Title

Permalink
https://escholarship.org/uc/item/93k1t6sg

Author
Hosek, Adrienne Laura

Publication Date
2014

Peer reviewed|Thesis/dissertation

By

Adrienne Laura Hosek

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Political Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Eric Schickler, Chair
Professor Jas Sekhon
Professor Sean Gailmard
Professor Rob Van Houweling
Professor Sean Farhang

Spring 2014

Copyright 2014
by
Adrienne Laura Hosek
Abstract


by

Adrienne Laura Hosek

Doctor of Philosophy in Political Science

University of California, Berkeley

Professor Eric Schickler, Chair

This dissertation considers how income inequality affects attitudes towards tax policies and income redistribution. In the first paper, I adopt traditional models of the relationship between income inequality and policy preferences to incorporate income volatility. I find that greater income volatility can lead voters to prefer less progressive taxes depending on the distribution of the income shocks. Because income volatility and inequality are positively correlated, with each potentially having an opposite effect on public opinion, my model predicts that support for increasing taxes on upper incomes may not rise in the face of growing income inequality. In the second paper, I estimate the casual effect of short term, temporary changes in household income on individuals’ tax progressivity preferences through a series survey experiments. I find evidence that volatility does in fact diminish preferences for tax progressivity. Respondents preferred significantly less progressive taxes when described households that had more volatile incomes regardless of the pattern of volatility over time. For the third and final paper, I develop a new, more powerful permutation procedure for analyzing two-way factorial experiments. The method is a non-parametric alternative to traditional analysis of variance (ANOVA). In Monte Carlo simulations, the procedure was better able to disentangle main factor and interaction effects than regression-based ANOVA tests, particularly when the design was imbalanced. I developed the procedure for use in the analysis of the income volatility experiments.
# Contents

List of Figures iii  
List of Tables v  
Acknowledgments vi  

## 1 Introduction  

### 2 Income Inequality, Income Volatility, and Redistribution  
2.1 Background 4  
2.2 Model of Inequality, Income Uncertainty and Redistribution 7  
2.3 Government allocation for redistribution 10  
2.4 Implications of Risk Aversion 12  
2.5 Countercyclical skewness 14  
2.6 Implications for consumption smoothing 17  
2.7 Conclusion 18  

## 3 Tax Preference on the Income Roller Coaster  
3.1 Context: Trends in Income Volatility and Income Inequality in the United States 22  
3.2 Survey Design 25  
3.3 Findings 30  
3.4 Survey Extension 37  
3.5 Results of survey extension 41  

## 4 Permutation Inference for Two-way Factorial Designed Experiments  
4.1 Two-way Factorial Designed Experiments 50  
4.2 Permutation Inference for Two Factor ANOVA 51  
4.3 Calculation of the Test Statistic 52  
4.4 Method of Permutation 54  
4.5 Uncorrelatedness 60  
4.6 Unbiasedness 62  
4.7 Consistency 63
# Contents

4.8 Simulations ................................................................. 65
4.9 Conclusion ................................................................. 78

Bibliography ................................................................. 79

A Map of vignette portion of the survey ................................. 83

B Survey questions ............................................................ 85

C Construction of ANOVA tests from OLS regression estimates 89
  C.1 One-way ANOVA ......................................................... 89
  C.2 Two-way ANOVA ......................................................... 90

D Additional simulation results ........................................... 95
List of Figures

2.1 Trust in government and income inequality, 1980-2010 ............................... 11
3.1 Perceptions of federal income taxes, 1992-2012 ............................................ 21
3.2 Design of survey vignette .................................................................................. 26
3.3 Preferred tax rates ............................................................................................. 32
3.4 Differences in preferred tax rates ..................................................................... 33
3.5 Differences in preferred tax rates ..................................................................... 34
3.6 Sample preferred tax rate schedules ................................................................. 34
3.7 Differences in preferred tax rates ..................................................................... 35
3.8 Differences in preferred level of tax progressivity ........................................... 36
3.9 Income patterns for survey extention ............................................................... 39
4.1 Simulation results: model (1), balanced design with normally distributed errors 70
4.2 Simulation results: model (2), balanced design with normally distributed errors 71
4.3 Simulation results: model (3), balanced design with normally distributed errors 72
4.4 Simulation results: model (4), balanced design with normally distributed errors 73
4.5 Simulation results: model (1), unbalanced design with normally distributed errors 74
4.6 Simulation results: model (2), unbalanced design with normally distributed errors 75
4.7 Simulation results: model (3), unbalanced design with normally distributed errors 76
4.8 Simulation results: model (4), unbalanced design with normally distributed errors 77
A.1 Decision tree of survey vignette ................................................................. 84
D.1 Simulation results: model (1), balanced design with exponentially distributed errors 96
D.2 Simulation results: model (2), balanced design with exponentially distributed errors 97
D.3 Simulation results: model (3), balanced design with exponentially distributed errors 98
D.4 Simulation results: model (4), balanced design with exponentially distributed errors 99
D.5 Simulation results: model (1), unbalanced design with exponentially distributed errors 100
D.6 Simulation results: model (2), unbalanced design with exponentially distributed errors 101
D.7 Simulation results: model (3), unbalanced design with exponentially distributed errors 102
D.8 Simulation results: model (4), unbalanced design with exponentially distributed errors .......................... 103
List of Tables

3.1 Table of means by treatment group ........................................ 31
3.2 Hypotheses for responses to pattern of income volatility ............. 38
3.3 Mean responses by treatment group ........................................ 43
3.4 Extended survey results ..................................................... 43
3.5 Estimates of income volatility pattern effects ............................. 44
3.6 Average federal tax rates, 2007 ........................................... 46

4.1 Examples of two-way unbalanced regular factorial designs .............. 48
4.2 $2 \times 3$ regular unbalanced factorial design .............................. 53
4.3 Calculation of partial test statistics ....................................... 53
4.4 The true model for a hypothetical $2 \times 2$ ‘get-out-the-vote’ experiment .... 67

C.1 One-way ANOVA table ..................................................... 90
C.2 Two-way ANOVA table ..................................................... 92
Acknowledgments

I thank Eric Schickler, Jas Sekhon, Sean Gailmard, Rob Van Houweling, Sean Farhang and Jon McCauliffe for their service on my political science dissertation and/or statistics master’s committees. They provided invaluable support and guidance on this project, as well as practical advice for how to navigate the dissertation and job market process. I have much gratitude to the entire faculty of the UC Berkeley Political Science Department, many of whom were willing to read drafts, attend talks and provide suggestions for how to improve my research despite not having an official position on my committee. In particular, I would like to thank Jack Citron and the staff at the Institute of Governmental Studies, Paul Pierson, Robert Powell, Gabriel Lenz, Terri Bimes, Henry Brady, Ernesto Dal Bo, Rui De Figueiredo, and Laura Stoker. Though not a member of UC Berkeley faculty, I also owe Betsy Sinclair tremendous gratitude for her guidance during the final stages of my dissertation. In addition, I gratefully acknowledge my current and former colleagues in graduate school: Erin Hartman, Chloe Thurston, Devin Caughey, and Rocio Titiunik for their continued support and inspiration. I thank Sebastian Hurtado Salazar for his love and encouragement. Finally, above all others, I am eternally grateful to my wonderful parents, Jim and Sue Hosek. They will always be my favorite teachers and most enthusiastic fans. The author is responsible for all errors.

This dissertation was typeset using the ucastrothesis \LaTeX{} template.
Chapter 1

Introduction

This dissertation considers how income inequality affects attitudes towards tax policies and income redistribution. In the first paper, I adopt traditional models of the relationship between income inequality and policy preferences to incorporate income volatility. In the second paper, I estimate the casual effect of short term, temporary changes in household income on individuals’ tax progressivity preferences through a series survey experiments. For the third and final paper, I develop a new non-parametric procedure for analyzing two-way factorial experiments. This procedure is also used in the analysis of the income volatility experiment.

The big puzzle in Political Science is why has the rise in income inequality in the U.S. not given rise to an increasing demand for redistribution and high taxes on the rich. Models in Economics based on the classic Meltzer-Richard model predict that as income inequality increases, public support to tax the rich and redistribute will rise (Meltzer and Richard 1981a). The incentive for the median voter to increase taxes on the rich increases as his income—the median income—falls below the mean income in the population. Instead, tax rates for top earners have gone down over the past 40 years, while the share of income going to the top one percent of income earners has doubled over the past forty years (Alvaredo et al. 2013). If the models are correct, why have not large numbers of middle class voters rebelled against current tax policies and pushed for substantial steps to redress the growing disparity between average working Americans and the top income earners? Most Americans are aware of the significant rise in income inequality and express concern that income has become increasingly concentrated. Yet a systematic examination of the data has not revealed a shift in public attitudes on taxation in response to increasing income inequality.

Standard explanations focus on elite preferences, specifically, the rising influence of business in politics, which has lobbied both parties effectively to reduce taxes on upper incomes. Hacker and Pierson and McCarty, Poole and Rosenthal, as well as Bartels provide evidence of elite capture and its contribution to income inequality (Hacker and Pierson 2005; Bartels 2008; McCarty, Poole and Rosenthal 2006). This does not explain, however, why preferences among individuals on issues of tax progressivity do not seem to be changing in response to growing income inequality. McCarty, Poole and Rosenthal point out that the influx of
immigrants, particularly illegal immigrants from Mexico, shifted the income distribution so that the median voter now earns more than the median income earner. As a result, the median voter no longer has the incentive to increase redistribution.

I argue that another important missing part of the story is the role of income volatility. Income volatility, or year-to-year changes in household income, has contributed significantly to the rise in inequality. The dissertation provides both theoretical and experimental evidence that income volatility tends to reduce public support for progressive taxation.

Models of income inequality and tax preferences overlook income volatility. In the first paper of my dissertation, I extend the standard Meltzer-Richard model to consider how temporary changes in income affect the decision to support a progressive tax policy. I find that greater income volatility can lead voters to prefer less progressive taxes depending on the distribution of the income shocks under conditions similar to those in the United States. Additionally, I find that risk aversion under the assumption of a quadratic utility function has an ambiguous effect on support for the progressive tax policy, indicating that increased income volatility may not lead voters to support a redistributive tax and transfer system. Because income volatility and inequality are positively correlated, with each potentially having an opposite effect on public opinion, my model predicts that support for increasing taxes on upper incomes may not rise in the face of growing income inequality.

For some intuition behind the result consider two possible extremes: one where there is high income inequality but no income volatility and the other where income inequality accounts for all income inequality. In the first extreme, the only way the middle class can gain ground relative to the upper class is to redistribute income through taxation. Now the other extreme: a world where there is high income volatility but all households have the same expected income. In this scenario, pressure to tax the rich would derive from individuals’ risk aversion and preferences for consumption smoothing. The remarkable aspect of existing models of income inequality and tax preferences is that they assume the first of these two extremes.

My experimental research explores possible hypotheses about the relationship between individuals’ income and their attitudes and behavior, including how changes in household income over time may affect tax preferences. I find evidence that volatility does in fact diminish preferences for tax progressivity. In the experiment, the respondents see a series of screens showing households experiencing greater versus less income volatility over a three year period; the degree of income volatility constitutes the treatment. Respondents preferred significantly less progressive taxes when described households that had more volatile incomes. The relationship persisted across a series of experiments conducted over two years. The decline in support for tax progressivity among the high volatility group occurred regardless of the pattern of volatility over time. Respondents’ tax attitudes did not depend on whether the hypothetical household performed better or worse in recent years, but according to the degree to which its income changed over the entire period. These findings suggest that to the extent that income volatility contributes to conditions of growing income inequality, it is possible that support for tax progressivity among the middle class will not rise as income inequality rises.
The demands of the experimental design led me to develop a new non-parametric method for the analysis of factorial experiments. For the third paper of my dissertation, I introduce a permutation procedure for analyzing experiments with multiple, interacting treatments. The method is a non-parametric alternative to traditional analysis of variance (ANOVA). The main advantage of permutation methods is that they do not have the normality assumptions of their parametric counterparts, and thus perform well even in experiments with few observations where the responses do not have well-behaved distributions. Another advantage of my method is that it accommodates experiments with unequal numbers of observations per treatment group and produces consistent estimates of the effects of the main factors even when there are significant interaction effects. The procedure is similar to one proposed by Basso et al. (2009), but with changes to increase the power of the test, i.e., the ability to correctly reject the null hypothesis. In addition, I consider the performance of the procedure for unbalanced factorial designs with the inclusion of proofs for uncorrelatedness, unbiasedness and consistency of the tests under conditions of unbalance.

Monte Carlo simulations comparing the non-parametric method I develop to the standard regression (ANOVA) method demonstrate substantial gains even under conditions when the assumptions of regression are met, namely when errors are independently and identically distributed normal. This is a result of the uncorrelatedness of the non-parametric tests for the main factor and interaction effects. The gains are also a direct result of improvements in the permutation procedure which enhance its ability to detect significant and small effects compared to alternate non-parametric procedures. Under conditions where we would expect parametric methods to perform poorly, namely when the errors do not have well-behaved distributions and when the design is unbalanced and has factors with multiple (> 2) levels, the gains from using the permutation method over the traditional parametric approach increase further.

Altogether, my research suggests that income volatility among ordinary citizens provokes a desire for lower taxes (on oneself and the affluent) rather than for government to protect citizens against increased risk (with new revenues from greater taxes on high earners). The theoretical model predicts an ambiguous relationship between income inequality and support for a progressive tax policy in the presence of income volatility. Under reasonable conditions similar to those observed in the United States, income volatility may mitigate public demands for redistribution in the face of growing income inequality. Through a series of experiments that test the relationship between transitory changes in household income and individuals preferences for tax progressivity, I find that experience with higher income volatility consistently led respondents to prefer significantly less progressive taxes. A subsequent experiment confirmed the initial finding and ruled out the possibility that the pattern of volatility, not the level, determined preferences for tax progressivity. To analyze the results, I develop a non-parametric permutation procedure as an alternative to traditional analysis of variance (ANOVA).
Chapter 2

Income Inequality, Income Volatility, and Redistribution

How do the dynamics that give rise to economic inequality alter the relationship between majority voting and income distribution? The factors that generate economic inequality in the United States today are not the same as those acting thirty years ago. Increasingly, individuals experience greater income shocks from one year to the next. This volatility in earnings, which explains a significant fraction of the increase in income inequality, has different welfare consequences from derived in traditional models of inequality, which assume individuals face no economic uncertainty. The long term consequences of economic risk depend on the nature of these shocks, whether they are temporary or permanent, and the extent to which public and private insurance markets can enable households to weather unforeseen events. I develop a model of income inequality and redistribution that incorporates uncertainty in future earnings and find that expectations about future earnings may moderate demands for redistribution when economic shocks are transitory in nature. I consider how factors such as government investment in redistribution and risk aversion affect model predictions and then evaluate the model based on recent evidence on the countercyclical behavior of shocks from the Economics literature.

2.1 Background

Income inequality has risen steadily in the United States over the past thirty years. The dynamics of income inequality, however, have not remained constant over this period. Heathcote, Giovanni and Violante (2009) summarize the trends across multiple measures of income. They find that the particular factors generating greater dispersion in household earnings have changed from decade to decade. In the 1970s, the United States experienced a rapid decline in traditional blue-collar jobs. Growing unemployment and a reduction in relative hours worked among low income workers contributed to rising inequality up to the mid 1980s, particularly at the bottom of the income distribution. At the same time, labor
demand shifts in favor of workers with white-collar skills drove up the college educated premium and contributed to growth in income inequality in the 1980’s and 1990’s. Since the 1990’s, however, individuals have experienced greater transitory shocks to income from year to year. Heathcote, Giovanni and Violante (2009) estimate that “a significant fraction of the overall increase in wage inequality was transitory" and argue that transitory shocks in income likely explain the divergence between income and consumption inequality over the past thirty years.

Several studies using data from the Consumer Expenditure survey (CEX) have found an increasing divergence in the trends for income and consumption inequality (Slesnick 2001; Primiceri and vanRens 2009; Heathcote, Giovanni and Violante 2009). Krueger and Perri (2003) estimate that approximately sixty percent of Americans experienced a welfare loss of 0.5% from 1972-2000. However, the welfare losses from increased inequality when based on earnings data alone grew much more rapidly, with a welfare loss of 10% for 80%-100% of households. Fisher and Johnson (2006) trace individuals over several periods by merging CEX data with panel data from the Panel Study of Income Dynamics (PSID) to investigate the contribution of income mobility in mitigating welfare disparities. They conclude that significant movement within the income distribution by individuals as they age has mitigated the social welfare loss due to increased inequality. Heathcote, Giovanni and Violante (2009) confirm this result and show that income dispersion within cohorts spikes in middle age, while dispersion in consumption rises steadily over the working years. Consumption appears unaffected by greater volatility in earnings, an indication that households have some ability to smooth these fluctuations through a mixture of borrowing and saving. To the extent that changes in income are temporary and consumption smoothing is possible, an increase in the frequency and severity of such shocks can increase income inequality without greatly affecting consumption inequality.

Households face an increased likelihood of experiencing an increase or decrease in earnings from one year to the next (Gottschalk and Moffitt 1994; Katz and Autor 1999; Dynan, Elmendorf and Sichel 2008). As a result, previous year income has become a less consistent predictor of current income. The causes of increased earnings volatility are less clear. The risk of involuntary job loss has actually declined since the 1970s (Cunha and Heckman 2006; Dynan, Elmendorf and Sichel 2008; Gorbachev 2007), until the 2008 recession. Rather, changes in income appear to be related to increased wage instability and greater fluidity in the labor market (Davis and Kahn 2008; Krueger and Perri 2003; Heathcote, Giovanni and Violante 2009). Blundell, Pistaferri, Preston (2008) find that increases in the transitory, non-persistent components of income explain much of the recent growth in the variance of log income. The stability of consumption inequality relative to income inequality is consistent with evidence suggesting that shocks to earnings are temporary in nature (Heathcote, Storesletten and Violante 2008; Primiceri and vanRens 2009).

Economic inequality can arise through a number of different factors, as noted. Most models of redistribution assume the distribution of income and individuals’ relative positions within that distribution are known. However, this approach may have less explanatory power in a country like the United States, where unpredictable, transitory shocks to income play
an important role in the dynamics of inequality. The focus of this paper is to understand how shocks in income affect tax policy preferences and demand for redistribution. In the next section I will discuss some earlier models of inequality and majority voting and then propose an extension of these models that incorporates income uncertainty.

My work draws from an extensive literature on income distribution and majority voting on tax policies. In his foundational paper, Romer (1975) models the choice between tax rates and productivity, where individuals trade off consumption and leisure in reaction to the generosity of government programs. Meltzer and Richard (1981) extend this work to make the size of government endogenous. These studies conclude that the demand for redistribution depends on the relationship between per capita income and the income of the median voter. As the ratio of mean income to median income rises, both models predict that tax policy will become more progressive. In addition, as public subsidies rise, more citizens decide to drop out of the labor force and subsist on government payments. Empirical tests of these models, however, have not confirmed the relationship between the tax rate and productivity. Rather, there are many examples of countries with higher tax rates that also have higher labor force participation rates.

The Scandinavian exception has led several scholars to develop models of endogenous growth that examine how taxation decisions affect capital investment and economic development and vice versa. These models attempt to explain the phenomenon known as the Kuznets curve, which posits that income inequality initially increases then decreases with economic development. They suggest how income distributions can change over generations depending on distributive politics and the initial level of economic inequality. Perotti (1993) shows how distributional considerations affect society’s choice of investment in human capital, specifically in education. Alesina and Rodrik (1994) consider the relationship between capital investment and distribution. In a society with ex-ante equitable distribution of capital, demand for taxes will be low as individuals benefit from greater capital investment in the economy. They find that more unequal societies have higher redistribution, but experience a lower growth rate as capital is extracted and redistributed through government subsidies. Persson and Tabellini (1994) reach a similar conclusion in their study of distribution effects on income dynamics. Disparities in income give rise to political support for large-scale market interventions that stymie growth. These models of endogenous growth share a common focus on explaining why some countries never escape from inequality despite relatively progressive tax regimes, while at the same time other countries adopt similar policies with apparent success.

Few studies explore how economic conditions affect individual perceptions about future economic well-being and the benefits of distributive programs. Alesina and Angeletos (2005) discuss how social perceptions of public subsidies and the individuals who rely on them determine support for progressive tax policies. They note that Americans attribute success to an individual’s work ethic, while Europeans are far less likely to associate poverty with poor life decisions. As a result, Europeans perceive aggressive redistribution as necessary to counter inequalities that arise from social inequity.

In this paper, I focus on individual perceptions of future income as economic uncertainty
2.2. MODEL OF INEQUALITY, INCOME UNCERTAINTY AND REDISTRIBUTION

rises. To the extent that individuals believe that income shocks are temporary and do not have lasting consequences for economic well-being, shifts in the distribution of income may not impact preferences for tax policy. Unlike previous models, the relationship between the median and mean income is not necessarily the most salient factor in determining the level of provision of public subsidies.

2.2 Model of Inequality, Income Uncertainty and Redistribution

The model is based primarily on those developed by Meltzer and Richard (1981a) and Perotti (1993) with the addition of uncertainty. Each individual has an annual income, $y_{i,t}$, which includes both a deterministic component, $x$, and an idiosyncratic stochastic component, $\varepsilon$.

$$y_{i,t} = x_i + \varepsilon_{i,t} \quad (2.1)$$

An individual receives $x_i$, a time-invariant stream of income that is determined by fixed characteristics about the individual associated with productivity. In addition to this predictable component of income, the individual receives a random shock, $\varepsilon_{i,t}$, that is distributed according to some continuous distribution $h(\cdot)$ with expectation $\bar{\varepsilon}$ and median $\tilde{\varepsilon}$. The density is independent and identically distributed for all individuals and uncorrelated across time.

Across individuals, the density of $x$ is denoted by $g(x)$. If there is no uncertainty, then we can simplify (1) to $y_{i,t} = x_i$ and the density of income across individuals is $f(y) = g(x)$. Otherwise, the density of income is denoted by the convolution of $g(x)$ and $h(\varepsilon)$:

$$f(y) = \int_{-\infty}^{\infty} g(x) h(y - x) dx$$

The expectation of individual income is $y_{i,t}$ is $E(y|x) = x + \bar{\varepsilon}$, while the expectation of income across individuals, or per capita income is:

$$\bar{y}_t = \bar{y} = E(y) = \int_{-\infty}^{\infty} y f(y) dy = E(E(y_i)) = \int_{-\infty}^{\infty} (x + \bar{\varepsilon}) g(x) dx = \bar{x} + \bar{\varepsilon}$$

This result follows from the fact that $x$ and $\varepsilon$ are independent and $\varepsilon$ is independently and identically distributed.

Income is taxed according to a flat tax rate, $\tau \in [0, 1]$. The tax revenue is collected by the government and redistributed in a lump sum payment to individuals, $k$. Assume a balanced budget requirement and $k = \tau \bar{y}$. As long as $\tau_i > 0$, the tax system is progressive, since the relative size of the subsidy decreases in income, that is, the fixed subsidy is a higher percent of income for low-income individuals than it is for high-income individuals. Additionally,
2.2. MODEL OF INEQUALITY, INCOME UNCERTAINTY AND REDISTRIBUTION

assume that individuals’ consumption equals their disposable incomes and there is no saving or borrowing.\(^1\) Thus, consumption is:

\[
c_{i,t} = (1 - \tau_t) y_{i,t} + \tau_t \bar{y}, \quad c \geq 0
\]  

(2.2)

where \( y_{i,t} = x_i + \epsilon_{i,t} \)

The game consists of two periods. The public votes on a tax policy in period one, which will apply to the income they earn over the upcoming year. The tax policy consists of a tax rate and lump sum subsidy, \( k\), that is applied uniformly to all citizens. Individuals know \( x\), which is the fixed component of their income, but not \( \epsilon\), the variable shock. They only know the distribution of shocks, \( h(\epsilon)\).

