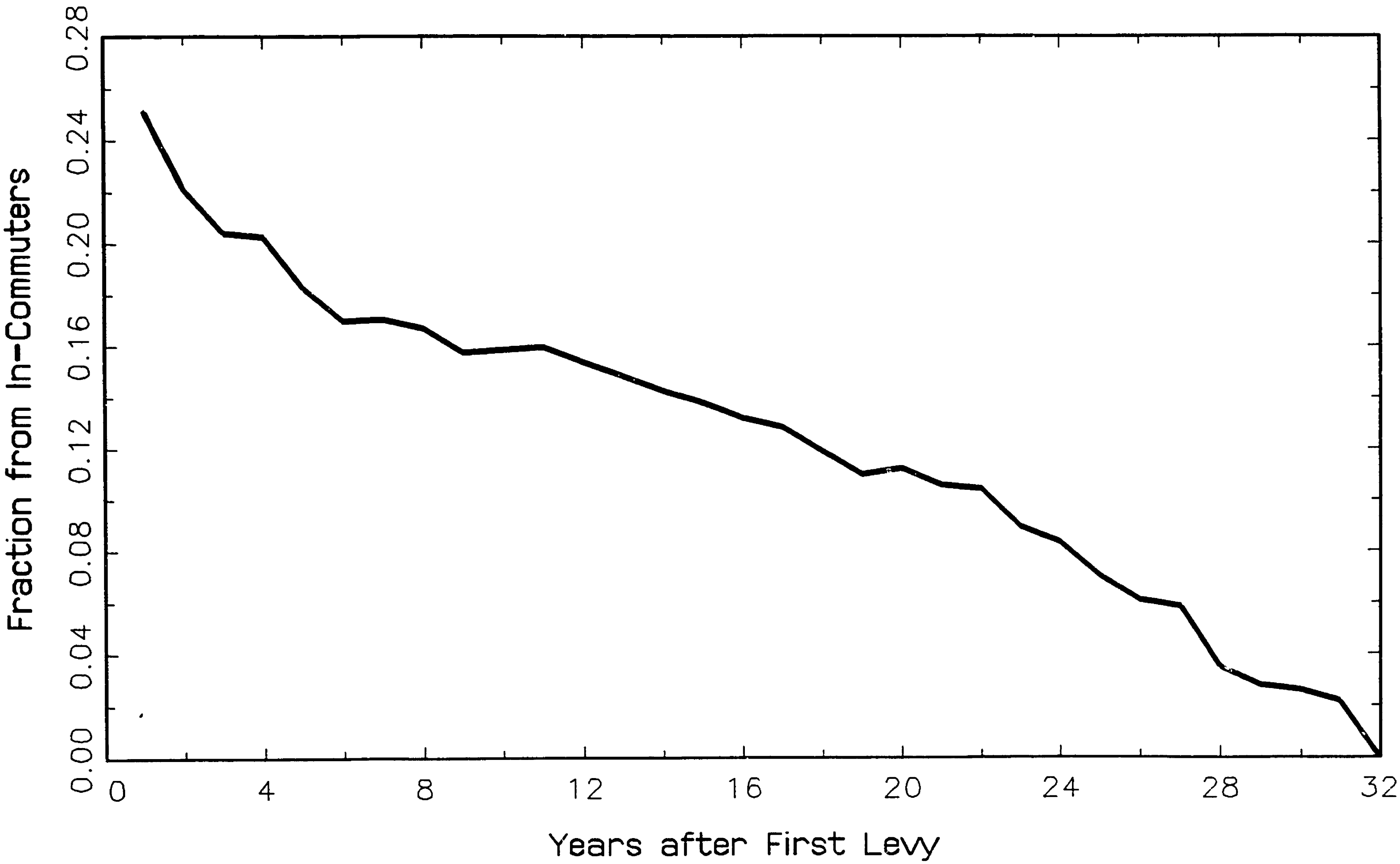


Figure 3: Percent Paying Tax Abroad Plus Seniors

in Year Before Levy

