

ESSAYS IN LABOR ECONOMICS AND INDUSTRIAL ORGANIZATION

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Stephen Michael Ciccarella

January 2013

ESSAYS IN LABOR ECONOMICS AND INDUSTRIAL ORGANIZATION

Stephen Michael Ciccarella, Ph.D.

Cornell University 2013

The first essay of my dissertation examines whether educational institutions respond to the threat of financial sanctions as the result of student underperformance on high-stakes testing. Federal education legislation passed in 2001 required states to implement a series of annual examinations, with the goal of having all students from a broad set of demographic and socioeconomic categories attain proficiency on those exams by 2014. Schools whose students did not meet proficiency cutoffs in mathematics and English for a single year were threatened with a set of financial sanctions, such as paying for transportation costs for students who transfer from the school. Since assignment to treatment (exposure to sanctions) is a discontinuous function of students' prior exam performance, a regression discontinuity design allows one to estimate the treatment effect of the threat of sanctions on future exam performance. Using a panel of school-level data for elementary school students in New Jersey and California, the econometric results indicate that in both states there is no statistically significant evidence that the threat of sanctions affects elementary student performance in the subsequent school year.

The second essay contributes to the demand estimation literature by applying a discrete choice analysis to the demand for wine in the United States. Wine consumption in the United States has increased continuously over the last twenty years, with the value of total wine sales increasing by more than 200% over this time period. The goal of the essay is to understand better the demand for wine by investigating wine consumers' preferences and decision-making processes. This is achieved by implementing a nested logit model of consumer

demand that takes into consideration product differentiation in the wine industry. Using nesting structures based on wine quality and origin, the correlation of individual tastes within constructed categories are estimated. Under a quality nesting structure, price and varietal offerings have a strong impact on the total brand market share relative to the share of an outside good. In addition, American wine drinkers who segment along quality tend to prefer European wines, and exhibit high correlation of tastes at both extremes of the quality spectrum. The origin-based nesting structure indicates similar effects from prices and varietals, but at a lower magnitude. Wine drinkers who segment based on origin have strongly heterogeneous preferences towards domestic wines.

The final essay is a methodological implementation of an imputation model that will be used by the U.S. Census Bureau to assign quarterly hours and earnings to part-time federal workers in the Longitudinal Employer-Household Dynamics (LEHD) programs's infrastructure files. An imputation procedure is necessary because the LEHD infrastructure source data provided by the Office of Personnel Management reports only the annualized salary for these workers and does not include information on the number of hours worked. The Bayesian statistical method implemented in the paper specifies a prior distribution and likelihood function for several demographic characteristics of federal workers. These characteristics, along with hours worked, can be identified in Current Population Survey data. A Dirichlet posterior predictive distribution is derived and used to generate cell probabilities for each combination of characteristics. Finally, draws are made from the posterior distribution to assign quarterly hours and earnings to part-time workers. A test implementation of the imputation strategy is performed, producing internally consistent results. The results are also compared to external public use data.

© 2013 Stephen Michael Ciccarella

ALL RIGHTS RESERVED

BIOGRAPHICAL SKETCH

Stephen Ciccarella holds a B.S.Econ. from the Wharton School, University of Pennsylvania and an M.P.A. from Columbia University. Before earning his Ph.D. in economics at Cornell University, he held a staff position on Capitol Hill and worked as a researcher at a think tank. He currently works in the financial services industry in New York City.

ACKNOWLEDGMENTS

I'd like to pay special thanks to my committee chair, John Abowd, for his constant support and guidance, and to Rajeev Dehejia for his mentorship and advice. I'd also like to thank Ron Ehrenberg and Claudio Lucarelli for serving on my committee and providing helpful comments on my research. Finally, I thank my wife and my family for their support.

Chapter three of this dissertation uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by the National Science Foundation Grants SES-9978093, SES-0339191, and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation. Any opinions and conclusions expressed are my own and do not necessarily represent the views of the U.S. Census Bureau, its program sponsors, or data providers. All results have been reviewed to ensure that no confidential information is disclosed.

TABLE OF CONTENTS

Biographical Sketch	iii
Acknowledgments	iv
Table of Contents	v
List of Tables	vii
List of Figures	viii
1 The Treatment Effect of Financial Sanctions on Low-Performing Schools: A Regression Discontinuity Approach	1
1.1 Abstract	1
1.2 Introduction	2
1.3 NCLB-based School Accountability in New Jersey and California	5
1.4 Data and Variable Construction	7
1.4.1 Assignment Variable Construction	9
1.4.2 Outcome Variable Construction	10
1.5 The Regression Discontinuity Approach	11
1.5.1 Assessing the Validity of the RD Design for the New Jersey and California Data	11
1.5.2 Identification of the Treatment Effect	12
1.6 Econometric Model	14
1.6.1 Nonparametric RD Estimation: Local Linear Regression	14
1.6.2 Choosing the Optimal Bandwidth	16
1.7 Estimation Results	17
1.7.1 Numerically Significant Subgroups	19
1.7.2 Changes in Per-Pupil Spending	24
1.8 Conclusion	27
Bibliography	29
2 A Discrete Choice Analysis of the U.S. Demand for Wine	31
2.1 Abstract	31
2.2 Introduction	32
2.3 Previous Studies of Wine Demand	33
2.4 Discrete Choice Analysis Under Product Differentiation	35
2.4.1 The Nested Logit Discrete Choice Model	35
2.4.2 The Nested Logit Model Applied to Wine Choice	37
2.5 Wine Data and Variable Construction	42
2.5.1 Outside Good, Market Share, and Product Characteristics Variable Construction	44
2.5.2 Instrumental Variables Construction	45
2.6 Empirical Results	46
2.6.1 Estimated Demand Parameters	46
2.6.2 Elasticities	49

2.7	Conclusion	50
2.8	Appendix. Derivation of Own-Price Elasticity for Wine Brand j : $\varepsilon_{j,j}$	52
	Bibliography	54
3	A Bayesian Imputation Algorithm of Quarterly Earnings for Part-Time Federal Workers in the U.S. Census Bureau Longitudinal Employer- Household Dynamics Infrastructure Files	56
3.1	Abstract	56
3.2	Introduction	57
3.3	Probability Model for OPM Hours Imputation	59
	3.3.1 Identifying Assumption	59
	3.3.2 Bayesian Inference	60
3.4	Data	63
	3.4.1 Estimation Sample	63
	3.4.2 Imputation Sample	66
3.5	Empirical Hours/Earnings Imputation Procedure	67
	3.5.1 Likelihood	67
	3.5.2 Prior Distribution	68
	3.5.3 Posterior Distribution	68
	3.5.4 Hours/Earnings Imputation	69
3.6	Quality Assessment	70
	3.6.1 Internal Consistency	70
	3.6.2 External Consistency	71
3.7	Conclusion	73
3.8	Appendix. Imputed Hours Frequency Graphs	74
	Bibliography	90

LIST OF TABLES

1.1	New Jersey AYP Targets	6
1.2	California AYP Targets	6
1.3	Descriptive Statistics, New Jersey Elementary Schools, 2003 . . .	8
1.4	Descriptive Statistics, California Elementary Schools, 2003	9
1.5	Effect of Threat of Sanctions in Year t on Overall Elementary School Student Performance in Year $t+1$, New Jersey, 2002-2008 .	18
1.6	Effect of Threat of Sanctions in Year t on Overall Elementary School Student Performance in Year $t+1$, California, 2002-2010 . .	18
1.7	Effect of Threat of Sanctions in Year t on Probability of Numerically Insignificant Subgroup in Year $t+1$, California, 2002-2010 . .	24
1.8	Effect of Threat of Sanctions in Year t on Per-Pupil Spending in Year $t+1$, New Jersey, 2002-2008	26
2.1	Price-Based Wine Quality Levels	39
2.2	Wine Brands by Country of Origin	40
2.3	Descriptive Statistics, Wine Scanner Data, 2002	43
2.4	Estimated Demand Parameters, Quality Nesting, IV Regression, 2001-2003	47
2.5	Estimated Demand Parameters, Origin Nesting, IV Regression, 2001-2003	48
2.6	Wine Market Share Own-Price Elasticity, Quality Nesting, 2001- 2003	50
2.7	Wine Market Share Own-Price Elasticity, Origin Nesting, 2001-2003	50
3.1	Public-Use OPM Employment, 1998-2011	58
3.2	Imputed OPM Hours/Earnings and MORG Data Comparisons, 2000-2003	72

LIST OF FIGURES

1.1	NCLB Sanctions	3
1.2	Assessing Continuity of the Conditional Density of X, New Jersey	12
1.3	Assessing Continuity of the Conditional Density of X, California	13
1.4	Effect of Threat of Sanctions in Year t on Overall Elementary School Student Performance in Year t+1, New Jersey, 2002-2008 .	20
1.5	Effect of Threat of Sanctions in Year t on Overall Elementary School Student Performance in Year t+1, New Jersey, 2002-2008 .	21
1.6	Effect of Threat of Sanctions in Year t on Overall Elementary School Student Performance in Year t+1, California, 2002-2010 . .	22
1.7	Effect of Threat of Sanctions in Year t on Overall Elementary School Student Performance in Year t+1, California, 2002-2010 . .	23
2.1	Total Wine Sales in the United States: 1991 - 2011 (Source: Wine Institute)	33
2.2	Wine Choice Model - Quality Nesting	38
3.1	Imputed Hours Frequency, all PTSI Workers, 2000Q1	74
3.2	Imputed Hours Frequency, all PTSI Workers, 2000Q2	74
3.3	Imputed Hours Frequency, all PTSI Workers, 2000Q3	75
3.4	Imputed Hours Frequency, all PTSI Workers, 2000Q4	75
3.5	Imputed Hours Frequency, all PTSI Workers, 2001Q1	76
3.6	Imputed Hours Frequency, all PTSI Workers, 2001Q2	76
3.7	Imputed Hours Frequency, all PTSI Workers, 2001Q3	77
3.8	Imputed Hours Frequency, all PTSI Workers, 2001Q4	77
3.9	Imputed Hours Frequency, all PTSI Workers, 2002Q1	78
3.10	Imputed Hours Frequency, all PTSI Workers, 2002Q2	78
3.11	Imputed Hours Frequency, all PTSI Workers, 2002Q3	79
3.12	Imputed Hours Frequency, all PTSI Workers, 2002Q4	79
3.13	Imputed Hours Frequency, all PTSI Workers, 2003Q1	80
3.14	Imputed Hours Frequency, all PTSI Workers, 2003Q2	80
3.15	Imputed Hours Frequency, all PTSI Workers, 2003Q3	81
3.16	Imputed Hours Frequency, all PTSI Workers, 2003Q4	81
3.17	Imputed Hours Frequency, all PT Workers, 2000Q1	82
3.18	Imputed Hours Frequency, all PT Workers, 2000Q2	82
3.19	Imputed Hours Frequency, all PT Workers, 2000Q3	83
3.20	Imputed Hours Frequency, all PT Workers, 2000Q4	83
3.21	Imputed Hours Frequency, all PT Workers, 2001Q1	84
3.22	Imputed Hours Frequency, all PT Workers, 2001Q2	84
3.23	Imputed Hours Frequency, all PT Workers, 2001Q3	85
3.24	Imputed Hours Frequency, all PT Workers, 2001Q4	85
3.25	Imputed Hours Frequency, all PT Workers, 2002Q1	86
3.26	Imputed Hours Frequency, all PT Workers, 2002Q2	86

3.27	Imputed Hours Frequency, all PT Workers, 2002Q3	87
3.28	Imputed Hours Frequency, all PT Workers, 2002Q4	87
3.29	Imputed Hours Frequency, all PT Workers, 2003Q1	88
3.30	Imputed Hours Frequency, all PT Workers, 2003Q2	88
3.31	Imputed Hours Frequency, all PT Workers, 2003Q3	89
3.32	Imputed Hours Frequency, all PT Workers, 2003Q4	89

CHAPTER 1
THE TREATMENT EFFECT OF FINANCIAL SANCTIONS ON
LOW-PERFORMING SCHOOLS: A REGRESSION DISCONTINUITY
APPROACH

1.1 Abstract

The No Child Left Behind (NCLB) Act of 2001 brought about a set of nationwide educational standards in an effort to improve school accountability across the states. Key components of this legislation are financial sanctions imposed on schools whose students fail in two consecutive years to make adequate yearly progress (AYP) on exams in English and mathematics. In this paper, I test whether the threat of these sanctions leads schools to improve student outcomes after failure to make AYP in the first year. Using school-level data from New Jersey and California, I implement a regression discontinuity (RD) design that exploits the fact that threatened exposure to sanctions is a discontinuous function of student performance at the assignment threshold. The local linear regression results at the optimal RD bandwidth are not estimated with precision. The point estimates are between 0.112 (NJ) and 0.129 (CA) standard deviations in English, and between 0.041 (NJ) and 0.043 (CA) standard deviations in mathematics. There is statistically significant evidence at larger bandwidths of a positive treatment effect in both subjects, although these estimates are generally larger than those from the economics literature on NCLB. Since only the underperformance of "numerically significant" subgroups may result in sanctions for schools, I also test whether assignment to treatment in the first year leads to schools' attempting to reassign students from subgroups that have the lowest exam performance. There is evidence in these data to support this hypothe-

sis, although it isn't clear that gaming by school administrations is driving the results. Sanction threats are also shown to lead to increased per-pupil spending, although quantitatively this is quite limited and not focused on classroom instruction or other measures that might directly improve performance.

1.2 Introduction

The *No Child Left Behind (NCLB) Act of 2001*¹ imposed a set of nationwide educational standards on individual states, including a regimen of high-stakes testing in mathematics and language arts at the elementary school, middle school, and high school levels. Although the goal of achieving 100% proficiency in math and reading for a broad range of student subgroups² by 2014 is a federal aim of the legislation, each state was given the flexibility to set its own yearly benchmarks with regard to the NCLB-designated exams in each subject area. Schools that meet these subject proficiency benchmarks for all relevant student subgroups (in addition to a set of auxiliary goals, such as sufficient levels of student participation in exams, maximum dropout rates, and minimum graduation rates), are deemed to have achieved *adequate yearly progress* (AYP) for the academic year. Likewise, schools that fail to meet these state benchmarks do not achieve AYP for that year,³ and face a set of financial and organizational sanctions for each year that they continue to underperform. In the first year

¹See <http://www2.ed.gov/policy/elsec/leg/esea02/index.html>.

²In all states, these subgroups include: 1) the total student population; 2) white students; 3) African-American students; 4) Hispanic students; 5) Native American students; 6) Asian students; 7) students from other demographic categories; 8) special education students; 9) Limited English Proficient (LEP) students; and 10) economically disadvantaged students. At least 95% of the students in each subgroup must participate in the NCLB-designated exam.

³If a participating student subgroup does not meet the AYP benchmark, schools may appeal to "safe harbor" rules that allow them to achieve AYP for that subgroup if they decrease the number of students receiving lower than proficient scores (the lowest achievement level) by 10% and satisfy a set of secondary requirements.

Figure 1.1: NCLB Sanctions

No Child Left Behind Sanctions Chart

Consecutive Years of Missing AYP	Sanctions
First Year	<ul style="list-style-type: none"> ▶ Placed on “watch” list ▶ Required to develop a school improvement plan
Second Year	<ul style="list-style-type: none"> ▶ Listed as “needs improvement” school. ▶ District must provide any student attending the “needs improvement” school the option of attending another school that has met adequate yearly progress. The district pays transportation costs.
Third Year	<ul style="list-style-type: none"> ▶ Listed as “needs improvement” school. ▶ District must provide any student attending the “needs improvement” school the option of attending another school that has met adequate yearly progress. The district pays transportation costs. ▶ The school district must offer “supplemental educational services” to any student who qualifies for free or reduced lunch. One option for supplemental services must be from an outside provider.
Fourth Year	<ul style="list-style-type: none"> ▶ Listed as “needs improvement” school. ▶ District must provide any student attending the “needs improvement” school the option of attending another school that has met adequate yearly progress. The district pays transportation costs. ▶ The school district must offer “supplemental educational services” to any student who qualifies for free or reduced lunch. One option for supplemental services must be from an outside provider. ▶ The school must change its staffing or make a “fundamental change” such as restructuring the school.
Fifth Year	<ul style="list-style-type: none"> ▶ Listed as “needs improvement” school. ▶ District must provide any student attending the “needs improvement” school the option of attending another school that has met adequate yearly progress. The district pays transportation costs. ▶ The school district must offer “supplemental educational services” to any student who qualifies for free or reduced lunch. One option for supplemental services must be from an outside provider. ▶ The school must convert into a charter school, turn management over to a private management company or be taken over by the state.

of failing to make AYP, a school receives a warning that they are in danger of being sanctioned, with sanctions being first implemented in the second consecutive year (see Figure 1.1 for a yearly breakdown of penalties that apply to all states).

A number of studies have been conducted to examine whether sanction *threats* induce increases in school accountability, mainly focusing on the Florida and North Carolina school systems.⁴ In this paper, I use school-level assessment data from New Jersey and California for elementary school students⁵ to

⁴See Greene (2001), Jacob (2005), Figlio and Rouse (2006), West and Peterson (2006), Chakrabarti (2007), Chakrabarti (2008), and Chiang (2009).

⁵While middle school and high school data are available, the nature of the financial sanctions

estimate the treatment effect of the threat of sanctions in year t on overall student performance in year $t + 1$. Since I am mainly interested in the incentive effects from this threat, I do not estimate the effects of the first- and second-year sanctions themselves. Sanctions in both years can be nontrivial and the reduction in resources imposed on the schools may mask changes in student outcomes that are a direct response to the initial threat and are the focus of this paper.