An individual’s most preferred tax policy is one that maximizes his expected utility, which is strictly increasing in consumption.

\[
\max_{\tau \in [0,1]} EU(c) = \max_{\tau \in [0,1]} EU[(1 - \tau)y + \tau \bar{y}]
\]

The most basic model assumes no risk aversion and utility is equal to consumption. An individual will support a tax policy, \( \tau\), if:

\[
EU[(1 - \tau)y + \tau \bar{y}] \geq EU(y)
\]

\( x + \bar{\epsilon} \leq \bar{y} \)

An individual’s voting decision depends upon their expected income, which is a function of the time-invariant portion of their income \( x\) and the expectation of the shock, \( \bar{\epsilon}\). They will favor a tax if their expected income is less than per capita income. If \( x + \bar{\epsilon} > \bar{y} \) the optimal tax rate is clearly zero. Individuals at the top of the income distribution do not benefit from a progressive tax policy. Only those who expect to earn less than per capita income desire taxes. Since \( \bar{\epsilon}\) is the same for all individuals, the level of taxes preferred is monotonically decreasing in \( x\) and does not depend on the realized income shock, \( \epsilon\). The same relationship holds in the basic model with no uncertainty. When there is no uncertainty, \( \epsilon = 0 \) and individuals favor the imposition of a tax if \( x \leq \bar{y} \).

As in previous models of majority voting, I apply the median voter theorem to determine the equilibrium tax policy. This well known theorem states that governments will propose policies that appeal to the median voter to win elections when the voting rule is simple majority. The median voter is the individual whose vote preference is the median of the distribution of all voter preferences. In both the model with and without uncertainty, support for taxation weakly increases with income and the median voter is the voter whose time-invariant income \( x\) is the median (For a proof of this result, see Roberts (1977)). Denote the median of \( x\) across all individuals as \( \tilde{x}\).

\(^1\)It is also easy to assume that the tax imposes a dead weight loss such that \( k = (\tau - \tau^2) \bar{y}\). Since the dead weight loss increases with \( tau\), it’s inclusion diminishes support for taxation in the public and limits the level of taxation preferred by supporters.
2.2. MODEL OF INEQUALITY, INCOME UNCERTAINTY AND REDISTRIBUTION

Exactly who that individual is depends on the initial distribution of incomes and the distribution of shocks for the period. In the basic model the median income \( \tilde{y} \) in society is equal to \( \bar{x} \). The addition of an idiosyncratic stochastic component to income complicates the picture. It is no longer true that \( \tilde{y} = \bar{x} \) in a world with uncertainty. An individual whose predictable income is low, but receives a significant positive shock in the current period, may be the decisive voter. Conversely a wealthy individual experiencing a temporary loss in income may end up influencing policy.

To understand how income volatility changes the relationship between income inequality and preference for redistribution, it is helpful to consider the following thought exercise. In a simple model, an individual’s income is entirely determined by a fixed, non-varying factor \( \bar{x} \) such that \( y = x \) and \( f(y) = g(x) \). From this assumption, it follows that per capita income \( \tilde{y} = \bar{x} \) and median income \( \tilde{y} = \bar{x} \).

Suppose income inequality is rising. Specifically, the mean income is increasing while median income remains fixed. Mathematically, this conditions can be written:

\[
\tilde{y}_2 > \tilde{y}_1 \text{ and } \tilde{y}_2 = \tilde{y}_1 \tag{2.3}
\]

where \( y_1 \) and \( y_2 \) denote income in time one and time two. Denote \( \tilde{f}(y) \) as the new distribution of income across society, changed from \( f(y)_1 = g(x)_1 \).

In the simple model with no income uncertainty, all of the growth in income inequality is generated by changes in the persistent component of income, \( x \), such that \( y_2 = x_2 \). This condition implies that \( \tilde{f}(y_2) = \hat{g}(x_2) \) and \( \tilde{y}_2 = \bar{x}_2 \). Also the assumption that mean income increases but median income is unchanged implies that \( \tilde{y}_2 = \bar{x}_2 = \bar{x}_1 \). Support for taxation increases as a result of rising income inequality, so long as mean income increases relative to median income from time one to time two and, in both periods, preference for the tax policy depends on whether \( x \leq \bar{y} \). In particular, the median voter is more likely to prefer the tax policy in period two because \( \bar{x}_2 = \bar{x}_1 \) but \( \tilde{y}_2 > \tilde{y}_1 \).

Now consider the model with income uncertainty and the case where all of the rise in income inequality is generated by changes in idiosyncratic variation in \( \varepsilon \). In this case, \( \bar{x}_1 = \bar{x}_2 \), \( \bar{x}_1 = \bar{x}_2 \), \( \bar{\varepsilon}_1 < \bar{\varepsilon}_2 \), and \( \bar{\varepsilon}_1 = \bar{\varepsilon}_2 \). The median voter favors taxation if \( \bar{x}_2 + \bar{\varepsilon}_2 \leq \tilde{y}_2 \), or if \( \bar{x}_2 \leq \bar{x}_2 \). But \( \bar{x}_2 = \bar{x}_1 \) and \( \bar{x}_2 = \bar{x}_1 \) so the median voter condition is equivalent to \( \bar{x}_1 \leq \bar{x}_1 \). Let \( d \) equal the difference between the mean and median such that \( d_2 = \bar{x}_2 - \bar{x}_2 \). Similarly, let \( d_1 = \bar{x}_1 - \bar{x}_1 \). It follows that \( d_2 = d_1 \) so there is no change in demand for the tax policy within the polity between time one and time two.

While not a necessary condition, an increase in income inequality such that the mean income rises with respect to the median income due to a change in the density of the distribution of shocks can arise if \( \bar{\varepsilon}_2 < \bar{\varepsilon}_2 \). I leave it for a future proof that this condition is necessary in order to reliably produce a change in the income distribution that conforms to (3) over multiple draws of the vector, \( \varepsilon \).

This thought experiment suggests that if the increase in income inequality is in part due to an increase in income volatility (without a reduction in the inequality of persistent income), then demand for taxation will be weaker than in the case where there is no increase
in income volatility and all the income inequality is generated by changes in persistent income. Under these conditions, a change in the overall distribution of income that increases the mean income relative to the median income as in (3) can be caused by a change in the distribution of income shocks, which is also right-skewed.

\section{Government allocation for redistribution}

The early model assumes that the government divides equally the entire tax revenue among all citizens in the population. Suppose the government only allocates a portion of the tax revenue towards redistribution in the form of a lump sum subsidy. Let $\alpha \in [0, 1]$ indicate the proportion of revenue that is dedicated to redistribution, such that $k = \alpha \tau \bar{y}$. Thus, an individual's consumption will be:

$$c = (1 - \tau)y + \alpha \tau \bar{y}$$

An individual will support taxation if:

$$x + \bar{\epsilon} \leq \alpha \bar{y} = \alpha (\bar{x} + \bar{\epsilon})$$

Demand for taxation is strictly increasing in $\alpha$. The less the government allocates towards redistribution, the less the public stands to benefit from taxation, and support for a tax policy will diminish. It is clear that support will decrease further if a portion of the increased income inequality is due to income volatility under conditions described in the previous section. Specifically, if for a fixed distribution in persistent income, $g(x)$, income inequality rises in the form of (3) due to a rise in income volatility, we should see weaker support for taxation than if all of the rise in income inequality was due to changes in persistent income.

Government allocation for redistribution is connected to ideas of social trust in the political science literature. Rothstein and Uslaner (2005) examine the relationship between social trust and income inequality. They find that in societies with lower social trust, individuals have less faith in government to redistribute income fairly, i.e. in ways which benefit their own group. This arises in countries with high levels of income inequality, because individuals in these societies attribute disparities in income to a overall system, including the government, which favors the rich over the poor. As a result, inequality causes an erosion in social trust which in turn diminishes the desire to contribute to redistributive programs that would alleviate income inequality.

We can re-conceptualize $\alpha$ in terms of social trust to represent a general belief among all individuals in society that the government will fairly allocate tax revenue for redistribution. As levels in government trust decline for all individuals in society, $\alpha$ shifts closer to zero. Moreover, we can make $\alpha$ a function of income inequality so that as income inequality grows, social trust weakens. Under this framework where $\alpha$ varies inversely with growing income inequality, a rise in income inequality may not generate increased demand for the tax policy. Support for the final tax depends both on the change in the difference between mean and median income versus and the change in social trust:
\[ \bar{x} + \varepsilon - \Delta \alpha (\Delta \bar{x} + \varepsilon) = (\bar{x} - \Delta \alpha \Delta \bar{x}) + \varepsilon (1 - \Delta \alpha) \]

The decrease in \( \alpha \) can neutralize the increase in \( \bar{x} \) relative to \( \tilde{x} \) arising from worsening income inequality. Figure 2.1 plots the relationship between trust in government and the share of income going to the top 1 percent of income earners, a common indicator of income inequality. While we see a general trend towards increased income inequality between 1980 and 2008, we see changes in the trend that correspond to changes in government trust, particularly in the late eighties and mid-2000’s. As income inequality rises, levels of trust in government drop. However, we do not see the same trend hold for the late nineties, when both income inequality and levels of trust in government increase. A regression of social trust on income inequality reveals a negative but not significant relationship, providing some evidence that the two may be inversely related in the United Sates.

Figure 2.1: The figure shows the relationship between trust in government and the share of income held by the top 1 percent of income earners, a common measure of income inequality. The sources include the American National Election Studies’ trust in government index and IRS individual tax return tables prepared by Piketty and Saez (2003, updated 2012).
2.4 Implications of Risk Aversion

It is interesting to consider how the addition of risk aversion under conditions of income volatility from transitory shocks affects support for taxation. Risk-averse individuals prefer less uncertainty in their income and are willing to accept a lower level of expected income for lower income volatility. In the model with income uncertainty presented above, taxation and redistribution act to dampen the impact of annual income shocks. To see this, recall that in this model an individual has consumption

\[ c = (1 - \tau)y + \alpha \tau \bar{y} = (1 - \tau)(x + \bar{\epsilon} + \alpha \tau \bar{y}) \]

Consider two possible income shocks the individual may experience, \( \varepsilon_a < \varepsilon_b \). If shocks are independent of permanent income, a higher realized shock has no effect on mean income, \( \bar{y} \). Evaluating the equation for consumption at \( \varepsilon_a \) and \( \varepsilon_b \), we see that the difference in the individual’s consumption between the higher and lower realized income shocks with taxation and redistribution is smaller than the difference in the individual’s consumption without taxation and redistribution:

\[ (1 - \tau)(\varepsilon_b - \varepsilon_a) < (\varepsilon_b - \varepsilon_a) \]

This is true for individuals at all levels of time-invariant income \( x \) and for all redistribution policies \( \alpha \). Therefore, if risk aversion is present the tax-and-redistribute policy increases the probability that voters will support a tax and redistribution policy.

How would adding risk aversion to the model alter its predictions for the effect of increasing income inequality on voter support for redistribution funded through an income tax? As before, consider the two extreme cases, where growing inequality results entirely from increasing inequality in permanent income or entirely from higher income volatility. In the first case, higher income inequality is not associated with an increase in income uncertainty and the earlier finding that support for taxation rises with income inequality is unchanged.

However, turning to the second case, where increasing income volatility is the cause of increasing inequality, it is unclear a priori how risk aversion affects support for taxation. Intuitively, this is because the individual’s expected income and expected redistribution payment are both subject to greater uncertainty.

I will use an example to illustrate how risk aversion can affect the fraction of voters willing to support a redistribution policy. For tractability, the example assumes a quadratic utility function and as before there is no saving or borrowing and all income after possible tax and transfer is consumed.

\[ u(c) = c + b c^2 \quad c \in [0, \frac{1}{2b}], \quad b > 0 \]

The function is limited to the range where utility is increasing with consumption. Absent tax and transfer, consumption equals income, \( c_i = y_i = x_i + \varepsilon_i \), and under tax and transfer \( c_i = (1 - \tau)y_i + \tau \bar{y} = (1 - \tau)(x_i + \varepsilon_i) + \tau \bar{y} \).
I assume that the income stocks have a linear density over a finite support, i.e. \( \varepsilon \sim h(\varepsilon) = \theta_0 + \theta_1\varepsilon \), with \( \varepsilon \) falling within a finite range and normalized so that \( \varepsilon_i \in [0, 1] \).\(^2\) Here, \( \theta_0 \) and \( \theta_1 \) are positive, which implies that the density is upward sloping. At \( \varepsilon_i = 0 \), the value of the density is \( \theta_0 \), and at \( \varepsilon_i = 1 \), it is \( \theta_0 + \theta_1 \). The fact that the integral under the density must integrate to one imposes a relationship between \( \theta_1 \) and \( \theta_0 \):

\[
1 = \int_{0}^{1} (\theta_0 + \theta_1\varepsilon)d\varepsilon = \theta_0 + \frac{1}{2}\theta_1, \text{ so } \theta_1 = 2(1 - \theta_0)
\]

The median of \( h(\varepsilon) \) occurs at \( \varepsilon = \frac{1}{2} \) and the density at the median is \( \theta_0 + 2(1 - \theta_0)(\frac{1}{2}) = 1 \), i.e., the density at the median is independent of the parameter \( \theta_0 \) and \( \theta_1 \). However, a decrease in \( \theta_0 \) requires an increase in \( \theta_1 \) to keep the density proper.

Now we compare expected utility with and without the tax policy to assess the net gain of the tax policy. Expected utility in the absence of the policy is:

\[
EU(c_i) = E[x_i + \varepsilon_i - b(x_i + \varepsilon_i)^2] = x_i\varepsilon - bE[x_i^2 + 2x_i\varepsilon_i + \varepsilon_i^2]
\]

\[
= x_i + \varepsilon - b(x_i^2 + 2x_i\varepsilon + E\varepsilon_i^2) \equiv x_i + \varepsilon - B_0
\]

Expected utility in the presence of the tax policy is:

\[
EU(c_i|\tau) = E[(1 - \tau)(x_i + \varepsilon_i) + \tau\bar{y} - b((1 - \tau)^2(x_i^2 + 2x_i\varepsilon_i + \varepsilon_i^2) + 2(1 - \tau)\tau(x_i + \varepsilon_i)\bar{y} + \tau^2\bar{y}^2)]
\]

\[
= (1 - \tau)(x_i + \varepsilon) + \tau\bar{y} - b((1 - \tau)^2(x_i^2 + 2x_i\varepsilon + E\varepsilon_i^2) + 2(1 - \tau)\tau(x_i + \varepsilon)\bar{y} + \tau^2\bar{y}^2)
\]

\[
\equiv (1 - \tau)(x_i + \varepsilon) + \tau\bar{y} - B_1
\]

\( EU(c_i|\tau) \) can be expressed as \( EU(c_i) \) plus a term related to the change in the risk aversion term in the presence of the policy:

\[
EU(c_i|\tau) = EU(c_i) + \tau(\bar{y} - (x_i + \varepsilon)) - b((-2\tau + \tau^2) + 2(1 - \tau)\tau(x_i + \varepsilon)\bar{y} + \tau^2\bar{y}^2)
\]

Per capita income is the sum of the mean deterministic component of income and the mean shock:

\[
\bar{y} = \bar{x} + \varepsilon
\]

We can express the net gain from the policy as:

\(^2\)Suppose the raw range of \( \varepsilon_i \) is from \( a \) to \( b \), such that \( b > a \). Then starting with \( y_i = x_i + \varepsilon_i \), we transform \( \varepsilon_i \) so that its range falls between 0 and 1. Set \( \eta_i = \frac{\varepsilon_i - a}{b - a} \) and then substitute \( \eta \) for \( \varepsilon \) in \( y_i = x_i + \varepsilon_i \) to get \( \frac{y_i - a}{b - a} = \frac{x_i - a}{b - a} + \eta_i \). Now we can define \( y_i \) and \( x_i \) according to this transformation without affecting the assessment of the net benefit of the tax policy.
2.5. COUNTERCYCLICAL SKEWNESS

\[ EU(c_i|\tau) - EU(c_i) = \tau(-x_i + \bar{x}) - b(-2\tau + \tau^2 + 2(1 - \tau)\tau(x_i + \bar{\varepsilon})(\bar{\varepsilon} + \bar{x}) + \tau^2(\bar{\varepsilon} + \bar{x})^2) \]

Express \( x_i \) as a deviation \( \bar{x} \) such that \( x_i = \bar{x} + \Delta \) to obtain:

\[ EU(c_i|\tau) - EU(c_i) = -\tau\Delta - b(-2\tau + \tau^2 + \tau^2(\bar{\varepsilon} + \bar{x})^2 + 2(1 - \tau)\tau(\bar{\varepsilon} + \bar{x})(\Delta + \bar{\varepsilon} + \bar{x})) \]

Isolate the terms involving the deviation (\( \Delta \)):

\[ EU(c_i|\tau) - EU(c_i) = -\tau\Delta - b(2(1 - \tau)\tau(\bar{\varepsilon} - \bar{x})\Delta) - b(-2\tau + \tau^2 + \tau^2(\bar{\varepsilon} + \bar{x})^2 + 2(1 - \tau)\tau(\bar{\varepsilon} + \bar{x})(\bar{\varepsilon} + \bar{x})) \]

This expression can be simplified to:

\[ EU(c_i|\tau) - EU(c_i) = -\tau\Delta - b(2(1 - \tau)(\bar{\varepsilon} - \bar{x})\Delta) - b((2 - \tau)(-1 + \bar{\varepsilon} + \bar{x})(1 + \bar{\varepsilon} + \bar{x})) \]

For \( \Delta > 0 \), the first, second and third terms of the expression are negative. This implies that for a person with permanent income (\( x_i \)) above the mean permanent income (\( \bar{x} \)), the presence of risk aversion unambiguously decreases support for the policy relative to the world in the absence of the policy. But for a person with \( x_i < \bar{x} \) where \( \Delta < 0 \), the first two terms are positive and the third term is negative. Depending on the values of the parameters, the overall expression could be positive. Further, the lower \( x_i \) is relative to the mean permanent income (\( \bar{x} \)), the more likely that the net gain of the tax policy (\( EU(c_i|\tau) - EU(c_i) \)) will be positive, that is, the more likely the person will support the policy.

We can also evaluate how the net gain changes with risk aversion by taking the derivative of the net gain with respect to the risk aversion parameter \( b \).

\[ \frac{\partial(EU(c_i|\tau) - EU(c_i))}{\partial b} = 2(-1 + \tau)\tau\Delta(\bar{\varepsilon} + \bar{x}) + (-2 + \tau)(-1 + \bar{\varepsilon} + \bar{x})(1 + \bar{\varepsilon} + \bar{x}) \]

The sign of this expression depends on whether \( \Delta \) is positive or negative, i.e., whether the person’s permanent income is above or below the mean permanent income. For \( \Delta > 0 \), the derivative is negative and an increase in the risk aversion parameter decreases the willingness to support the policy. However, we know that this person would have not supported the policy in any case. For \( \Delta < 0 \), the first term is positive and the second term is negative, so the effect is ambiguous. As \( x_i \) becomes smaller, however, the first term increases in size relative to the second term and the potential support for the policy increases.

2.5 Countercyclical skewness

Recent research suggests that the skewness of transitory income shocks depends on whether the economy is expanding or receding, in part explaining why income inequality increases with \( b \).
2.5. COUNTERCYCLICAL SKEWNESS

increases during economic expansions and contracts during recessions. Guvenen, Ozkan and Song (2012) show that the distribution of transitory income shocks are heterogeneous across the population and respond to changes in the economy. The authors use a panel dataset on adult male earnings from a ten percent random sample of the Master Earnings File of Social Security records. The benefit of this data is that it contains observations of individuals over several years and the data on earnings is not top-coded.

They find that transitory shocks as a percentage of income are largest for lower income households and the very wealthy (above the 90th percentile). During economic recessions, the transitory shocks become more left skewed, particularly for lower income households. In contrast, during economic expansions, the skewness shifts right and the median income shock becomes relatively closer to the mean. In particular, during economic expansions upper income households have a much higher chance of seeing a large positive change in income. This improved outlook for upper income households leads to an increase in income inequality. Thus, we see a relationship between the skewness of income shocks and increases in income inequality, which results from changes in macroeconomic performance. As income inequality worsens, the distribution of shocks becomes less left skewed, thus mitigating the increased demand for raising taxes that we might expect from increased income inequality.

We can relax the assumption of homogeneity and independence of income shocks and adapt the model to incorporate the insights of Guvenen, Ozkan and Song (2012).

As before each individual has an annual income, \( y_{i,t} \), which now includes a deterministic component, \( x \), an aggregate shock, \( \lambda_t \) and an idiosyncratic stochastic component, \( \varepsilon \).

\[
y_{i,t} = x_i + \lambda_t + \varepsilon_{i,t}
\]  

(2.4)

The aggregate shock represents annual changes in economic performance that affect all individuals in the society. Unlike the earlier model, the distribution of shocks varies depending on the fixed characteristics of the individual that determine income as well as the overall performance of the economy. Hence, \( \varepsilon_{i,t} \sim H(\varepsilon|x_i, \lambda_t) \) which is conditionally independent and identically distributed for all individuals and uncorrelated across time. The conditional expectation, \( E[\varepsilon|x_i, \lambda_t] = \bar{\varepsilon}_t \) and the conditional median is denoted \( \text{Med}[\varepsilon|x_i, \lambda_t] \). Thus, the mean of \( \varepsilon_{i,t} \) is the same for all individuals in a given year, while the second order and higher moments of the distribution depend on \( x_i \) and \( \lambda_t \). This simplifying assumption improves the ease of calculations without affecting the substantive implications of the model. In addition, it conforms to the model of income volatility in Guvenen, Ozkan and Song (2012).

The skewness of the distribution of shocks can be represented by the difference between the mean and median:

\[
E[\varepsilon|x_i, \lambda_t] - \text{Med}[\varepsilon|x_i, \lambda_t]
\]

A positive shift indicates that the distribution is relatively more right skewed and visa versa. We can impose restrictions on this relationship that conform to the findings of Guvenen, Ozkan and Song (2012). Specifically, they find that as the economy improves, the
distribution of shocks becomes less left skewed for all groups of individuals in the population. Mathematically, this restriction can be written:

\[
\frac{\partial \{E[\varepsilon|x_i] - \text{Med}[\varepsilon|x_i]\}}{\partial \lambda_t} > 0, \quad \forall x_i
\] (2.5)

As before, the density of \( x \) is denoted by \( g(x) \) and the density of income, \( f(y) \) is the convolution of \( g(x) \) and \( h_{\varepsilon|x}(\varepsilon) \):

\[
f(y) = \int_{-\infty}^{\infty} g(x) h_{\varepsilon|x,\lambda}(y-x) dx
\]

The expectation of individual income, \( y_{i,t} \), is \( E(y_{i,t}|x,\lambda_t) = E[x + \lambda_t + \varepsilon_{i,t}|x,\lambda] = E[x + \lambda_t + \varepsilon_{i,t}] = x + \lambda_t + \bar{\varepsilon}_t \), while the expectation of income across individuals, or per capita income is:

\[
\bar{y}_t = E(y) = \int_{-\infty}^{\infty} y f(y) dy = E(E(y_i)) = \int_{-\infty}^{\infty} (x + \lambda_t + \bar{\varepsilon}_t) g(x) dx = \bar{x} + \lambda_t + \bar{\varepsilon}_t
\]

Following the same logic in the simple model above, an individual will support a tax policy \( \tau \), if:

\[
EU[(1-\tau)y_{i,t} + \tau \bar{y}_t] \geq EU(y_{i,t}) \quad x + \lambda_t + \bar{\varepsilon}_t \leq \bar{y}_t = \bar{x} + \lambda_t + \bar{\varepsilon}_t
\]

Since \( \bar{\varepsilon}_t \) and \( \lambda_t \) are the same for all individuals in a given year, the level of taxes preferred is again solely a function of the deterministic component of income such that \( \tau \) is monotonically decreasing with \( x \). Thus, the median voter is the individual with the median time-invariant portion of income \( \bar{x} \). This is the same result as the simple model due to the assumption that the expectation of shocks is the same for all individuals in a given year, though the conditional distribution of the shock varies.

Although the identity of the pivotal voter does not depend on \( \lambda_t \), economic performance factors into the relationship between economic inequality and the tax policy approved by the voting majority. Returning to Guvenen, Ozkan and Song (2012), they find that as the economy improves, the distribution of shocks becomes less left-skewed. Conditional on the distribution of \( x_i \), an improvement in the economy, which changes the distribution of shocks according to (5), should induce growth in income inequality. Thus restriction (5) implies:

\[
\frac{\partial \{E[y|x_i] - \text{Med}[y|x_i]\}}{\partial \lambda_t} > 0, \quad \forall x_i
\] (2.6)

The connection between economic expansion and growing income inequality stems from sensitivity of top income shares, which depend more heavily on capital gains, to changes in the performance of markets (Piketty and Saez 2013). The important point of Guvenen,
2.6. IMPLICATIONS FOR CONSUMPTION SMOOTHING

Ozkan and Song (2012) is that changes in inequality over the business cycle derive in part from changes in the distribution of shocks. During economic expansions the probability of a large increase in income becomes more likely. Note that under this model, the final tax policy \( \tau^* \) still depends on the size of the difference between mean income in the population \( \bar{x} \) and median income \( \tilde{x} \) (assuming \( \bar{y}_t \geq \tilde{x} + \lambda_t + \varepsilon_t \)) and does not vary with \( \varepsilon \). Thus, to the extent that changes in income inequality arise from changes in the distribution due to a right shift in the skewness of the shocks, we may not see an increased demand for higher taxes as a result of income inequality.