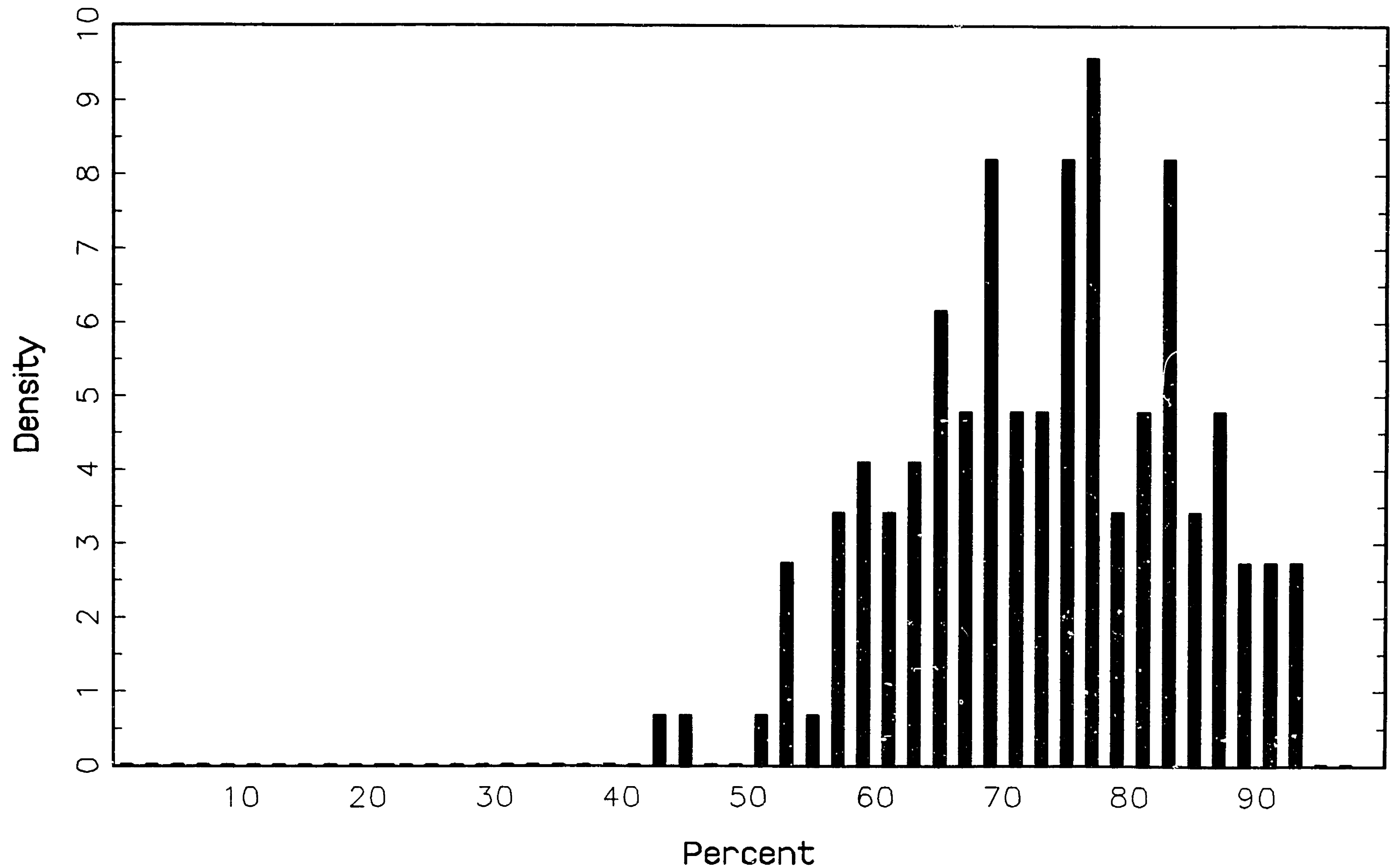


Figure 4: ln[Overhead Spending] as Function of ln[Expenditures]