Both New Jersey and California have diverse student populations that provide us with information on the performance of a broad range of students. I exploit the fact that assignment to treatment is a discontinuous function of student achievement at the AYP benchmark by using a regression discontinuity (RD) design. Estimating the effects of this treatment are an important component in determining the relative benefits of punitive vs. non-punitive student accountability policy measures, as well as the heterogeneity of effects that may arise across different states under NCLB.

Since only exam scores of students in subgroups that meet enrollment cutoffs are counted towards achieving AYP, I also test whether schools threatened by sanctions may try to game the examination system by manipulating the test enrollment of low-performing subgroups. I examine this possibility by estimating the probability that the lowest-performing subgroup in year t goes from being numerically significant to numerically insignificant in year $t + 1$. Finally, I also examine whether the threat of sanctions leads schools to take more substantive reform actions such as increasing per-pupil spending in areas that may directly increase student performance on exams.

The remainder of the paper is organized as follows: Section 1.3 reviews the imposed (such as paying transportation costs for travel to alternative within-district schools) would be expected to have less impact due to the presence of fewer of these school types in many school districts.

NCLB-based school accountability system in New Jersey and California; Section 1.4 discusses data sources and construction of variables; Section 1.5 discusses validity of the RD design for these data and identification of the treatment effect; Section 1.6 discusses my RD econometric model; Section 1.7 reviews empirical estimates; and Section 1.8 concludes.

1.3 NCLB-based School Accountability in New Jersey and California

Both New Jersey and California require their primary and secondary school students to pass grade-level specific high-stakes exams to comply in part with *No Child Left Behind*.⁶ Students may receive scores of "partially proficient", "proficient", or "advanced proficient" on each exam in the subject areas of mathematics and language arts literacy (see Tables 1.1 and 1.2 for yearly benchmarks by subject area and grade-specific exams). If each of ten student subgroups (including students as a whole) receives scores of "proficient" or "advanced proficient", such that the combined percentage meets or exceeds the state-implemented cutoffs for each subject area, then a school will be deemed to have achieved AYP (with the possible exceptions noted above).

In New Jersey, at least 40 students from each subgroup must be represented to be counted towards AYP proficiency goals (and to be deemed a "numerically significant subgroup"), and in California, 100 students from each subgroup must

⁶New Jersey elementary school students in grades 3 and 4 were required to take the New Jersey Assessment of Skills and Knowledge (NJASK) from 2003-2007 to satisfy NCLB requirements. All students in grades 3-8 were required to take a grade-specific version of the NJASK after 2008. Beginning in 1999, California elementary school students in grades 3-6 were required to take a series of exams in the Standardized Testing and Reporting (STAR) program. See <http://www.nj.gov/education/assessment> and <http://www.cde.ca.gov/ta/tg/sr/cefstar.asp> for additional information on the NCLB-designated exams.

Table 1.1: New Jersey AYP Targets

Content/Grade Level	AYP Targets (Percent Proficient)				
	2002-2003	2004-2007	2008-2010	2011-2013	2014
<i>Language Arts</i>					
Elementary Schools	68	75	59	79	100
Middle Schools	58	66	72	86	100
High Schools	73	79	85	92	100
<i>Mathematics</i>					
Elementary Schools	53	62	66	83	100
Middle Schools	39	49	61	80	100
High Schools	55	64	74	86	100

Table 1.2: California AYP Targets

Content/Grade Level	AYP Targets (Percent Proficient)				
	2002-2004	2005-2007	2008	2009	2010
<i>Language Arts</i>					
Elementary Schools	13.6	24.4	35.2	46	56.8
Middle Schools	13.6	24.4	35.2	46	56.8
High Schools	11.2	22.3	33.4	44.5	55.6
<i>Mathematics</i>					
Elementary Schools	16	26.5	37	47.5	58
Middle Schools	16	26.5	37	47.5	58
High Schools	9.6	20.9	32.2	43.5	54.8
	2011	2012	2013	2014	
<i>Language Arts</i>					
Elementary Schools	67.6	78.4	89.2	100	
Middle Schools	67.6	78.4	89.2	100	
High Schools	66.7	77.8	88.9	100	
<i>Mathematics</i>					
Elementary Schools	68.5	79	89.5	100	
Middle Schools	68.5	79	89.5	100	
High Schools	66.1	77.4	88.7	100	

be represented, or at least 50 students that make up 15% or more of total enrollment at the school. States are given broad flexibility in determining whether a particular school will face sanctions after not achieving AYP in a given year. For example, in New Jersey, schools that fail to achieve AYP in different subjects over two years will face immediate sanctions, whereas in California, schools

only face sanctions if they do not achieve AYP in the same subject area.

1.4 Data and Variable Construction

I use elementary school-level data from the New Jersey Department of Education (NJDOE) for 2002-2008 and from the California Department of Education (CDE) for 2002-2010 to perform the RD estimation.⁷ Each observation records aggregate results from test-takers as a whole, and in certain subgroups, at a specific elementary school. Since New Jersey changed the assessment exams at the elementary-school level in the 2008-2009 academic year, I restrict my analysis to prior years to ensure comparability in exam results. In each year, the NJDOE and the CDE provide the following information for each school: 1) county where school is located; 2) school district; 3) attendance rate for all students; 4) assessment scores in mathematics and language arts literacy for all students and for students in NCLB-specific subgroups; 5) the percentage of students participating by subgroup and overall (in NJ, cells with symbols rather than numbers indicate that a subgroup was not numerically significant, and did not count towards AYP); 6) AYP status; 7) the number of consecutive years a school has not achieved AYP; and 8) the overall graduation rate at the high school level. California also provides information on the number of students from each subgroup that sit for the exam.

Tables 1.3 and 1.4 show descriptive statistics for the NJDOE and CDE data for 2003. In these tables, we see information regarding participation rates and student exam performance as a whole and for NCLB-specific subgroups. For

⁷See <http://education.state.nj.us/rc/nclb/archive.html> for additional information regarding the New Jersey NCLB assessment data, and <http://www.cde.gov/ta/ac/ay/aypdatafiles.asp> for more information regarding the California data.

Table 1.3: Descriptive Statistics, New Jersey Elementary Schools, 2003

Variables	(1)	(2)	(3)	(4)	(5)
	N	Mean	Std. Dev.	Min.	Max.
<i>Overall, Language Arts</i>					
Participation Rate (%)	2252	99.5	1.5	72.3	100
Proficient (%)	2250	70.6	15.8	0	97.7
Advanced Proficient (%)	2250	6.5	8.2	0	71.8
<i>Overall, Mathematics</i>					
Participation Rate (%)	2252	99.7	1.3	73.7	100
Proficient (%)	2249	42.9	12.2	0	84.6
Advanced Proficient (%)	2249	21.8	15.3	0	92.3
<i>African American, Language Arts</i>					
Proficient (%)	629	53.9	17.0	0	100
Advanced Proficient (%)	629	2.0	4.1	0	47.6
<i>African American, Mathematics</i>					
Proficient (%)	630	29.4	13.9	0	86.4
Advanced Proficient (%)	630	5.5	7.1	0	49.2
<i>Hispanic, Language Arts</i>					
Proficient (%)	532	59.2	16.7	10	100
Advanced Proficient (%)	532	2.4	4.6	0	54.1
<i>Hispanic, Mathematics</i>					
Proficient (%)	535	36.2	13.5	0.9	77.8
Advanced Proficient (%)	535	8.8	10.2	0	70.8
<i>White, Language Arts</i>					
Proficient (%)	1615	77.8	9.9	28.6	100
Advanced Proficient (%)	1615	8.8	9.1	0	71.4
<i>White, Mathematics</i>					
Proficient (%)	1614	48.4	8.9	9.1	77.1
Advanced Proficient (%)	1614	27.4	14.3	0	90.9

example, in New Jersey, we see that there were 630 schools in which African American students participated in the NCLB mathematics exam. Among those schools, the average mathematics exam score at the proficient level for African American students was 29.4%, and the highest percentage of African American students achieving a proficient score in mathematics at any of these schools was 86.4%.

Table 1.4: Descriptive Statistics, California Elementary Schools, 2003

Variables	(1)	(2)	(3)	(4)	(5)
	N	Mean	Std. Dev.	Min.	Max.
<i>Overall, Language Arts</i>					
Participation Rate (%)	3889	99.0	2.1	18.8	100
Proficient or Advanced Proficient (%)	3858	31.4	15.3	0	93.3
<i>Overall, Mathematics</i>					
Participation Rate (%)	3889	98.7	2.4	8.6	100
Proficient or Advanced Proficient (%)	3857	39.3	14.9	0	95.6
<i>African American, Language Arts</i>					
Proficient or Advanced Proficient (%)	2024	25.0	13.5	0	88.2
<i>African American, Mathematics</i>					
Proficient or Advanced Proficient (%)	2020	28.7	14.3	0	87.7
<i>Hispanic, Language Arts</i>					
Proficient or Advanced Proficient (%)	3613	23.2	10.9	0	85.8
<i>Hispanic, Mathematics</i>					
Proficient or Advanced Proficient (%)	3611	32.5	11.8	0	94.6
<i>White, Language Arts</i>					
Proficient or Advanced Proficient (%)	3128	44.1	15.6	0	100
<i>White, Mathematics</i>					
Proficient or Advanced Proficient (%)	3127	48.8	15.9	0	96.9

1.4.1 Assignment Variable Construction

In an RD design, different values of the assignment variable, X_j , determine whether a unit j under consideration is exposed to the treatment. For a given threshold value, $X_j = c$, we can define assignment to treatment as: $D_j = 1\{X_j \geq c\}$, where D_j is a binary treatment variable for all units j . In the NCLB context,⁸ we have a "sharp" RD design, whereby treatment is determined solely by the value of the assignment variable (i.e., the probability of treatment is one for $X_j \geq c$ and 0 for $X_j < c$).⁹ We can reduce the dimensionality of

⁸This excludes the relatively few cases in which a school would have achieved AYP by student subgroup assessment scores but does not due to failing a secondary criterion such as participation; it also excludes the cases in which a school would not achieve AYP by assessment performance, but does so through other means such as "safe harbor".

⁹Since treatment (exposure to the threat of financial sanctions) is defined for failing to reach a cutoff in the NCLB context, I define the converse relationships for assignment to treatment, i.e.,

the assignment problem to a univariate problem by choosing the scores for a specific subgroup each year the exams are administered, thereby constructing a single assignment variable. This procedure allows us to estimate the joint average treatment effect for both language arts literacy and mathematics, rather than the average treatment effect for each individual subject area.

Following Gill et al. (2007), I use a normalized version of the *minimum subgroup assessment score* as the univariate assignment variable. This variable comprises the assessment score for the lowest-performing student subgroup i for a given subject area s , school j , and year t . Since we are using the lowest-performing subgroup, assignment to treatment binds for this group (i.e., as long as $X < c$ for the minimum subgroup, we could raise the performance of all other subgroups with no consequent change in AYP status). I subtract the AYP cutoff from the assessment score for each subject area and student subgroup to center the threshold value, c , at zero:

$$X_{min} = \min_{i,s} \{assessment_score_{is} - AYP_benchmark_s\} \quad (1.1)$$

where $i \in \{0, 1, \dots, 9\}$ is the student subgroup (I define subgroup 0 as that containing all students), and $s \in \{M, L\}$ is the subject component.

1.4.2 Outcome Variable Construction

I define the outcome variable, $Y_{ijs(t+1)}$ as the standardized assessment score for student subgroup $i \in \{0, 1, \dots, 9\}$, in school j , subject area $s \in \{M, L\}$, and year $t + 1$ when students take the relevant NCLB-designated exam in year t . This variable is defined for years 2003-2008 in New Jersey and years 2003-2010 in California. The standardization is performed over statewide means and stan-

assignment occurs for values of the assignment variable below the cutoff. I write the standard conditions here to be consistent with the prototypical RD design.

dard deviations. Because the outcome variable does not appear in the data for subgroups that are not numerically significant, low sample sizes make estimation of outcomes for these subgroups infeasible. Therefore, in the subsequent analysis I will focus on the performance of students as a whole (i.e., subgroup 0).

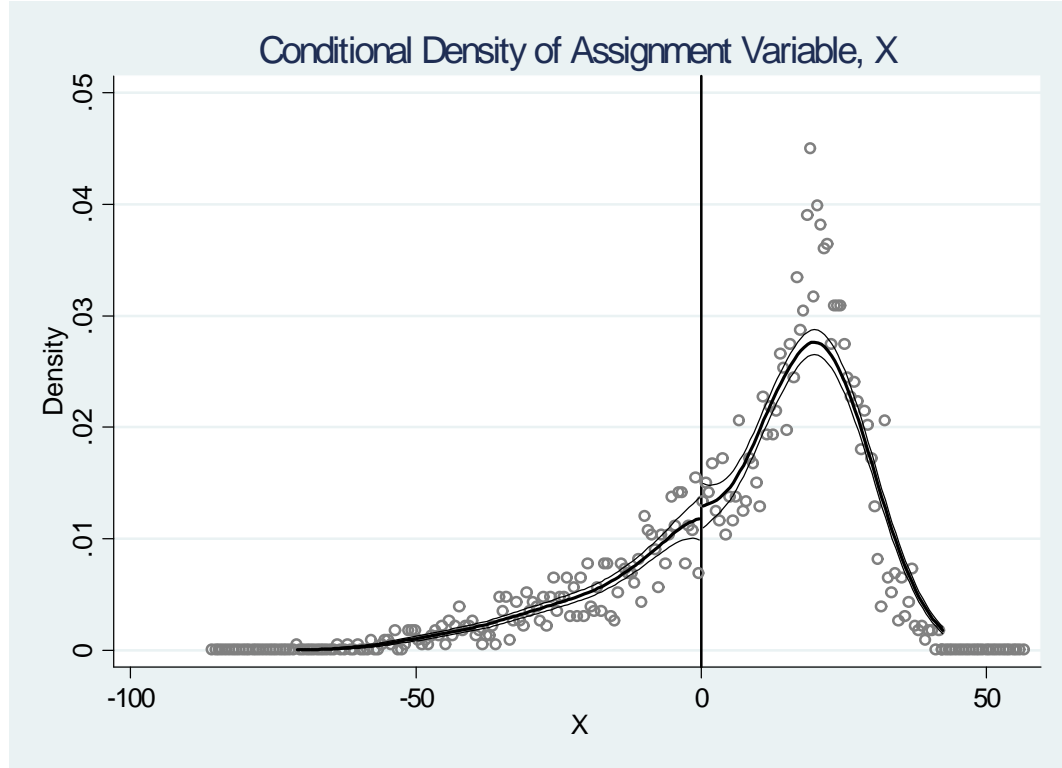
1.5 The Regression Discontinuity Approach

1.5.1 Assessing the Validity of the RD Design for the New Jersey and California Data

The results from an RD design in a neighborhood of the threshold $\{c - h \leq X_i \leq c + h\}$, for some bandwidth, h , can be considered "as good as" those from a randomized experiment when individuals cannot precisely manipulate the assignment variable, X (Lee and Lemieux, 2010). It can be shown that imprecise manipulation of treatment assignment is equivalent to the continuity of the conditional density of the assignment variable at the cutoff. I implement the local linear density estimator proposed by McCrary (2008) to determine whether the RD design is valid for the New Jersey and California assessment data. See Figures 1.2 and 1.3 for graphs of the assignment variable conditional density function for elementary schools.

As we can see from the graphs, there is a slight discontinuity in the conditional density at the cutoff. For the New Jersey data, the discontinuity point estimate is 0.0869 with a standard error of 0.1175, producing a t-statistic of 0.739. We cannot reject the null hypothesis that the point estimate is statistically significantly different from zero at traditional significance levels, which provides

Figure 1.2: Assessing Continuity of the Conditional Density of X, New Jersey



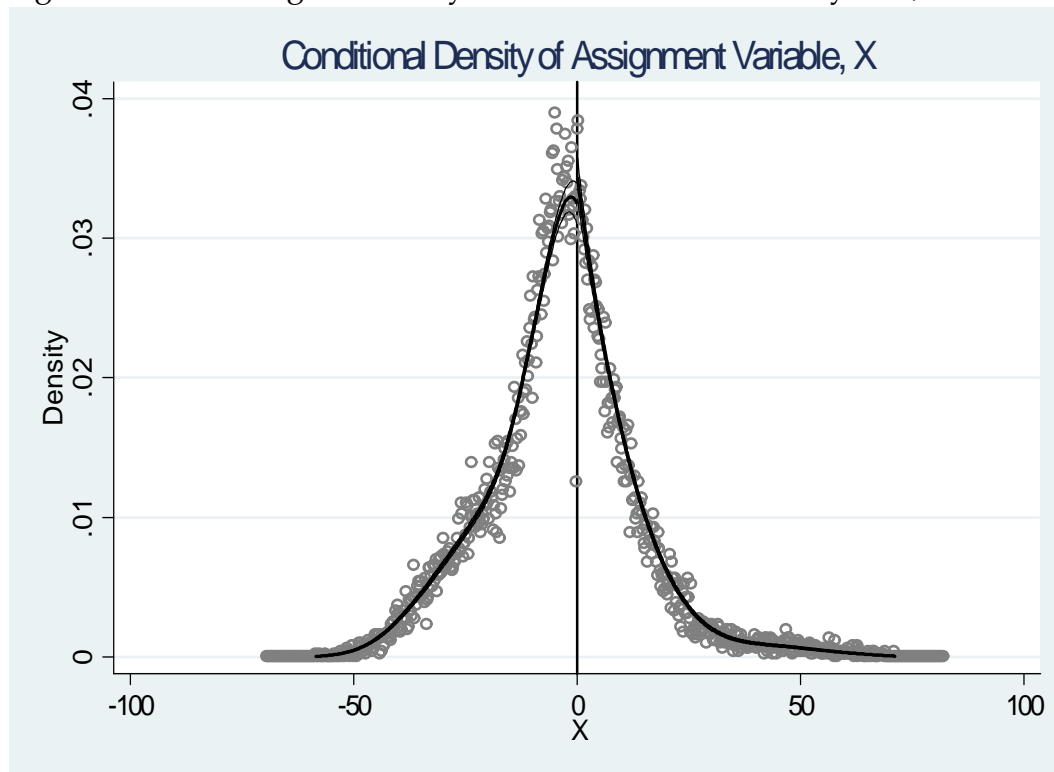
evidence that the RD design is valid for these data. For the California data, the discontinuity point estimate is 0.0599 with a standard error of 0.0318, producing a t-statistic of 1.88 and a p -value of 0.0602. The value of this point estimate is more ambiguous than that for the New Jersey data. It cannot be rejected at traditional significance levels, although there may be concern that there is potential manipulation of the assignment variable.¹⁰

1.5.2 Identification of the Treatment Effect

A consequence of the continuity of the conditional density of the assignment variable is that the treatment effect of interest, τ , is identified at the threshold. Following Lee and Lemieux (2010), we consider a set of equations for an RD

¹⁰The choice of different kernels or bandwidths affects the smoothness of the density, but does not introduce discontinuities to the graph.