Note that the reverse is also true. During times of recession, Guvenen, Ozkan and Song (2012) find that income inequality lessens and the distribution of shocks becomes more left skewed. Incorporating the countercyclical nature of income volatility into our model of income inequality and redistribution suggests that volatility may act as a counterweight to changes in income inequality over the business cycle. Income volatility under these conditions stabilizes tax policy preferences against cyclical fluctuations of income inequality.

2.6 Implications for consumption smoothing

Individuals may be able to self-insure against temporary income shocks through consumption smoothing, i.e., by saving a portion of income in years when they experience positive income shocks and dissaving in years when they experience negative income shocks. Krueger and Perri (1999) explore the trade-off between private and public insurance schemes in an economy where agents are debt-constrained. As the generosity of public subsidies increases, individuals are less likely to take steps on their own to guard against unforeseen economic losses. As fewer individuals seek insurance in private credit markets, these markets have less capacity to provide insurance. Thus, Krueger and Perri argue that a more generous public redistribution program may on net reduce overall social welfare. In the Krueger and Perri model, tax policy is exogenous so the generosity of public subsidies is not counterbalanced by changes in the tax rate. If the cost to the individual of higher subsidies is considered, individual preferences for public subsidies over self-insurance at different income levels is less obvious. As income volatility increases and individuals face larger income shocks, individuals may look to public redistribution to complement their own consumption smoothing.

In the future, a more comprehensive model of inequality and redistribution could consider the allocation of tax dollars to public subsidies, the tax cost associated with higher subsidies, the net cost of self-insuring, and risk aversion. Such a model is likely to be inconclusive without empirical research to establish key relationships. In a related paper, I conduct a series of Internet experiments that employ vignettes to determine how individual tax-rate preferences are affected by different levels of income volatility. The results support the basic insight from the model and discussion in this paper, namely, that to the extent that the growth in income inequality increasingly comes from transitory income volatility, voter support for policies aimed at taxation and redistribution will receive less support than they would have had if the growth in income inequality had been driven by divergence in
permanent income. That is, the growing prominence of transitory shocks in income inequality acts to restrain the political impetus for more progressive tax-and-redistribute policies.

### 2.7 Conclusion

The United States has experienced increased income volatility over the last thirty years. This increase in volatility explains some of the divergence in income and consumption inequality (Fisher and Johnson 2006). We also know that households have shown an ability to insure against negative shocks to income as long as they are temporary. Those at the bottom of the distribution are especially sensitive to business cycle fluctuations, but there is evidence that even they are able to weather economic downturns in part through a mixture of borrowing and saving (Heathcote, Giovanni and Violante 2009). As a result, the data show that lower incomes during periods of economic downturn do not fully translate into lower consumption.

This paper begins to examine theoretically how the emerging role of income volatility in increasing income inequality during the past two decades in the United States might alter voter preferences for redistribution. Because of temporary income shocks, an individual can expect to move up and down the income distribution over the course of their lifetime. As a result of this shift in income dynamics, the ratio of mean income to median income may no longer be as good a predictor of society’s demand for larger government.

A theoretical model incorporating income volatility reveals important differences in how individuals adjust their preferences to short-term versus long-term changes in relative incomes, and thus income inequality generated by increasing the variance of the volatility of income may have very different effects on preferences for taxation than the traditional model of rising income inequality without increasing volatility. In addition, I find that incorporating risk aversion into the model with income volatility does not necessarily increase support for the tax policy, because individuals face both uncertainty in their income and in the lump sum redistribution subsidy. Finally, adopting the model to incorporate insights from research about the nature of income shocks suggests that under existing conditions, income volatility may diminish increasing support for progressive tax policies that we might expect from rising income inequality. Future models of economic inequality and taxation should distinguish between persistent and temporary components of income when making predictions about voter preferences for redistribution in response to changes in the income distribution.

In a related paper, I present evidence from a series of Internet experiments that income volatility may decrease voter support for income tax progressivity, another mechanism for redistribution. Together, the two papers provide one explanation why increasing income equality in the United States has not led to a demand for government redistribution.
Chapter 3

Tax Preference on the Income Roller Coaster

Income inequality has increased in the United States over the past thirty years. Between 1973 and 2004, the proportion of earnings held by the top 1% of households doubled from 6.5% to 13% of total earnings (Kopczuk, Saez and Song, 2010). Models in political science predict that as income inequality rises so should support for progressive tax reforms (Meltzer and Richard, 1981; Perotti, 1993). Yet, public opinion on tax progressivity has remained relatively stable over this period in spite of worsening income inequality. In fact, the percentage of Americans who believe the upper class do not pay their fair share in taxes has declined, while the percentage that believe they pay the right amount has increased even as the tax rate on upper incomes has fallen for the most part (Gallup polls from 1990-2012. See Figure 3.1).

Neither a lack of awareness nor salience of the issue can explain the model’s failure to reflect reality. The majority of Americans are aware of growing income inequality and believe it is a serious problem. However, public opinion is divided on whether the government should redistribute income more equitably through taxation. The American National Election Survey recently conducted a series of online surveys assessing American’s evaluation of government and society, which included questions related to income inequality and government redistribution. While 76% of respondents correctly believed that the difference between incomes for the top and bottom 20% had increased over the past 30 years, only 41% of those aware of the problem believed the government should take steps to redress income inequality. Support for government intervention among those that did not believe income inequality had worsened stood at only 22% (ANES, EGSS 2).

These findings are consistent with other nationally-representative surveys. In a series of polls dating back to the late nineties, Gallup has asked Americans whether they believed wealth should be more equitably distributed in society and whether it was appropriate for the government to facilitate a more equitable distribution through taxation. In 1998, 63% of Americans agreed with the statement that wealth should be more equitably distributed, but only 45% believed that the government should redistribute wealth through taxation.
compared to 51% who were opposed to such interventions (Gallup). These beliefs have not changed in the past decade despite growing income inequality. In 2011, 61% of Americans agreed in a more equitable distribution of wealth, but only 47% supported government redistribution through taxation as compared to 49% opposed.

Americans are clearly divided over whether taxes should be used as a tool to reduce income inequality. This division plays out in elections, where tax policy consistently emerges as a central issue of the debate. While both parties support tax cuts for the middle class—a reassuring validation of the median voter theorem—they disagree over how much to tax the rich with Democrats generally in favor of raising taxes on upper incomes. The party whose message consonates with majority opinion changes from election to election. In some years, voters appear more concerned about cutting taxes and less concerned about whether the privileged are contributing their fair share. More recently, surveys find widespread support for policies that raise taxes for incomes at the very top. The same ANES survey on government and society found that 63% of Americans favor raising taxes on households earning greater than $250,000 in order to reduce the deficit, a key policy goal of Obama’s tax reform. However, the same percentage of Americans (63%) disapprove of Obama’s handling of tax issues. Furthermore, 50% of Americans would prefer to reduce the deficit through spending cuts alone, while only 37% would like to achieve reductions through a mixture of spending cuts and tax increases. Taken altogether, these surveys reveal at most an ambivalence among Americans towards policies to redistribute income more equitably through taxation.

It makes common sense that support for tax progressivity should rise with growing income equality. Yet, as income inequality worsens, public opinion towards taxes remains stable. Over the past thirty years, large numbers of middle class voters have not rebelled against current tax policies. Voters have not overwhelmingly pushed government to take significant steps to redress the growing disparity between average working Americans and the top income earners. Thus, we are left with the puzzle: why are the models as well as our intuition wrong.

In this paper I explore one possible explanation for why we fail to see a response in tax attitudes to increasing income inequality. Models based on the Meltzer-Richard model of income inequality and tax preferences overlook one important source of income inequality: income volatility, or year-to-year changes in household income. While income volatility and inequality are positively correlated, they potentially have opposite effects on demand for tax progressivity. Theoretically, income volatility could weaken support for progressive taxes and thus public opinion may not change on net if some portion of growing income inequality is due to growing income volatility.

Models assume either a constant level of income volatility over time or that voter preferences are independent of expected income. Neither assumption appears to be reasonable in the recent American context. Income volatility has risen in the past 30 years, with households now experience far more frequent and larger changes in income from one year to the next, and the rise in volatility has contributed to the rise in income inequality (Heathcote and Violante 2009). But to what degree does income volatility alter the relationship between growing income inequality and demand for tax progressivity? Under what circumstances
In a series of polls dating back to the 1990s, Gallup asked whether respondents believed that the upper class pays a fair amount or too little in taxes. The percentage of respondents that stated the upper class pays too little in taxes declined from 77% in 1992 to 62% in 2012, while the percentage that stated the upper class pays a fair amount increased from 16% to 25%. The percentage that indicated the upper class pay too much increased from 4% to 10% over the same period (not shown). At the same time, the share of income accrued to households at the top decile of the income distribution increased by 10 percentage points between 1991 and its pre-recession high in 2007 as indicated in the lower bar graph.

are voters sensitive to changes in income that are not permanent, but nonetheless, have substantial consequences for current prosperity and future stability?

In a series of survey experiments conducted online, I find evidence that supports theoretical predictions regarding the relationship between income volatility and tax preferences in the context of growing income inequality. On average, survey respondents prefer significantly less progressive taxes after being shown vignettes that describe household income changing more over time. The greater the income volatility, the lower respondents set tax rates for upper incomes. Furthermore, the pattern of income volatility did not affect their preference for progressivity. When described households facing greater changes in income from year to year, respondents preferred less progressive taxes regardless of whether the households experienced a decline or rise in income during the final year.
The effect is strongest for Independents and married respondents. This finding is consistent with a large body of research in political science that suggests preferences of partisans are more fixed compared to those of Independents. The results for married respondents may reflect differences between how volatility is perceived between one and two income households. Married respondents have a greater capacity to weather changes in income, and may have a more positive perception of income volatility. Indeed, I find that respondent attitudes towards risk impacts the size of the effect. While both risk seeking and risk averse groups prefer significantly less progressive taxes after exposure to high volatility households, the decline is much larger for risk seeking individuals.

Altogether, these findings suggest that greater income volatility may lead voters to support tax reforms that reduce taxes for high income households, even if they themselves are unlikely to personally benefit from such reforms. To the extent that income volatility contributes to conditions of growing income inequality, it is possible that support for tax progressivity among the middle class will not rise as income inequality rises. The presence of income volatility may counteract an increase in support for tax progressivity that arises from growing income inequality.

3.1 Context: Trends in Income Volatility and Income Inequality in the United States

Household incomes are more volatile today than they were thirty years ago and this year to year variance contributes to the overall variance between households in income in any given year. To understand how income volatility relates to income inequality, consider two scenarios. In the first, imagine ten households at the same level of income. Suppose there is no income volatility so all ten households continue to earn the same income in the following years. Because earnings are stable, income inequality, or the variance in income between households is zero in every year. Now imagine for the second scenario the same ten households, but suppose their incomes are volatile. There is some probability that each household will earn more or less than their average income in the following years. While all households have the same expected income, their earnings will differ from each other in each year, because individual incomes are unstable. Thus, the introduction of income volatility gives rise to income inequality, and the degree of inequality depends on the extent to which incomes fluctuate. If households experience large shocks in income from year to year, there will be large variance in incomes between households in any given year.

Inserting income volatility into the relationship between income inequality and individual preferences for tax progressivity can induce changes in public opinion, which deviate from those predicted by standard models. Specifically, the presence of income volatility can dampen any increased support for progressive tax reform that arises from rising income inequality. Consider two possible extremes. In one extreme, there is no income volatility and high income inequality. Households have no possibility of moving up or down the income distribution without government intervention to redistribute income more equitably between
households. As upper income households become better off relative to the rest of society, the majority of voters will increasingly support progressive tax reforms. Thus, growing income inequality in the absence of income volatility should lead to greater demand for more tax progressivity.

In the other extreme, all of the income inequality is due to income volatility. Every household has the same expected income over the long run, but in some years households earn more or less than the average. As income volatility rises, it is unclear how its rise will alter individuals’ tax preferences. In expectation, all households earn the same income and support for government policies to redistribute income will depend on factors such as risk aversion and preferences for consumption smoothing. As a result, growing income inequality due to income volatility is likely to be weaker than when there is growing income inequality without income volatility.

Traditional models of income inequality and demand for public redistribution assume the first of the two extremes. They predict that as the difference between the incomes of the top earner and the median voter grows, the voting majority will become more supportive of progressive tax reforms (Meltzer and Richard 1981; Perotti 1993). This prediction rests upon the strong assumption that the relative economic standing of households in society is fixed. As wealth concentrates within a smaller segment of the population, and the middle class becomes comparatively worse off, the model predicts that we should expect broader public support for tax reforms which increase the tax burden on the top income earners. But these predictions do not consider conditions where income inequality rises with income volatility. Rather, they assume that disparities between wealthy and middle class households are permanent in the sense that the same set of households consistently gain in earnings relative to the rest of society. These persistent disparities in society are often linked to experiences or characteristics, such as gender, race and education, that define different groups within society.

Persistent, group-based income disparities have contributed to growing income inequality in the United States (Heathcote and Violante 2009). One of the largest drivers has been the education gap. The difference in expected earnings between high-school and college educated individuals has diverged considerably since the 1970s, a result of multiple factors including the decline of labor unions, technological innovation, and the expansion of the service sector economy. Alternatively, some group-based disparities have improved. In particular, there has been a substantial narrowing of the gender gap. Women have made significant gains in earnings compared to men since the 1980s, contributing to a reduction in income inequality. However, the mass entry of women into the labor force increased the number of married households with two income earners, helping to worsen income disparities between married and single households.

Income volatility, or greater instability in earnings for individual households, has also factored into the rise in income inequality. Earnings instability increased particularly for men in the early 1980s, leveled off during the last part of the 1980s and 1990s and rose again in this past decade (Dynan, Elmendorf and Sichel 2008; Hacker et al. 2011). This trend has been verified by multiple sources, although the degree of income volatility measured
depends considerably on the methodology and source of the data. Using data from the March Current Population Survey, Ziliak and Hardy estimate that income volatility rose by 38 percent between 1973 and 2008, affecting all American workers regardless of race, education level, and family structure (Ziliak and Hardy 2010). A common source of data on income volatility is the Panel Study of Income Dynamics, the most comprehensive panel data set on income and consumption patterns in the United States. Using this data, Hacker and Jacobs estimate that the frequency and size of shocks to both individual and household incomes have increased over the past three decades (Jacobs and Hacker 2008). The proportion of workers who experience a drop in income of 50 percent or more in any given year is between 10 and 15 percent of workers, up from only 4 percent in 1970 (Jacobs and Hacker 2008; CBO 2007).

Hacker and Jacobs’s research focuses on trends in the average volatility across all households over the past thirty years. In a series of papers, Jensen and Shore disaggregate these trends in volatility by groups within the population and find evidence that the distribution of volatility shocks is heterogenous. Specifically, the rise in the average volatility is driven by a minority of individuals in the population, the self-employed (Jensen and Shore 2011, 2009). Using data from the PSID, they estimate that half of all income changes are relatively modest, between 11 and 14 percent, while 10 percent of shocks result in change of approximately 70 percent. Married and more educated men are less likely to have volatile incomes, while the data show a higher degree of volatility among older men in their fifties (their sample consisted of men between the ages of 22 to 60). Interestingly, while increased volatility spans the income distribution, the variance in persistent shocks has increased inordinately for households with above-median income, while the variance in transitory shocks has increased more for households with below-median income. The rise in volatility among the most volatile households is not driven solely by either low or high income households, but is evident across income groups.  

Incorporating income volatility into the standard models of income inequality and demand for redistribution can either amplify or temper support for tax progressivity depending on the nature of the volatility. Transitory shocks to income that have little impact on longer term earnings may not greatly affect public opinion, especially since American households have enhanced capacities to smooth consumption over time with the expansion of credit markets (Krueger and Perri 2006). On the other hand, the possibility of large changes in income, whether negative or positive, may change how voters perceive the fairness of existing tax policies even if the shocks are temporary in nature. For instance, the prospect of a significant windfall may cause voters to prefer to keep rates applied to the top income brackets low, allowing households to pocket a greater portion of temporary gains. Alternatively, the potential of a significant loss in income may drive voters to seek greater contributions from the wealthy to finance enhancements to programs and policies, which help families weather temporary declines in fortune.

The latter theory has been explored more thoroughly in the political science literature

---

1Households whose incomes are above the median consistently for four years are less likely to have volatility. But by definition less volatility means household income is less likely to fall below the median.
3.2. SURVEY DESIGN

related to income volatility, notably by Jacob Hacker. Both reactions—a demand for more or less progressive tax policies—are possible and this research is a first attempt to measure empirically how public opinion on taxes responds to income volatility. How individuals perceive changes in income is key to understanding how these changes will in turn shape preferences for tax progressivity. It is likely that our perceptions of income volatility and the implications it has for how we assess our economic wellbeing depends on the nature of the volatility, including whether shocks appear temporary or permanent. In addition, it may depend on the degree to which individuals are short or farsighted when evaluating their own earnings and the earnings of others. In a series of survey experiments, I assess how the introduction of income volatility alters the connection between income and preferences on taxes. I find that respondents who are described households facing greater income volatility prefer less progressive taxes than those described households facing lower income volatility regardless of the pattern of income change over time. Particularly, respondents’ attitudes were not affected by whether the hypothetical household had relatively higher or lower incomes in more recent years compared to earlier years.

3.2 Survey Design

The purpose of the survey experiments is to probe whether income volatility, or changes in income from year to year for a single household, affects individual preferences for the degree of progressivity in federal income taxes. In the first half of the survey, respondents are described a vignette, where they are asked to imagine themselves as members of a household. They are shown the earnings of a single imaginary household over a three year period and asked what percentage in income should go towards federal taxes in each year. While the average level of income accrued over the three years is the same in every scenario, the degree of income volatility the household experiences during the period will vary at random.

At the beginning of the survey respondents receive basic information about taxes in the context of income volatility. The introduction reads:

For many years, US family incomes were very stable. More recently, family incomes move around more from year to year. The amount of money you pay in taxes is based solely on the earnings you made over the past year regardless of how much was earned in previous years. In other words, the tax rate this year does not depend on whether you earned more or less income two years ago, or expect to earn more or less income in the upcoming years. The next questions ask what you think the percentage in federal income tax rate should be when income is higher or lower in different years.

Respondents are then selected at random for one of two possible vignettes, which describe a household’s earnings over a three year period. In one vignette, households experience a high level of income volatility. I will refer to respondents assigned to this vignette as the high volatility group. The household’s income shifts $30,000 during the period. In the other,
households experience smaller changes in income, in total shifting only $10,000. The high volatility household reports earning incomes of $45,000, $60,000, $75,000 over the course of three years. In the other vignette, households experience a relatively low level of income volatility. I will refer to respondents assigned to this vignette as the low volatility group. The low volatility household earns incomes of $55,000, $60,000, $65,000 over the three year period. In both cases, the average income for the household is $60,000 and deviations from the average are symmetric: income deviates by $15,000 for the high volatility household and $5,000 for the low volatility household.

Figure 3.2: Survey respondents were assigned to one of two possible types of households. One type of household has highly variable income, ranging from $45,000 to $75,000 over a three year period. The other experiences less variability, ranging only from $55,000 to $65,000. Respondents were then asked to state how much they thought the federal income tax rate should be in each year given the household’s annual income. The income trajectories of the high volatility and low volatility households are indicated in red and blue respectively.

The vignette begins by asking respondents to imagine that their household’s income was
3.2. SURVEY DESIGN

a certain amount over the past year. They are asked what they believe the fair tax rate should be given this income. Respondents can select a percentage from 0 to 100. Then the vignette proceeds to the second period. In period two, respondents are told the household’s annual income for year two and asked if the household should pay the same or a higher or lower rate in taxes compared to year one. The survey reminds the respondent of the household’s income in year one, the tax rate he selected for that income, as well as the dollar amount associated with this tax rate. If respondents answer “the same”, respondents skip to period three and the rate for period two is recorded as the same rate as period one. Only if the respondent answers “higher” or “lower”, does the survey prompt him to indicate what percentage in taxes the household should pay in year two. This way, the respondent is first provided the chance to consider whether the rate should change and in what direction as income changes from year one to year two before being asked to select a new rate.

As a final check, the survey then confirms whether this percentage is in fact higher/lower than the rate provided in period one. If the answer is consistent, the respondent skips to period three. If the respondent indicated that the household should pay a higher (lower) rate, but in fact writes in a percentage that is lower (higher) than the rate selected in period one, then the answer is flagged as inconsistent. In cases of inconsistency, the survey asks the respondent whether he would like a chance to change his previous answers. If the respondent indicates “yes”, then the respondent is allowed to select new tax rates for period one and two. They are reminded of the household’s annual income for both years when making the selection. If the respondent decides to change the rates they had previously entered for year 1 and 2, the new rates are recorded and the respondent proceeds to year three.

The same process for period two is repeated again in period three. Respondents are told the household’s income in the third year. They are asked if the tax rate should be higher, lower or the same as it was in year two. If the respondent indicates that the household should pay a higher or lower amount, then the survey prompts the respondents to indicate the exact percentage. Finally, their answers in period two are checked for consistency. The vignette ends after three periods.

Once exposed to a greater or lesser degree of volatility, the respondents are then asked what they think the federal tax rate should be for DIFFERENT households with DIFFERENT incomes in the same year. The conclusion of the vignette in the first stage and subsequent transition to this second stage of the survey is made clear to respondents in the text of the survey. This question is used for the analysis of the treatment effect—not the responses to questions asked during the vignette. The dependent variable is the average difference in the preferred tax rates between respondents described high volatility households versus those described low volatility households. In addition to analyzing the preferred tax rates separately for each income level, the set of preferred tax rates that respondents select for this question are aggregated to estimate the degree of progressivity desired by respondents in the federal tax schedule.

Lastly, respondents are asked standard socioeconomic and political questions for background. In addition, respondents answer questions regarding their level of risk aversion. These questions were initially developed for use in the Panel Study of Income Dynamics.
Americans overwhelming favor flat or progressive tax schedules, although they may have difficulty selecting tax rates for various levels of income that are consistent with this preference. The survey is designed to make the selection easier by breaking down the choice of tax rate in periods two and three into two parts, and by allowing respondents to change their previous answers when these answers appear inconsistent. In practice, asking respondents first if they would like the rate to be the same, higher or lower than the previous year was sufficient to prevent the vast majority of wild answers. The consistency check flagged very few cases: only five of the 677 respondents were caught having written a tax rate that was not consistent with their earlier expressed preference for a higher or lower rate. However, four of these five changed their responses to a more plausible answer (shifting from a regressive to a progressive preference), and the remaining individual confirmed that his initial selection—an already progressive choice—was correct.\(^2\)

The number of questions the respondent encounters in this first stage of the survey ranges from as few as three to as many as eight. This design helps ensure that respondents have an opportunity to consider the relationship between tax rates and income and become practiced in expressing their preferences on this issue while not being overburdened with too many questions. At each stage, a summary table reminds respondents of their previous answers, which allows them to more easily make comparisons from year to year to determine the appropriate rate of taxation. In addition, these tables emphasize the degree of income volatility experienced by the household. They also provide the equivalent dollar amount in taxes the household would pay based on the tax rates selected by the respondent.

The vignettes provide respondents with an opportunity to consider their tax preferences and how rates should change as income changes. Individuals are not necessarily accustomed to thinking about the specific structure of tax schedules, but may have vague notions of the general principles that they believe should inform their construction (Campbell 2010). Most Americans pay little attention to policy in general, including reforms related to redistributive policy, and base voting decisions more on political ideology and partisanship than policy preferences (Lenz 2009). They express great ambivalence when asked about the role of government in redressing income inequality through tax reform. Polls show broad support for increasing taxes on the rich, but at the same time the public believes there should be limits to the level of taxation, even for the wealthy, and on the larger question of whether the tax system should be used to redistribute wealth within society, Americans are divided. According to a 2011 Gallup poll, conducted April 7-11, 47 percent of respondents believe the government should facilitate redistribution by “heavy taxes on the rich”, while 49 percent disagree with this statement. Sizable majorities, however, favor the idea that wealth should be more evenly distributed and only 1 in 3 Americans would agree with the argument that the current distribution is fair (Gallup 2011).