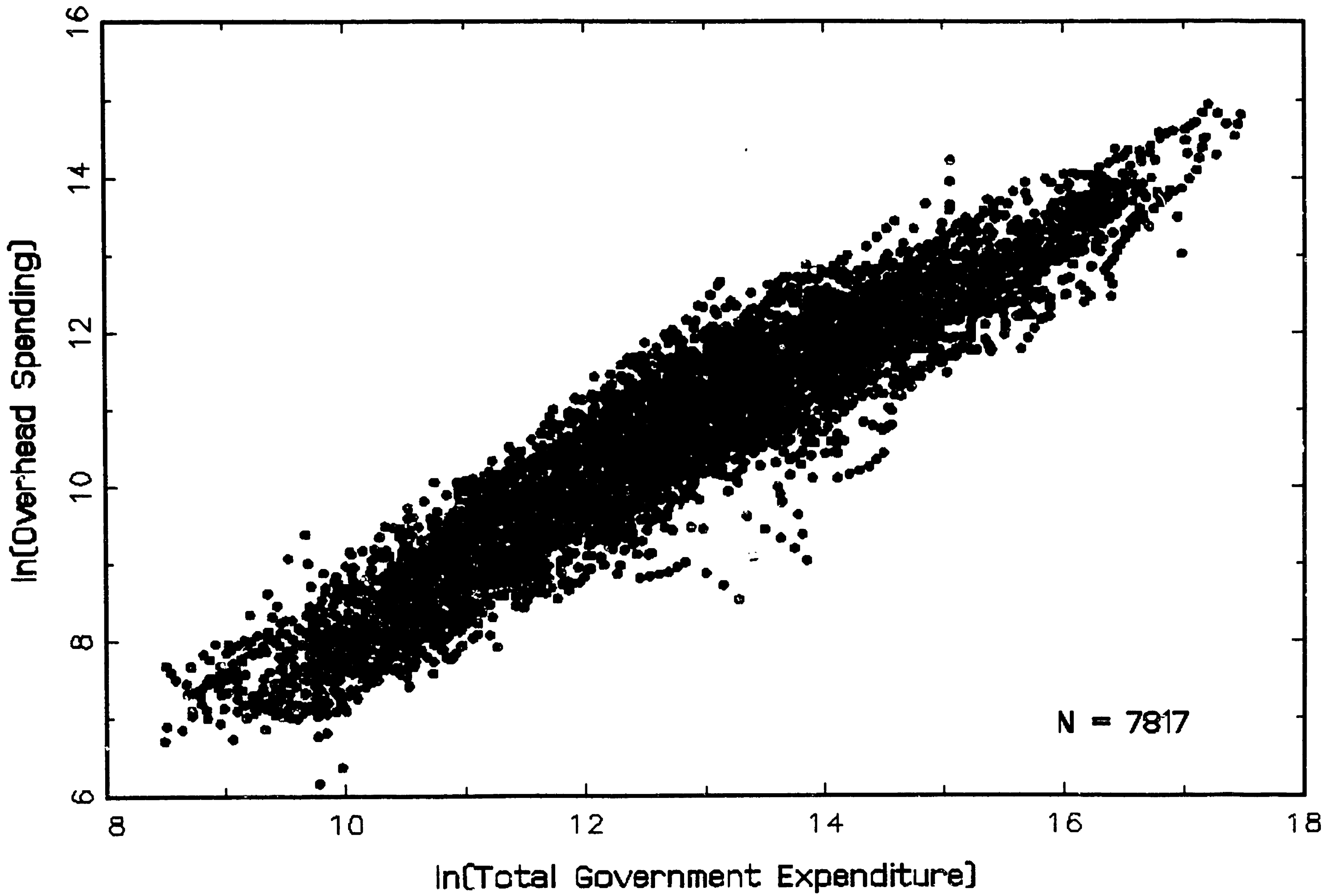


Figure 5: Distribution of Overhead Index for All Communities