Figure 1.3: Assessing Continuity of the Conditional Density of X , California



design:

$$Y = D\tau + W\delta_1 + U \quad (1.2)$$

$$D = 1\{X \geq c\} \quad (1.3)$$

$$X = W\delta_2 + V \quad (1.4)$$

where Y is an outcome variable, D is an indicator variable for the treatment, and W is a vector of observable characteristics that may affect the assignment variable, X . For given values of W and U (w and u , respectively), consider the conditional probability of W and U given X , $P(W = w, U = u | X = x)$. By Bayes' Rule, we have:

$$P(W = w, U = u | X = x) = f(x | W = w, U = u) \frac{P(W = w, U = u)}{f(x)}. \quad (1.5)$$

Therefore, the continuity of $P(W = w, U = u | X = x)$ depends on the continuity of the conditional density of X , $f(x | W = w, U = u)$.

Denote $Y_i(1)$ as the "potential" outcome if unit i is exposed to the treatment, and $Y_i(0)$ as the "potential" outcome if unit i is not exposed to the treatment. We define the average treatment effect in an ε -neighborhood of the assignment cutoff, c , as:

$$\lim_{\varepsilon \downarrow 0} E[Y_i | X_i = c + \varepsilon] - \lim_{\varepsilon \uparrow 0} E[Y_i | X_i = c + \varepsilon] = E[Y_i(1) - Y_i(0) | X = c] \quad (1.6)$$

Then, we have:

$$\begin{aligned} & \lim_{\varepsilon \downarrow 0} E[Y_i | X_i = c + \varepsilon] - \lim_{\varepsilon \uparrow 0} E[Y_i | X_i = c + \varepsilon] & (1.7) \\ &= \tau + \lim_{\varepsilon \downarrow 0} \sum_{w,u} (w\delta_1 + u) \cdot P(W = w, U = u | X = c + \varepsilon) \\ & \quad - \lim_{\varepsilon \uparrow 0} \sum_{w,u} (w\delta_1 + u) \cdot P(W = w, U = u | X = c + \varepsilon) \\ & \quad \quad \quad = \tau \end{aligned}$$

where the last equality follows by the continuity of $P(W = w, U = u | X = c + \varepsilon)$. This, in turn, follows from the continuity of the conditional density of X , as shown above. Thus the treatment effect, τ , is identified when estimating a valid RD design.

1.6 Econometric Model

1.6.1 Nonparametric RD Estimation: Local Linear Regression

I estimate the extent to which an elementary school's exposure to the threat of financial sanctions in year t affects student performance in year $t + 1$. Since the RD design depends on local estimates of the regression function at the assignment cutoff point, RD estimates of the treatment effect, τ , may be biased when performing linear parametric estimation if the functional form is misspecified.

While using linear estimation when the true functional form is nonlinear will minimize specification errors globally, there may still be large specification errors locally, such as at the cutoff point, leading to biased estimates (Lee and Lemieux, 2010). To reduce the importance of the bias, I use a local linear regression estimation procedure. This estimation procedure amounts to running separate linear regressions within bins in a small neighborhood of the cutoff point, where the approximation to the true regression line is more plausibly linear. The procedure also generates consistent estimates of the treatment effect (Hahn, Todd, and van der Klaauw, 2001).

More formally, let $Y_{ijs(t+1)}$ denote the standardized assessment score for student subgroup i , in school j , subject area s , and year $t + 1$. Let the assignment variable $X_{min} - c$ denote the centered student performance at school j in year t for the lowest-performing student subgroup i , as described above. Let α_T denote the expected value of year $t + 1$ performance for a school that is exposed to the treatment and has an assignment value within a bandwidth, h , below the cutoff and let α_{NT} denote the expected value of year $t + 1$ performance for a school that is not exposed to the treatment and has an assignment value exceeding the cutoff within the same bandwidth. For schools that are exposed to the treatment, I estimate:

$$Y_{ijs(t+1)} = \alpha_T + \beta_T(X_{min} - c) + \varepsilon_{ijs(t+1)} \text{ s.t. } -h \leq X_{min} - c < 0 \quad (1.8)$$

and for schools that are not exposed to the treatment, I similarly estimate:

$$Y_{ijs(t+1)} = \alpha_{NT} + \beta_{NT}(X_{min} - c) + \varepsilon_{ijs(t+1)} \text{ s.t. } 0 \leq X_{min} - c \leq h \quad (1.9)$$

Then the estimated treatment effect, $\hat{\tau}$, is precisely the estimated difference between the two intercepts from these regressions: $\hat{\alpha}_T - \hat{\alpha}_{NT}$. In practice, I use a pooled regression, so that the estimates and standard errors for the treatment

effect can be computed directly:

$$\begin{aligned}
 Y_{ijs(t+1)} &= \alpha_{NT} + \tau D + \beta_{NT}(X_{min} - c) + (\beta_T - \beta_{NT})D(X_{min} - c) + \varepsilon_{ijs(t+1)}, \\
 \text{s.t. } -h &\leq X_{min} - c \leq h
 \end{aligned} \tag{1.10}$$

where D is a binary treatment indicator and $\tau = \alpha_T - \alpha_{NT}$ is the treatment effect. By including interaction terms between D and X , the model allows the slope of the regression function to differ on both sides of the cutoff. Requiring $(\beta_T = \beta_{NT})$ would use data from the non-treatment side of the cutoff to estimate the intercept from the treatment side (and vice-versa), violating the spirit of the RD design.

1.6.2 Choosing the Optimal Bandwidth

Choice of the optimal bandwidth, h , above involves a tradeoff between bias and efficiency. By using a larger bandwidth, the treatment effect can be estimated more precisely, since more data are available for the estimation. However, as we move farther from the cutoff point, a linear model may poorly approximate the underlying regression function, introducing bias. I apply the asymptotically optimal bandwidth choice rule proposed by Imbens and Kalyanaraman (2009), which constructs an estimator to the optimal bandwidth, h_{opt} , that minimizes the mean squared error criterion:

$$h_{opt} = \operatorname{argmin} MSE(h) = \operatorname{argmin} E [(\hat{\tau} - \tau)]^2 \tag{1.11}$$

where $\hat{\tau}$ is the estimated treatment effect.¹¹ This bandwidth is optimal in the sense that it attempts to find mutually exclusive, equally-sized neighborhoods

¹¹Imbens and Kalyanaraman (2009) devise a data-dependent way in which to find the optimal bandwidth and provide Stata programs to implement their model. See http://www.economics.harvard.edu/faculty/imbens/software_imbens

on either side of the regression discontinuity threshold that best fit a local linear regression function. In more convention cross-validation procedures, optimality is defined in terms of the fit over the entire support of the data (see Imbens and Kalyanaraman, 2009). Since the expression in equation (11) is infeasible, Imbens and Kalyanaraman introduce the following feasible estimator defined at the RD threshold, c :

$$\hat{h}_{opt} = \left(\frac{2 \cdot \hat{\sigma}^2(c) / \hat{f}(c)}{\left(\hat{m}_+^{(2)}(c) - \hat{m}_-^{(2)}(c) \right)^2 + (\hat{r}_+ - \hat{r}_-)} \right)^{1/5} \cdot N^{-1/5}, \quad (1.12)$$

where N is the sample size, $\hat{f}(x)$ is the estimator for the density function of the assignment variable, $\hat{\sigma}^2(x)$ is the estimator for the conditional variance function of the assignment variable, and $\hat{m}_-^{(2)}(x)$ and $\hat{m}_+^{(2)}(x)$ are estimators for the left and right second derivatives of the assignment variable, respectively. The authors implement this optimal bandwidth by constructing an algorithm for estimating these functions.

1.7 Estimation Results

The results from the local linear regression procedure for schools that are threatened with financial sanctions (i.e., those schools that fail to achieve AYP for one year) are in Tables 1.5 and 1.6. The dependent variable is the standardized overall performance of elementary school students in year $t + 1$ for assignment to treatment occurring in year t . In the first column of the table are a set of specifications that vary by the length of the neighborhood of the cutoff in which the regression is performed. The preferred specification is that using the optimal bandwidth, h , derived by Imbens and Kalyanaraman (2009). Standard errors are clustered at the county level.

Table 1.5: Effect of Threat of Sanctions in Year t on Overall Elementary School Student Performance in Year $t+1$, New Jersey, 2002-2008

Bandwidth/Specification (percentage points)	Standardized Test Score in Year $t + 1$	
	Language Arts (1)	Mathematics (2)
5	0.063 (0.108) [342]	-0.073 (0.115) [342]
7.6 (LA), 7.2 (M) (optimal)	0.112 (0.086) [527]	0.041 (0.095) [504]
10	0.055 (0.075) [685]	0.072 (0.081) [685]
15	-0.015 (0.062) [1079]	0.016 (0.071) [1079]

Table 1.6: Effect of Threat of Sanctions in Year t on Overall Elementary School Student Performance in Year $t+1$, California, 2002-2010

Bandwidth/Specification (percentage points)	Standardized Test Score in Year $t + 1$	
	Language Arts (1)	Mathematics (2)
5	-0.009 (0.148) [287]	0.008 (0.123) [287]
6.3 (LA), 5.9 (M) (optimal)	0.129 (0.140) [341]	0.043 (0.119) [333]
10	0.277 (0.136)** [409]	0.288 (0.120)** [409]
15	0.166 (0.129) [429]	0.181 (0.107)* [429]

Notes: Heteroskedasticity-robust standard errors clustered at the county level in parentheses. Sample sizes in brackets. Optimal bandwidth determined by the Imbens/Kalyanaraman (2009) MSE criterion. Analysis performed on schools without any prior exposure to sanctions.

As we see in the table, the treatment effect on student performance in both language arts literacy and mathematics is mainly positive, but the point estimates are not estimated with precision. For language arts, the model indicates that a one percentage point increase in the assignment variable leads to an increase of 0.112 standard deviations of students' exam performance in New Jersey and an increase of 0.129 standard deviations in California.

In mathematics, it leads to an increase of 0.041 standard deviations in New Jersey, and an increase of 0.043 standard deviations in California, although again the results are not statistically significant. As a robustness check, I also run the local linear regression at a set of bandwidths that are smaller and larger than the

optimal one. In both New Jersey and California, the treatment effects are almost uniformly positive irrespective of the bandwidth chosen. At bandwidths larger than the optimal one, we see positive and statistically significant results in California. At a bandwidth of 10, the model indicates an increase of 0.277 standard deviations in student performance in language arts literacy and an increase of 0.288 standard deviations in mathematics. At a bandwidth of 15, there is an increase of 0.181 standard deviations in mathematics. These latter results are larger than those from recent work done on accountability pressure and student achievement at the elementary school level in Florida (Chiang, 2009).

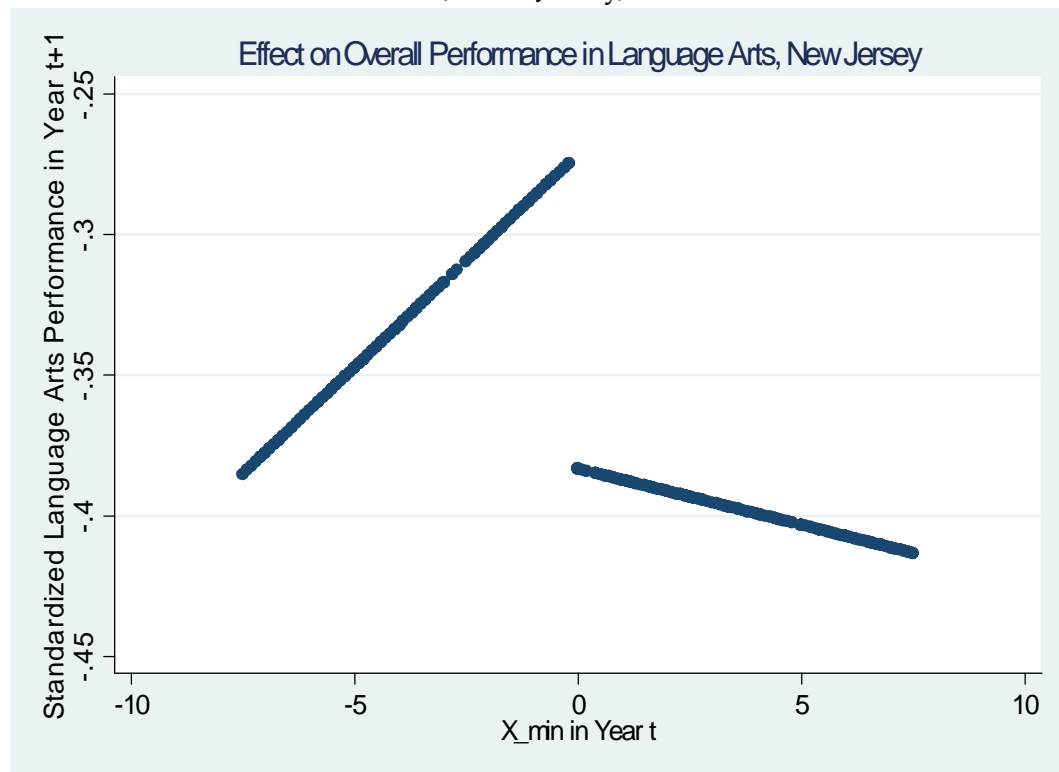
In Figures 1.4 through 1.7, we also see the resulting discontinuity of the treatment at the AYP threshold. Each of the figures shows a more-steeply sloped regression line (in absolute value) to the left of the cutoff (indicating the region in which treatment occurs), and a less-steeply sloped line (in absolute value) to the right of the cutoff. The difference in intercepts between these lines indicates the estimated treatment effect, $\hat{\tau}$.¹²

1.7.1 Numerically Significant Subgroups

As mentioned above, in both New Jersey and California, only scores from numerically significant subgroups are counted toward AYP proficiency results. It is reasonable to postulate that schools may try to game the testing system by encouraging students in low-performing and underrepresented subgroups, at the margin, not to sit for the NCLB-designated exams. For example, there is evidence that states have reclassified low-performing students into special education so that such students would take non-standard exams that do not count

¹²The differences in slopes for each of the four main sets of results were also statistically insignificant at the optimal bandwidth.