The contradictions in existing polls demonstrate the difficulty of measuring people’s true preferences on taxes using simple binary response questions. Survey results, which ask for a yes/no response to the question of raising taxes on the wealthy, provide at best crude...\(^2\)The inconsistency rate could be much higher on large-scale nationally-representative surveys, where participant fatigue is common.
estimates of the desired degree of progressivity. A respondent’s answer refers implicitly to his own belief about the current tax rate for the top income bracket, which is unknown to the researcher. It is unclear how closely this rate mirrors the actual rate, and therefore how likely the response would change if the respondent was informed of the actual rate (by either the researcher or by a politician). In addition, an individual expressing support for tax increases for the wealthy may not prefer a more progressive tax policy in general, since progressivity depends on the steepness of the tax schedule across all income levels. One must collect information about what the tax rate should be for various incomes along the income distribution to ascertain whether the respondent would support a more or less progressive tax policy.

At the same time, surveys that solicit the preferences of voters regarding different tax rates are prone to respondent error, especially when the issue is not immediately salient to the individual. It is likely that most people, even the highly educated, do not have well-formed preferences about what should be the ideal federal tax rate for themselves and others. Surveys suggest that from a fourth to a third of respondents are unsure about their own preferences on whether in general they support a progressive or a flat tax (Brady and Frisby 2011). For example, there is a tendency among individuals who embrace progressive tax policies in the abstract to prefer either flat or regressive rates when provided with concrete choices (Roberts, Hite and Bradley 1994). Asking respondents for actual preferred tax rates along an income distribution requires considerably more consideration from respondents. A common technique for reducing error that arises when the complexity of the information sought from respondents is high, has been to repeat questions on the same topic. But it is not clear how best to aggregate multiple responses, particularly when they yield contradictory information, and repeated questioning might aggravate the respondent and affect the answers given or willingness to continue with the survey.

The approach in this survey addresses these limitations by directing follow-up questions only to respondents who exhibit inconsistent preferences, allowing these individuals the chance to clarify their intent and provide more considered responses. This should increase the reliability of the response, and it eliminates the need for the researcher to select a method of aggregating multiple, perhaps conflicting responses. An adaptive approach thus strikes a better balance between the need for parsimony and accuracy in survey design. Follow-up questions appear only when there is information to be gained, and every attempt is made to solicit respondent preferences reliably with the fewest number of questions.

The fact that respondents learn over the course of the vignette does not pose an issue for the experiment. I do not use their answers from this first stage in my analysis since these questions as a group constitute the “treatment” of the survey experiment. Recall that the purpose of this stage is to expose respondents to scenarios of high or low income volatility. Providing opportunity for respondents to consider their preferences on how much the hypothetical household should pay in taxes as the household’s income changes from year to year helps them to internalize the concept of income volatility. They have a chance to think about the relationship between income and taxes, as well as how this relationship relates to income volatility. Merely describing the income trajectories of the two households
without asking respondents for their preferred tax rates or simply informing respondents that incomes have become more volatile over time would be too weak of an intervention to induce any treatment effect. The process of asking respondents to select tax rates for the household, checking these rates for consistency, and then showing the respondents their selections in summary tables is key to ensuring that the respondents internalize the treatment.

3.3 Findings

The survey experiment was designed and implemented through Survey Gizmo. Respondents were recruited through Amazon’s Mechanical Turk (MTurk), a web-based platform where workers sign up to perform tasks, including but not limited to online surveys. The first survey was conducted on November 21, 2011 with 200 respondents. The second was conducted February 7-8, 2012 with 200 respondents and the last was conducted on February 13-15, 2012 with 300 respondents. To be eligible for the survey, workers had to originate from the United States and have a demonstrated record of performing tasks on MTurk successfully. Upon completion of the survey, respondents received $.60. On average, the surveys took approximately 4 minutes to complete. Twenty-three of the total 700 respondents were dropped from the sample, because they were not residents of the United States.

As anticipated, the randomization produced treatment and control groups with similar socio-economic and political characteristics in the three experiments (Figure 3.1). Respondents were asked to indicate their household income as reported on their 2011 Federal Income Tax form (if applicable). The bulk of respondents in all three samples earned incomes between $20,000 and $59,000 and were slightly more likely to be female, childless and have completed at least four years of college. About fifty percent of respondents identified as Democrat or leaning Democrat. The remaining respondents were split equally between Republican (including leaning Republican) and independent. In addition to their household income, I asked respondents to indicate their filing status on their 2011 Federal Income Tax form as a proxy for marriage status (as well as an indication of the number of heads of households). Respondents were equally divided between having filed a single return or a married joint/separate return. Interestingly, very few respondents (less than 5% of the combined sample) reported that they did not know their own filing status.

Respondents exposed to vignettes describing households experiencing high volatility differed in their preferences for tax rates affecting the upper income brackets compared to those exposed to vignettes describing households experiencing low volatility. Specifically, respondents in the “high volatility” group wanted significantly lower taxes for households earning $100,000 and $200,000. Figures 3.3 and 3.4 plot the average preferred tax rate for each of the five income levels for the treatment and control groups individually, as well as the difference between these responses. The difference for the top income group $200,000 amounts to a sizable three percentage point drop in the preferred tax rate: from rate of 23% to 20%. The samples from all three experiments are pooled together for this analysis.

Expressed preferences for tax rates for each of the five income levels provides only a rough
### Table of Means

<table>
<thead>
<tr>
<th></th>
<th>$T_0$</th>
<th>$T_1$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Socioeconomic Status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>$37,000$</td>
<td>$34,800$</td>
<td>0.29</td>
</tr>
<tr>
<td>Age</td>
<td>36.78</td>
<td>35.94</td>
<td>0.40</td>
</tr>
<tr>
<td>Female</td>
<td>0.50</td>
<td>0.56</td>
<td>0.14</td>
</tr>
<tr>
<td>White</td>
<td>0.75</td>
<td>0.83</td>
<td>0.02</td>
</tr>
<tr>
<td>Years of schooling</td>
<td>14.57</td>
<td>14.82</td>
<td>0.04</td>
</tr>
<tr>
<td>Number of children</td>
<td>0.55</td>
<td>0.53</td>
<td>0.85</td>
</tr>
<tr>
<td><strong>Political Identification</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Republican</td>
<td>0.27</td>
<td>0.29</td>
<td>0.48</td>
</tr>
<tr>
<td>Democrat</td>
<td>0.52</td>
<td>0.51</td>
<td>0.77</td>
</tr>
<tr>
<td>Independent</td>
<td>0.20</td>
<td>0.18</td>
<td>0.61</td>
</tr>
<tr>
<td><strong>2011 Federal Income Tax</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Filing status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>0.47</td>
<td>0.55</td>
<td>0.04</td>
</tr>
<tr>
<td>Married Filing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joint/Separate Return</td>
<td>0.40</td>
<td>0.36</td>
<td>0.25</td>
</tr>
<tr>
<td>Other</td>
<td>0.05</td>
<td>0.04</td>
<td>0.61</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0.05</td>
<td>0.05</td>
<td>0.62</td>
</tr>
<tr>
<td>Don’t know</td>
<td>0.02</td>
<td>0.01</td>
<td>0.14</td>
</tr>
</tbody>
</table>

| N                              | 330   | 347   |

Table 3.1: The table reports mean characteristics of the treatment ($T_1$) and control ($T_0$) groups from three combined experiments. The p-values report the results of two-sided t-tests of the difference in means. Reported income means are approximate.
Figure 3.3: The figure reports the tax rates selected by higher and lower volatility groups for respondents from the three experiments combined.

A common measure of tax progressivity, the Tax Burden Inequality Index, was developed by economists to measure the degree to which the structure of tax systems reduces income inequality (Suits 1977). The index compares the cumulative distribution of income in society to the cumulative distribution of the tax burden. This index is similar to the one used to calculate the Gini index, a common measure of income inequality. A progressive tax schedule is one where an income decile’s cumulative share of the tax burden is less than its cumulative share of income. In other words, the distribution of income net taxes is less unequal than the distribution pre taxes in a progressive tax system. The degree to which taxes reduce the degree of inequality determines the level of progressivity in the tax structure. A flat tax has a progressivity measure of zero since it leaves the distribution of income unchanged. The extent that tax revenues are then redistributed within society is not factored into the degree of tax progressivity in the tax schedule, even though ultimately both the distribution of the tax burden and the distribution of government expenditures within the population determines the overall level public redistribution of income in society.

This index is one method for measuring the level of progressivity in an existing tax system and is calculated through analysis of data on income and tax payments. The measure cannot be used for the purposes of this survey, since the data in the survey is of a different nature,
3.3. FINDINGS

Figure 3.4: The figure on the right reports the difference in tax rates between the treatment and control groups, or the average treatment effect (ATE) and their 95% confidence intervals. The higher volatility group prefers significantly lower tax rates for households earning $100,000 and $200,000, indicating that on average exposure to descriptions of greater income volatility lead respondents to prefer lower taxes for high income households.

but its underlying principle can help to construct a meaningful measure of the desired level of progressivity in the sample.

There are two critical factors that determine the level of progressivity in a tax system: 1) the range and 2) convexity of the schedule. First, the range of the tax schedule is a measure of the difference between the bottom and top income brackets. The greater this difference, the more progressive the tax system. Second, the convexity of the tax schedule considers the marginal rate of change in the tax rate across income levels. As the marginal rate of change in the tax rate increases, the level of progressivity increases. A convex tax schedule is relatively more progressive than a linear one, which in turn is more progressive than a concave one given that the schedules are monotonically increasing. It is worth noting that measures of tax progressivity, such as the Tax Burden Inequality Index, do not consider the overall level of taxation. Progressivity does not depend on the amount of revenue raised, and thus a fixed tax rate of 9% may be equally as progressive as one at 18%, even though the two policies have different implications for the funding of government redistributive programs.

To measure the preferences of the survey respondents for tax progressivity, I develop a simple index which captures the critical factors of existing measures used to evaluate the progressivity of existing tax policies. The index calculates the sum of the differences between the preferred tax rate for the top income level and the rates at all other income levels. The
3.3. FINDINGS

Figure 3.5: The graph demonstrates how the progressivity would be measured for potential responses by respondents in the survey. The index is highest for the respondent, who exhibits a convex relationship in their preferred tax rates and zero for the respondent, who expresses a preference for a flat tax.

\[
\sum_{i=1}^{4} (\tau_5 - \tau_i) = 0.33
\]

Figure 3.6: A sample of actual tax rate schedules preferred by individuals in the survey.

The index expressed in mathematic terms is: 
\[
\sum_{i=1}^{4} (\tau_5 - \tau_i).
\]
Where $\tau_i$ is the preferred tax rate for the household at the $i^{th}$ income level. This is one method for measuring tax progressivity that respects the general principles underlying standard definitions of progressivity, but it is not the only one possible. The advantage of this measure is that it is simple and intuitive. A key assumption of the index, that preferences are weakly monotonically increasing, holds for all respondents in the survey samples.

By this measure, respondents exposed to vignettes describing households experiencing greater income volatility preferred significantly less progressive taxes than those that were exposed to vignettes describing households facing less income volatility. Figure 3.7 depicts a scatterplot of progressivity scores for treatment and control groups. Note that a handful of individuals in the control group (those exposed to less volatility) had comparatively high progressivity scores. To avoid drawing inferences based on a small group of outliers, I use a Wilcoxon rank sum test to calculate the p-values for the average treatment effect. The advantage of this test is that it considers only the relative scores of respondents, rather than their absolute differences. This quality is important given that the progressivity index used to calculate scores is useful only as a measure of relative progressivity and does not provide a meaningful measure of the absolute degree of progressivity.

Figure 3.7: The figure is a scatterplot of the progressivity index for treated and control respondents. The red and blue dots in the upper left corner indicate those that preferred a flat tax schedule.

Figure 3.8 plots the standardized differences in the effect for high and low volatility respondents as well as the corresponding p-values from two sets of Wilcoxon rank sum tests. The results are reported for the full sample as well as for important subgroups. In every instance, exposure to high volatility vignettes leads to a decline the the level of preferred tax progressivity. The average treatment effect for the full sample is both statistically and
substantially significant: a difference of close to one standard deviation with a p-value of .004.

While the randomization does not permit causal inference for the treatment effects associated with subgroups within the survey sample, they suggest that three groups are particularly responsive to exposure to greater volatility: Independents, married respondents,
and respondents who report low levels of risk aversion. I observe a relatively greater decline in preference for progressivity among high income earners (defined as those with reported incomes of greater than $60,000 in 2011) as compared to low income earners. In addition, females are more likely to see a decline in preference for progressivity compared to males, however this may reflect the fact that females are more likely to be married in the sample. In fact, when I run a regression of progressivity scores on an indicator for treatment with interaction terms for both marriage status and gender, only the estimate for the treatment effect associated with marriage status is substantively meaningful and statistically significant (see Table ?? in the Appendix). Thus, the larger decline in preference for tax progressivity among females as compared to males seems to be mostly explained by differences in marital status between the two groups.

The results show comparable rates of decline for individuals with higher levels of schooling (defined as greater than two years of college) than those with lower levels of schooling. Interestingly, Republicans and especially Democrats see relatively small declines in preferences for progressivity between treatment and control groups (variables for party ID include those who report leaning Democrat or Republican). This may reflect an ideological stickiness on the part of party ID, where individuals with stronger partisan biases are less likely to change their preferences for tax progressivity in response to descriptions of income volatility. Independents, on the other hand, appear highly influenced by exposure to descriptions of higher income volatility, which would be consistent with the hypothesis that groups with weaker partisan biases have policy preferences that are more adaptable to changing economic circumstances.

Measures for risk aversion were included in only two of the three surveys and therefore their results cannot be directly compared with those exhibited by the other subgroups in Figure 3.8. Respondents were given a series of questions to assess their level of risk aversion. Among those respondents, who exhibited It does appear that there was a larger decline in the preference for progressivity for respondents, who reported higher levels of risk acceptance compared to more risk averse respondents. The high risk respondents in the control group prefer a more progressive tax schedule than the low risk respondents in the control group. This evidence is intriguing, since individuals who report having higher levels of risk acceptance are more likely to experience frequent and sizable changes in income (Jensen and Shore 2009).

3.4 Survey Extension

So far we have seen that respondents prefer significantly less progressive taxes when they are described households facing greater degrees of income volatility. This relationship, however, may change depending on the pattern of volatility described in the vignettes. In a extension of the survey, I randomly assign respondents to one of six possible patterns of income over time (Figure 3.9). Like the original experiment, respondents are described household incomes over three years. These incomes range from $55,000 to $65,000 for the
low volatility group and $45,000 to $75,000 for the high volatility group. In addition to the
degree of income volatility, the pattern of income volatility over the three year period also
differs at random across respondents. One condition corresponds to the original experiment
design, the others permute the years in which income is high and low to create all six possible
combinations.

Three possible characteristics of these patterns (apart from the degree of volatility itself)
may be important in shaping respondents’ perception of income volatility and how it in turn
shapes their preferences for tax progressivity. In forming their perception of income volatility,
respondents may weigh the household’s income in the final year above the proceeding two,
specifically whether income in year 3 was higher or lower than or the same as average income
for the three year period. Alternatively, the change in income between the second and third
year may be important in determining respondents perceptions of volatility: whether the
difference was negative or positive, or large or small. Finally, the respondent’s perception
of volatility may take into consideration the trend over the entire three year period. The
trend in income may be either constant (increasing or decreasing) or variable across the three
years. The latter trend is suggestive of temporary changes in income, while the former is
suggestive of a persistent longterm change in income (Table 3.2).

<table>
<thead>
<tr>
<th>Key Attributes</th>
<th>LMH</th>
<th>MLH</th>
<th>HLM</th>
<th>HML</th>
<th>MHL</th>
<th>LHM</th>
</tr>
</thead>
<tbody>
<tr>
<td>yr3</td>
<td>high</td>
<td>high</td>
<td>middle</td>
<td>low</td>
<td>low</td>
<td>middle</td>
</tr>
<tr>
<td>yr2 to yr3</td>
<td>(&gt; )</td>
<td>(&gt;&gt;)</td>
<td>( &gt;)</td>
<td>( &gt;)</td>
<td>(&gt;&gt;)</td>
<td>( &gt;)</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3yr trend</td>
<td>↑</td>
<td>↘↗</td>
<td>↘↗</td>
<td>↓</td>
<td>↗</td>
<td>↘↗</td>
</tr>
</tbody>
</table>

Table 3.2: The table depicts the change in preferences for progressivity between respondents who saw high
volatility vignettes and those who saw low volatility vignettes. The treatment effect may vary according
to the pattern of income described in the vignettes. Specifically, respondents may reflect upon different
characteristics of the household’s income. They may attach greater importance to the household’s income
in the final year, the change between year 2 and 3 or the overall three year trend. Which characteristic
respondents find more important (on average) can be analyzed from the pattern of results.

There is evidence to suggest that individuals are myopic in their perceptions of past
events. This behavior would lead respondents in the survey to focus on the income earned
in the final third year of the vignette above the preceding two years. Andrew Healy and
Gabriel Lenz’s work on economic voting finds that survey respondents place greater emphasis
Figure 3.9: Survey respondents were asked to imagine themselves to be part of one of two possible types of households. One type of household has highly variable income, ranging from $45,000 to $75,000 over a three year period. The other experiences less variability, ranging only from $55,000 to $65,000. Respondents are then asked to state how much they think the federal income tax rate should be in each year given the household’s annual income. Respondents are assigned to high or low volatility scenarios at random. The high and low volatility income trends for both types of households are drawn in red and blue respectively for each of the six possible patterns of income over the three year period.
on the performance of the economy during the finals years of a presidential term when assessing the President’s efficacy (Healy and Lenz 2012). While myopic behavior can be related to a failure to remember past events clearly, memory is not necessarily the sole cause of short-sided reasoning. Consistent with Kahneman and Tversky’s influential work on prospect theory, Healy and Lenz find evidence that myopic behavior results from a desire to simplify complicated evaluations (Kahneman and Tversky 1979). In addition, individuals may emphasize more recent events due to a belief that the near past contains information of greater relevance for decision-making in the present (Mackuen, Erikson and Stimson 1992).

Both heuristic and information motivated behaviors with respect to myopia could be potential factors in the context of this survey experiment. Respondents may place greater emphasis on the final year if they struggle with the concept of income volatility and its significance for longterm household economic health. Alternatively, respondents may view the third year as more informative of future potential earnings, and thus ultimately demand a level of progressivity in tax policy consistent with their most recent income. Memory-based factors are less likely given that the survey repeatedly reminds respondents of the hypothetical household’s income for all preceding years throughout the vignette, as well as the tax rates they selected in each of those years.

Alternatively, respondents might care about the difference between the household’s second and third year earnings, specifically whether the change was positive or negative and the size of this shift. The tendency to evaluate changing circumstances with respect to a reference point, typically the individual’s current circumstances, brings to mind the concept of loss aversion: the idea that greater value is attached to the loss of a good than to its gain (Kahneman and Tversky 1979). Experiments have shown, however, that loss aversion applies to losses of goods and not money, and it is difficult to see how it might apply to our perceptions of income volatility (Kahneman and Novemsky 2005). It seems plausible, however, that individuals may evaluate their current earnings relative to previous earnings. In the survey experiment, consider two possible vignettes. One where the household earns $45,000, $75,000 and $60,000 respectively over three years and the other where the household earns $75,000, $45,000 and $60,000. Both households earn the same amount—$60,000—in the third year, but in the first scenario this income represents a $15,000 loss from the previous year, while in the second, it amounts to a $15,000 gain. Respondents’ feelings about the overall level of progressivity in the federal tax schedule may respond differently to negative versus positive shocks in income, even though the final outcome is the same.

Finally, volatility can affect respondents’ longterm perceptions of their economic well-being. If shocks are persistently negative or positive, they might lead individuals to change their beliefs about their own longterm earnings expectations. Three periods of income decline or growth may not necessarily be indicative of a longer term trend, but nonetheless, respondents may view these periods of consistent decline or growth as trends that are likely to continue into the future. This type of behavior may cause individuals to be overly optimistic or pessimistic if the trend is eventually reversed and the changes in income are only temporary. In addition, increasing volatility may negatively affect respondents’ perceptions of their economic wellbeing if the shocks cannot be anticipated. Greater uncertainty leads to
lower utility among risk averse individuals. To the extent that households can weather losses in income through increased borrowing or drawing upon savings, the uncertainty associated with greater volatility is somewhat mitigated. On the other hand, individuals that are risk accepting, may view greater volatility positively and see it as an opportunity to possibly earn a higher income. This idea is supported by evidence that the individuals who experience the greatest degree of income volatility are typically self-employed and self-identify as risk tolerant (Jensen and Shore 2011). This is a potential source of bias in the survey experiment, since it is possible that workers on MTurk are possibly more likely to be self-employed.

The survey experiment is designed to potentially be able to distinguish between which of these perceptions of income volatility respondents find most salient when assessing their tax preferences. Table 2 shows the results of the first series of preliminary survey experiments, which will be described in the next section. The results of subsequent experiments will complete the table. The pattern that emerges across columns can help to determine which characteristics of the households’ income over the past three years is most predictive of the relationship between income volatility and preferences for tax progressivity. It is reasonable that preferences for taxes may not just depend on the size of the income volatility described in the vignettes, but the pattern of income change over the three years.

3.5 Results of survey extension

The experiment employed a two-way layout factorial design, with involves two treatment factors with fixed levels. One factor corresponds to the degree of income volatility. It has two levels: 1) high income volatility ranging from $45,000 to $75,000 over three years and 2) low income volatility ranging from $55,000 to $65,000. The second factor corresponds to the pattern of income volatility. There are six possible patterns. The experiment was carried out through Amazon’s Mechanical Turk between September 9-12, 2012. The total number of respondents was 1842, of which 1698 observations were included in the sample. Ineligible respondents included 62 individuals, who prematurely exited the survey, 72 individuals, who were located outside the United States, and 10 individuals, who exhibited highly irregular tax preferences.

The linear model for the two-way layout is:

3Nine of the 10 flagged cases had non-monotonic tax preferences—the respondents’ preferred rates did not rise (or remained flat) with income. Instead, their chosen schedules contained a dip where the rates dropped after initially rising. The progressivity index I develop assumes monotonic preferences. Naively including these responses with the sample does not change the results of the experiment, but it remains a question about how to properly assess progressivity in such irregular cases. The final respondent preferred a tax rate of 100% for all incomes. I am currently adopting rank-based ANOVA testing procedures, which can produce estimates that are robust to outliers in the data.
3.5. RESULTS OF SURVEY EXTENSION

\[
Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}
\]
\[
\begin{align*}
    i &= 1, 2 \\
    j &= 1, 2, 3, 4, 5, 6 \\
    k &= 1, \ldots, n_{ij}
\end{align*}
\]

given the usual zero-sum constraints for the main effects and interaction:

\[
\sum_i \alpha_i = \sum_j \beta_j = \sum_i \alpha\beta_{ij} = \sum_j \alpha\beta_{ij} = 0
\]

where \(y_{ijk}\) is the progressivity score of the \(k^{th}\) respondent of the \(i^{th}\) level of factor A and of the \(j^{th}\) level of factor B. Factor A is associated with the degree of income volatility and has \(I = 2\) levels, representing low and high income volatility. \(\alpha_i\) is the \(i^{th}\) main effect for factor A. Factor B is associated with the pattern of income volatility and has \(J = 6\) levels. \(\beta_j\) is the \(j^{th}\) main effect for factor B. \(\alpha\beta_{ij}\) represents the twelve possible interaction effects between the two factors.

The main hypotheses of interest are:

- \(H_{0A}: \alpha_i = 0\), \(H_{0B}: \beta_j = 0\), \(H_{0AB}: \alpha\beta_{i,j} = 0\), \forall i, j

Factorial experiments such as this one are typically analyzed within a parametric framework, but if the errors are not normally distributed, the standard approach to analysis of variance (ANOVA) may not be preferable. Consistent with the original experiment, the response variable is the progressivity index described earlier. Plots of the residuals for 12 treatment combinations reveal that the data is not normally distributed, though it does have constant variance across treatments. As a result, I use both the standard ANOVA approach, as well as a non-parametric permutation procedure that I developed and describe in another paper (Hosek 2013). The method uses a restricted permutation procedure similar that proposed by Basso et al. (2009), but with changes that make it easier to accommodate large numbers of observations (multiple replications), as well as to compute weights to adjust for imbalance.