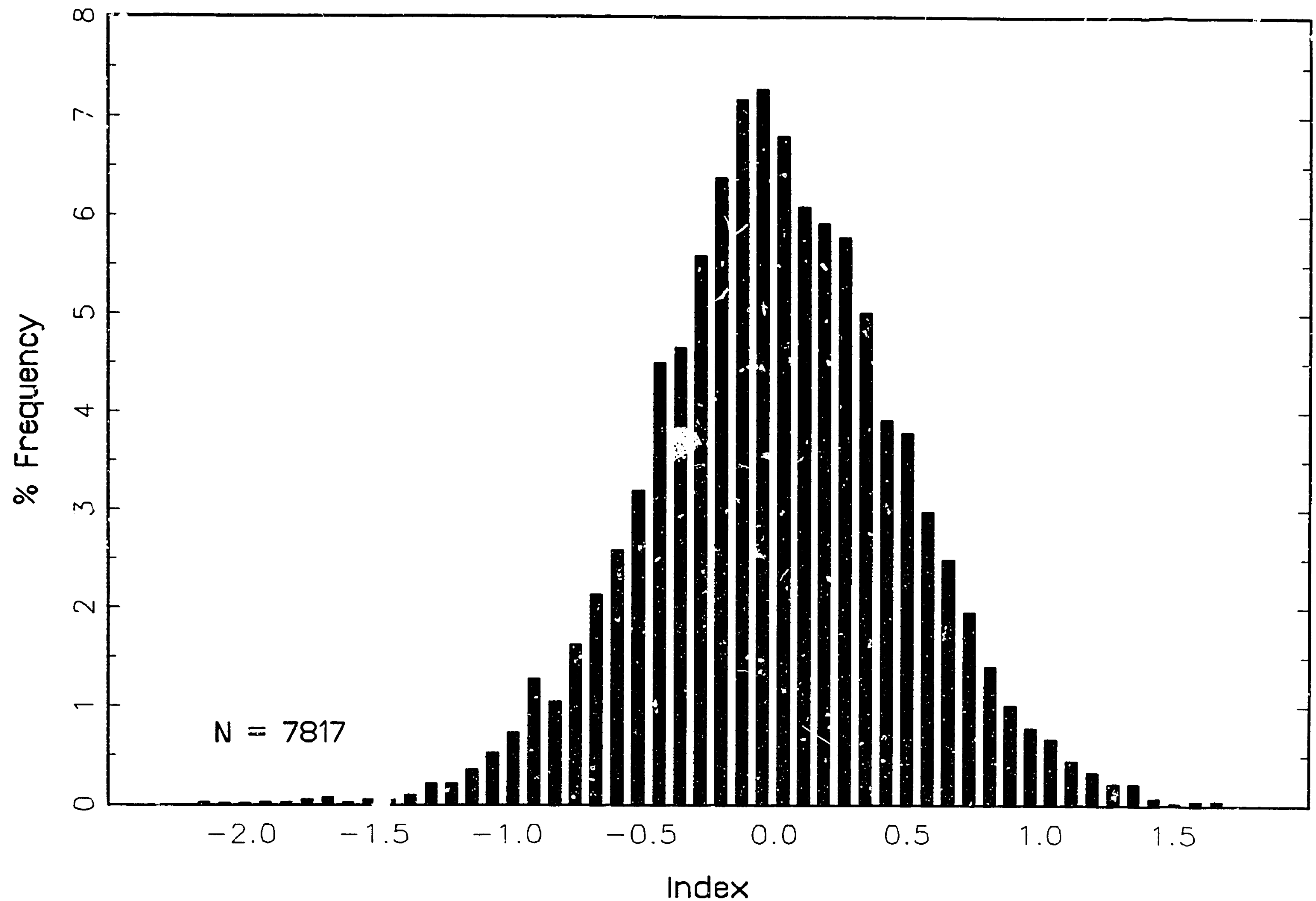


Figure 6: Distribution of Overhead Index in Year Before Levy