Figure 1.4: Effect of Threat of Sanctions in Year t on Overall Elementary School Student Performance in Year $t+1$, New Jersey, 2002-2008



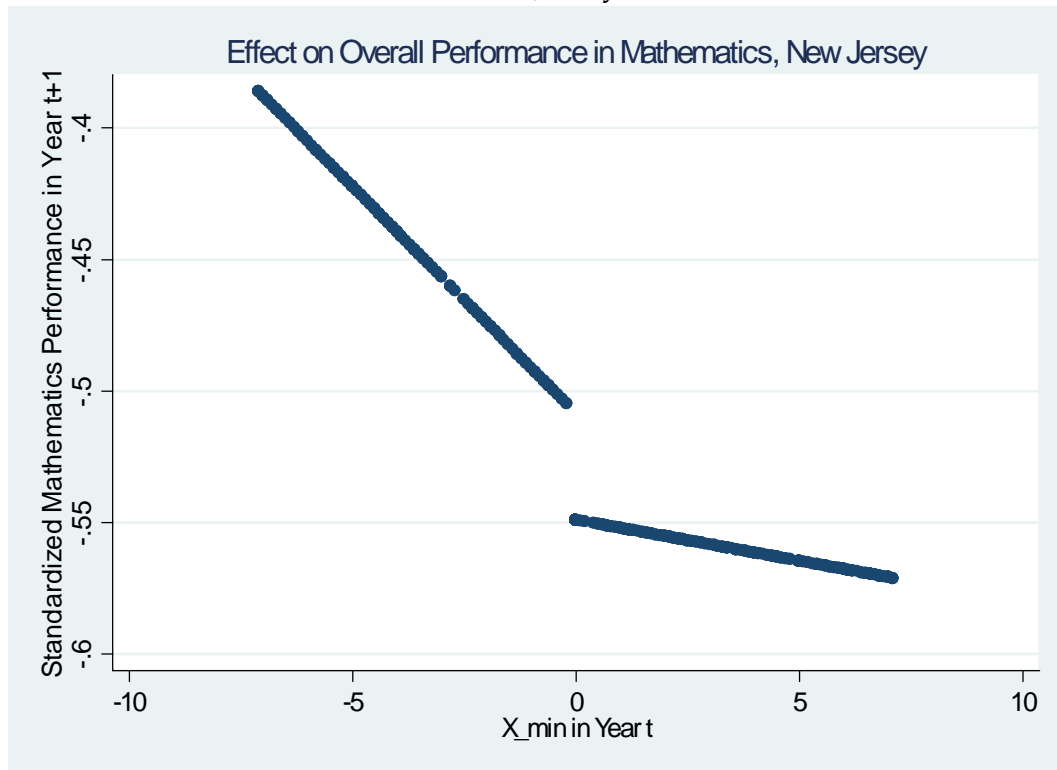
Notes: The results in the graph correspond to the optimal bandwidth specifications in Table 1.5 for language arts literacy. The slope of the regression line on either side of the AYP threshold is that computed within the optimal bandwidth.

towards AYP.¹³

Since California provides data on the number of students from each subgroup that sit for each exam, I can test whether this type of gaming behavior occurs due to exposure to the treatment. In particular, I test to see whether the subgroup that contains the worst-performing students in year t becomes numerically insignificant in year $t + 1$. In this way, schools can try to make AYP in the following year by excluding these students from the schoolwide results.

¹³See Deere and Strayer, 2001; Figlio and Getzler, 2002; Jacob, 2005; and Cullen and Reback, 2006.

Figure 1.5: Effect of Threat of Sanctions in Year t on Overall Elementary School Student Performance in Year t+1, New Jersey, 2002-2008



Notes: The results in the graph correspond to the optimal bandwidth specifications in Table 1.5 for mathematics. The slope of the regression line on either side of the AYP threshold is that computed within the optimal bandwidth.

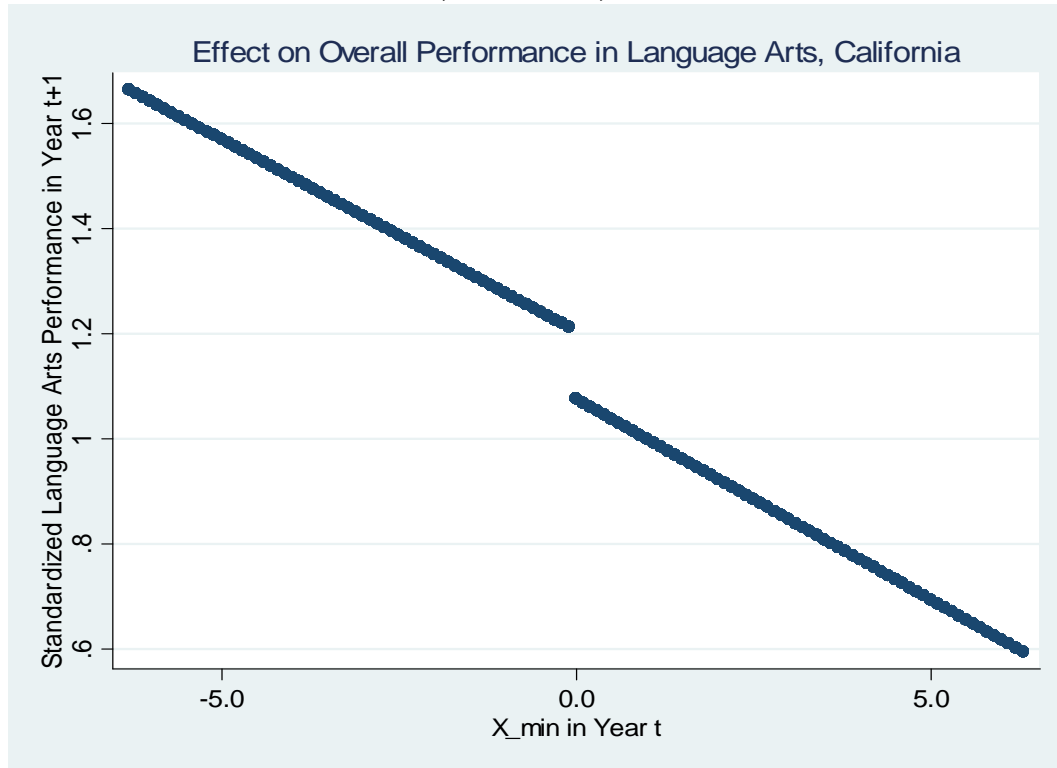
I estimate the following model:

$$NSS_{ijs(t+1)} = \alpha_{NT} + \tau D + \beta_{NT}(X_{min} - c) + (\beta_T - \beta_{NT})D(X_{min} - c) + \varepsilon_{ijs(t+1)},$$

$$\text{s.t. } -h \leq X_{min} - c \leq h \quad (1.13)$$

where $NSS_{ijs(t+1)}$ is a binary variable that takes a value of 1 when students in subgroup i , in school j , and subject area s , are not numerically significant in year $t + 1$ after being exposed to the treatment in year t . All other variables are as defined above. The results for each subgroup that could be tested (low sample sizes made regressions for Whites, Asians, and Native Americans infeasible) are found in Table 1.7. These regressions were run at the average of the subject-

Figure 1.6: Effect of Threat of Sanctions in Year t on Overall Elementary School Student Performance in Year t+1, California, 2002-2010



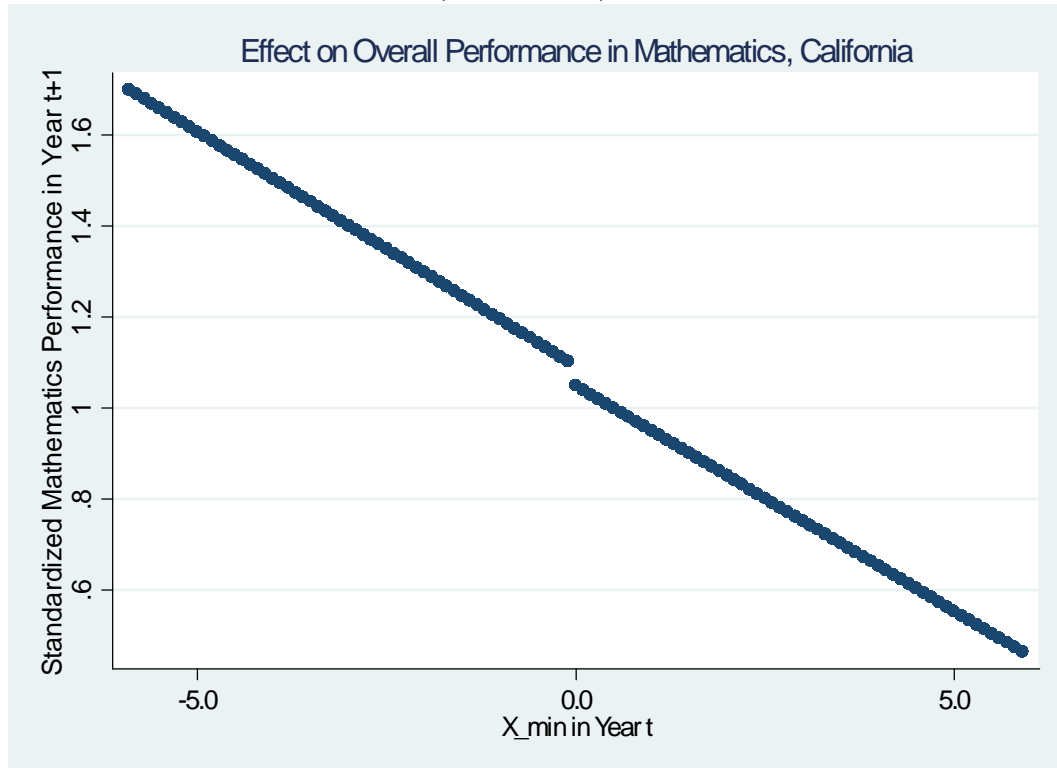
Notes: The results in the graph correspond to the optimal bandwidth specifications in Table 1.6 for language arts literacy. The slope of the regression line on either side of the AYP threshold is that computed within the optimal bandwidth.

specific optimal bandwidths computed above, with standard errors clustered at the county level.

As we can see in the table below, all of the results have the expected signs based on the average performance of each subgroup on the NCLB exams. The subgroups containing African American, Hispanic, Limited English Learner, and disabled students are more likely to become numerically insignificant in the subsequent year when these groups have the lowest exam performance in the current year and a school is threatened with sanctions.

The point estimate for Limited English Learner students is strongly statisti-

Figure 1.7: Effect of Threat of Sanctions in Year t on Overall Elementary School Student Performance in Year t+1, California, 2002-2010



Notes: The results in the graph correspond to the optimal bandwidth specifications in Table 1.6 for mathematics. The slope of the regression line on either side of the AYP threshold is that computed within the optimal bandwidth.

cally significant, suggesting that there is an approximately 12% probability that such students as a whole become numerically insignificant after the school is exposed to the threat of sanctions. These results are only suggestive, since we cannot ascertain for certain why these students are excluded in the subsequent year. It is less likely that Limited English Learners are reclassified so as not to sit for the NCLB exam, because these students could be moved into regular classes without much effect on AYP performance (due to the low number of these students). Since Limited English Learner students are the lowest performing subgroup in approximately 80% of the schools that are exposed to the

Table 1.7: Effect of Threat of Sanctions in Year t on Probability of Numerically Insignificant Subgroup in Year $t+1$, California, 2002-2010

Student Subgroup	Numerically Insignificant in Year $t + 1$
<i>African American</i>	0.282 (0.234) [28]
<i>Hispanic</i>	0.057 (0.062) [32]
<i>Limited English Learner</i>	0.115 (0.039)*** [274]
<i>Students with Disabilities</i>	0.250 (0.329) [24]

Notes: Heteroskedasticity-robust standard errors in parentheses. Sample sizes in brackets. Each row shows results in year $t + 1$ when the student subgroup had the lowest performance in year t for schools without any prior exposure to sanctions. Regression results at optimal bandwidth determined by the Imbens/Kalyanaraman (2009) MSE criterion.

treatment, it may more likely be the case that parents who are dissatisfied with their child's performance are simply moving them to other schools in the district or moving out of the district. The data indicate that approximately 57% of schools in which Limited English Learner students become numerically insignificant are in multiple school districts, further suggesting that parents may be moving their children to other within-district schools.

1.7.2 Changes in Per-Pupil Spending

An additional way to check whether positive increases in student assessment scores in year $t + 1$ are due to substantive policy changes, rather than gaming of the testing system, is to estimate the extent to which the threat of sanctions affects per-pupil spending levels. If spending levels are augmented in areas that may directly increase performance when schools do not meet AYP, this may suggest that positive test score increases are the result of these spending increases.

Since school finance data from New Jersey break out per-pupil spending into categories such as classroom costs, administrative costs, and opera-

tions/maintenance costs, I use these data to test the response in exam scores to student spending at the district level.¹⁴ I am particularly interested in the relative spending response that may occur in areas that most directly improve student performance, such as classroom costs.

I estimate the following model:

$$S_{j(t+1)} = \alpha_{NT} + \tau D + \beta_{NT}(X_{min} - c) + (\beta_T - \beta_{NT})D(X_{min} - c) + S_{jt} + \varepsilon_{j(t+1)},$$

$$\text{s.t. } -h \leq X_{min} - c \leq h \tag{1.14}$$

where $S_{j(t+1)}$ is a per-pupil spending variable in district j in year $t + 1$, S_{jt} is the lagged value of the relevant spending variable, and all other variables are as defined above. All spending variables are expressed in 2002 dollars and then logged. As in earlier models, the variable of interest is D , exposure to the treatment. The regression results are found in Table 1.8.

As we see in the table, there are statistically significant total per-pupil spending increases of approximately 1% due to exposure to the treatment, resulting in a \$113 increase. Total classroom costs, while not estimated precisely, increase by less than 1%, or \$50 per student. Within this category, classroom services (such as amounts paid for speech therapy, classroom equipment, or assembly speakers) decline by approximately 13%, which translates into a reduction of \$12.

While most of the spending categories show increases after exposure to the treatment, the point estimates for classroom spending (which would be expected to directly increase student performance) are quantitatively limited relative to other categories. For example, total administrative costs increased by approximately 2%, food services costs increased by about 5%, and personal services costs (an employee benefits category) increased by almost 8%. Therefore,

¹⁴See <http://nj.gov/education/guide/2011/ind.shtml> for descriptions of each category of per-pupil spending.

Table 1.8: Effect of Threat of Sanctions in Year t on Per-Pupil Spending in Year $t+1$, New Jersey, 2002-2008

Dependent Variable (\$ per pupil)	Estimated Effect (1)	Mean (\$ per pupil) (2)	Computed Effect (\$ per pupil) (3)
ln(Total Costs)	0.011 (0.004)**	\$10,274	\$113
ln(Total Classroom Costs)	0.008 (0.010)	\$6,214	\$50
ln(Classroom Salaries)	0.008 (0.011)	\$5,882	\$47
ln(Classroom Supplies)	0.055 (0.047)	\$239	\$13
ln(Classroom Svcs)	-0.133 (0.112)	\$92	-\$12
ln(Total Support Svcs)	0.025 (0.014)*	\$1,475	\$37
ln(Support Salaries)	0.020 (0.011)*	\$1,292	\$26
ln(Total Admin. Costs)	0.023 (0.019)	\$1,154	\$27
ln(Admin. Salaries)	0.024 (0.023)	\$931	\$22
ln(Plant Oper./Maint.)	-0.007 (0.009)	\$1,183	-\$8
ln(Oper./Maint. Salaries)	-0.0002 (0.018)	\$617	-\$0.12
ln(Food Services Costs)	0.048 (0.154)	\$32	\$2
ln(Extracurricular Costs)	-0.00004 (0.026)	\$186	-\$0.01
ln(Personal Services)	0.078 (0.043)*	\$0.23	\$0.02
ln(Total Equip. Costs)	0.103 (0.197)	\$53	\$5

Notes: Heteroskedasticity-robust standard errors in parentheses. Sample size for all rows is 503. Each row is a separate regression and shows spending in year $t + 1$ for schools exposed to the treatment in year t , and without any prior exposure to sanctions. All spending variables are in 2002 dollars. Mean of spending variables are exponentiated and averaged over all years. Regression results at optimal bandwidth determined by the Imbens/Kalyanaraman (2009) MSE criterion.

while per-pupil spending increases in response to the treatment, this spending does not appear to be primarily focused on those categories that would directly increase student performance. This leaves open the possibility that other factors are responsible for increases in student test scores on the NCLB-designated exams. In multiple school districts, it may also be the case that increases in funding may arise from shifts in spending from non-failing schools to those that do not meet AYP, which would be an unintended consequence of NCLB. Since detailed data on within-district funding transfers are not publicly available, it is

not possible to directly test this hypothesis without access to unpublished data.

1.8 Conclusion

The No Child Left Behind Act of 2001 set out to raise school accountability across the states by imposing financial sanctions on schools whose students do not achieve specific levels of proficiency in English and mathematics. This paper contributes to the school accountability literature by testing whether schools, when threatened with these sanctions in the subsequent academic year, increase student performance so as to avoid being sanctioned. Using a regression discontinuity approach that exploits the discontinuity of treatment as a function of the AYP proficiency cutoff, I find no statistically significant evidence at the optimal bandwidth that student performance in both English and mathematics change after sanctions are threatened. There is statistically significant evidence at larger bandwidths that the treatment effect is positive in both subjects. However, these point estimates are generally larger than those found in more recent studies on student accountability in other states.

I also examine whether schools try to improve the likelihood of achieving AYP by encouraging low-performing subgroups not to sit for a particular NCLB-designated exam. The point estimates in this test provide evidence to support the hypothesis that these subgroups are more likely to become numerically insignificant when they have the lowest exam performance in the prior year. In particular, Limited English Learners, who comprise nearly 80% of the lowest-performing subgroups across all schools, are about 12% less likely to be numerically significant after exposure to the treatment. However, we cannot conclude from this evidence that gaming of the system is occurring. The consistently low performance of Limited English Learners in particular may lead

parents simply to remove these students from the failing school district.

Finally, examining responses to per-pupil spending show that total spending costs increase by about 1% after exposure to the treatment. Total spending on classroom instruction increases by a similar percentage, although this is less than the response levels for spending categories that would not be expected to directly increase student performance. This suggests that other factors may be responsible for increases in student performance on NCLB exams. If increases in funding in multiple school districts result from shifts in spending from non-failing to failing schools, then this may be an unintended consequence of the NCLB legislation. Additional research should be performed across other states to examine both the effects of punitive vs. non-punitive school accountability policies, as well as the potential for gaming of the NCLB examination system by schools.