Unlike traditional parametric ANOVA methods that stand on normality assumptions, permutation methods assume only that errors are exchangeable and come from the same unknown distribution. Less restrictive than independence, the exchangeability assumption is justified by randomization in the experimental design. The method I propose, which works for any regular factorial experiment with two factors can accommodate imbalance in the design without making any additional assumptions on the model. In comparison, the two most common adjustments to parametric ANOVA, known as ANOVA “Type-II” and “Type-III”, require further assumptions and perform poorly in the presence of possible interaction effects (Langsrud 2003).

The permutation method uses a synchronized permutation structure to test the three main hypotheses associated with the two main factors and their interaction. The synchronization is necessary since responses for different combinations of levels for the factor of interest might have different expected values across levels of the other factor. We want a permutation structure that tests the null hypothesis for each main effect (and interaction)
irrespective of whether the null hypothesis for the other factor is true. For example, the null hypothesis $H_{0A}$, states that the degree of volatility, high or low, described in the vignette does not affect respondents’ preferences for tax progressivity. This assumption implies that observations are exchangeable with respect to factor A, but does not make any assumption about exchangeability with regards to factor B, the pattern of income volatility. In constructing a test for $H_{0A}$, the permutation procedure must respect that respondents may differ in their responses according to the pattern of income volatility.

Both the parametric ANOVA and permutation methods reach the same conclusion. High volatility respondents preferred less progressivity taxes than low volatility respondents. The pattern of the volatility over time, however, did not significantly affect respondents preferences for tax progressivity. Table 3.3 shows the mean progressivity score for each of the 12 different treatment combinations. The second and third largest differences in progressivity scores occurred for respondents who either saw steadily increasing or steadily decreasing income throughout the period.

Table 3.3: Mean responses for each of the 12 treatment combinations.

<table>
<thead>
<tr>
<th>Treatment Combinations</th>
<th>Mid-low-high</th>
<th>Mid-high-low</th>
<th>Low-mid-high</th>
<th>High-mid-low</th>
<th>Low-high-mid</th>
<th>High-low-mid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low volatility</td>
<td>0.39</td>
<td>0.36</td>
<td>0.36</td>
<td>0.39</td>
<td>0.41</td>
<td>0.39</td>
</tr>
<tr>
<td>High volatility</td>
<td>0.38</td>
<td>0.34</td>
<td>0.29</td>
<td>0.32</td>
<td>0.33</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Table 3.4: The table presents the results from traditional and non-parametric ANOVA tests of the experiment. The degree of income volatility was significant, while the pattern of volatility and the interaction between degree and pattern of volatility was not.

<table>
<thead>
<tr>
<th>p-value:</th>
<th>Regression ANOVA</th>
<th>Permutation Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>degree of volatility</td>
<td>.005</td>
<td>.005</td>
</tr>
<tr>
<td>pattern of volatility</td>
<td>.282</td>
<td>.286</td>
</tr>
<tr>
<td>interaction</td>
<td>.604</td>
<td>.621</td>
</tr>
</tbody>
</table>

The results of the ANOVA, however, indicate that there were no significant differences in preferences between the six patterns of income change. Neither is the interaction between the degree and pattern of income volatility significant. A regression analysis confirms that respondents attitudes on tax progressivity remaining unaffected by that pattern of volatility across time with one exception (Figure 3.5). Among respondents in the high volatility group, those who were described households whose incomes rose steadily throughout the three year period desired the least progressive taxes compared to all other groups. This finding suggests that if individuals believe that household income is rising, they might prefer to keep taxes for upper incomes low. This preference may be motivated by self interest. If respondents identify with the household described in the vignette and believe their own incomes could
rise substantially, they may want to keep tax rates at upper incomes low in order to pocket a greater share of the gains when income rises.

|                  | Estimate | Std. Error | Pr(>|t|) |
|------------------|----------|------------|----------|
| (Intercept)      | 0.4533   | 0.0300     | 0.0000   |
| high degree of volatility | -0.0461  | 0.0159     | 0.0038***|
| pattern of volatility |         |            |          |
| mid-high-low     | -0.0385  | 0.0280     | 0.1702   |
| low-mid-high     | -0.0598  | 0.0262     | 0.0227** |
| high-mid-low     | -0.0289  | 0.0269     | 0.2831   |
| low-high-mid     | -0.0166  | 0.0291     | 0.5686   |
| high-low-mid     | -0.0097  | 0.0266     | 0.7159   |

Table 3.5: The table presents the results of a regression of the respondents’ progressivity scores on the degree and pattern of income volatility described in vignette. Interaction terms were omitted after the ANOVA results indicate that they are not significant. Respondents in the high volatility group preferred significantly less progressive taxes than those in the low volatility group. None of the different patterns of income over the three year period are significant with the exception of the constantly increasing pattern (low-mid-high).

**Discussion**

The results from the survey experiment suggest that increasing income volatility may lead individuals to prefer less progressive taxes. Specifically, greater income volatility may be associated with a rise in support for tax cuts for top income earners. In general, higher income volatility induces individuals to prefer a less progressive tax schedule. This decline in preference for tax progressivity occurs regardless of the pattern of volatility over time. Respondents tax preferences were not affected by whether the household described to them in the vignette earned relatively more or less in the final year. Additionally, respondents’ tax preferences did not change depending on whether household income rose or fell between the second and third year. What appears to matter is how income changes over the entire three year period. Large shifts up or down, regardless of whether these shifts occurred in year 1, 2, or 3, caused respondents to choose lower tax rates for upper incomes. This effect is particularly strong for individuals who are married, as well as to those who identify as Independents.

Altogether, the series of surveys suggests that there is a negative relationship between income volatility and support for progressive taxes. The question remains why do respondents prefer relatively lower taxes for upper incomes after being described households that face higher income volatility. The survey results are consistent with two possible explanations. One possibility is that respondents in the high volatility group were more likely to believe that their own incomes could increase at some point in the future. Under this belief, they
would prefer to keep taxes at upper incomes low in order to collect a greater share of these gains. In addition, respondents may not see the connection between tax revenue and public assistance programs. While the potential for large gains may lead individuals to support tax cuts for the rich, the potential for large losses should have the opposite effect of engendering support for raising top tax rates in order to fund programs that help low income households. The connection between tax revenue and government expenditures is not as straightforward as the link between tax rates and net household earnings. Respondents may not link higher taxes for the rich with the idea of expanding social safety net. Alternatively, they may not believe that the government will spend the revenue collected from high income households on public assistance programs. In comparison, households know that lower taxes means that they will be able to pocket a greater share of their earnings at the end of the year.

It is also possible that respondents’ tax preferences are not entirely shaped by self-interest. Those in the high volatility group may believe that all families are more likely to experience sizable changes in income. Thus, households earning top incomes today may not have earned as much in the past or may see their incomes drop in the future. As a result, respondents who were described households facing greater income volatility may want to keep tax rates at upper incomes lower out of a sense of fairness to top income households. They believe that all households, including those at the top, should be able to keep more of their income when times are good in order to protect against future losses in earnings. There is some evidence to support this idea. Respondents described households facing greater income volatility wanted significantly lower tax rates for incomes greater than $100,000 compared to those in the low volatility group. This income is much higher than the majority of respondents in the survey, whose median income was around $35,000. In addition, it is much higher than the incomes earned by the households described in the vignette. Thus, regardless of whether respondents are using the hypothetical household’s or their own income as a reference point, changes in rates for incomes above $100,000 are unlikely to apply to most respondents. Believing that all households face greater income volatility, respondents in the high volatility group may prefer lower taxes for upper incomes on average out of a sense of fairness. They may reason that these upper income households may not always be at the top, and the tax rate should reflect the fact that households at the top may have lower expected incomes.

Measuring these subtle relationships, especially concerning complex policy issues, requires a strong survey design, which facilitates the expression of more considered preferences from respondents. This survey experiment required a great deal from respondents. They were asked to select their preferred tax rates for multiple income levels. Given the complexity of the task, their responses appear reasonable. In fact, when you compare the average rates for the five income levels in the survey to actual federal tax rates they align to a great extent. Table 3.6 compares the sample average to the average federal tax rate (including pay roll

---

4 This reasoning also extends to lower incomes. While not significant, respondents in the high volatility group also wanted higher tax rates at incomes below $60,000 compared to the low volatility group. This choice is consistent with the idea that under conditions of high income volatility lower income households may not always have low incomes and they should pay a rate in taxes that reflects this potential for future gains.
3.5. RESULTS OF SURVEY EXTENSION

taxes) and the average income tax rate. The average rates in the survey sample align fairly closely to average federal tax rates (including pay roll tax), but even more surprisingly, the tax rates increase for all three at similar rates from one income level to the next. This provides some evidence that the level of progressivity desired by voters may be close to the level found in the current federal tax system.

<table>
<thead>
<tr>
<th>Income</th>
<th>Survey Average Preferred Tax Rate</th>
<th>Income</th>
<th>Total Average Federal Tax Rate</th>
<th>Average Individual Income Tax Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>20,000</td>
<td>6.0</td>
<td>18,400</td>
<td>4</td>
<td>-6.8</td>
</tr>
<tr>
<td>40,000</td>
<td>9.6</td>
<td>42,500</td>
<td>10.6</td>
<td>-.4</td>
</tr>
<tr>
<td>60,000</td>
<td>12.6</td>
<td>64,500</td>
<td>14.3</td>
<td>3.3</td>
</tr>
<tr>
<td>100,000</td>
<td>16.7</td>
<td>94,100</td>
<td>17.4</td>
<td>6.2</td>
</tr>
<tr>
<td>200,000</td>
<td>21.4</td>
<td>264,700</td>
<td>25.1</td>
<td>14.4</td>
</tr>
</tbody>
</table>

Table 3.6: Average Federal Tax Rates for All Households, by Comprehensive Household Income Quintile, 2007. The total average federal tax rate includes payroll taxes, while the average individual income tax rate only reflects income tax rates (including credits). Source: IRS data

Altogether, these findings suggest that greater income volatility may lead voters to support tax reforms that reduce taxes for the rich, even if they themselves are unlikely to personally benefit from such reforms. To the extent that income volatility contributes to conditions of growing income inequality, it is possible that support for tax progressivity among the middle class will not rise as income inequality rises as predicted by standard models. In future work, I plan to conduct similar survey experiments to determine whether income volatility increases support for expanding government assistance to low income households. It is possible that Americans do not make the connection between tax progressivity and funding for programs to help low income families. In a world with greater income volatility, voters may demand both low tax rates for upper incomes as well as increased government expenditures on programs that help families weather losses in income.
Chapter 4

Permutation Inference for Two-way Factorial Designed Experiments

Different concerns arise in experiments conducted for the social sciences than those found in science and industry and thus the methods developed for the latter are not necessarily well adapted for use in social science research. The standard textbooks on experiments focus their discussion on applications in the laboratory, an environment that affords researchers significant control over the design, measurement precision, sample selection, and consistency of the experiment. In addition, responses typically have well-behaved distributions with low variance producing large effects with tight confidence intervals. Under these conditions, relatively few observations are needed to justify the normality assumptions of the parametric methods used in the analysis.

This level of control is rare in the social sciences, and more generally in experiments involving human subjects. Units are heterogeneous and assumptions of normality with regard to human behavior are questionable, especially when sample sizes are small. Experiments conducted outside of the laboratory suffer from additional problems. Survey and field experiments often have unequal numbers of units per treatment group, since researchers do not have the ability to apply sophisticated randomization procedures. In addition, randomness in the willingness of subjects to participate introduces imbalance in the design. Thus, traditional parametric methods that work for science and industrial experiments are not necessarily well suited for those in the social sciences.

In this paper, I introduce a non-parametric method for analyzing regular factorial experiments that can handle unbalanced designs, i.e. the types of experiments often encountered in political science. A factorial experiment is one where the units (participants in the study) are assigned to two or more treatments. Each treatment may have multiple levels (conditions). The design is unbalanced if there are unequal numbers of units assigned to each combination of treatments. For example, consider a ‘get out the vote’ experiment where the researcher is interested in both the effects of telephone and in-person contact. The design has four conditions. Individuals are randomly assigned to receive a phone call, an in-person visit, both or neither. The last condition is often referred to as ‘the control’ condition since these
Table 4.1: Examples of two-way unbalanced regular factorial designs from hypothetical 'get-out-the-vote' experiments. The first design is a basic two factor design with each factor having two levels. In the second design, another treatment condition is added to factor B, expanding the original $2 \times 2$ experiment to a $2 \times 3$ design.

<table>
<thead>
<tr>
<th>2 × 2 regular unbalanced factorial design, $N = 200$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No phone call</td>
<td>Phone call</td>
</tr>
<tr>
<td>No in-person visit</td>
<td>$N_{11} = 53$</td>
</tr>
<tr>
<td>In-person visit</td>
<td>$N_{21} = 45$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2 × 3 regular unbalanced factorial design, $N = 300$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No phone call</td>
<td>Phone call with partisan message</td>
</tr>
<tr>
<td>No in-person visit</td>
<td>$N_{11} = 56$</td>
</tr>
<tr>
<td>In-person visit</td>
<td>$N_{21} = 45$</td>
</tr>
</tbody>
</table>

individuals were not contacted during the study, though this classification is not quite right. If we characterize the experiment in terms of a factorial design, there is a treatment and control condition associated with both the phone and in-person interventions (See Table 4.1).

This experiment has a simple $2 \times 2$ design (two factors each with two levels), but it is easy to imagine a similar experiment having a $2 \times 3$ design. Suppose in the previous 'get-out-the-vote' experiment, there are two different types of phone messages. One is a simple reminder to vote and the other has a partisan message. As before, there are two factors: phone and in-person contact, but now phone contact has three levels (or conditions). More complex factorial experiments can encompass more than two treatments each with multiple, varying numbers of levels. In this paper, I propose a non-parametric method for analyzing any two-way (regular) factorial experiment where each factor has two or more levels. In the future, I plan to extend the procedure to handle any regular factorial design.

Factorial experiments abound in the social sciences, though these experiments are not always identified or analyzed as such. Their main purpose is to test for interaction effects between the treatments. The failure to identify the experiment as a factorial experiment, however, can lead researchers to apply inappropriate analysis methods that assume no interaction effects. Consider the simple $2 \times 2$ designed ‘get out the vote’ experiment (Table 4.1). To assess the effect of phone contact, one could perform a t-test comparing the voting rate between participants who were assigned to receive a phone call and those who were not, regardless of whether they were assigned to receive a house visit. This is a valid estimate of the effect of calling only if there is no interaction effect between in-person contact and telephone contact. Another strategy might be to test the difference in the effect of calling for
only those individuals who were not assigned to an in-person visit. While a valid estimate of the treatment effect, this method would unnecessarily throw out half of the data.

Regression analysis also has drawbacks. In a factorial design, there is a natural hierarchy of hypotheses tests, which begins with a more general test for main factor and interaction effects and then proceeds to estimates of pairwise differences between treatments. The standard parametric method for analyzing a factorial experiment begins with analysis of variance (ANOVA), which can be derived from the standard OLS beta estimates in a regression that includes interaction effects (See appendix). ANOVA simultaneously tests the separate null hypotheses of zero effect associated with the main effects and their interactions. The method provides information on which factors and interactions are relevant, guiding subsequent estimation of effect sizes and confidence intervals, as well as heterogeneous treatment effects.

In this paper, I propose a non-parametric alternative to ANOVA that is tailored to accommodate features common to experiments in political science and the social sciences in general: unknown distribution of the errors and $f$ in the experimental design. The main advantage of permutation methods is that they do not have the normality assumptions of their parametric counterparts, and thus perform well even in experiments with few observations where the responses do not have well-behaved distributions (Keele, McConnaughy and White 2012; Pesarin and Salmaso 2010). In the context of factorial experiments, permutation methods have further advantages over parametric methods. Whereas, parametric tests for main factors and interactions are positively correlated under the normality assumptions for the errors (Pesarin and Salmaso 2010), the permutation method produces three separate and uncorrelated tests of the main hypotheses by synchronizing exchanges of observations across treatments. In addition, the standard adjustments to parametric ANOVA for unbalanced designs make new assumptions on the data, which are highly sensitive to interactions between the main factors (Langsrud 2003).

The permutation procedure I propose accommodates experiments with unequal numbers of observations per treatment group and produces consistent estimates of the effects of the main factors even when there are significant interaction effects. The procedure is similar to one proposed by Basso et al. (2009), but with changes to increase the power of the test or the ability to correctly reject the null hypothesis. This method applies to two-way designs, however, it could potentially be extended to all regular factorial experiments. In addition, I consider the performance of the procedure for unbalanced factorial designs with the inclusion of proofs for uncorrelatedness, unbiasedness and consistency of the tests under conditions of unbalance.

This method requires the assumption of homoskedasticity since the tests for the main factor and interaction effects depend on mean differences. There are non-parametric methods for rank based tests of unbalanced factorial experiments, which relax this assumption, but are less comparable to traditional ANOVA analysis (Akritas, Arnold and Brunner 1997). Anderson and Ter Braak (2003) discuss a full range of non-parametric methods for a variety of factorial designs, including procedures for two-way design methods that permute residuals from linear regressions of the full model. However there is no exact permutation test for an interaction effect under this permutation strategy. In addition, they do not provide
an analysis of different methods for cases of unbalanced designs and predates Basso et al. (2009)’s introduction of exact tests using restricted permutation methods.

Monte Carlo simulations comparing the non-parametric method I develop to the standard regression (ANOVA) method demonstrate substantial gains even under conditions when the assumptions of regression are met, namely when errors are independently and identically distributed normal. This is a result of the uncorrelatedness of the non-parametric tests for the main factor and interaction effects. The gains are also a direct result of improvements in the permutation procedure which enhance its ability to detect significant and small effects compared to alternate non-parametric procedures. Under conditions where we would expect parametric methods to perform poorly, namely when the errors do not have well-behaved distributions and when the design is unbalanced and has factors with multiple (> 2) levels, the gains from using the permutation method over the traditional parametric approach increase further.

4.1 Two-way Factorial Designed Experiments

The permutation method I propose applies to any factorial experiment with two factors of multiple levels. This type of design is also known as a two-way layout. I have already discussed two examples of factorial designs for a hypothetical ‘get-out-the-vote-experiment’ (See Table 4.1). I will refer to the $2 \times 3$ design throughout the paper in order to illustrate the permutation method. Mathematically, the standard linear model for the two-way layout is written:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

where $y_{ijk}$ is the response variable of the $k$th respondent for the $i$th level of factor A and of the $j$th level of factor B. Factors A and B correspond to the two treatments. Each factor has two or more levels. The simplest design is one where each factor has two levels corresponding to the ‘treated’ and ‘control’ conditions, as in the $2 \times 2$ ‘get out of the vote’ experiment. The term, $\alpha_i$, represents the marginal effect of level $i$ of factor A and $\beta_j$ represents the marginal effect of level $j$ of factor B. The $\alpha\beta_{ij}$’s capture the possible interaction effects between the two main factors. The errors, $\epsilon_{ijk}$, have mean zero and come from the same unknown distribution. Unlike traditional parametric ANOVA methods that stand on normality assumptions, the non-parametric permutation procedure assumes only that errors are exchangeable and identically distributed. Less restrictive than independence, the exchangeability assumption is justified by randomization in the experimental design.

The model is over-parameterized and therefore we assume the standard zero-sum constraints for the main effects and their interactions:
\[ \sum_i \alpha_i = \sum_j \beta_j = \sum_i \alpha \beta_{ij} = \sum_j \alpha \beta_{ij} = 0 \]

A common question of interest is whether the responses of units differ across treatment levels within a single factor and whether there are interaction effects. Testing for pairwise differences between treatment groups using a t-test or non-parametric alternative would raise multiple comparison issues. In addition, this approach is biased if there are interaction effects. Analysis of variance is the traditional method for analyzing factorial designed experiments, providing a valid test of differences between multiple treatment groups with possible interaction effects. In a two-way layout, the main hypotheses of interest in the standard analysis of variance approach are the following:

- \( H_{0A} : \alpha_i = 0, \forall i \),
- \( H_{0B} : \beta_j = 0, \forall j \),
- \( H_{0AB} : \alpha \beta_{i,j} = 0, \forall i, j \)

The first hypothesis states that none of the treatments that comprise the levels of factor A has an effect on the responses. For example, in the 2 × 3 ‘get out the vote experiment’, \( H_{0A} \) posits that assignment to receive an in-person visit had no effect on an individual’s decision to vote. Similarly, \( H_{0B} \) is the hypothesis that assignment to receive a phone call, with or without a partisan message, had no effect on turnout. Finally, \( H_{0AB} \) states that there are no interaction effects. We want a permutation procedure that allows us to test each hypothesis separately against its corresponding alternative hypothesis, independent of whether the other two hypotheses are true. The alternative hypotheses are \( H_{0A} : \alpha_i \neq 0 \), \( H_{0B} : \beta_j \neq 0 \), \( H_{0AB} : \alpha \beta_{i,j} \neq 0 \), for some \( i, j \). Other alternative hypotheses are possible, including one-sided alternatives.

## 4.2 Permutation Inference for Two Factor ANOVA

All permutation procedures proceed by exchanging observations between treatments in order to estimate the distribution of the test statistic under the assumption that the null hypothesis is true. A common test statistic is the difference in the responses between treatments, which provides a valid test of the standard sharp null hypothesis. The sharp null hypothesis assumes zero treatment effect for every individual. Differences in the responses across treatment groups arise by chance, because the null hypothesis assumes that the observed data come from the same unknown population distribution. Thus, responses are exchangeable with respect to treatment groups. This exchangeability condition justifies the permutation procedure used to estimate the treatment effect. In practice, permutation procedures are carried out by shuffling responses between treatments repeatedly and, with each iteration, recalculating the test statistic from the permuted data. The collection of
4.3. CALCULATION OF THE TEST STATISTIC

The test statistic calculated from the observed data is compared with those generated from the permutations to estimate a p-value under the assumption of the null hypothesis. The p-value estimate is just the percent of test statistics at least as extreme as the one calculated from the observed data. This estimate is exact if all possible values of the test statistic are calculated from the permuted data given the permutation procedure.\(^1\) Keele, McConnaughy and White (2012) provide a good introduction to permutation inference and procedures.

In a factorial designed experiment, only certain exchanges between treatments are possible and the permutation procedure across pairs of treatments must be synchronized. The synchronization is necessary since responses for different levels of the factor of interest might have different expected values across levels of the other factor. The permutation procedure must be able to distinguish between effects linked to factor A and effects linked to factor B (as well as their interaction) even though all units receive treatment with respect to both factors. We want a permutation structure that enables us to test each hypothesis independently, such that the likelihood of rejecting or accepting one hypothesis does not depend on whether or not the other two null hypotheses are true. In the ‘get-out-the-vote’ example, the null hypothesis \(H_{0A}\), states that being assigned to an in-person visit has no effect on turnout. This assumption implies that observations are exchangeable with respect to factor A, but does not make any assumption about exchangeability with regards to factor B, assignment to phone call. This requirement imposes restrictions on how data is exchanged between treatments.

First I will describe the calculation of the test statistic, followed by a detailed discussion of the permutation procedure. The constrained permutation procedure is similar to procedures proposed by Basso et al. (2009), but with changes that increase the power of the test to correctly reject the null hypothesis when there are significant effects. This method applies to two-factor experimental designs; however, it could potentially be extended to all regular factorial experiments.

4.3 Calculation of the Test Statistic

Under the permutation framework, first we calculate partial test statistics for the main factors. The test statistics are weighted differences in sums between two levels of the factor of interest at a fixed level for the other factor. The weights adjust for the imbalance in the design. All pairwise combinations of levels for each factor are compared.

\[
\alpha T_{i|ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk} - \frac{1}{n_{sj}} \sum_{k=1}^{n_{sj}} y_{sjk}
\]

\[
\beta T_{j|i} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk} - \frac{1}{n_{ih}} \sum_{k=1}^{n_{ih}} y_{ihk}
\]

\(^1\)Note that exactness does not necessarily require all possible permutations of the data to be performed.
4.3. CALCULATION OF THE TEST STATISTIC

Consider the 2 × 3 ‘get-out-the vote’ experiment. Suppose $S_{ij} = \sum_{k}^{n_{ij}} y_{ijk}$, the sum of the responses for the $i^{th}$ level of factor A and of the $j^{th}$ level of factor B. Factor A is assignment to in-person visit and factor B is assignment to phone call (Table 4.2).