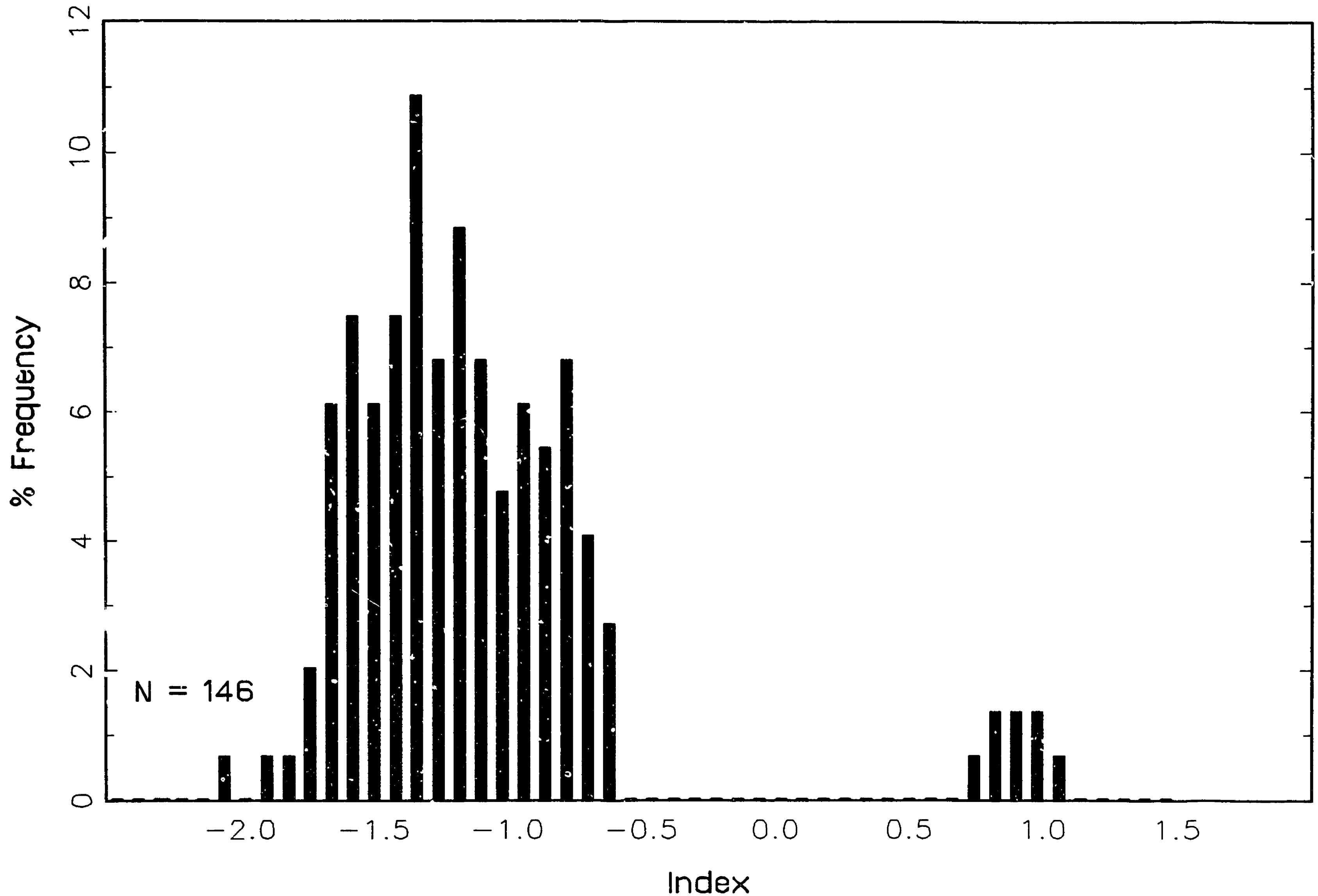


Figure 7: Residuals from Spending Regression (3.4)