BIBLIOGRAPHY

- [1] Chakrabarti, R. (2007). "Vouchers, Public School Response and the Role of Incentives: Evidence From Florida", Staff Report Number 306, Federal Reserve Bank of New York.
- [2] Chakrabarti, R. (2008). "Impact of Voucher Design on Public School Performance: Evidence From Florida and Milwaukee Voucher Programs", Staff Report Number 315, Federal Reserve Bank of New York.
- [3] Chiang, H. (2009). "How Accountability Pressure on Failing School Affects Student Performance", *Journal of Public Economics*, 93: 1045-1057.
- [4] Cullen, J. and R. Reback (2006). "Tinkering Toward Accolades: School Gaming under a Performance Accountability System", In: Gronberg, T. and D. Jansen (Eds.), *Improving School Accountability: Check-Ups or Choice*, *Advances in Applied Microeconomics*, 14 Elsevier Science, Amsterdam, pp. 1-34.
- [5] Deere, D. and W. Strayer (2001). "Competitive Incentives: School Accountability and Student Outcomes in Texas", Working Paper. Department of Economics, Texas A&M University.
- [6] Figlio, D. and L. Getzler (2002). "Accountability, Ability, and Disability: Gaming the System", Working Paper 9307. National Bureau of Economic Research.
- [7] Figlio, D. and C. Rouse (2006). "Do Accountability and Voucher Threats Improve Low-Performing Schools?", *Journal of Public Economics* 90: 239–255.
- [8] Gill, B., J.R. Lockwood III, P. Razquin, K. Booker, and R. Zimmer (2007). "State and Local Implementation of the No Child Left Behind Act: Volume I—Title I School Choice, Supplemental Educational Services, and Student Achievement." U.S. Department of Education, Office of Planning, Evaluation and Policy Development.
- [9] Greene, J. (2001). "An Evaluation of the Florida A-Plus Accountability and School Choice Program", Manhattan Institute for Policy Research, New York.
- [10] Hahn, J., P. Todd, and W. van der Klaauw (2001). "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design", *Econometrica*, 46(4): 201-209.

- [11] Imbens, G. and T. Lemieux (2008). "Regression Discontinuity Designs: A Guide to Practice", *Journal of Econometrics*, 142(2): 615-635.
- [12] Imbens, G. and K. Kalyanaraman (2009). "Optimal Bandwidth Choice for the Regression Discontinuity Estimator", National Bureau of Economic Research Working Paper No. 14726.
- [13] Jacob, B. (2005). "Accountability, Incentives, and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools", *Journal of Public Economics* 89: 761-796.
- [14] Jacob, B. and L. Lefgren (2004). "The Impact of Teacher Training on Student Achievement: Quasi-experimental Evidence from School Reform Efforts in Chicago", *Journal of Human Resources*, 39(1): 50-79.
- [15] Lee, D. and T. Lemieux (2010). "Regression Discontinuity Designs in Economics", *Journal of Economic Literature*, 48: 281-355.
- [16] Papay, J., J. Willett and R. Murnane (2011). "Extending the Regression-Discontinuity Approach to Multiple Assignment Variables", *Journal of Econometrics*, 161(2): 203-207.
- [17] West, M. and P. Peterson (2006). "The Efficacy of Choice threats within School Accountability Systems: Results from Legislatively Induced Experiments. *Economic Journal* 116, C46-C62.
- [18] Wong, V., P. Steiner, and T. Cook (2011). "Analyzing Regression Discontinuity Designs with Multiple Assignment Variables: A Comparative Study of Four Estimation Methods", Mimeo., Northwestern University.

CHAPTER 2

A DISCRETE CHOICE ANALYSIS OF THE U.S. DEMAND FOR WINE

2.1 Abstract

Using the Berry (1994) nested logit discrete choice model that directly incorporates features of product differentiation, I analyze the behavioral processes of American wine drinkers and estimate the demand for wine in the United States. Two separate nesting structures that segment the wine market by price-based quality levels and country of origin are implemented. Recognizing the endogeneity of wine prices and within-group market shares, I apply a two-stage least squares (2SLS) instrumental variables estimation procedure. The estimation of demand parameters reveals several facts about the U.S. demand for wine. Under a quality-level nesting structure, price and varietal offerings have a strong impact on the total brand market share relative to the share of the outside good. In addition, American wine drinkers who segment along quality tend to prefer European wines, and exhibit high correlation of tastes at both extremes of the quality spectrum. The origin-based nesting structure indicates similar effects from prices and varietals, but at a lower magnitude than the quality model. Wine drinkers who segment based on origin have strongly heterogeneous preferences towards domestic wines, perhaps reflecting greater product differentiation among U.S. brands. Further, drinkers of European wines show high correlation of preferences, while those who drink Australian wines show low correlation. This latter result may again be due to higher levels of product differentiation among Australian wine brands. Finally, using the demand parameters, the market share own-price elasticity is derived and is shown to be monotone increasing under a quality level nesting structure.

2.2 Introduction

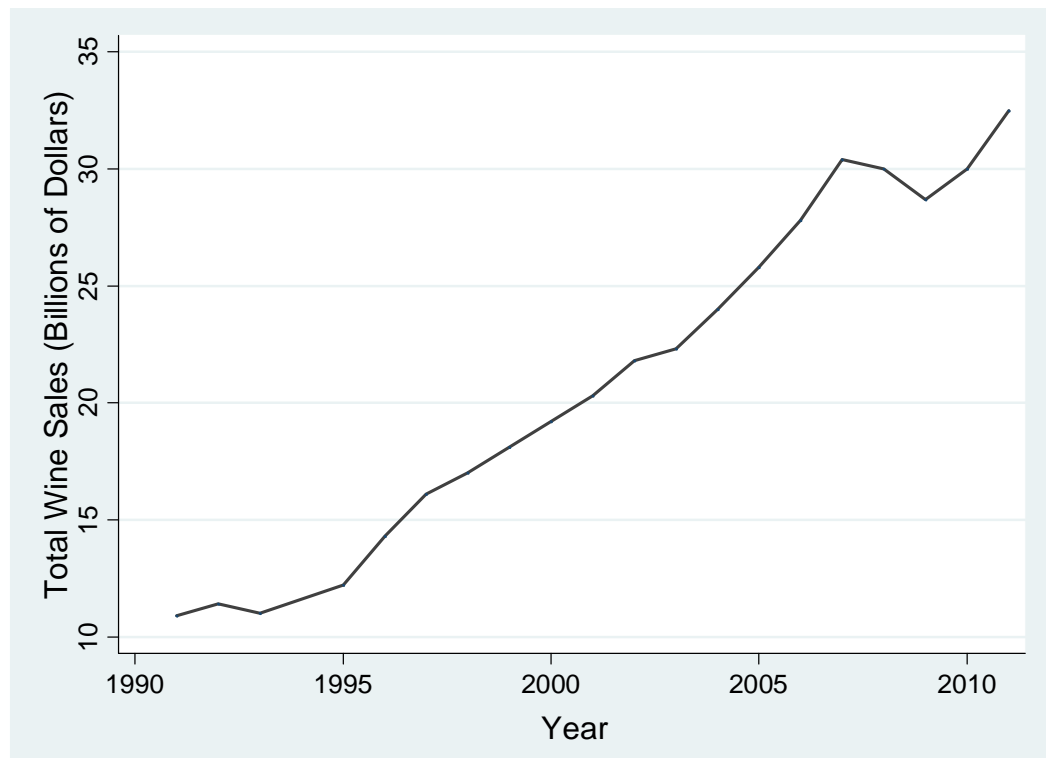
The demand for wine in the United States has been growing continuously for the last twenty years, and has recently seen a rebound after the financial crisis of 2008-2009 (see Figure 2.1). Beverage preferences of American consumers have clearly been undergoing a change during this time, and in a milestone, Americans as a whole now drink more wine than the French.¹ In this paper, my goal is to better understand the demand for wine in the United States by investigating the preferences and decision-making processes of American wine drinkers.

The paper contributes to the wine demand literature by applying a discrete choice analysis that takes into consideration the significant level of product differentiation in the wine industry. In the nested logit model that I implement (Berry, 1994), individual wine brands are assigned to groups within different nesting structures based on quality levels and country of origin. This assignment allows me to understand the level of heterogeneity of preferences for wine among consumers in those groups through estimation of a correlation parameter embedded in the model. Significantly, the model allows for use of market-level data, and accounts for the endogeneity of prices and within-group market shares by deriving a linear estimating equation that allows one to apply instrumental variables methods such as two-stage least squares (2SLS).

Once consumer preferences towards wine are understood in more detail through the application of the present discrete choice model, it is my hope that this knowledge can be put to several uses. Researchers can extend these models to produce more salient research of wine demand, while at the same time pri-

¹See <http://articles.latimes.com/2011/jun/01/business/la-fi-wine-rebound-20110601> for more information on the recent rebound in American consumption of wine. As noted in the article, the French still drink more than Americans do on a per capita basis.

Figure 2.1: Total Wine Sales in the United States: 1991 - 2011 (Source: Wine Institute)



vate wine producers can have a better understanding of consumers' tastes and provide more targeted offerings.

The remainder of the paper is organized as follows: Section 2.3 reviews previous studies of the demand for wine; Section 2.4 discusses the discrete choice theoretical model and derives the wine demand estimating equation; Section 2.5 discusses data sources and construction of variables; Section 2.6 reviews empirical estimates; and Section 2.7 concludes.

2.3 Previous Studies of Wine Demand

There is a large literature on the empirical estimation of wine demand with two primary areas of focus. The first area focuses on applying a demand system

such as the Almost Ideal Demand System (AIDS) to estimate demand parameters and elasticities.² These studies generally focus on the demand for alcoholic beverages, of which wine is a component in the system of demand equations. While demand system models are well-suited to learning about aggregate features of consumer demand theory, such as testing for homogeneity and symmetry restrictions, they cannot be used to learn about consumer preferences towards differentiated products such as wine. In particular, these systems do not allow specific attributes of products to enter the model and are thus unsuited to understanding consumers' tastes for such attributes. Therefore, while the present paper does use aggregated data, I am most interested in estimating the heterogeneity of preferences across different nesting structures and thus apply a nested logit model rather than more general demand systems used in previous work.

The second literature focuses mainly on single-equation analyses of demand for alcoholic beverages.³ While more flexible than demand systems, they suffer from the same inapplicability towards features of demand under product differentiation due to their inability to focus on specific attributes of wine products at the brand level.

There are two previous studies of which I am aware that feature a discrete choice approach to wine demand in the United States. The first⁴ applies this approach to the study of white wine demand. While using a discrete choice model, the analysis is still carried out at an aggregate level and does not specifically take product differentiation into account.

The second study⁵ applies a product differentiation model to wine demand

²See, for example, Heien and Pompelli (1989); Jones (1989); Andrikopoulos, Brox and Carvalho (1997); Blake and Nied (1997); and Chang, Griffeth, and Bettington (2002).

³See, for example, Lee and Tremblay (1992); and Blaylock and Blisard (1993).

⁴Pompelli and Heien, 1991.

⁵Davis, Ahmadi-Esfahani and Iranzo, 2008.

in the United States, but potentially suffers from biased estimates due to sample selection bias. It focuses on the top 50 wine brands in the U.S. and includes in the outside good the remaining wine brands in the data set. Since there are thousands of domestic and foreign wine brands in the U.S. directly competing with the top 50 brands that may be included in the data, it isn't immediately obvious why they should be excluded from the sample.

2.4 Discrete Choice Analysis Under Product Differentiation

2.4.1 The Nested Logit Discrete Choice Model

Discrete choice analysis describes the behavioral process by which individual agents make choices among a countable set of options.⁶ For any product j , we denote the mean utility level for consumer i as:

$$\delta_j \equiv x_j \beta - \alpha p_j + \xi_j \quad (2.1)$$

where x_j are observed product demand characteristics, ξ_j represent the mean valuation of unobserved (by the econometrician) demand characteristics (e.g., regarding product quality), p_j is the price, and β is the mean valuation of the taste parameter.

The standard logit discrete choice model makes the assumption that variation in consumer preferences is embodied in the distribution of unobserved consumer preferences, which is assumed to have an independently and identically distributed extreme value distribution $\exp(-\exp(-\epsilon))$, where ϵ is the distribution of unobserved consumer preferences about ξ_j . The *i.i.d.* assumption

⁶See McFadden (1974) for an early, seminal contribution to the field. Train (2009) provides a comprehensive overview of discrete choice models, including the nested logit model.

imposes strong restrictions on demand substitution patterns by requiring product differences to depend only on the mean utility level for the product, δ_j . This implies that cross-price elasticities are not affected by individual product characteristics or prices.

For example, two products, a and b , that have the same market share will have the same cross-price elasticity with any other product, whether a and b are close substitutes or of very different quality and price. This is known as the *independence from irrelevant alternatives* (IIA) property and appears to be quite unreasonable in the case of wine, for which individual product characteristics are likely to place a large role in differentiating products.

To allow for more flexible correlation patterns, I apply the *nested logit* discrete choice model, which assigns products to a set of groups (or "nests") for which it is assumed that the IIA property holds for products within a group, but does not hold for products across groups. More formally, let there be $j = 1, \dots, N$ products and assign each of the products to $G + 1$ mutually exclusive and exhaustive groups, $g = 0, 1, \dots, G$. The set of products j in group g is referred to as J_g . Group 0 contains one element, $j = 0$, known as the "outside good". The outside good can be purchased instead of one of the products, $j = 1, \dots, N$ and does not respond to price changes among these N grouped goods.

If we let $\mathbf{1}\{j \in J_g\}_{jg}$ be the indicator function that takes a value of one if $j \in J_g$, and zero otherwise, then we can write the utility function for the nested logit model as:

$$u_{ij} = \delta_j + \sum_g [\mathbf{1}\{j \in J_g\}_{jg} \zeta_{ig}] + (1 - \sigma_g) \epsilon_{ij} \quad (2.2)$$

where $0 \leq \sigma_g < 1$ is a correlation parameter for group g , ζ_{ig} is a taste parameter that is common to all goods in group g , and ϵ_{ij} follows an extreme value distribution. Cardell (1997) shows that if ϵ_{ij} follows an extreme value distribution,

then $[\zeta_{ig} + (1 - \sigma_g) \epsilon_{ij}]$ does as well. Equation (2) shows how the nested logit model allows for within-group correlation of tastes for common products as a function of σ_g . As σ_g approaches one, the within-group correlation of utility approaches one; as σ_g approaches zero, the correlation approaches zero. As we can see, this model allows the degree of heterogeneity for consumer preferences to differ across groups. Similar discrete choice models (See Bresnahan, 1981, 1987) are more restrictive in this regard, requiring σ to be the same across groups. Since demand for wine depends on both observable factors (e.g., varietal, country of origin), as well as unobservable factors (e.g., brand quality, perceived brand prestige), it would be seem more realistic that consumers' preferences would differ across different nesting structures. Therefore, we will apply the approach indicated above in the subsequent analysis.

2.4.2 The Nested Logit Model Applied to Wine Choice

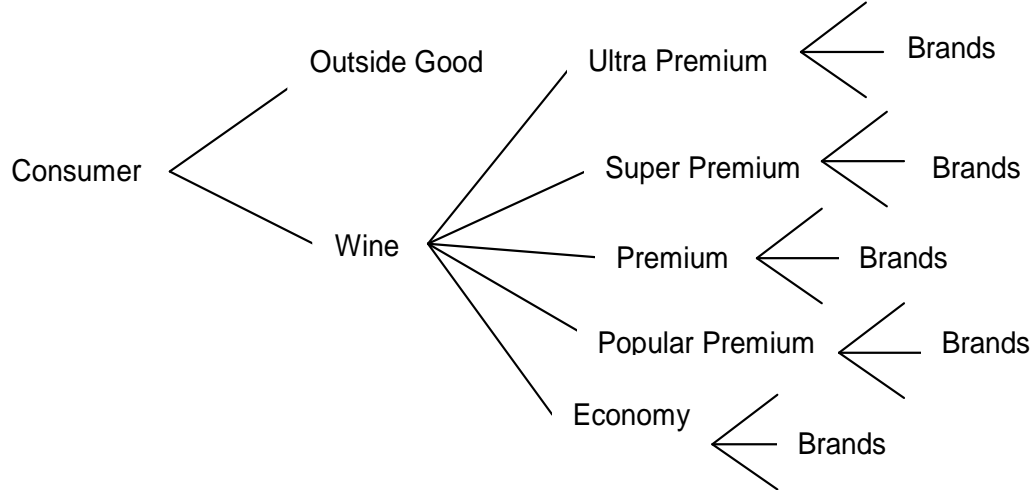
Quality Nesting Structure

Applying the nested logit discrete choice model above to wine choice, the consumer's decision to purchase a particular brand of wine can be modeled as a series of nests, as shown in Figure 2.2. At the top level, the consumer chooses between the outside good and wine. After deciding to purchase wine, at the second level the consumer chooses among a series of product groups.

In this particular nesting structure, I group wine brands according to five price-based quality levels that are commonly used in the wine industry,⁷ $g = 1, 2, 3, 4, 5$, where the groups correspond to Economy, Popular Premium, Pre-

⁷While there are different pricing conventions used in the wine industry, the five quality categories noted above follow a standard segmentation. See wine industry newsletter http://www.winespiritsdaily.com/publications_daily.php?id=12 for the price-based quality categories used in this paper.