Table 4.2: 2 × 3 regular unbalanced factorial design

<table>
<thead>
<tr>
<th>No phone call</th>
<th>Phone call</th>
<th>Phone call with partisan message</th>
</tr>
</thead>
<tbody>
<tr>
<td>No in-person visit</td>
<td>$S_{11} = \sum_{k}^{n_{11}} y_{11k}$</td>
<td>$S_{12} = \sum_{k}^{n_{12}} y_{12k}$</td>
</tr>
<tr>
<td>In-person visit</td>
<td>$S_{21} = \sum_{k}^{n_{21}} y_{21k}$</td>
<td>$S_{22} = \sum_{k}^{n_{22}} y_{22k}$</td>
</tr>
</tbody>
</table>

As an example, consider the computation of the test statistic for factor B, assignment to phone call. For each of the $\binom{3}{2} = 3$ different pairwise combinations of levels for factor B, we calculate a difference in sums separately for individuals assigned to an in-person visit and those who were not. This step produces six partial test statistics (See Table 4.3).

The test statistics for factors A and B are then calculated from their partial statistics as follows:

$$\alpha T_A = \sum_{i<s} \left[ \sum_{j}^{T_{is|j}} \right]^2 \quad \beta T_B = \sum_{j<h} \left[ \sum_{i}^{T_{jh|i}} \right]^2$$

Returning to the previous example, to calculate the test statistic for factor B, we sum the rows of Table 4.3, and then square and sum down the column. I will not demonstrate the calculation of the test statistics for factor A; the computation is straightforward since the factor only has two levels. The partial test statistics for factor A are simply the weighted differences between individuals who were assigned to receive an in-person visit and those who were not, calculated separately for each level of factor B. The three partial test statistics are squared and then added together to produce the final test statistic.

Table 4.3: The partial test statistics for factor B, assignment to phone call, are the differences in the average responses over all pairs of levels for factor B, calculated separately for each level of factor A.

<table>
<thead>
<tr>
<th>No in-person visit</th>
<th>In-person visit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{a}T_{12</td>
<td>1} = \frac{1}{n_{21}} S_{21} - \frac{1}{n_{11}} S_{11}$</td>
</tr>
<tr>
<td>$^{a}T_{13</td>
<td>1} = \frac{1}{n_{31}} S_{31} - \frac{1}{n_{11}} S_{11}$</td>
</tr>
<tr>
<td>$^{a}T_{23</td>
<td>1} = \frac{1}{n_{31}} S_{31} - \frac{1}{n_{21}} S_{21}$</td>
</tr>
</tbody>
</table>
4.4 Method of Permutation

We also use the partial statistics of the two factors to derive the test static for the interaction effect as follows:

$$\alpha_{T_{AB}} = \sum_{i<s} \sum_{j<h} \left[ \alpha_{T_{is|j}} - \alpha_{T_{is|h}} \right]^2$$

$$\beta_{T_{AB}} = \sum_{j<h} \sum_{i<s} \left[ \beta_{T_{jh|i}} - \beta_{T_{jh|h}} \right]^2$$

$$T_{AB} = \alpha_{T_{AB}} + \beta_{T_{AB}}$$

$\alpha_{T_{AB}}$ and $\beta_{T_{AB}}$ are equivalent terms, even though the former was obtained by exchanging units between levels of factor A and the latter was obtained through permutations of units within factor B. Considered individually or added together they provide an exact test for the interaction effects. In practice, it is helpful to compute both to confirm that there are no errors in your code.

I will now discuss the permutation procedure to estimate the distributions of the test statistics under the null hypotheses.

4.4 Method of Permutation

Data is permuted across all pairs of treatment levels for the factor of interest within each level of the other factor. When conducting the permutation, we must choose both the number of exchanges and the units to be exchanged between treatment pairs. The units are the experimental subjects. Data is permuted across treatments to estimate an empirical distribution of the test statistic for a given factor assuming the relevant null hypothesis is true. The permutation occurs only between levels for the factor of interest, and does not occur across levels of the other factor. Returning to the 2 × 3 ‘get-out-the-vote’ experiment, to calculate the test statistic for factor A, assignment to in-person visit, data is permuted between levels of factor A, but never between levels for factor B, assignment to phone call. We use an asterisk to distinguish permuted data from the observed data. Following this notation, the partial test statistics computed from the permuted data are:

$$\alpha_{T_{is|j}}^* = \frac{1}{n_{ij}} \sum_k y_{ijk}^* - \frac{1}{n_{sj}} \sum_k y_{sjk}^*$$

$$\beta_{T_{jh|i}}^* = \frac{1}{n_{ij}} \sum_k y_{ijh}^* - \frac{1}{n_{ih}} \sum_k y_{ih}^*$$

I propose a method of synchronized permutations that ensures that the test statistics for the two main factors and the interaction effect are uncorrelated and consistent even when the number of observations per cell differs at random, producing an unbalanced design. The proofs are provided in the following section. My method for the synchronized permutation procedure is similar to one proposed by Basso et al. (2009), but with changes to increase the power of the three tests and reduce the computational complexity for relatively large experiments.

The critical steps to the procedure are in selecting the number of exchanges made in each permutation of the data and in choosing which units to exchange between treatment pairs. Let $\nu_{is|j}$ denote the number of units to be exchanged between level $i$ and $s$ of factor A within
level $j$ of factor B. Similarly, let $\eta_{j|h|i}$ denote the number of units to be exchanged between level $j$ and $h$ of factor B within level $i$ of factor A.

The process is illustrated with a $2 \times 3$ factorial design similar to the ‘get-out-the-vote’ example, but where there are four observations per treatment group.
4.4. METHOD OF PERMUTATION

1. Randomly permute observations between all pairwise combinations of levels of factor A at each level of factor B. In practice, the vector of treatment assignments for each pair is sampled without replacement giving each possible permutation equal probability.

\[ \begin{bmatrix}
\nu_{12|1} = 2 & \nu_{12|2} = 1 & \nu_{12|3} = 2 \\
& & \\
\end{bmatrix} \]

<table>
<thead>
<tr>
<th>(A_1)</th>
<th>(B_1)</th>
<th>(B_2)</th>
<th>(B_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Y_{112})</td>
<td>(Y_{121})</td>
<td>(Y_{131})</td>
<td></td>
</tr>
<tr>
<td>(Y_{113})</td>
<td>(Y_{122})</td>
<td>(Y_{132})</td>
<td></td>
</tr>
<tr>
<td>(Y_{114})</td>
<td>(Y_{123})</td>
<td>(Y_{134})</td>
<td></td>
</tr>
</tbody>
</table>

\[ \rightarrow \]

<table>
<thead>
<tr>
<th>(A_2)</th>
<th>(B_1)</th>
<th>(B_2)</th>
<th>(B_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Y_{211})</td>
<td>(Y_{221})</td>
<td>(Y_{231})</td>
<td></td>
</tr>
<tr>
<td>(Y_{212})</td>
<td>(Y_{222})</td>
<td>(Y_{232})</td>
<td></td>
</tr>
<tr>
<td>(Y_{213})</td>
<td>(Y_{223})</td>
<td>(Y_{233})</td>
<td></td>
</tr>
<tr>
<td>(Y_{214})</td>
<td>(Y_{224})</td>
<td>(Y_{234})</td>
<td></td>
</tr>
</tbody>
</table>

2. Calculate the test statistics \(\alpha T_A^*\) and \(\alpha T_{AB}^*\) from the permuted dataset, based on the following three differences in means (the asterisk indicates that the data has been permuted):

\[ \alpha T_{12|1}^* = \frac{1}{n_{11}} \sum_{k=1}^{n_{11}} y_{11k}^* - \frac{1}{n_{21}} \sum_{k=1}^{n_{21}} y_{21k}^* \]

\[ \alpha T_{12|2}^* = \frac{1}{n_{12}} \sum_{k=1}^{n_{12}} y_{12k}^* - \frac{1}{n_{22}} \sum_{k=1}^{n_{22}} y_{22k}^* \]

\[ \alpha T_{12|3}^* = \frac{1}{n_{13}} \sum_{k=1}^{n_{13}} y_{13k}^* - \frac{1}{n_{23}} \sum_{k=1}^{n_{23}} y_{23k}^* \]

\[ \alpha T_A^* = (\alpha T_{12|1}^* + \alpha T_{12|2}^* + \alpha T_{12|3}^*)^2 \]

\[ \alpha T_{AB}^* = (\alpha T_{12|1}^* - \alpha T_{12|2}^*)^2 + (\alpha T_{12|1}^* - \alpha T_{12|3}^*)^2 + (\alpha T_{12|2}^* - \alpha T_{12|3}^*)^2 \]
3. Randomly permute observations between all pairwise combinations of levels of factor \( B \) at each level of factor \( A \). In practice, the vector of treatment assignments for each pair is sampled without replacement giving each possible permutation equal probability.

\[
\begin{align*}
\eta_{12|1} &= 2 & \eta_{12|2} &= 3 \\
\begin{pmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{114} \end{pmatrix} & \rightarrow \\
\begin{pmatrix} y_{211} \\ y_{212} \\ y_{213} \\ y_{214} \end{pmatrix} & \begin{pmatrix} y_{111} \\ y_{122} \\ y_{114} \\ y_{123} \end{pmatrix} & \begin{pmatrix} y_{222} \\ y_{221} \\ y_{214} \\ y_{223} \end{pmatrix} \end{align*}
\]

\[
\eta_{13|1} = 2 & \eta_{13|2} = 2 \\
\begin{pmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{114} \end{pmatrix} & \rightarrow \\
\begin{pmatrix} y_{211} \\ y_{212} \\ y_{213} \\ y_{214} \end{pmatrix} & \begin{pmatrix} y_{111} \\ y_{134} \\ y_{114} \\ y_{132} \end{pmatrix} & \begin{pmatrix} y_{211} \\ y_{231} \\ y_{214} \\ y_{234} \end{pmatrix} \end{align*}
\]

\[
\eta_{23|1} = 3 & \eta_{23|2} = 1 \\
\begin{pmatrix} y_{121} \\ y_{122} \\ y_{123} \\ y_{124} \end{pmatrix} & \rightarrow \\
\begin{pmatrix} y_{221} \\ y_{222} \\ y_{223} \\ y_{224} \end{pmatrix} & \begin{pmatrix} y_{131} \\ y_{132} \\ y_{133} \\ y_{134} \end{pmatrix} & \begin{pmatrix} y_{212} \\ y_{232} \\ y_{233} \\ y_{234} \end{pmatrix} \end{align*}
\]
4.4. METHOD OF PERMUTATION

4. Calculate the test statistics $\beta T_A$ and $\beta T_{AB}$ from the permuted dataset.

$$\beta T_{12|1} = \frac{1}{n_{11}} \sum_{k=1}^{n_{11}} y_{11k}^* - \frac{1}{n_{12}} \sum_{k=1}^{n_{12}} y_{12k}^*$$

$$\beta T_{13|1} = \frac{1}{n_{11}} \sum_{k=1}^{n_{11}} y_{11k}^* - \frac{1}{n_{31}} \sum_{k=1}^{n_{31}} y_{31k}^*$$

$$\beta T_{23|1} = \frac{1}{n_{21}} \sum_{k=1}^{n_{21}} y_{21k}^* - \frac{1}{n_{31}} \sum_{k=1}^{n_{31}} y_{31k}^*$$

$$\beta T_{12|2} = \frac{1}{n_{21}} \sum_{k=1}^{n_{21}} y_{21k}^* - \frac{1}{n_{22}} \sum_{k=1}^{n_{22}} y_{22k}^*$$

$$\beta T_{13|2} = \frac{1}{n_{12}} \sum_{k=1}^{n_{12}} y_{12k}^* - \frac{1}{n_{32}} \sum_{k=1}^{n_{32}} y_{32k}^*$$

$$\beta T_{23|2} = \frac{1}{n_{22}} \sum_{k=1}^{n_{22}} y_{22k}^* - \frac{1}{n_{32}} \sum_{k=1}^{n_{32}} y_{32k}^*$$

$$\beta T_B = (\beta T_{12|1} + \beta T_{12|2})^2 + (\beta T_{13|1} + \beta T_{13|2})^2 + (\beta T_{23|1} + \beta T_{23|2})^2$$

$$\beta T_{AB} = (\beta T_{12|1} - \beta T_{12|2})^2 + (\beta T_{13|1} - \beta T_{13|2})^2 + (\beta T_{23|1} - \beta T_{23|2})^2$$

5. Repeat these steps over many iterations (> 1000)

6. Collect and sort the test statistics calculated from the permuted data sets to estimate distributions of $\alpha T_A$, $\beta T_B$ and $T_{AB}$ under the set of null hypotheses.

7. Estimate the p-value from the fraction of test statistics greater than the test statistic calculated from original data:

$$\Pr\{T^* \geq T^0 | X \} = \frac{\#T^* \geq T^0}{\#T^*}$$

The procedure respects the exchangeability restrictions imposed by the dependencies of the data under the set of null hypotheses and provides consistent estimates of the intermediate test statistics. The data is first permuted within each cell, generating a permuted data set $y^*$, and then two sets of exchanges are made: one to calculate the test statistic for factor A and the other to calculate the test statistic for factor B. The test statistic for the interaction effect is derived from partial test statistics for the main factors.

The imbalance in the design restricts the number of possible exchanges to the smallest of the two paired cells, such that $\nu_{islj} \in \{0, \cdots, n_{islj}\}$ where $n_{islj} \equiv \min\{n_{ij}, n_{sj}\}$ and $\eta_{jhli} \in \{0, \cdots, n_{jhli}\}$ where $n_{jhli} \equiv \min\{n_{ij}, n_{ih}\}$. In a $I \times J$ design, the total number of $\nu_{islj}$’s are $\binom{I}{2} \times J$ and the total number of $\eta_{jhli}$’s are $\binom{J}{2} \times I$. Allowing the number of exchanges to vary across all paired cells increases the power of the test and represents the main difference between this permutation method and the one proposed by Basso et al. (2009). Their method holds the number of exchanges constant for each iteration of the permutation procedure, guaranteeing that the test statistics for factor A and B only depend

\[\text{For } 2 \times 2 \text{ factorial experiments, estimates are unbiased. In the more general case, I demonstrate uncorrelatedness among the test statistics and consistency.}\]
4.4. METHOD OF PERMUTATION

on $\alpha_i$ and $\beta_j$ respectively (i.e. $\nu = \nu_{ij} = \eta_{j|i} = \eta$ for $i, s, j, h$). However, restricting the number of exchanges significantly reduces power of the test, particularly in cases when the design is unbalanced. Both permutation procedures generate test statistics for the main factor and interaction effects that are consistent and mutually uncorrelated.

The number of exchanges between each paired cell, $\nu_{is|j}$, has a hypergeometric distribution with parameters that depend on $n_{ij}$ and $n_{sj}$ conditional on the number of observations per cell. This distribution of $\nu_{is|j}$ guarantees that all possible values of the test statistic for factor A are equally likely, providing a uniform distribution of the test statistic under the null hypotheses. If $\nu_{is|j}$ were chosen uniformly at random, values of the test statistic associated with very small and large values of $\nu$ would occur more frequently than those for mid ranges of $\nu$, producing a biased permutation distribution. Instead, we want to choose $\nu$ such that the probability of any value of the test statistic calculated from the permuted data is equally likely.

Consider the exchange of observations between $A_1$ and $A_2$ within $B_1$. There are $n_{11}$ and $n_{21}$ observations in each cell and thus there are $n_{11} + n_{21}$ choose $\min\{n_{11}, n_{21}\}$ number of possible permutations of data across the two cells. The distribution of $\nu_{12|1}$ is chosen so that each one of these possible permutations is equally likely. In order to determine the appropriate distribution, one can think of the distribution of $\nu_{12|1}$ in terms of a matching problem. Suppose we have $n_{11}$ white balls in a white box and $n_{21}$ black balls in a black box. We throw all the balls into a single bin and mix them together. Then we draw $n_{11}$ balls from the bin and place them in the black box. The remaining $n_{21}$ balls are placed in the white box. $\nu_{12|1}$ is the number of mismatches or “failures” in each box (i.e. the number of white balls in the black box and visa versa). Thus, $\nu_{is|j}$ and $\eta_{j|h|i}$ are both independently distributed hypergeometric for all $i, s, j, h$ conditional on the number of observations across cells.

Given that each pair of cells has $\binom{n_{ij} + n_{sj}}{\nu_{is|j}}$ possible permutations, the total number of possible values of $\alpha_{T_A}$ is:

$$\prod_j \prod_{i<s} \binom{n_{ij} + n_{sj}}{\nu_{is|j}}$$

which simplifies under a balanced design to $\binom{2n}{n}^{J \times I (I-1)/2}$. The total number of possible exchanges determines the minimum significance levels: greater cardinality increases the power of the test. In contrast, the cardinality of the most powerful permutation procedure proposed by Basso et al. (2009) is $\sum_{\nu=0}^{n} \binom{n}{\nu}^{J \times I (I-1)}$. To put this difference into perspective, if the number of observations per cell is 5 and the design is $2 \times 2$, the total number of possible

---

3 $\eta_{j|h|i}$ also has a conditional hypergeometric distribution with parameters that depend on $n_{ij}$ and $n_{ih}$.

4 Both $\alpha_{T_A}$ and $\beta_{T_B}$ are sums of the differences between cells. For example, in a $2 \times 3$ design, the test statistic for factor A is the sum of three differences in means: the difference in levels 1 and 2 of factor A for levels 1, 2 and 3 of factor B (see item two describing the permutation procedure).
permutations under my method is 63504 compared to 21252 for their method. The difference grows significantly as \( n \) increases.

In addition, the computation of the distribution of the number of exchanges under their procedure is not straightforward, since its distribution does not follow a standard distribution such as the hypergeometric. As \( n \), \( I \) and \( J \) increase, the cardinality rapidly grows making it difficult to calculate the distribution of the directly. Furthermore, the range of the number of exchanges is limited to the smallest cell size in the experiment, whereas the range of exchanges under my procedure is only limited to the smallest of each paired cell. This flexibility allows my procedure to better accommodate imbalance in the design.

The next three sections prove the uncorrelatedness, unbiasedness and consistency of the test statistics associated with the non-parametric permutation method. The proofs assume imbalance in the design, i.e. differing numbers of observations per cell. Specifically, they assume that the imbalance in the design is random. Random differences in the number of observations per cell often arise in political science experiments. Two common reasons are random attrition (departure is independent of treatment assignment) and the use of a randomization procedure, which assigns treatments with equal probability but without restrictions to ensure balance, such as a simple “roll of the dice” procedure. The proofs do not apply to conditions where imbalance is non-random, which often arises because of non-random attrition (departure is correlated with treatment assignment) and in hierarchical designs. In the latter case, hierarchical designs deviate from the standard factorial design and therefore cannot be analyzed as such (Gelman et al. 2005). Non-random attrition introduces selection bias, which invalidates the assumption of independence of treatment assignment and thus raises more serious issues than simple imbalance.

### 4.5 Uncorrelatedness

One of the benefits of the proposed method is that the test statistics for the main factors and interaction effects are uncorrelated, unlike the regression/ANOVA parametric alternatives. This result is a consequence of the constrained permutation method, where observations are exchanged between treatment levels of the factor of interest, but within the same level of the other factor. I will prove that \( T_A^* \) and \( T_{AB}^* \) are uncorrelated for the simple \( 2 \times 2 \) design and then provide evidence that uncorrelatedness extends to more complicated designs through simulation exercises.

The covariance of \( T_A^* \) and \( T_{AB}^* \) is:

\[
Cov(T_A^*, T_{AB}^*) = E(T_A^* T_{AB}^*) - E(T_A^*) E(T_{AB}^*)
\]

where \( T_A^* = (\alpha T_{12|1} + \alpha T_{12|2})^2 \) and \( T_{AB}^* = (\alpha T_{12|1} - \alpha T_{12|2})^2 \). The squares do not affect the calculation of the covariance in the \( 2 \times 2 \) case and can be dropped. Substituting these values into the covariance formula, we have:
\[ Cov(\alpha T_A^*, \alpha T_{AB}^*) = E[(\alpha T_{121}^* + \alpha T_{122}^*)(\alpha T_{121}^* - \alpha T_{122}^*)] - E(\alpha T_{121}^* + \alpha T_{122}^*)E(\alpha T_{121}^* - \alpha T_{122}^*) \]

\[ = E[(\alpha T_{121}^*)^2] - E[(\alpha T_{122}^*)^2] - [E(\alpha T_{121}^*)]^2 - [E(\alpha T_{122}^*)]^2 = V(\alpha T_{121}^*) - V(\alpha T_{122}^*) \]

The vector \( n = (n_{11}, n_{21}, n_{12}, n_{22}) \) represents the number of observations in each cell. It is a random variable with a multinomial distribution with parameters \( N = \sum_i \sum_j n_{ij} \), \( p = p_{ij} = \frac{1}{I \times J} \) \( \forall i, j \) and \( E(n_{ij}) = Np \) \( \forall i, j \). Its distribution follows from the assumption that the imbalance in the design is generated through random chance, with each observation having an equal probability of being assigned to any of the \( I \times J \) treatment combinations. Starting with the first term, we can assess the overall variance through a decomposition of the variance that conditions on \( n \) as follows:

\[ V(\alpha T_{121}^*) = V_n[E(\alpha T_{121}^*|n)] + E_n[V(\alpha T_{121}^*|n)] \]

The partial test statistics are differences in sums of the independent random variables \( y_{ijk}^* \), which have mean \((\alpha_i + \beta_j + \alpha\beta_{ij})\) and variance \(\sigma^2\). By assumption, the \( y_{ijk}^* \)'s are a function of the fixed components \(\alpha_i, \beta_j\) and \(\alpha\beta_{ij}\) and a random component \(\varepsilon_{ijk}\), which has mean zero and variance \(\sigma^2\). Thus, we have:

\[ V_n[E(\alpha T_{121}^*|n)] = V_n \left[ E \left( 1 - \nu \left( \frac{1}{n_{11}} + \frac{1}{n_{21}} \right) \right) (\alpha_1 - \alpha_2 + \alpha\beta_{11} - \alpha\beta_{21}) \right] \]

Note that \(\nu_{i\alpha j}\) is a random variable that depends on \( n \) and has a conditional distribution that is hypergeometric with \(E[\nu_{i\alpha j}|n] = \frac{n_{ij}n_{\alpha j}}{n_{ij} + n_{\alpha j}}\) and \(V[\nu_{i\alpha j}|n] = \left( \frac{n_{ij}n_{\alpha j}}{n_{ij} + n_{\alpha j}} \right)^2 \frac{1}{n_{ij} + n_{\alpha j} - 1} \). Simplifying further, we have:

\[ V_n[E(\alpha T_{121}^*|n)] = V_n \left[ E \left( \alpha_1 - \alpha_2 + \alpha\beta_{11} - \alpha\beta_{21} \right) \right] = 0 \]

Solving for the second term of the decomposition we find:

\[ E_n[V(\alpha T_{121}^*|n)] = E_n[V(\frac{1}{n_{11}} \sum_{j=1}^2 y_{11}^* - \frac{1}{n_{21}} \sum_{j=1}^2 y_{21}^*|n)] \]

\[ E_n[V(\alpha T_{121}^*|n)] = E_n \left[ \left( \frac{1}{n_{11}} \right)^2 n_{11}\sigma^2 + \left( \frac{1}{n_{21}} \right)^2 n_{21}\sigma^2 \right] = E_n \left[ \left( \frac{1}{n_{11}} + \frac{1}{n_{21}} \right)\sigma^2 \right] = \frac{2}{Np}\sigma^2 \]

Similarly \( V_n[E(\alpha T_{122}^*|n)] = 0 \) and \( E_n[V(\alpha T_{122}^*|n)] = \frac{2}{Np}\sigma^2 \) and thus \( Cov(\alpha T_A^*, \alpha T_{AB}^*) = V(\alpha T_{121}^*) - V(\alpha T_{122}^*) = 0 \). We conclude that \(\alpha T_A^*\) and \(\alpha T_{AB}^*\) are uncorrelated, which implies \(\alpha T_A^*, \beta T_B^*\), and \(T_{AB}^*\) are uncorrelated.
The sampling errors of the permuted data set.