Includes Overhead Index

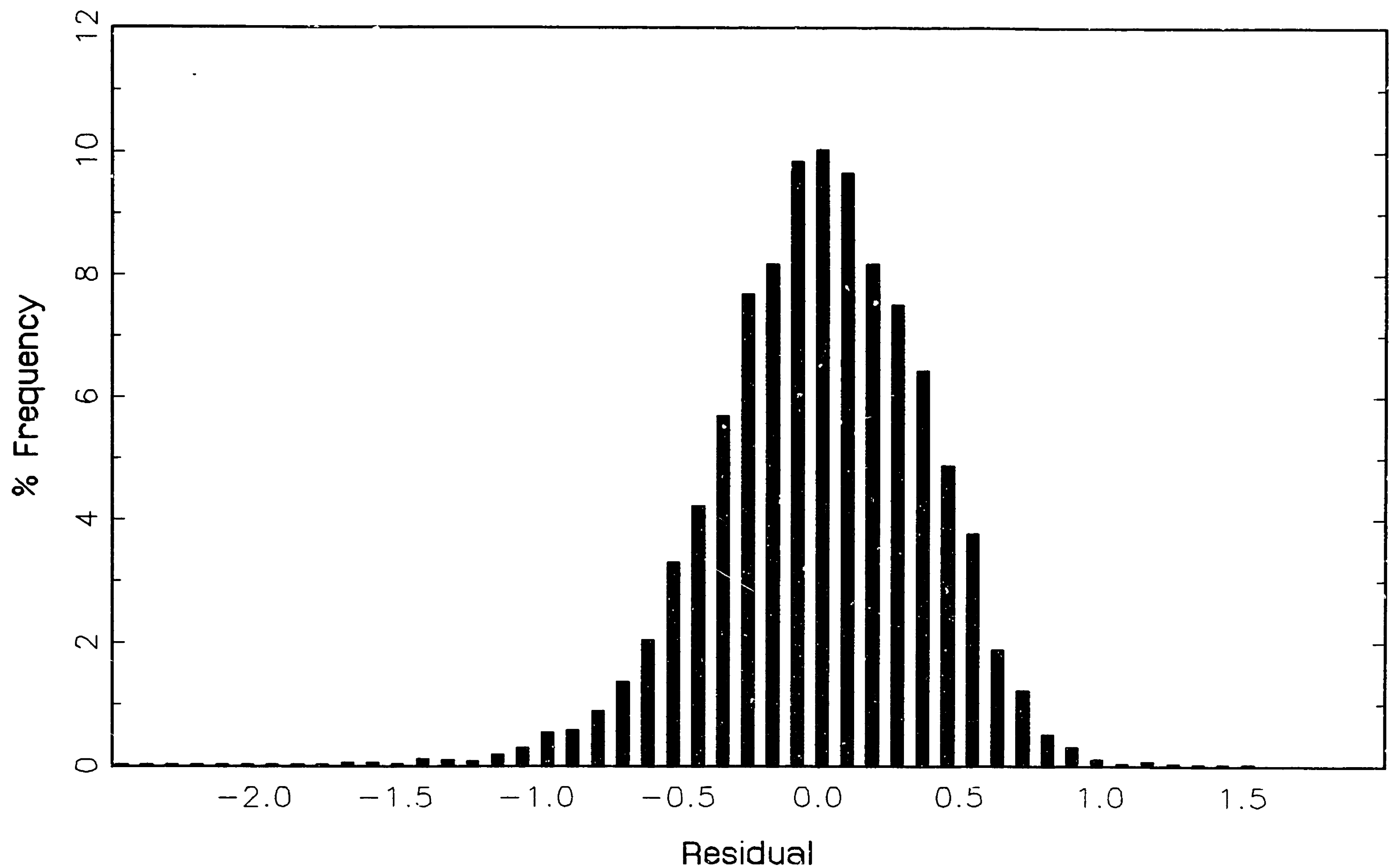


Figure 8: Residuals from Spending Regression (3.4)

No Overhead Index