Figure 2.2: Wine Choice Model - Quality Nesting



mium, Super Premium, and Ultra Premium, respectively. Table 2.1 describes these quality levels in more detail. As mentioned above, there are $G + 1$ groups in total and the outside good is the only element of group 0. At the bottom nesting level, consumers choose a specific brand of wine from a specific quality category.⁸

Origin Nesting Structure

A second nesting structure that I use to model wine choice is the country or continent of origin. Conditional on purchasing wine, at the second level the consumer chooses among five product groups, $g = 1, 2, 3, 4, 5$, where the groups in this case correspond to the United States, Europe, South America, Australia, and South Africa, respectively. At the bottom nesting levels, consumers again choose among specific wine brands from these groups.⁹

⁸Although this framework sets up a seemingly ordered decision tree, no sequential decision-making process should be inferred. The nested logit model only assumes that individuals choose among specific wine brands; the tree merely reflects correlation patterns among products. See Hensher, Rose, and Greene (2005, chap. 13).

⁹As indicated below, I only focus on regions that appear in the data as producing table wines (i.e., non-fortified, non-sparkling wines).

Table 2.1: Price-Based Wine Quality Levels

Quality Level	Price Range for 750 mL Bottle
Economy	< \$5.00
Popular Premium	\$5.00 - \$7.99
Premium	\$8.00 - \$10.99
Super Premium	\$11.00 - \$14.99
Ultra Premium	\$15 and above

Notes: Specific price ranges from wine industry newsletter winespiritsdaily.com.

Table 2.2 lists the frequency of brands by origin. As we see in the table, the United States and Europe comprise approximately 86% of the data in any given year.

The Berry Product Differentiation Model and Demand Estimation

The nested logit model is frequently estimated at the level of individual consumer choice using consumer microdata (see Train, 2009). However, most wine consumption data, including the data used in this paper, are aggregated up to the state or national level, requiring an aggregate nested logit model based on wine brand market shares. In addition, since wine prices and within-group market shares (discussed below) are endogenously determined, and enter the discrete choice demand framework nonlinearly, this prevents a direct application of linear instrumental variables methods to the wine product choice model.

I therefore apply the theoretical work done by Berry (1994), which shows how to transform a nonlinear nested logit discrete choice model into a linear form that can be estimated with an instrumental variables procedure using market-level data. In particular, for wine brand j in group g , define the within-

Table 2.2: Wine Brands by Country of Origin

Country/State of Origin	Number of Brands		
	2001	2002	2003
Argentina	38	40	44
Australia	140	160	168
California (U.S.)	879	928	931
Chile	84	97	90
France	456	519	466
Germany	101	102	101
Greece	11	14	12
Hungary	12	13	12
Italy	474	527	508
New Zealand	39	48	54
New York (U.S.)	24	27	24
Oregon/Washington (U.S.)	153	159	164
Other U.S.	222	230	213
Portugal	25	28	26
South Africa	54	57	55
Spain	128	136	132
Total	2,840	3,085	3,000

Notes: The following countries had fewer than ten brands in any given year: Austria, Bulgaria, Canada, Israel, Lebanon, Mexico, Morocco, the Netherlands, Romania, Russia, and the former Yugoslavia. Variations in sample sizes due to number of brands in each year with zero recorded sales.

group market share function as:

$$s_{j/g}(\delta, \sigma_g) = \frac{e^{\delta_j/(1-\sigma_g)}}{D_g} \quad (2.3)$$

where the denominator, also known as the "inclusive value" is defined as:

$$D_g \equiv \sum_{j \in J_g} e^{\delta_j/(1-\sigma_g)} \quad (2.4)$$

and can be interpreted as the the total utility from all wines j in group g .

Moving up one level in the nesting structure, the probability of choosing group g (i.e., the group market share function) is defined as:

$$s_g(\delta, \sigma_g) = \frac{D_g^{(1-\sigma_g)}}{\left[\sum_g D_g^{(1-\sigma_g)} \right]}. \quad (2.5)$$

Thus, the total market share function for wine j is the product of the within-group share, $s_{j/g}$, and the probability of choosing group g , s_g :

$$s_j(\delta, \sigma_g) = s_{j/g}(\delta, \sigma_g) s_g(\delta, \sigma_g) = \frac{e^{\delta_j/(1-\sigma_g)}}{D_g^{\sigma_g} \left[\sum_g D_g^{(1-\sigma_g)} \right]}. \quad (2.6)$$

For the outside good, as the only member of $g = 0$, we define the mean utility level as $\delta_o \equiv 0$, implying $D_o = 1$, so that when the outside good is chosen, we have:

$$s_o(\delta, \sigma_g) = \frac{1}{\left[\sum_g D_g^{(1-\sigma_g)} \right]}. \quad (2.7)$$

If we define the total market share of wine j relative to the outside good as $\frac{s_j(\delta, \sigma_g)}{s_o(\delta, \sigma_g)}$ and take logs of this expression, we get:

$$\ln(s_j) - \ln(s_o) = \frac{\delta_j}{(1-\sigma_g)} - \sigma_g \left[\frac{\ln(s_g) - \ln(s_o)}{(1-\sigma_g)} \right] \quad (2.8)$$

where the last term in brackets is $\ln(D_g)$, derived from equation (7). Substituting for δ_j from equation (3) and rearranging terms, we arrive at our estimating equation:

$$\ln(s_j) - \ln(s_o) = x_j \beta - \alpha p_j + \sigma_g \ln(s_{j/g}) + \xi_j. \quad (2.9)$$

I estimate the demand parameters $\theta = (\beta, \alpha, \sigma_g)$ for all $g \in G$ (for both the quality and origin nesting structures) by regressing differences in log market shares on x_j , p_j , and $\ln(s_{j/g})$ using a 2SLS procedure that uses distance to the United States, exchange rates, and a world crop index as instruments for market price and the within-group market share.¹⁰

¹⁰I also fit a three-level nested logit model that comprised the cross of quality levels and countries of origin, producing 25 categories. While the correlation coefficients for the 25 categories are difficult to interpret, the effects of price and varietal on the relative overall market share were similar to the two-level model presented here.

Once the demand parameters are estimated, we can write the own-price elasticity for wine j in group g as:¹¹

$$\varepsilon_{j,j} = \frac{\partial s_j}{\partial p_j} \frac{p_j}{s_j} = \alpha p_j \left[\frac{\sigma_g}{(1 - \sigma_g)} s_{j/g} + s_j - \frac{1}{(1 - \sigma_g)} \right]. \quad (2.10)$$

Using the estimated demand parameters for price and within-group preferences, and the relevant market shares, I calculate the own-price elasticity at the group level for each nesting structure using revenue-weighted averages of the individual brand elasticities in each group.

2.5 Wine Data and Variable Construction

To perform the estimation, I use market-level supermarket and drug store point-of-sale scanner data from a proprietary data set constructed by *IRI InfoScan Reviews* of domestic and foreign wine sales in the entire United States.¹² Each observation in the data is an individual brand of wine. The data cover the period 2001-2003 and include the brand name, the type of wine¹³, dollar sales, average retail price per 750mL bottle, country of origin (including state of origin for U.S. states CA, OR/WA, and NY), a quality description¹⁴ and a varietal/non-varietal¹⁵ indicator.

The data do not include sales at retail wine stores, wineries, or restaurants,

¹¹The equation for own-price elasticity is not indicated in Berry (1994). See the appendix for a detailed derivation.

¹²The data set was originally purchased from *IRI InfoScan Reviews* by the late Frank Vannerson of Princeton, NJ.

¹³The wine types are: 1) Table; 2) Fruit Varietal; 3) Dessert; 4) Fortified; 5) Vermouth-Apertif; and 6) Sake/Plum.

¹⁴As mentioned above, the wine quality types are: 1) Economy; 2) Popular Premium; 3) Premium; 4) Super Premium; and 5) Ultra Premium.

¹⁵A "varietal" is a wine made from a single grape variety such as Merlot or Chardonnay. By U.S. law, a wine must contain at least 75% of the grape variety to allow use of the variety on the wine label (see <http://www.ttb.gov/labeling/index.shtml> for more information on federal labeling requirements). Wines made from more than one grape variety are denoted "non-varietal".

Table 2.3: Descriptive Statistics, Wine Scanner Data, 2002

Variables	(1)	(2)	(3)	(4)	(5)
	N	Mean	Std. Dev.	Min.	Max.
Revenue (2003 dollars)	3,085	1,181,945	7,664,858	15.07	1.56 x 10 ⁸
Price (2003 dollars)	3,085	13.13	9.69	2.24	146.14
Varietal	3,085	0.78	0.42	0	1
Total Market Share, s_j	3,085	0.00003	0.00008	1.46 x 10 ⁻⁹	0.004
<i>Quality Nests</i>					
Economy	3,085	0.03	0.17	0	1
Within-Group, $s_{j/g}$	89	0.01	0.06	4.72 x 10 ⁻⁷	0.56
Popular Premium	3,085	0.24	0.43	0	1
Within-Group, $s_{j/g}$	741	0.001	0.008	3.89 x 10 ⁻⁹	0.11
Premium	3,085	0.29	0.45	0	1
Within-Group, $s_{j/g}$	882	0.001	0.006	5.69 x 10 ⁻⁹	0.11
Super Premium	3,085	0.19	0.39	0	1
Within-Group, $s_{j/g}$	585	0.002	0.01	2.23 x 10 ⁻⁸	0.29
Ultra Premium	3,085	0.25	0.43	0	1
Within-Group, $s_{j/g}$	777	0.001	0.008	6.48 x 10 ⁻⁸	0.19
<i>Origin Nests</i>					
United States	3,085	0.44	0.50	0	1
Within-Group, $s_{j/g}$	1,344	0.001	0.004	2.32 x 10 ⁻⁹	0.05
Europe	3,085	0.43	0.50	0	1
Within-Group, $s_{j/g}$	1,339	0.001	0.005	1.23 x 10 ⁻⁸	0.09
South America	3,085	0.04	0.21	0	1
Within-Group, $s_{j/g}$	137	0.01	0.03	1.15 x 10 ⁻⁷	0.30
Australia	3,085	0.07	0.25	0	1
Within-Group, $s_{j/g}$	208	0.005	0.03	3.04 x 10 ⁻⁸	0.25
South Africa	3,085	0.02	0.13	0	1
Within-Group, $s_{j/g}$	57	0.02	0.04	5.81 x 10 ⁻⁶	0.19

Notes: Quality levels defined in Table 2.1.

and as such only cover approximately 40% of all wine sales in the United States¹⁶. I restrict the sample to brands identified as "table wine", to isolate the U.S. demand for non-fortified, non-sparkling wines. Table 2.3 provides descriptive statistics for the wine scanner data in 2002.

¹⁶See <http://www.symphonyiri.com/SolutionsandServices/Detail.aspx?ProductID=181> for more information on the IRI InfoScan data.

2.5.1 Outside Good, Market Share, and Product Characteristics

Variable Construction

I follow Berry (1994) in computing the outside good for the wine market. Since the nesting structure used in this paper is exhaustive, I assume that when consumers do not choose to purchase wine, they purchase beer or distilled spirits instead. Therefore, I take total alcohol sales in the United States as the base for computing the relevant market.

In examining national-level sales data compiled by the Wine Institute and the Economic Research Service of the U.S. Department of Agriculture,¹⁷ I have determined that the data used in this paper comprise approximately 4% of the total sale of alcohol (wine, beer, and spirits) during the time period studied. I therefore construct the observed total market share, s_j , in the relevant year by dividing the sales for a specific wine brand j by total alcohol sales in that year.

In computing the within-group market share, $s_{j/g}$, for each nesting structure, I use the total revenue for the specific group g as the denominator. The within-group shares are then interacted with a set of nesting dummies to perform the estimation. The share of sales outside of the nesting structure that makes up the remainder of alcohol sales in the United States is used to construct the share of the outside good.

The product characteristics for each wine brand are indicators for varieties and country of origin. I construct a set of origin dummies for use in the quality nesting estimation. The average price per 750 mL bottle is also converted to 2003 dollars for each year to ensure consistent quality assignments.

¹⁷See <http://www.wineinstitute.org/resources/statistics/article639> for historical statistics on table wine sales and http://www.ers.usda.gov/briefing/cpi/foodandexpenditures/Data/Expenditures_tables/table4.htm for historical statistics on total alcohol expenditures in the United States.

2.5.2 Instrumental Variables Construction

Since wine prices, and therefore within-group market shares, are likely to be correlated with unobservable product characteristics that also affect relative total market shares, we must use an instrumental variables procedure to counteract the potential bias caused by this endogeneity. I use three sets of variables to instrument for price and within-market shares: 1) exchange rates;¹⁸ 2) world crop indices;¹⁹ and 3) distance from country of origin to the United States.²⁰ Each observation in all of the instrumental variable data sets is a country-year, covering all of the countries and years included in the *IRI* supermarket data set described above.

Intuitively, these variables would be expected to be correlated with movements in wine prices. I ran first-stage regressions for both the price and within-market share variables and tested the joint significance of the instruments. The F -statistic for the price regression was 30.50 and the statistic for the within-group market share regression was 49.04. Therefore, these instruments satisfy the instrument relevance requirement. The first two are plausibly uncorrelated with unobservable product characteristics, satisfying instrument exogeneity, while crop production may have an indirect effect on wine quality, although this is unclear. The instruments are also interacted with a set of nesting dummies to perform the estimation.

¹⁸Exchange rates were downloaded from the International Monetary Fund's exchange rate archives: http://www.imf.org/external/np/fin/data/param_rms_mth.aspx.

¹⁹World crop production indices by country were downloaded from the World Bank: <http://data.worldbank.org/indicator/AG.PRD.CROP.XD>.

²⁰Distances were calculated using: <http://www.distancefromto.net/countries.php>.

2.6 Empirical Results

2.6.1 Estimated Demand Parameters

Quality Nesting Structure

Table 2.4 shows the estimated demand parameter results of the 2SLS regressions for the quality nesting structure. As we see in the table, wine price has a negative and strongly significant effect on total relative market share, which is expected. Consumers in the U.S. also appear to prefer varietal wines over nonvarietals to a large degree during this time period. The origin dummies suggest that wines from France²¹, Germany, and Italy are preferred by Americans, with wines from Australia slightly behind this grouping. The coefficients on the within-group share variables indicate a large degree of heterogeneity between consumers of wines in differing quality segments. Preferences appear to be highly correlated at both extremes of the quality spectrum, with consumers of Ultra Premium wines having the highest level of homogeneity of tastes.

At the lower end of the spectrum, it may be reasonable to believe that correlation of preferences may be due to infrequent purchases of wine. For example, individuals in this category may have very little knowledge of wine and may follow common cues about varietals or the countries that produce the "best" wines. At the other end of the spectrum, individuals may be very knowledgeable about wine, and may have a common understanding regarding the types of attributes that produce enjoyable wines.

If knowledge is continuous in quality levels, then in the middle of the qual-

²¹One may recall in the popular press a U.S. boycott of French wine shortly after 9/11. There is evidence to suggest that seasonality, rather than changing preferences towards French wine, were responsible for temporary declines in French wine consumption during this period (see Ashenfelter, Ciccarella, and Shatz, 2007).

Table 2.4: Estimated Demand Parameters, Quality Nesting, IV Regression, 2001-2003

Dependent Variable: $\ln(s_j) - \ln(s_o)$	Estimated Parameters
Price (α)	-0.186 (0.066) ***
β	
Varietal	0.433 (0.160) ***
Australia	0.229 (0.128) *
United States	0.169 (0.074) **
Chile	0.057(0.276)
France	0.402 (0.203) **
Germany	0.330 (0.177) *
Italy	0.283 (0.125)**
2002	0.115 (0.092)
2003	0.024 (0.090)
<i>Correlation of Within-Group Preferences (σ_g)</i>	
$\ln(s_{j/g})$: Economy	0.728 (0.145) ***
$\ln(s_{j/g})$: Popular Premium	0.883 (0.116) ***
$\ln(s_{j/g})$: Premium	0.565 (0.180) ***
$\ln(s_{j/g})$: Super Premium	0.779 (0.216) ***
$\ln(s_{j/g})$: Ultra Premium	0.925 (0.185) ***

Notes: 2SLS regression performed using the following instruments: distance to United States, exchange rates and world crop index. Price and within-group market share treated as endogenous variables. Heteroskedasticity-robust standard errors in parentheses. Analysis performed on wine brands with positive sales. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

ity spectrum, it is reasonable to surmise that individuals may be just beginning to acquire information regarding wines and are thus developing more heterogeneous preferences. There may be a greater selection of wines at this level as well, and indeed, wines in the Premium category make up almost a third of all the brands sold in the data (see Table 2.3).

Origin Nesting Structure

Table 2.5 shows the estimated demand parameters for the origin nesting structure. In this structure, the price of wine has a similarly negative effect on to-

Table 2.5: Estimated Demand Parameters, Origin Nesting, IV Regression, 2001-2003

Dependent Variable: $\ln(s_j) - \ln(s_o)$	Estimated Parameters
Price (α)	-0.171 (0.075) **
β	
Varietal	0.222 (0.124) *
Popular Premium	0.300 (0.132) **
Premium	0.266 (0.126) **
Super Premium	0.243 (0.041) ***
Ultra Premium	0.079 (0.114)
2002	0.087 (0.063)
2003	0.185 (0.084) **
<i>Correlation of Within-Group Preferences (σ_g)</i>	
$\ln(s_{j/g})$: United States	0.538 (0.052) ***
$\ln(s_{j/g})$: Europe	0.771 (0.037) ***
$\ln(s_{j/g})$: South America	0.832 (0.055) ***
$\ln(s_{j/g})$: Australia	0.762 (0.069) ***
$\ln(s_{j/g})$: South Africa	0.636 (0.026) ***

Notes: 2SLS regression performed using the following instruments: distance to United States, exchange rates and world crop index. Price and within-group market share treated as endogenous variables. Heteroskedasticity-robust standard errors in parentheses. Analysis performed on wine brands with positive sales. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

tal relative market share. Varietals also are favored by U.S. consumers who segment by country of origin, but not as strongly as in the previous model. The quality dummies suggest that market share relative to the outside good increases in a similar fashion for American wine drinkers of popular premium, premium and super premium quality levels. There is less of a positive effect on market share from consumption of ultra premium brands. The correlation coefficients indicate that American drinkers of South American wine have the lowest heterogeneity of tastes, while those who consume wine produced in the U.S. have the greatest. The latter finding may be explained by the variety of American wine brands available domestically. This is corroborated in Table 2.2,

in which we see that American-produced wines make up approximately 45% of the data. On the other hand, only about 4% of the brands are from South America, which may lead to strong convergence of tastes.