4.6 Unbiasedness

I will prove unbiasedness in the $2 \times 2$ case. Recall the partial test statistic:

$$\alpha T_{is|j}^* = \frac{1}{n_{ij}} \sum_{k} y_{ijk}^* - \frac{1}{n_{sj}} \sum_{k} y_{skj}^*$$

Consider the test statistic:

$$\alpha T_A = \sum_{j=1}^{2} \alpha T_{is|j}^*$$

Suppose we want to test the hypothesis $H_{0A} : \alpha_1 - \alpha_2 = 0$ against the one-sided alternative that $H_{1A} : \alpha_1 - \alpha_2 > 0$. I refer to these hypotheses as $y(0)$ and $y(\alpha)$ respectively. The test statistic is unbiased if the probability that the test statistic under the alternate hypothesis lies in the rejection region is greater than or equal to $\alpha$, the probability that the test statistic under the null hypothesis lies within the rejection region, $P_{\alpha}(T \in R(\theta)) \geq \alpha = P_{H_0}(T \in R(\theta))$.

If the alternative hypothesis, $H_{1A}$, is true, the test statistic calculated from the permuted data set can be written:

$$\alpha T_A = \left(1 - \nu_{12|1} \left(\frac{1}{n_{11}} + \frac{1}{n_{21}}\right)\right) \left(\alpha_1 - \alpha_2 + \alpha \beta_{11} - \alpha \beta_{21}\right) + \left(1 - \nu_{12|2} \left(\frac{1}{n_{12}} + \frac{1}{n_{22}}\right)\right) \left(\alpha_1 - \alpha_2 + \alpha \beta_{12} - \alpha \beta_{22}\right) + \frac{1}{n_{11}} \sum_{k=1}^{n_{11}} \varepsilon_{11k}^* - \frac{1}{n_{21}} \sum_{k=1}^{n_{21}} \varepsilon_{21k}^* + \frac{1}{n_{12}} \sum_{k=1}^{n_{12}} \varepsilon_{12k}^* - \frac{1}{n_{22}} \sum_{k=1}^{n_{22}} \varepsilon_{22k}^*$$

Define $\varepsilon_{is|j}^* = -\frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} \varepsilon_{ijk}^* - \frac{1}{n_{sj}} \sum_{k=1}^{n_{sj}} \varepsilon_{skj}^*$ and the equation further simplifies to:

$$\alpha T_A = \left(2 - \nu_1 \left(\frac{1}{n_{11}} + \frac{1}{n_{21}}\right) - \nu_2 \left(\frac{1}{n_{12}} + \frac{1}{n_{22}}\right)\right) \left(\alpha_1 - \alpha_2\right) - \left(\nu_1 \left(\frac{1}{n_{11}} + \frac{1}{n_{21}}\right)\right) (\alpha \beta_{11} - \alpha \beta_{21}) - \left(\nu_2 \left(\frac{1}{n_{12}} + \frac{1}{n_{22}}\right)\right) (\alpha \beta_{12} - \alpha \beta_{22}) + \sum_{j=1}^{2} \varepsilon_{12|j}^*$$

$$\alpha T_A = 2(\alpha_1 - \alpha_2) - \sum_{j=1}^{2} \nu_j \left(\frac{1}{n_{1j}} + \frac{1}{n_{2j}}\right) (\alpha_1 - \alpha_2) - \sum_{j=2}^{2} \nu_j \left(\frac{1}{n_{1j}} + \frac{1}{n_{2j}}\right) (\alpha \beta_{1j} - \alpha \beta_{2j}) + \sum_{j=1}^{2} \varepsilon_{12|j}^*$$
because the zero sum constraints restrict \( \sum_{j=1}^{2}(\alpha_1 \beta_j - \alpha_2 \beta_j) = 0 \). It follows that the test statistic calculated from the non-permuted data where \( \nu_j = 0 \) is:

\[
\alpha^T A = 2(\alpha_1 - \alpha_2) + \sum_{j=1}^{2} \bar{\varepsilon}_{12j}
\]

The observed test statistic is compared to the distribution of test statistics generated from multiple permutations of the data in order to produce a p-value. Assuming the alternate hypothesis is true, then the p-value for factor A is written as:

\[
\Pr[\alpha^T A \geq \alpha^T A | y(\alpha)] = \Pr\left[\sum_{j=1}^{2} \bar{\varepsilon}_{12j} \geq \sum_{j=2}^{2} \nu_j\left(\frac{1}{n_{1j}} + \frac{1}{n_{2j}}\right)(\alpha_1 - \alpha_2) + \sum_{j=2}^{2} \nu_j\left(\frac{1}{n_{1j}} + \frac{1}{n_{2j}}\right)(\alpha \beta_1 j - \alpha \beta_2 j) + \sum_{j=1}^{2} \bar{\varepsilon}_{12j}\right]
\]

If \( H_{0A} \) is true and \( \alpha_1 - \alpha_2 = 0 \), the p-value simplifies to:

\[
\Pr[\alpha^T A \geq \alpha^T A | y(0)] = \Pr\left[\sum_{j=1}^{2} \bar{\varepsilon}_{12j} \geq \sum_{j=2}^{2} \nu_j\left(\frac{1}{n_{1j}} + \frac{1}{n_{2j}}\right)(\alpha \beta_1 j - \alpha \beta_2 j) + \sum_{j=1}^{2} \bar{\varepsilon}_{12j}\right]
\]

Given that \( \Pr(\nu_j < 1) < 1 \) and \( \alpha_1 - \alpha_2 > 0 \) under the alternate hypothesis, we conclude:

\[
\Pr[\alpha^T A \geq \alpha^T A | y(\alpha)] \leq \Pr[\alpha^T A \geq \alpha^T A | y(0)]
\]

Therefore, the p-values under \( H_{1A} \) are no greater than the p-values under \( H_{0A} \), and the test statistic will correctly reject the null hypothesis at the confidence level set by \( \alpha \). A similar proof can be constructed to prove unbiasedness of \( \beta^T B \) and the proof for \( T_{AB} \) follows, because it depends on the partial test statistics for factors A and B.

### 4.7 Consistency

It is now straightforward to prove consistency for \( \alpha^T A \). I show consistency for the \( 2 \times J \) case, and the same logic can be extended to show consistency for all \( I \times J \) designs. A test statistic \( T \) is consistent if \( \lim_{N \to \infty} \Pr[T(y) \in R_{\alpha}(T)] = 1 \). From the proof of unbiasedness, \( \Pr[\alpha^T A | y(\alpha)] \) is:

\[
\Pr\left[\sum_{j=1}^{2} \bar{\varepsilon}_{12j} \geq \sum_{j=1}^{2} \nu_j\left(\frac{1}{n_{1j}} + \frac{1}{n_{2j}}\right)(\alpha_1 - \alpha_2) + \sum_{j=2}^{2} \nu_j\left(\frac{1}{n_{1j}} + \frac{1}{n_{2j}}\right)(\alpha \beta_1 j - \alpha \beta_2 j) + \sum_{j=1}^{2} \bar{\varepsilon}_{12j}\right]
\]

Again, simplify the notation for \( \nu_{isl} \) to \( \nu_j \) since \( I = 2 \). By assumption, \( E(\varepsilon_{ij}) = 0 \) and
V(\varepsilon_{ij}) = \sigma^2 < \infty$. The expectation of the permuted error terms is:

\[
E\left[ \frac{1}{N} \sum_{j=1}^{2} \tilde{\varepsilon}_{12lj} \right] = E_n\left[ E\left[ \frac{1}{N} \sum_{j=1}^{2} \tilde{\varepsilon}_{12lj} | n \right] \right] = 0
\]

Note that \(V(\tilde{\varepsilon}_{12lj}) = V(\varepsilon_{ij}^\ast - \varepsilon_{ij}^\ast) = V(\varepsilon_{ij}^\ast + \varepsilon_{ij}^\ast)\) and thus \(V(\sum_{j=1}^{2} \tilde{\varepsilon}_{12lj}) = V(\frac{1}{N} \sum_{i} \sum_{j} \sum_{k=1}^{2} \varepsilon_{ij}^\ast)\)

or more simply stated \(V(\tilde{\varepsilon}_{12lj}) = V(\frac{1}{N} \sum_{k=1}^{N} \varepsilon_{k})\). As \(N\) approaches infinity, this variance becomes zero:

\[
\lim_{N \to \infty} V\left[ \frac{1}{N} \sum_{k=1}^{N} \varepsilon_{k} \right] = 0
\]

Now I show:

\[
\lim_{N \to \infty} Pr\left[ \sum_{j=1}^{2} \nu_j \left( \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \right)(\alpha_1 - \alpha_2) + \sum_{j=2}^{2} \nu_j \left( \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \right)(\alpha_\beta_{1j} - \alpha_\beta_{2j}) > 0 \right] = 1
\]

The law of iterative expectations can be used to show that the expectation of sum is positive by first taking the expectation conditional on \(n\) as follows:

\[
E_n\left[ E\left[ \sum_{j=1}^{2} \nu_j \left( \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \right)(\alpha_1 - \alpha_2) + \sum_{j=2}^{2} \nu_j \left( \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \right)(\alpha_\beta_{1j} - \alpha_\beta_{2j}) | n \right] \right]
\]

Now the expectation over all values of \(n\) is:

\[
= E\left[ \sum_{j=1}^{2} n_{1j}n_{2j} \frac{n_{1j} + n_{2j}}{n_{1j}n_{2j}} (\alpha_1 - \alpha_2) + \sum_{j=2}^{2} n_{1j}n_{2j} \frac{n_{1j} + n_{2j}}{n_{1j}n_{2j}} (\alpha_\beta_{1j} - \alpha_\beta_{2j}) \right] = E_n[2(\alpha_1 - \alpha_2)]
\]

\[
= 2(\alpha_1 - \alpha_2)
\]

Thus, the expectation is positive given \(\alpha_1 - \alpha_2 > 0\) under the alternate hypothesis. As \(N \to \infty\), the variance approaches zero. This is shown through the decomposition of the variance as follows:

\[
V\left[ \sum_{j=1}^{2} \nu_j \left( \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \right)(\alpha_1 - \alpha_2) + \sum_{j=2}^{2} \nu_j \left( \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \right)(\alpha_\beta_{1j} - \alpha_\beta_{2j}) \right]
\]

\[
= V_n\left[ E\left[ \sum_{j=1}^{2} \nu_j \left( \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \right)(\alpha_1 - \alpha_2) + \sum_{j=2}^{2} \nu_j \left( \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \right)(\alpha_\beta_{1j} - \alpha_\beta_{2j}) | n \right] \right] + E_n\left[ V\left[ \sum_{j=1}^{2} \nu_j \left( \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \right)(\alpha_1 - \alpha_2) + \sum_{j=2}^{2} \nu_j \left( \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \right)(\alpha_\beta_{1j} - \alpha_\beta_{2j}) \right] \right]
\]
\[
\frac{1}{n_{2j}}(\alpha_1 - \alpha_2) + \sum_{j=2}^{2} \nu_j \left( \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \right) (\alpha \beta_{1j} - \alpha \beta_{2j}) | \mathbf{n} \right]\]
\[
= V_n \left[ 2(\alpha_1 - \alpha_2) \right] + E_n \left[ \sum_{j=1}^{2} \left( \frac{n_{1j}n_{2j}}{n_{1j} + n_{2j}} \right)^2 \frac{1}{n_{ij} + n_{sj} - 1} \left( \frac{n_{1j} + n_{2j}}{n_{1j}n_{2j}} \right)^2 (\alpha_1 - \alpha_2)^2 \right]
\]
\[
+ \sum_{j=2}^{2} \left( \frac{n_{1j}n_{2j}}{n_{1j} + n_{2j}} \right)^2 \frac{1}{n_{1j} + n_{2j} - 1} \left( \frac{n_{1j} + n_{2j}}{n_{1j}n_{2j}} \right)^2 (\alpha \beta_{1j} - \alpha \beta_{2j})^2 \right]
\]
\[
= \frac{2(\alpha_1 - \alpha_2)^2}{2Np - 1} + \frac{1}{2Np - 1} \sum_{j=1}^{2} (\alpha \beta_{1j} - \alpha \beta_{2j})^2
\]

and in the limit:
\[
\lim_{N \to \infty} \left\{ \frac{2(\alpha_1 - \alpha_2)^2}{2Np - 1} + \frac{1}{2Np - 1} \sum_{j=1}^{2} (\alpha \beta_{1j} - \alpha \beta_{2j})^2 \right\} = 0
\]

Thus, I have proven that \( \lim_{N \to \infty} \mathbb{P}(\sum_{j=1}^{2} \nu_j \left( \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \right) (\alpha_1 - \alpha_2) + \sum_{j=2}^{2} \nu_j \left( \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \right) (\alpha \beta_{1j} - \alpha \beta_{2j}) > 0 | \mathbf{n} \) = 1 since the expectation of the sum is constant and positive and its variance goes to zero in the limit. As a result, \( \lim_{N \to \infty} \mathbb{P}(T_A^* \geq T_A | y(\alpha)) = 0 \). This follows from the fact that the \( y_{ijk} \) are independent and identically distributed random variables that are not correlated with either \( \nu_j \) or \( \mathbf{n} \). Thus, as the number of observations increases, the p-value tends to zero when the alternate hypothesis is true. Thus, \( \lim_{N \to \infty} \mathbb{P}(T(y) \in R_\alpha(T)) = 1 \) and the test statistic is consistent. The sum of consistent tests is also consistent and thus \( T_A^* \) is consistent. The proof of the consistency of \( T_B^* \) uses similar logic, and the consistency of \( T_{AB}^* \) follows from the proofs of the two main factors.

### 4.8 Simulations

I conducted a series of simulations to compare the performance of the non-parametric permutation method against parametric ANOVA methods derived from regression estimates (see appendix for discussion of the link between ANOVA and regression). Each simulation consisted of 1000 runs, with 1000 permutations of the data conducted for each run. The examples include both \( 2 \times 2 \) and \( 2 \times 3 \) designs across a range of sample and effect sizes. Errors were sampled independently from four distributions: normal, exponential, and a normal mixture model. The distributions were standardized to have mean zero and variance \( \sigma^2 = 1 \).

Simulations were run for both balanced and unbalanced designs. Imbalance in the design was generated through the use of a simple randomization procedure. The total number of observations was fixed at \( N = I \times J \times n \) where \( n \) is the average number of replications per treatment group. For example, for a \( 2 \times 2 \) design with \( I = J = 2 \) and \( n = 5 \) the total
number of observations is $N = 20$. At the start of each run, units are assigned to treatment
groups sequentially whereby each unit is assigned at random independently of the other
units. Even though units have an equal probability of being assigned to any given treatment
group, some treatment groups end up with greater or fewer numbers of observations by
chance. On average, the experiments in the simulations have unbalanced designs, although
there are a few instances where the designs have equal numbers of observations per cell
by chance. Similar randomization procedures are a common source of imbalance in social
science experiments, where researchers have less control over experimental design and are
forced to rely on more rudimentary randomization procedures to make assignments.

For all the simulations the effect sizes are proportional to the standard deviation of the
error (i.e., $k\sigma = k$) and are set to satisfy the zero-sum constraints:

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_i \alpha \beta_{ij} = \sum_j \alpha \beta_{ij} = 0$$

There are four basic models tested:

1. The design is $2 \times 3$ under a model where $H_{0A}$, and $H_{0B}$ and $H_{0AB}$ are false. The effects
are $\alpha = (k, -k)$, $\beta = (-k, 0, k)$, and $\alpha \beta = (k, 0, -k, -k, 0, k)$

2. The design is $2 \times 3$ under a model where $H_{0A}$ is true, and $H_{0B}$ and $H_{0AB}$ are false.
The effects are $\alpha = (0, 0)$, $\beta = (-k, 0, k)$, and $\alpha \beta = (k, 0, -k, -k, 0, k)$

3. The design is $2 \times 2$ under a model where $H_{0A}$, and $H_{0B}$ and $H_{0AB}$ are false. The effects
are $\alpha = (k, -k)$, $\beta = (-k, k)$, and $\alpha \beta = (k, -k, -k, k)$

4. The design is $2 \times 2$ under a model where $H_{0A}$ is true, and $H_{0B}$ and $H_{0AB}$ are false.
The effects are $\alpha = (0, 0)$, $\beta = (-k, k)$, and $\alpha \beta = (k, -k, -k, k)$

Simulations were run for $k \in \{.2, .4, .6, .8\}$, i.e., for $k$ equal to .2 times the standard
deviation of the error, .4 times the standard deviation of the error, etc. Increasing the scalar
$k$ increases the effect size.

This configuration of models was designed to assess the behavior of the testing procedures
under a full range of cases, some more extreme than others. It may be counterintuitive to
imagine a model with interaction effects but where the effects of a main factor are zero as
is the case in models (2) and (5). Recall that $H_{0A}$ is an assumption about the marginal
effect of factor A and not the overall effect. Factor A does have an effect on the responses,
but this effect is entirely mediated through factor B. To make this clear, consider the $2 \times 2$

---

5 Designs were restricted to at least one observation per treatment group.

6 I also performed simulations for two other models: 1) The design is $2 \times 3$ under a model where $H_{0A}$
and $H_{0B}$ are false, and $H_{0AB}$ is true. The effects are $\alpha = (k, -k)$, $\beta = (-k, 0, k)$, and $\alpha \beta = (0, 0, 0, 0, 0, 0)$.
2) The design is $2 \times 2$ under a model where $H_{0A}$ and $H_{0B}$ are false, and $H_{0AB}$ is true. The effects are
$\alpha = (k, -k)$, $\beta = (-k, k)$, and $\alpha \beta = (0, 0, 0, 0)$. There were no significant differences in performance
between the non-parametric and parametric methods. In the interest of space, I do not include them in the
paper.
‘get out the vote’ experiment. Suppose, in the true model, $\eta = 0$, $\alpha = (0, 0)$, $\beta = (-.5,.5)$, $\alpha_1 \beta = (.5,-.5)$ and $\alpha_2 \beta = (-.5,.5)$. The errors are independently and identically distributed according to some unknown distribution with expectation zero (See Table 4.4).

Table 4.4: The true model for a hypothetical $2 \times 2$ ‘get-out-the-vote’ experiment

<table>
<thead>
<tr>
<th></th>
<th>No phone call</th>
<th>Phone call</th>
</tr>
</thead>
<tbody>
<tr>
<td>No in-person visit</td>
<td>$y_{11k} = 0 - .5 + .5 + \epsilon_{11k}$</td>
<td>$y_{12k} = 0 + .5 - .5 + \epsilon_{12k}$</td>
</tr>
<tr>
<td>In-person visit</td>
<td>$y_{21k} = 0 - .5 - .5 + \epsilon_{21k}$</td>
<td>$y_{22k} = 0 + .5 + .5 + \epsilon_{22k}$</td>
</tr>
</tbody>
</table>

What circumstances in the real world could give rise to such a model? Suppose individuals become engaged in the election after speaking with a volunteer in person. They have a candidate whom they support, but are unsure if he will win the election. If these individuals then receive a phone call from the candidate’s campaign, they are reassured that their candidate is strong and has financial resources and support from the community. Believing that the candidate has a real chance of winning, they are more likely to turn out for the election. If on the other hand, these individuals do not receive a call from the campaign, they may take this omission as a sign that their candidate is weak. These individuals believe their cause is lost and are less likely to turn out to vote than they would have been had they not received the initial home visit. Results from experiments on voter turnout show that this model does not describe how registered voters generally behave in elections, but the example provides some intuition about behavior that could in principle occur and helps to motivate the models used in the simulation exercises.

I discuss only a selection of results from the simulation study here and include a fuller range of the results in the Appendix. In particular, I focus on examples where the distribution of the errors is normal, since this condition represents the best case scenario for linear regression based ANOVA estimates. In general, non-parametric methods are preferred to parametric methods because they do not assume a particular, parametric distribution such as a normal distribution. The data need only to be identically and independently distributed according to some unknown distribution. It is well known that linear regression estimates are optimal among all linear unbiased estimators when the normality and independence assumptions are met. However, problems arise when using regression estimates in ANOVA to construct F-tests. The parametric tests for main factor and interaction effects are positively correlated and thus are sensitive to interaction effects (Pesarin and Salmaso 2010). In addition, the standard adjustments to parametric ANOVA for unbalanced designs make new assumptions on the data, which worsen their ability to differentiate between main factor and interaction effects (Langsrud 2003).

Figures D.1- D.4 display the results of the simulations comparing the permutation procedure to regression for balanced designs with normal distribution of the errors. The graphs compare the power of the two methods to correctly reject (or accept) the null hypothesis...
when it is false (or true) over a range of effect sizes from \( .2\sigma \) to \( .8\sigma \). The number of observations per cell ranges from 5 to 10. The y-axis plots the percentage of p-values that fall below the .05 significance level and thus measures the percentage of simulations in which one would reject the null hypothesis at the standard 95% confidence level. In general, there is no significant difference in power between the non-parametric and parametric methods even though the errors are normally distributed. In fact, the non-parametric method outperforms regression for designs with greater than three levels when there are interaction effects. Regression tests are less effective at correctly rejecting the null hypothesis for interaction effects when effect sizes are small relative to error variance. Remarkably, this remains true as the number of observations increases.

Figures D.5- D.8 display similar simulations, but in this case the design is unbalanced (i.e. the number of observations per cell varies). This is a similar pattern of results as compared to the balanced designs, however the ability of regression to differentiate between main factor and interaction effects worsens. The non-parametric method performs equally well to regression and in some circumstances outperforms regression. Again, when the design has more than three levels for one treatment factor, regression does a poorer job at detecting relatively small interaction effects. In addition, regression incorrectly rejects the null hypothesis when it is true when there are interaction effects. Again, both problems remain as the number of observations grows. Under models (2) and (4) when the marginal effect of a treatment is zero, but the interaction effect is positive, regression appears less able to parse out interaction effects from the marginal effect of the main factor. This model represents an extreme example of heterogeneous treatment effects, where the effect of one factor depends only on the other main factor.

The identification of interaction effects is a critical goal of ANOVA. The results of ANOVA are used to guide subsequent estimation of treatment effects and confidence intervals. If the analyst assumes there are no interaction effects when they exist will bias estimates of the main effects. The simulations suggest that traditional parametric ANOVA methods may be deficient in identifying interaction effects, as well as differentiating them from estimates of marginal main effects. This problem worsens in experiments with unbalanced designs where the number of levels for a main factor is greater than two. The permutation tests were better at distinguishing the marginal effect of factor B from effects associated with its interaction with factor A. In other cases, the non-parametric method performed similarly to the parametric method in terms of power.

These simulations compare the non-parametric permutation method to regression-based ANOVA tests when the errors are normally distributed. In the Appendix, I include the results from models where the errors are generated from exponential and normal mixture models. The results are comparable. Theoretically, regression should perform worse when the errors do not have normal distributions and the size of the experiment is small (fewer than 10 observations per cell). That regression performs no worse than under the normal model may reflect the fact that the errors in the simulations are generated from standard well-behaved distributions, but this requires additional research. In a few cases, regression outperforms the non-parametric method when the number of observations is 5, but these
may be random outcomes. It is well known that p-values derived from theoretical probability distributions may not hold, especially in small samples (Bowers and Panagopoulos 2009).

By comparison, with permutation tests the minimum achievable p-value is solely a function of the number of observations and the permutation procedure. Permutation methods can often produce exact p-values in small samples because it is possible to carry out all permutations of the data. This difference between the two tests is reflected in their distributions of p-values near zero. The parametric tests report extremely low p-values at a frequency unobtainable by the permutation tests, whose p-values are limited by the size of the data. The advantage of parametric procedures is that they overcome issues related to data sparsity and can generate estimates even when some treatment groups have relatively few observations. But their ability to make inferences depends on normality assumptions that may not be justified in small samples. By design, the power of permutation tests is limited by the size of the sample and cannot test hypotheses in the absence of data. In addition, permutation tests depend on the design of the permutation procedure.