Drinkers of wines from Europe have low heterogeneity of preferences, whereas drinkers of Australian wine appear to exhibit more heterogeneity. While European wines comprise approximately 41% of the data, there may be less overt product differentiation among these wines, relative to wine from the United States or Australia. This may lead to stronger correlation of preferences among these wine drinkers. For the same reason, even though Australian wines comprise about 5% of the data, there may be a greater variety of wine types for American consumers to sample.²²

2.6.2 Elasticities

Tables 2.6 and 2.7 show market-share own-price elasticities for quality and origin segmentations, respectively. As mentioned earlier, these results are revenue-weighted averages of individual brand elasticities that are elements of each specific group. This may be driving the generally larger magnitudes of the elasticities in both tables.

In Table 2.6, we see that own-price elasticities are increasing monotonically in quality levels. This may suggest that drinkers of relatively more expensive wines are more adept at finding substitutes for brands that have increased in price. This would coincide with the idea noted above that these drinkers tend to be more knowledgeable about wine.

In Table 2.7, we find own-price elasticities for drinkers of wines from particular regions. The results suggest that American drinkers of domestic wine ex-

²²See, for example, <http://www.vinodiversity.com/> and <http://www.aavws.com/public/>.

Table 2.6: Wine Market Share Own-Price Elasticity, Quality Nesting, 2001-2003

	Economy	Popular Premium	Premium
2001	-0.962	-1.83	-3.27
2002	-0.859	-1.55	-2.70
2003	-0.841	-1.59	-2.83
Overall	-0.896	-1.68	-2.98
	Super Premium	Ultra Premium	
2001	-4.33	-6.45	
2002	-3.54	-5.23	
2003	-3.74	-5.57	
Overall	-3.94	-5.85	

Table 2.7: Wine Market Share Own-Price Elasticity, Origin Nesting, 2001-2003

	United States	Europe	South America	Australia	South Africa
2001	-1.96	-5.07	-9.19	-4.22	-2.61
2002	-1.97	-5.09	-9.24	-4.25	-2.62
2003	-1.88	-4.85	-8.79	-4.04	-2.50
Overall	-1.94	-5.00	-9.08	-4.17	-2.58

Notes: See equation (2.10) for own-price elasticity formula and the appendix for a derivation.

hibit demand behavior that is the least price-elastic, whereas drinkers of South American wine exhibit behavior that is the most price-elastic.

2.7 Conclusion

In this paper, I have applied a nested logit discrete choice model with two separate nesting structures to examine the behavioral process by which American wine drinkers choose different brands of wine. The theoretical foundation for this model lies in recognizing the high degree of product differentiation inherent in the production of wine. Using market-level data on wine sales, average prices per unit, and product characteristics of individual wine brands, I have used a 2SLS regression procedure to estimate demand parameters from the dis-

crete choice model and have calculated market share own-price elasticities using these parameters.

Applying the quality level nesting structure, I have found that wine prices and varietal wine offerings have a strong impact on total relative market share. In addition, American wine drinkers who focus on quality segmentation tend to prefer European wines, and exhibit a strong correlation of preferences at both extremes of the quality spectrum. Own-price elasticities for this nesting structure are monotone increasing in quality levels, indicating perhaps an ability of more knowledgeable wine drinkers to readily find substitutes for higher-priced wine brands.

The origin-based nesting structure indicates similar effects from prices and varietals, but at a lower magnitude than the quality model. Wine drinkers who segment based on origin have a high level of preference heterogeneity towards domestic wines, perhaps reflecting greater product differentiation among U.S. brands. Further, drinkers of wines from Europe show high correlation of preferences, while those who drink Australian wines show low correlation. This latter result may again be due to higher levels of product differentiation among Australian wine brands.

It should be noted that it is difficult to distinguish between preferences for wine, as reflected in higher prices, and price changes that simply result from higher transport costs. I do not attempt to separate out these effects in this paper, although it could have implications for the correlation results that I find. More research on the wine industry using discrete choice models under price differentiation should be undertaken, so that researchers and wine producers alike can learn more about the behavioral processes of wine choice.

2.8 Appendix. Derivation of Own-Price Elasticity for Wine

Brand j : $\varepsilon_{j,j}$

Since total market share, s_j , is continuous in its arguments, we can use calculus to derive the own-price elasticity:

$$\varepsilon_{j,j} = \frac{\partial s_j}{\partial p_j} \frac{p_j}{s_j} = \alpha p_j \left[\frac{\sigma_g}{(1-\sigma_g)} s_{j/g} + s_j - \frac{1}{(1-\sigma_g)} \right]. \quad (2.11)$$

For convenience, we rewrite the total market share for wine j from equation (2.6) above:

$$s_j(\delta, \sigma_g) = \frac{e^{\delta_j/(1-\sigma_g)}}{D_g^{\sigma_g} \left[\sum_g D_g^{(1-\sigma_g)} \right]} \quad (2.12)$$

where, again:

$$D_g \equiv \sum_{j \in J_g} e^{\delta_j/(1-\sigma_g)} \text{ and } \delta_j \equiv x_j \beta - \alpha p_j + \xi_j. \quad (2.13)$$

Define $k_{jg} \equiv e^{\delta_j/(1-\sigma_g)}$ and $M_g \equiv D_g^{\sigma_g} \left[\sum_g D_g^{(1-\sigma_g)} \right]$. Applying the quotient rule to $\frac{\partial s_j}{\partial p_j}$, we have:

$$\frac{\partial s_j}{\partial p_j} = \frac{M_g \left[\frac{\partial k_{jg}}{\partial p_j} \right] - k_{jg} \left[\frac{\partial M_g}{\partial p_j} \right]}{[M_g]^2}. \quad (2.14)$$

Now, we apply the chain rule to $\frac{\partial k_{jg}}{\partial p_j}$:

$$\frac{\partial k_{jg}}{\partial p_j} = \frac{\partial k_{jg}}{\partial \delta_j} \frac{\partial \delta_j}{\partial p_j} = [e^{\delta_j/(1-\sigma_g)}] \left[\frac{-\alpha}{(1-\sigma_g)} \right] = \frac{-\alpha k_{jg}}{(1-\sigma_g)}. \quad (2.15)$$

Further define $M_{(1)g} \equiv D_g^{\sigma_g}$ and $M_{(2)g} \equiv \sum_g D_g^{(1-\sigma_g)}$. Applying the product rule to $\frac{\partial M_g}{\partial p_j}$, we have:

$$\frac{\partial M_g}{\partial p_j} = M_{(2)g} \left[\frac{\partial M_{(1)g}}{\partial p_j} \right] + M_{(1)g} \left[\frac{\partial M_{(2)g}}{\partial p_j} \right]. \quad (2.16)$$

Applying the chain rule to $\frac{\partial M_{(1)g}}{\partial p_j}$, we get:

$$\begin{aligned}\frac{\partial M_{(1)g}}{\partial p_j} &= \frac{\partial M_{(1)g}}{\partial \delta_j} \frac{\partial \delta_j}{\partial p_j} \\ &= [\sigma_g D_g^{\sigma_g - 1}] [e^{\delta_j / (1 - \sigma_g)}] \left[\frac{-\alpha}{(1 - \sigma_g)} \right] = \frac{-\alpha k_{jg} \sigma_g}{M_{(1)g}^{(1 - \sigma_g) / \sigma_g} (1 - \sigma_g)}.\end{aligned}\quad (2.17)$$

Since $j \in J_g$ for one and only one group g , applying the chain rule to $\frac{\partial M_{(2)g}}{\partial p_j}$ gives us:

$$\begin{aligned}\frac{\partial M_{(2)g}}{\partial p_j} &= \frac{\partial M_{(2)g}}{\partial \delta_j} \frac{\partial \delta_j}{\partial p_j} \\ &= [(1 - \sigma_g) D_g^{-\sigma_g}] [e^{\delta_j / (1 - \sigma_g)}] \left[\frac{-\alpha}{(1 - \sigma_g)} \right] = \frac{-\alpha k_{jg}}{M_{(1)g}}.\end{aligned}\quad (2.18)$$

Therefore, we have:

$$\begin{aligned}\frac{\partial M_g}{\partial p_j} &= M_{(2)g} \left[\frac{-\alpha k_{jg} \sigma_g}{M_{(1)g}^{(1 - \sigma_g) / \sigma_g} (1 - \sigma_g)} \right] + M_{(1)g} \left[\frac{-\alpha k_{jg}}{M_{(1)g}} \right] \\ &= \frac{-\alpha k_{jg} M_{(2)g} \sigma_g}{M_{(1)g}^{(1 - \sigma_g) / \sigma_g} (1 - \sigma_g)} - \alpha k_{jg}.\end{aligned}\quad (2.19)$$

Returning to the original equation for $\frac{\partial s_j}{\partial p_j}$, we have:

$$\begin{aligned}\frac{\partial s_j}{\partial p_j} &= \frac{M_{(1)g} M_{(2)g} \left[\frac{-\alpha k_{jg}}{(1 - \sigma_g)} \right] - k_{jg} \left[\frac{-\alpha k_{jg} M_{(2)g} \sigma_g}{M_{(1)g}^{(1 - \sigma_g) / \sigma_g} (1 - \sigma_g)} - \alpha k_{jg} \right]}{[M_{(1)g}]^2 [M_{(2)g}]^2} \\ &= \frac{-\alpha k_{jg}}{M_{(1)g} M_{(2)g} (1 - \sigma_g)} + \frac{\alpha k_{jg}^2 \sigma_g}{[M_{(1)g}]^{(1 + \sigma_g) / \sigma_g} [M_{(2)g}] (1 - \sigma_g)} + \frac{\alpha k_{jg}^2}{[M_{(1)g}]^2 [M_{(2)g}]^2} \\ &= \frac{-\alpha s_j}{(1 - \sigma_g)} + \frac{\alpha s_j k_{jg} \sigma_g}{[M_{(1)g}]^{1 / \sigma_g} (1 - \sigma_g)} + \alpha s_j^2.\end{aligned}\quad (2.20)$$

After rearranging terms and noting that $s_{j/g} = k_{jg} / [M_{(1)g}]^{1 / \sigma_g}$ (see equation (2.3) above), we get:

$$\begin{aligned}\varepsilon_{j,j} &= \frac{\partial s_j}{\partial p_j} \frac{p_j}{s_j} = \frac{p_j}{s_j} \left[\frac{-\alpha s_j}{(1 - \sigma_g)} + \frac{\alpha s_j k_{jg} \sigma_g}{[M_{(1)g}]^{1 / \sigma_g} (1 - \sigma_g)} + \alpha s_j^2 \right] \\ &= \frac{\alpha p_j k_{jg} \sigma_g}{[M_{(1)g}]^{1 / \sigma_g} (1 - \sigma_g)} + \alpha p_j s_j - \frac{\alpha p_j}{(1 - \sigma_g)} \\ &= \alpha p_j \left[\frac{\sigma_g}{(1 - \sigma_g)} s_{j/g} + s_j - \frac{1}{(1 - \sigma_g)} \right] \blacksquare\end{aligned}\quad (2.21)$$

BIBLIOGRAPHY

- [1] Andrikopoulos, A., J. Brox, E. Carvalho (1997). "The Demand for Domestic and Imported Alcoholic Beverages in Ontario, Canada: a Dynamic Simultaneous Equation Approach", *Applied Economics*, 29(7): 945-954.
- [2] Ashenfelter, O., S. Ciccarella, Jr., and H. Shatz (2007). "French Wine and the U.S. Boycott of 2003: Does Politics Really Affect Commerce?", *Journal of Wine Economics*, 2(1): 55-74.
- [3] Berry, S. (1994). "Estimating Discrete-Choice Models of Product Differentiation", *RAND Journal of Economics*, 25(2): 242-262.
- [4] Blake, D., and A. Nied (1997). "The Demand for Alcohol in the United Kingdom", *Applied Economics*, 29(12): 1655-1673.
- [5] Blaylock, J., and W. Blisard (1993). "Wine Consumption by U.S. Men", *Applied Economics*, 24: 645-651.
- [6] Bresnahan, T. (1981). "Departures from Marginal-Cost Pricing in the American Automobile Industry", *Journal of Econometrics*, 17, 201-227.
- [7] Bresnahan, T. (1987). "Competition and Collusion in the American Automobile Oligopoly: The 1955 Price War", *Journal of Industrial Economics*, 35, 457-482.
- [8] Cardell, N. (1997). "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models with Heterogeneity", *Econometric Theory* 13, 185-213.
- [9] Chang, H., G. Griffith, and N. Bettington (2002). "Demand for Wine in Australia using a Systems Approach: Industry Implications", *Australian Agribusiness Review*, 10: Paper No. 9.
- [10] Davis, T., F. Ahmadi-Esfahani, S. Iranzo (2008). "Demand under Product Differentiation: an Empirical Analysis of the US Wine Market", *Australian Journal of Agricultural and Resource Economics*, 52: 401-417.
- [11] Heien, D. and G. Pompelli (1989). "The Demand for Alcoholic Beverages: Economic and Demographic Effects", *Southern Economic Journal* 55(3): 759-771.

- [12] Hensher, D., J. Rose, and W. Greene (2005). *Applied Choice Analysis: A Primer*. New York: Cambridge University Press.
- [13] Jones, A. (1989). "A Systems Approach to the Demand for Alcohol and Tobacco", *Bulletin of Economic Research* 41(2): 85-101.
- [14] Lee, B. and V. Tremblay (1992). "Advertising and the U.S. Market Demand for Beer", *Applied Economics*, 1: 69-76.
- [15] McFadden, D. (1974). "Conditional Logit Analysis of Qualitative Choice Behavior", in P. Zarembka, ed., *Frontiers in Econometrics*. New York: Academic Press.
- [16] Pompelli, G. and D. Heien (1991). "Discrete/Continuous Consumer Demand Choices: An Application to the U.S. Domestic and Imported White Wine Markets", *European Review of Agricultural Economics* 18: 117-130.
- [17] Train, K. (2009). *Discrete Choice Methods with Simulation*. New York: Cambridge University Press.
- [18] Wine Institute. <http://www.wineinstitute.org/resources/statistics/article639>.

CHAPTER 3

A BAYESIAN IMPUTATION ALGORITHM OF QUARTERLY EARNINGS FOR PART-TIME FEDERAL WORKERS IN THE U.S. CENSUS BUREAU LONGITUDINAL EMPLOYER-HOUSEHOLD DYNAMICS INFRASTRUCTURE FILES

3.1 Abstract

In this paper, I describe a Bayesian imputation procedure of annual hours and quarterly earnings for part-time, seasonal, and intermittent (PTSI) federal workers in the OPM administrative data that are integrated into the LEHD infrastructure files. Using a template probability model composed of an informative Dirichlet prior, I illustrate the functional form for the posterior distribution and the parameters for this distribution. Construction of the estimation sample is discussed and the empirical Bayes procedure that applies the probability model to this sample is described. The resulting data set of imputed hours and earnings is tested, and is found to be both internally and externally consistent. Using 14 quarters of confidential OPM data, I find that the imputed hours distribution is qualitatively similar across all quarters, showing frequency spikes at common intervals with almost identical magnitudes. The modal hours worked in each of the distributions is approximately 35 hours/week. Weekly imputed hours and earnings data are also compared to the Merged Outgoing Rotation Groups (MORG) in the Current Population Survey. The percentage difference between imputed hours and earnings for PTSI workers in the OPM data and part-time workers in the public administration industry in the MORG are approximately 3 and 2 percent, respectively. The match for imputed earnings grows monotonically stronger over time for the years examined.

3.2 Introduction

Administrative records from the Office of Personnel Management's (OPM) Central Personnel Data File (CPDF) provide earnings and employment information for federal workers that are integrated into the existing Longitudinal Employer-Household Dynamics (LEHD) infrastructure files. In contrast to the state UI wage data that are the source for private-sector earnings information in these files, quarterly earnings for federal workers are not recorded by OPM. The OPM data provide only an annual measure of earnings, without additional information regarding the distribution of earnings across quarters. Under the assumption that *full-time* federal workers work approximately equal numbers of hours each quarter, quarterly earnings for these workers can be assigned by taking the annual earnings measure and dividing it equally across each quarter.

In the case of *part-time, seasonal, and intermittent* federal workers (hereafter referred to as PTSI workers), this procedure cannot be replicated. (See Table 3.1 for information on the number of PTSI workers in the public-use OPM data.)¹ OPM provides a full-time annual earnings equivalent for these workers, which does not allow one simply to spread the recorded annual earnings measure equally across quarters to arrive at quarterly earnings. To construct a quarterly earnings measure for PTSI workers, one must first determine the true annual earnings amount, rather than a full-time equivalent. Learning the true PTSI annual earnings in turn depends on determining the number of hours that are worked under these work schedules. With this information known, the true measure can be constructed by taking the proportion of hours worked relative to the number of hours that could have been worked by a full-time worker during the year and then multiplying this fraction by the recorded full-time annual

¹The OPM website www.fedscope.opm.gov/employment.asp contains detailed information on public-use OPM employment data.