The fact that inferences can be made in most cases, even with small sample sizes, is an advantage of my permutation procedure over previous methods. Basso et al. (2009) fix the number of exchanges across all cells for each iteration of the permutation procedure, whereas under my method allows the number of exchanges to vary between each paired exchange between cells. Basso et al.’s approach allows them to construct test statistics for the main factors and interaction effects that rely solely on the parameter of interest (i.e., $\alpha^T_A$ is a function of only $\alpha_i$ and not $\beta_j$ nor $\alpha\beta_{ij}$). However, fixing the number of exchanges to be common across cells comes at a great cost, particularly when the design is unbalanced. In simulations where the average cell size was ten, the minimum cell size was invariably less than ten and no greater than 5 in 30 percent of the cases for a $2 \times 3$ design. For instance, if the smallest cell contains 5 observations, the number of exchanges could only vary between 0 and 5 in their method, whereas in my method the maximum number of exchanges is only limited to five in exchanges that involve the smallest cell. This change, enabling greater use of the data, enhances the power of the permutation method while maintaining the uncorrelatedness and consistency of the tests.
Figure 4.1: The simulations employ a balanced $2 \times 3$ design under the model where $H_{0A}$, and $H_{0B}$ and $H_{0AB}$ are false and the errors are distributed **standard normal**. Effect sizes are scaled to between $0.2\sigma$ and $0.8\sigma$, while the number of observations per cell ranges from 5 to 30. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.
Figure 4.2: The simulations employ a balanced $2 \times 3$ design under the model where $H_{0A}$ is true, and $H_{0B}$ and $H_{0AB}$ are false and the errors are distributed standard normal. Effect sizes are scaled to between $2\sigma$ and $8\sigma$, while the number of observations per cell ranges from 5 to 30. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.

```
0.2\sigma  0.4\sigma  0.6\sigma  0.8\sigma
```

```
\begin{table}
\begin{tabular}{cccc}
\hline
$H_{0A}$ & $H_{0B}$ & $H_{0AB}$ & power \\
\hline
\end{tabular}
\end{table}
```
Figure 4.3: The simulations employ a balanced $2 \times 2$ design under the model where $H_{0A}$, and $H_{0B}$ and $H_{0AB}$ are false and the errors are distributed **standard normal**. Effect sizes are scaled to between $.2\sigma$ and $.8\sigma$, while the number of observations per cell ranges from 5 to 30. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.

![Graph showing power vs number of observations for different effect sizes](image-url)
Figure 4.4: The simulations employ a balanced $2 \times 3$ design under the model where $H_{0A}$ is true, and $H_{0B}$ and $H_{0AB}$ are false and the errors are distributed standard normal. Effect sizes are scaled to between $0.2\sigma$ and $0.8\sigma$, while the number of observations per cell ranges from 5 to 30. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.
Figure 4.5: The simulations employ an **unbalanced** $2 \times 3$ design under the model where $H_{0A}$, and $H_{0B}$ and $H_{0AB}$ are false and the errors are distributed **standard normal**. Effect sizes are scaled to between $.2\sigma$ and $.8\sigma$, while the number of total observations varies from 60, 120, and 180 with approximately 10, 20 and 30 observations per cell. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.
Figure 4.6: The simulations employ an **unbalanced** $2 \times 3$ design under the model where $H_{0A}$ is true, and $H_{0B}$ and $H_{0AB}$ are false and the errors are distributed **standard normal**. Effect sizes are scaled to between $.2\sigma$ and $.8\sigma$, while the number of total observations varies from 60, 120, and 180 with approximately 10, 20 and 30 observations per cell. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.
Figure 4.7: The simulations employ an unbalanced $2 \times 2$ design under the model where $H_{0A}$, and $H_{0B}$ and $H_{0AB}$ are false and the errors are distributed standard normal. Effect sizes are scaled to between $.2\sigma$ and $.8\sigma$, while the number of total observations varies from 40, 80, and 130 with approximately 10, 20 and 30 observations per cell. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.
Figure 4.8: The simulations employ an unbalanced $2 \times 2$ design under the model where $H_{0A}$ is true, and $H_{0B}$ and $H_{0AB}$ are false and the errors are distributed standard normal. Effect sizes are scaled to between $.2\sigma$ and $.8\sigma$, while the number of total observations varies from 40, 80, and 130 with approximately 10, 20 and 30 observations per cell. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.
4.9 Conclusion

The permutation procedure described here provides a useful tool for analyzing two-way regular factorial experiments. The method is non-parametric and therefore does not rely on the normality assumptions required for parametric ANOVA methods. It contributes to existing permutation methods for factorial experiments by taking greater advantage of the data through a more flexible permutation procedure. It is well suited to experiments of unbalanced design (where the imbalance is random). In addition, it can accommodate relatively large experiments comprised of factors with more than two levels and consisting of many observations with relative ease.

Monte Carlo simulations comparing the permutation procedure to traditional ANOVA tests demonstrate that the permutation method is superior to regression-based ANOVA methods at identifying interaction effects when factors have more than two levels. The improvements were even better when the design is unbalanced. Further power simulations are required to assess the behavior of the permutation procedure versus parametric ANOVA methods when errors have less well-behaved distributions than the ones used here.

The permutation procedure could be extended to cover all regular factorial designs, with the caveat that estimation of every interaction effect is feasible only if the number of observations by paired cells is sufficiently large. Similar methods could be developed for hierarchical designs and extended to the realm of Bayesian analysis. Extension to hierarchical designs would require assumptions that higher order interactions in the hierarchy are negligible, but the same assumptions are required for parametric ANOVA methods because this problem is one of sparse data. The permutation procedure could be tested against Bayesian analysis approaches and, more broadly, could be used in conjunction with Bayesian methods to estimate confidence intervals or conduct subgroup analyses as part of a comprehensive analysis plan (Gelman et al. 2005; Imai and Ratkovic 2013).
Bibliography


Brady, David and Tammy Frisby. 2011. “Do Americans Favor a Flat Tax?”


URL: www.stat.columbia.edu/ gelman


URL: www.columbia.edu/cu/economics/graduate/jobmarket/olga.html


Healy, Andrew and Gabriel S Lenz. 2012. “Substituting the End for the Whole: Why Voters Respond Primarily to the Election-Year Economy.”.


**URL:** [http://gatton.uky.edu/faculty/ziliak/Volatility_020810_all.pdf](http://gatton.uky.edu/faculty/ziliak/Volatility_020810_all.pdf)
Appendix A

Map of vignette portion of the survey
Figure A.1: The following decision tree maps the adaptive question portion of the survey, where respondents are asked to imagine themselves to be part of a household. In period one, respondents select a tax rate they believe is fair given the household’s income that year. In period two, they are told the household’s income for year two and asked if the household should pay the same or a higher or lower amount in taxes in year two as in year one. If they answer “the same”, respondents skip to period three. If they answer “higher” or “lower”, they select the new rate. The survey then confirms whether this percentage is in fact higher/lower than the rate provided in period one. If the answer is consistent, the respondent skips to period three. If the answer is inconsistent, then the respondent is asked whether he would like a chance to change his previous answers. If the respondent indicates “yes”, then the respondent is allowed to select new tax rates for period one and two. The same process is repeated again in period three.
Appendix B

Survey questions

Introduction
For many years, US family incomes were very stable. More recently, family incomes move around more from year to year. The amount of money you pay in taxes is based solely on the earnings you made over the past year regardless of how much was earned in previous years. In other words, the tax rate this year does not depend on whether you earned more or less income two years ago, or expect to earn more or less income in the upcoming years. The next questions ask what you think the percentage in federal income tax rate should be when income is higher or lower in different years.

Question 1
Suppose that your income is $60,000 this past year. What do you think is a fair amount to pay in federal taxes? Please enter a percentage between 0 and 100 in the box below.

\[ \% \]

---Page 2---

<table>
<thead>
<tr>
<th>Year</th>
<th>Income</th>
<th>Tax Rate</th>
<th>Tax Paid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>$60000</td>
<td>X%</td>
<td>${60000X}</td>
</tr>
</tbody>
</table>

Question 2A
The next year you earn $\{45,000\text{ or }55,000\}. Do you think it is fair to pay the same percentage, a higher percentage, or a lower percentage in taxes as you did the previous year when you earned $60,000?

○ same \( \Rightarrow \) Go to page 4
○ higher
○ lower
→ If answered higher or lower, the following question appears on this page:

**Question 2B**
What percentage in taxes do you think is a fair rate assuming you earned ${45,000 or 55,000}$ over the past year? Please enter a number between 0 and 100 in the box below.

- Enter a number between 0 and 100 in the box below.
- If percentage is consistent with answer to question 2A, go to page 4.
- If percentage is NOT consistent with answer to question 2A, go to page 3.

---

---Page 3---

You indicated that it is fair to pay a higher/lower percentage, but you entered a lower/higher percentage. Specifically, you indicated that the tax rate should be X% in the first year and Y% in the second year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Income</th>
<th>Tax Rate</th>
<th>Tax Paid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>$60000</td>
<td>X%</td>
<td>$60000X</td>
</tr>
<tr>
<td>Year 2</td>
<td>${45000 or 55000}</td>
<td>Y%</td>
<td>${45000Y or 55000Y}</td>
</tr>
</tbody>
</table>

**Question 3A**
Is this correct?
- Yes → Go to page 4
- No → If no, the following questions appear on this page:

**Question 3B**
In year one, you earn $60,000. What is a fair percentage in taxes to pay for this income? Please enter a new percentage in the box if you would like to change your initial response.

- Enter a new percentage in the box if you would like to change your initial response.

**Question 3C**
In year two, you earn ${45,000 or 55,000}. What is a fair percentage in taxes to pay for this income? Please enter a new percentage in the box if you would like to change your initial response.

---Page 4---

<table>
<thead>
<tr>
<th>Year</th>
<th>Income</th>
<th>Tax Rate</th>
<th>Tax Paid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>$60000</td>
<td>X%</td>
<td>$60000X</td>
</tr>
<tr>
<td>Year 2</td>
<td>${45000 or 55000}</td>
<td>Y%</td>
<td>${45000Y or 55000Y}</td>
</tr>
</tbody>
</table>
Question 4A
The next year you earn $\{75,000 \text{ or } 65,000\}$. Do you think it is fair to pay the same percentage, a higher percentage, or a lower percentage in taxes as you did the previous year when you earned $\{45,000 \text{ or } 55,000\}$?

- same → Go to page 6
- higher
- lower → If answered higher or lower, the following question appears on this page:

Question 4B
What percentage in taxes do you think is a fair rate assuming you earned $\{75,000 \text{ or } 65,000\}$ over the past year? Please enter a number between 0 and 100 in the box below.

\[
\boxed{\text{%}}
\]

→ If percentage is consistent with answer to question 4A, go to page 6.
→ If percentage is NOT consistent with answer to question 4A, go to page 5.

---Page 5---

You indicated that it is fair to pay a higher/lower percentage, but you entered a lower/higher percentage. Specifically, you indicated that the tax rate should be Y% in the second year and Z% in the third year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Income</th>
<th>Tax Rate</th>
<th>Tax Paid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>$60000</td>
<td>X%</td>
<td>${60000X}$</td>
</tr>
<tr>
<td>Year 2</td>
<td>${45000 \text{ or } 55000}$</td>
<td>Y%</td>
<td>${45000Y \text{ or } 55000Y}$</td>
</tr>
<tr>
<td>Year 2</td>
<td>${75000 \text{ or } 65000}$</td>
<td>Z%</td>
<td>${75000Z \text{ or } 65000Z}$</td>
</tr>
</tbody>
</table>

Question 5A
Is this correct?

- Yes → Go to page 6
- No → If no, the following questions appear on this page:

Question 5B
In year one, you earn $60,000. What is a fair percentage in taxes to pay for this income? Please enter a new percentage in the box if you would like to change your initial response.

\[
\boxed{\text{%}}
\]

Question 5C
In year two, you earn $\{45,000 \text{ or } 55,000\}$. What is a fair percentage in taxes to pay for this income? Please enter a new percentage in the box if you would like to change your initial response.

\[
\boxed{\text{%}}
\]
Question 5D
In year three, you earn $\{75,000 \text{ or } 65,000\}$. What is a fair percentage in taxes to pay for this income? Please enter a new percentage in the box if you would like to change your initial response.

\[ Z \text{%} \]

--- Page 6 ---

<table>
<thead>
<tr>
<th>Year</th>
<th>Income</th>
<th>Tax Rate</th>
<th>Tax Paid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>$60000</td>
<td>X%</td>
<td>${60000X}$</td>
</tr>
<tr>
<td>Year 2</td>
<td>${45000 \text{ or } 55000}$</td>
<td>Y%</td>
<td>${45000Y \text{ or } 55000Y}$</td>
</tr>
<tr>
<td>Year 3</td>
<td>${75000 \text{ or } 65000}$</td>
<td>Z%</td>
<td>${75000Z \text{ or } 65000Z}$</td>
</tr>
</tbody>
</table>

The chart above shows the tax rates and total amount of taxes paid based on the rates that you selected in the previous questions. These questions have asked what you think the federal tax rate for the SAME household should be if its income changes from year to year. The next question asks what you think the federal tax rate should be for DIFFERENT households who have different incomes in the same year.

--- Page 7 ---

Question 7
For a household at each of the following income levels, what percentage of their income should go towards federal income taxes:

**Tax Rate (0-100%)**

- $20,000
- $40,000
- $60,000
- $100,000
- $200,000
Appendix C

Construction of ANOVA tests from OLS regression estimates

The purpose of this appendix is to demonstrate how to construct standard ANOVA tables from linear regression estimates.

C.1 One-way ANOVA

Consider the following experiment where you have \( J \) treatments. The experiment is a replicated design, meaning that there are multiple observations per treatment. The number of observations per treatment is denoted \( n_j \) such that \( N = \sum_{j=1}^{J} n_j \), i.e., the sum of observations across treatments equals the total number of observations in the experiment. The experiment considers the effect of six different types of feed on the weight of chicks. Thus, \( J = 6 \). The dataset contains two variables, the type of feed and the weight of the chick, which respectively represent the independent and dependent variables. The design is unbalanced because the number of observations differs across treatments.

To construct an ANOVA table one needs the degrees of freedom associated with the treatment and residual, as well as two quantities:

1. The explained sum of squares (ESS)
2. The residual sum of squares (RSS)

All other values in the ANOVA table depend on these four quantities as shown in Table C.1.

An ANOVA table can be generated from linear regression regression estimates. First, consider the explained sum of squares (ESS). An alternative characterization based on coefficient estimates from linear regression is:

\[
ESS = \sum_{j=1}^{(J-1)} \left\{ n_j (\hat{\beta}_0 + \hat{\beta}_j - \bar{y})^2 \right\} + n_J (\hat{\beta}_0 - \bar{y})^2 = \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2
\]
Table C.1: One-way ANOVA table

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>Pr(&gt; F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>J - 1</td>
<td>$E_{\text{SS}} = \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2$</td>
<td>$E_{\text{SS}} / (J-1)$</td>
<td>$E_{\text{SS}} / (J-1)$</td>
</tr>
<tr>
<td>Residual</td>
<td>N - J</td>
<td>$R_{\text{SS}} = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$</td>
<td>$R_{\text{SS}} / (N-J)$</td>
<td>$R_{\text{SS}} / (N-J)$</td>
</tr>
</tbody>
</table>

The reference variable in the linear regression is set to treatment $J$. This is the variable dropped when running the regression and represented by the intercept term. To understand why these two equations are equivalent, consider $\hat{y}_i$. Suppose individual $i$ receives treatment $j = 1$, then $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots = \hat{\beta}_0 + \hat{\beta}_1$. All the other terms drop out of the equation since $x_1 = 1$ and the other $x$’s are zero. For individuals assigned to treatment $J$, $\hat{y}_i$ is simply equal to $\hat{\beta}_0$. The value of $\hat{y}_i$ is the same for all individuals receiving the same treatment, and thus:

$$
\sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n_j} (\hat{y}_{ij} - \bar{y})^2 = n_j (\hat{y}_j - \bar{y})^2 = n_1 (\hat{\beta}_0 + \hat{\beta}_1 - \bar{y})^2 + \ldots + n_J (\hat{\beta}_0 - \bar{y})^2
$$

Now consider the residual sum of squares (RSS). It is the square of the residual standard error (denoted $s$) multiplied by the degrees of freedom for the residual $(N - J)$. The formula for the residual standard error is:

$$s = \left( \frac{1}{N-J} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \right)^{\frac{1}{2}}$$

Thus:

$$RSS = (N - J)s^2$$

C.2 Two-way ANOVA

The conversion from linear regression to ANOVA becomes more complicated for a two-way design. In a two-way layout there are two treatments and an interaction term. Each treatment has multiple levels. Suppose treatment A has $I$ levels and treatment B has $J$ levels. The interaction term is denoted $A \times B$.

To demonstrate the equivalence between linear regression and ANOVA with a $2 \times 2$ design, with $I = 2$ and $J = 2$, assume that the design is balanced, i.e., there are equal numbers of observations per treatment group such that $N = nIJ$. To construct an ANOVA table one needs the degrees of freedom associated with the treatments, interaction term, and residual and four quantities:
C.2. TWO-WAY ANOVA

1. The sum of squares associated with treatment $A$
2. The sum of squares associated with treatment $B$
3. The sum of squares associated with the interaction term $A \times B$
4. The residual sum of squares (RSS)

All other values in the ANOVA table depend on these four quantities as shown in Table C.2.
### C.2. TWO-WAY ANOVA

**Table C.2: Two-way ANOVA table**

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>Pr(&gt; F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(I - 1)</td>
<td>(SS_A = nJ \sum_{i=1}^{I} (\bar{y}<em>{i..} - \bar{y}</em>{..})^2)</td>
<td>(MS_A = \frac{SS_A}{(I-1)})</td>
</tr>
<tr>
<td>B</td>
<td>(J - 1)</td>
<td>(SS_B = nI \sum_{j=1}^{J} (\bar{y}<em>{..j} - \bar{y}</em>{..})^2)</td>
<td>(MS_B = \frac{SS_B}{(J-1)})</td>
</tr>
<tr>
<td>A × B</td>
<td>((I - 1)(J - 1))</td>
<td>(SS_{A×B} = n \sum_{i=1}^{I} \sum_{j=1}^{J} (\bar{y}<em>{ij} - \bar{y}</em>{i..} - \bar{y}<em>{..j} + \bar{y}</em>{..})^2)</td>
<td>(MS_{A×B} = \frac{SS_{A×B}}{(J-1)})</td>
</tr>
<tr>
<td>Residual</td>
<td>(IJ(N - 1))</td>
<td>(RSS = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2)</td>
<td>(RMS = \frac{RSS}{IJ(N - 1)})</td>
</tr>
</tbody>
</table>
The design matrix for a linear regression will typically use baseline constraints. The following is an example of the design matrix for a $2 \times 2$ factorial experiment with $N = 8$ observations and $n = 2$ replicates per treatment combination:

$$
X_{8,4} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad y_{8,1} = \begin{pmatrix} y_{1,1,1} \\ y_{1,1,2} \\ y_{1,2,1} \\ y_{1,2,2} \\ y_{2,1,1} \\ y_{2,1,2} \\ y_{2,2,1} \\ y_{2,2,2} \end{pmatrix}
$$

The design matrix generates the estimates for the regression coefficients via the well-known equation $\hat{\beta}_{IJ,1} = (X'X)^{-1}X'y$. Applying the equation, the parameter estimates are:

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} y_{2,2,i} \\ \frac{1}{n} \sum_{i=1}^{n} y_{1,2,i} - \frac{1}{n} \sum_{i=1}^{n} y_{2,2,i} \\ \frac{1}{n} \sum_{i=1}^{n} y_{2,1,i} - \frac{1}{n} \sum_{i=1}^{n} y_{2,2,i} \\ \frac{1}{n} \sum_{i=1}^{n} y_{1,1,i} + \frac{1}{n} \sum_{i=1}^{n} y_{2,2,i} - \frac{1}{n} \sum_{i=1}^{n} y_{1,2,i} - \frac{1}{n} \sum_{i=1}^{n} y_{2,1,i} \end{pmatrix}$$

From these calculations it follows:

- $\sum_{i=1}^{n} y_{2,2,i} = \bar{y}_{22} = \hat{\beta}_0$
- $\sum_{i=1}^{n} y_{1,2,i} = \bar{y}_{12} = \hat{\beta}_0 + \hat{\beta}_1$
- $\sum_{i=1}^{n} y_{2,1,i} = \bar{y}_{21} = \hat{\beta}_0 + \hat{\beta}_2$
- $\sum_{i=1}^{n} y_{1,1,i} = \bar{y}_{11} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$

Now the regression coefficients can be substituted into the equations for the sum of squares used in the ANOVA table. To make the process clearer, rewrite the sum of squares equation for treatment A as:

$$SS_A = nJ \sum_{i=1}^{T} (\bar{y}_{i.} - \bar{y}_{..})^2 = \frac{n}{4} \left( \frac{1}{n} \sum_{i=1}^{2n} y_{2,i} - \frac{1}{n} \sum_{i=1}^{2n} y_{1,i} \right)^2 = \frac{n}{4} \left( \frac{1}{n} \sum_{i=1}^{n} y_{21,i} + \frac{1}{n} \sum_{i=1}^{n} y_{22,i} - \frac{1}{n} \sum_{i=1}^{n} y_{11,i} - \frac{1}{n} \sum_{i=1}^{n} y_{12,i} \right)^2 = \frac{n}{4} (\hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_0 - \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 - \hat{\beta}_0 + \hat{\beta}_1)^2 = \frac{n}{4} (2\hat{\beta}_1 + \hat{\beta}_3)^2$$
Similarly:

\[ SS_B = \frac{n}{4} (2\hat{\beta}_2 + \hat{\beta}_3)^2 \]

and

\[ SS_{A \times B} = \frac{n}{4} \hat{\beta}_3^2 \]

As in the one-way ANOVA example, the RSS is the square of the residual standard error (s) multiplied by the degrees of freedom for the residual \((N - J)\). Thus:

\[ RSS = (N - J)s^2 \]

Calculations for larger factorial designs with multiple levels and multiple treatments becomes complicated quickly and is not pursued here. The equations above work for a balanced \(2 \times 2\) factorial design. If there are varying numbers of observations per treatment combination, i.e., \(n_{11} \neq n_{12} \neq n_{21} \neq n_{22}\), adjustments need to be made to the conversion equations.
Appendix D

Additional simulation results
Figure D.1: The simulations employ a balanced $2 \times 3$ design under the model where $H_{0A}$, and $H_{0B}$ and $H_{0AB}$ are false and the errors have a standardized exponential distribution. Effect sizes are scaled to between $.2\sigma$ and $.8\sigma$, while the number of observations per cell ranges from 5 to 30. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.
Figure D.2: The simulations employ a balanced $2 \times 3$ design under the model where $H_{0A}$ is true, and $H_{0B}$ and $H_{0AB}$ are false and the errors have a standardized exponential distribution. Effect sizes are scaled to between $.2\sigma$ and $.8\sigma$, while the number of observations per cell ranges from 5 to 30. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.
Figure D.3: The simulations employ a balanced $2 \times 2$ design under the model where $H_{0A}$, and $H_{0B}$ and $H_{0AB}$ are false and the errors have a standardized **exponential** distribution. Effect sizes are scaled to between $0.2\sigma$ and $0.8\sigma$, while the number of observations per cell ranges from 5 to 30. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.
Figure D.4: The simulations employ a balanced $2 \times 3$ design under the model where $H_{0A}$ is true, and $H_{0B}$ and $H_{0AB}$ are false and the errors have a standardized exponential distribution. Effect sizes are scaled to between $0.2\sigma$ and $0.8\sigma$, while the number of observations per cell ranges from 5 to 30. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.
Figure D.5: The simulations employ an **unbalanced** $2 \times 3$ design under the model where $H_{0A}$, and $H_{0B}$ and $H_{0AB}$ are false and the errors have a standardized **exponential** distribution. Effect sizes are scaled to between $0.2\sigma$ and $0.8\sigma$, while the number of total observations varies from 60, 120, and 180 with approximately 10, 20 and 30 observations per cell. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.

- $H_{1A}$
- $H_{1B}$
- $H_{1AB}$
- Power
- Number of Observations
Figure D.6: The simulations employ an unbalanced $2 \times 3$ design under the model where $H_{0A}$ is true, and $H_{0B}$ and $H_{0AB}$ are false and the errors have a standardized exponential distribution. Effect sizes are scaled to between $0.2\sigma$ and $0.8\sigma$, while the number of total observations varies from 60, 120, and 180 with approximately 10, 20 and 30 observations per cell. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.
Figure D.7: The simulations employ an unbalanced $2 \times 2$ design under the model where $H_{0A}$, and $H_{0B}$ and $H_{0AB}$ are false and the errors have a standardized exponential distribution. Effect sizes are scaled to between .2σ and .8σ, while the number of total observations varies from 40, 80, and 130 with approximately 10, 20 and 30 observations per cell. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.
Figure D.8: The simulations employ an **unbalanced** $2 \times 2$ design under the model where $H_{0A}$ is true, and $H_{0B}$ and $H_{0AB}$ are false and the errors have a standardized **exponential** distribution. Effect sizes are scaled to between $.2\sigma$ and $.8\sigma$, while the number of total observations varies from 40, 80, and 130 with approximately 10, 20 and 30 observations per cell. The red dotted line displays the results from the non-parametric method, while the blue dotted line displays the ANOVA estimates from regression.