Table 3.1: Public-Use OPM Employment, 1998-2011

	All Federal Workers	PTSI Workers	Percentage (%)
Year	(1)	(2)	(3)
1998	1,745,553	115,871	6.64
1999	1,709,918	99,874	5.84
2000	1,697,753	99,810	5.88
2001	1,711,325	98,262	5.74
2002	1,757,725	99,235	5.65
2003	1,785,390	101,797	5.70
2004	1,794,798	111,197	6.20
2005	1,798,766	109,427	6.08
2006	1,802,228	113,221	6.28
2007	1,814,638	112,813	6.22
2008	1,889,459	116,557	6.17
2009	1,993,409	121,189	6.08
2010	2,061,320	125,064	6.07
2011	2,070,375	129,063	6.23

Source: OPM - <http://www.opm.gov/feddata/html/acpdf.asp>

earnings equivalent.

Unfortunately, the number of hours worked is not provided in the OPM data files, which requires one to impute these hours using statistical imputation methods. In this paper, I describe a Bayesian imputation algorithm that imputes hours worked for PTSI federal workers and that satisfies a set of quality assessment criteria. Once hours are imputed, the imputed PTSI annual earnings measure is constructed as described above. The quarterly earnings assignment procedure used for full-time workers can then be replicated by dividing imputed PTSI annual earnings across the quarters depending on the number of quarters worked by each PTSI worker.

The remainder of the paper is organized as follows: Section 3.3 discusses the Bayesian probability model for the OPM hours imputation; Section 3.4 illustrates construction of the estimation and imputation samples; Section 3.5 de-

scribes the empirical imputation procedure; Section 3.6 discusses quality assessment of the hours imputation; and Section 3.7 concludes.

3.3 Probability Model for OPM Hours Imputation

3.3.1 Identifying Assumption

Assume we have data², $\{H_i, Z_i, m_i\}_{i=1}^N$, where H_i is the annual hours worked for individual i , Z_i is a set of covariates for individual i , and m_i is an indicator variable that takes a value of one if the number of hours worked for individual i is observed in the data. $M = (m_1, \dots, m_N)$, known as the *inclusion vector*, can be modeled jointly with the data and describes the missing-data mechanism. The variable of interest is H_i and in imputing the missing outcomes in the data (i.e., those for which $m_i = 0$), we need to identify the probability:

$$\pi(H_i|Z_i, m_i = 0) = \Pr(H_i|Z_i, m_i = 0).$$

We make the identifying assumption that the observations for which $m_i = 0$ are *missing completely at random* (MCAR), such that:

$$\pi(H_i|Z_i, m_i = 0) = \pi(H_i|Z_i, m_i = 1).$$

That is, the probability that individual i works H_i hours, conditional on a set of covariates Z_i , is the same whether that person is observed in the data or not. This assumption allows us to bypass further modeling of the missing-data mechanism and implies that the distribution of M is completely independent of the data (i.e., the missing-data mechanism is *ignorable*). We can therefore use the *observed* data to make inferences regarding the number of hours worked for

²This probability model builds on prior work done by John Abowd, Melissa Bjelland, and Ian Schmutte for the LEHD Human Capital Estimates Project.

individuals who are *missing* in the data by sampling from the derived posterior distribution.

3.3.2 Bayesian Inference

Now assume we have data, $\{H_i, Z_i\}_{i=1}^N$, as defined above, and we would like to learn about the probability $\pi(H_i|Z_i)$ (where we have suppressed the missingness indicator, m_i). In the subsequent analysis, we want to focus on predicted probabilities, π , for hours-covariate cells that are indexed by the number of annual hours worked by a PTSI federal worker and the unique combination of covariates for that worker. More precisely, let $R = 35 \times 52 = 1,820$ be the number of possible hours that a PTSI worker can work annually (a maximum of 35 hours per week for 52 weeks) and let $C = 14,400$ be the number of possible covariate combinations (see section 3.4.1 below). Then matrix HZ with row dimension C and column dimension R has typical element π_{cr} , such that:

$$c \in Z = (Z_1, \dots, Z_C) \text{ and } r \in H = (H_1, \dots, H_R),$$

and contains $C \times R = 14,400 \times 1,820 = 26,208,000$ hours-covariate cells:

$$HZ = \begin{pmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1,1820} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2,1820} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{14400,1} & \pi_{14400,2} & \cdots & \pi_{14400,1820} \end{pmatrix}.$$

Since we will be interested in subsample counts defined by hours-covariate cells, we can rewrite the data vector as $\{H_i\}_{i=1}^{N_Z}$, where N_Z is the number of observations for a specific covariate combination, Z .

Likelihood

The number of annual hours worked by PTSI federal worker i , H_i , has discrete support H , with R possible outcomes. We can therefore characterize the sampling distribution, $\pi(H_i|Z_i)$, over these possible outcomes as *multinomial* with parameter $\theta = (\theta_1, \dots, \theta_R)$, such that $\sum_{r=1}^R \theta_r = 1$. If we let $\mathbf{1}(\cdot)$ be the indicator function that equals one when the condition expressed in the parentheses is true, and zero otherwise, and assume the observations are i.i.d., the likelihood function for $\pi(H_i|Z_i)$ can be written (up to a proportionality constant) as:

$$\begin{aligned}\pi(H|\theta, Z) &\propto \prod_{i=1}^{N_Z} \prod_{r=1}^R \theta_r^{\mathbf{1}(H_i=H_r)} \\ &= \prod_{r=1}^R \theta_r^{N_{Z_r}},\end{aligned}$$

where $N_{Z_r} = \sum_{i=1}^{N_Z} \mathbf{1}(H_i = H_r)$. We interpret N_{Z_r} as the count of individuals with covariate combination Z and annual number of hours worked equal to H_r .

Prior Distribution

The prior distribution reflects our beliefs regarding the distribution of the multinomial parameter, θ , prior to knowledge of the data. We select an informative prior for θ that ensures an analytic solution for the posterior distribution that follows the same parametric form as the prior distribution. This latter property is known as *conjugacy*, and the conjugate prior for the parameters of the multinomial likelihood is the *Dirichlet* distribution (a multivariate generalization of the beta distribution). We can write the Dirichlet prior distribution with hyperparameter $\alpha = (\alpha_1, \dots, \alpha_R)$ as:

$$\pi(\theta|\alpha, Z) = \frac{1}{B(\alpha)} \prod_{r=1}^R \theta_r^{\alpha_r - 1},$$

where $\alpha_r > 0, \forall r \in R$ and $\sum_{r=1}^R \alpha_r = 1$. The normalizing constant, $B(\cdot)$, is the multinomial beta function, which can be expressed in terms of the gamma

function as:

$$B(\alpha) = \frac{\prod_{r=1}^R \Gamma(\alpha_r)}{\Gamma\left(\sum_{r=1}^R \alpha_r\right)}, \text{ with } \Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt, \text{ for } z \in \mathbb{C}.$$

Posterior Distribution

The posterior distribution, $\pi(\theta|H, Z)$, summarizes all of the information regarding the multinomial parameter, θ , after using the data to update our prior beliefs. It is therefore a compromise between the data likelihood, $\pi(H|\theta, Z)$, and the prior distribution, $\pi(\theta|\alpha, Z)$. Using Bayes' Rule, we can write the posterior distribution of θ as:

$$\pi(\theta|H, Z) = \frac{\pi(H|\theta, Z)\pi(\theta|\alpha, Z)}{\int_{\theta} \pi(H|\theta, Z)\pi(\theta|\alpha, Z)d\theta'}$$

where the integral in the denominator is calculated over the support of θ . To derive the posterior distribution, we first derive the numerator:

$$\begin{aligned} \pi(H|\theta, Z) \pi(\theta|\alpha, Z) &= \prod_{r=1}^R \theta_r^{N_{Z_r}} \cdot \frac{1}{B(\alpha)} \prod_{r=1}^R \theta_r^{\alpha_r-1} \\ &= \frac{1}{B(\alpha)} \prod_{r=1}^R \theta_r^{N_{Z_r} + \alpha_r - 1}. \end{aligned}$$

Now we consider the denominator:

$$\begin{aligned} \int_{\theta} \pi(H|\theta, Z) \pi(\theta|\alpha, Z) d\theta &= \int_{\theta} \frac{1}{B(\alpha)} \prod_{r=1}^R \theta_r^{N_{Z_r} + \alpha_r - 1} d\theta \\ &= \frac{1}{B(\alpha)} \int_{\theta} \prod_{r=1}^R \theta_r^{N_{Z_r} + \alpha_r - 1} d\theta. \end{aligned}$$

Note that for any numbers $x, y \in \mathbb{R}^+$ we can write the multinomial beta function, $B(\cdot)$, as:

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt.$$

Replacing x and y with $\alpha = (\alpha_1, \dots, \alpha_R)$; t and $(1-t)$ with $\theta = (\theta_1, \dots, \theta_R)$; and noting that $\theta \in [0, 1]$, we have:

$$B(\alpha) = \int_{\theta} \prod_{r=1}^R \theta_r^{\alpha_r-1} d\theta, \text{ and therefore, } B(N_Z + \alpha) = \int_{\theta} \prod_{r=1}^R \theta_r^{N_{Z_r} + \alpha_r - 1} d\theta.$$

Thus the denominator of the posterior is $\frac{B(N_Z + \alpha)}{B(\alpha)}$, where $N_Z = (N_{Z_1}, \dots, N_{Z_R})$. We can therefore write the posterior distribution of the multinomial parameter, θ , as:

$$\begin{aligned} \pi(\theta|H, Z) &= \frac{\pi(H|\theta, Z)\pi(\theta|\alpha, Z)}{\int_{\theta} \pi(H|\theta, Z)\pi(\theta|\alpha, Z)d\theta} \\ &= \frac{1}{B(\alpha)} \prod_{r=1}^R \theta_r^{N_{Z_r} + \alpha_r - 1} \cdot \frac{B(\alpha)}{B(N_Z + \alpha)} \\ &= \frac{1}{B(N_Z + \alpha)} \prod_{r=1}^R \theta_r^{N_{Z_r} + \alpha_r - 1}. \end{aligned}$$

This posterior is Dirichlet with hyperparameter $N_Z + \alpha = (N_{Z_1} + \alpha_1, \dots, N_{Z_R} + \alpha_R)$. As noted above, this result is an implication of our choosing a conjugate prior for the parameters of the multinomial likelihood. We will sample from this posterior distribution to assign multinomial probabilities to hours-covariates cells, and subsequently use those probabilities to impute hours worked to PTSI federal workers.

3.4 Data

The data used for the hours imputation are broken down into two main categories: 1) the estimation sample and 2) the imputation sample.³

3.4.1 Estimation Sample

The estimation sample is used in the imputation procedure both to generate the likelihood, as well as to construct the parameters for the Dirichlet prior distribution. We construct this data set by linking information on reported hours

³Sample construction and the empirical Bayes procedure are built on prior work done by Ian Schmutte for the LEHD Human Capital Estimates Project.

worked for individuals found in survey data with a set of conditioning variables for those same individuals found in the LEHD infrastructure files.

Hours Worked Variables

Information on hours worked comes from two sources: 1) the 2000 Decennial Sample Census Edited File (SCEF)⁴ and 2) the Current Population Survey (CPS). The SCEF contains weighted 100-percent and sample characteristics for individuals in the Decennial long form sample and is used to generate sufficiently large cell counts in the estimation sample. The CPS is a monthly survey of households and is used to account for time-series variation in the reported number of hours worked. Each observation in the data is an individual worker.

Annual hours worked are reported in both the SCEF and the CPS as the product of hours worked per week and the number of weeks worked in the previous year. Hours per week are restricted to lie between 1 and 99 in these surveys, however we restrict the maximum number of hours for PTISI workers to lie between 1 and 35. Thus, as noted above, the maximum number of hours that can be imputed for these workers is $1,820 = 35 \times 52$.

Conditioning Variables

The conditioning variables, Z , are a set of demographic and earnings characteristics that stratify the set of PTISI workers. Specific combinations of characteristics for each worker imply the assignment of specific annual hours worked in the imputation procedure. As we will see below, the aggregate count of workers in each cell that is formed from the combinations of the conditioning

⁴See <http://www.census.gov/main/www/cen2000.html> for more information on demographic and earnings variables in the 2000 Decennial Census, as well as information on the SCEF.

variables will determine the likelihood and prior values in the construction of the posterior distribution.

In each year, there are six conditioning variables derived from demographic and earnings variables found in the LEHD infrastructure files (Abowd et al., 2009):

- male - an indicator variable derived from *sex*
- white - an indicator variable derived from *race*
- born_us - an indicator variable derived from *place of birth (pob)*
- nempl_cat - number of jobs held (3 maximum), derived from earnings histories
- sixqwindow - a binary variable of the six-quarter earnings history for the worker, determined by employment in the four quarters of the contemporaneous year and the quarters immediately before and after that year
- decile - the 1999 earnings decile of the worker's income (in 2000 dollars), derived from earnings histories.

The 'male', 'white', and 'born_us' variables are transformed from the indicated demographic variables (*sex*, *race*, *pob*), which are found in the LEHD Individual Characteristic File (ICF) and Person Characteristics File (PCF). The indicator variable transformations from the ICF/PCF to the estimation sample (ES) are as follows: 1) ICF ($sex = 1/sex = 0$) \iff ES (male = 1/male = 0); 2) ICF ($race = white/race \neq white$) \iff ES (white = 1/white = 0); 3) ICF ($pob = USA/pob \neq USA$) \iff ES (born_us = 1/born_us = 0). The 'nempl_cat', 'sixqwindow', and 'decile' variables are derived from earnings histories found in the Employment History File (EHF) and Person History File (PHF).

There are 14,400 combinations of the conditioning variables and 1,820 possible imputed values for hours worked, producing an annual matrix of 26,208,000 hours-covariate cells. Only workers between the ages of 14-85 who have positive hours reported in the SCEF and CPS, as well as positive earnings in the EHF/PHF, are selected for the estimation sample.

3.4.2 Imputation Sample

The imputation sample comprises the set of PTSI workers in the relevant OPM data file. We construct the set of conditioning variables in this file so that it can be merged with the file of covariate-combination-specific multinomial probabilities that are generated from the posterior distribution. Once these probabilities are attached to the imputation sample, they are used to impute annual hours for each worker.

The OPM data set used to construct the conditioning variables is the quarterly *status file*. The status file is a snapshot of the federal workforce. Each observation is an individual federal worker and includes demographic information (e.g., sex, race, ethnicity) and job information (e.g., occupation, pay grade, duty station) for each worker (see OPM, 1990-2012). Since the OPM data does not contain the same demographic variables as the ICF/PCF, we use different mappings to construct the covariates.⁵

Specifically, the 'male' and 'white' indicator variables are transformed from similar 'sex' and 'race' variables in the status file as those in the ICF described above. However, OPM does not record information on the place of birth variable used in the estimation sample. Instead, an indicator for whether a federal worker is a U.S. citizen (citizen), is used as a proxy for place of birth. If the

⁵See OPM: Guide to Data Standards, 2012 for a codebook of variables in the OPM Central Personnel Data File.

worker is a U.S. citizen (citizen = 1), then that individual is coded as being born in the United States. This variable is matched to the 'born_us' indicator variable in the estimation sample. The conditioning variables that are derived from earnings histories in the EHF/PHF are similarly constructed from the OPM status file data. Specifically, individuals who appear in the relevant status files over the six-quarter window are coded as having worked in the respective quarters.⁶

3.5 Empirical Hours/Earnings Imputation Procedure

We now describe how the probability model for the OPM hours imputation is applied to the data.

3.5.1 Likelihood

As noted above, the likelihood values, $N_{Z,r}$, are determined by aggregate counts of workers in each hours-covariate cell that is formed from different combinations of the conditioning variables in the estimation sample. Since the estimation sample draws from both the SCEF and the CPS, their weighted likelihood contributions to the posterior distribution are described by the following linear combination:

$$N_Z = \lambda \cdot L_{SCEF} + (1 - \lambda) \cdot L_{CPS},$$

where L_{SCEF} and L_{CPS} are the SCEF and CPS likelihood contributions, and the mixing parameter $\lambda = \frac{41}{66}$ is the weight placed on the SCEF contribution.

⁶An exception is in the construction of the `nempl_cat` variable, which records the number of jobs held by an individual. We make the simplifying assumption that the majority of federal workers hold one position, and do not have additional jobs in either the public or private sectors. Thus the `nempl_cat` variable is set equal to 1 for workers in the imputation sample.

3.5.2 Prior Distribution

We use an informative Dirichlet prior distribution with shape parameter values that equal the proportion of workers in the estimation sample appearing in a restricted set of hours-covariate cells. In particular, the proportions are based on worker counts in cells of white-male-nempl_cat combinations for all values of annual hours that can be imputed. Since there are twelve categories of this restricted set of conditioning variables, the total number of cells used for the prior is $21,840 = 12 \times 1,820$.

To smooth the posterior distribution, an uninformative uniform prior is added to the empirical prior, such that the complete Dirichlet shape parameter is the linear combination:

$$\alpha = 0.99A + 0.01B$$

where A is the empirical prior (i.e., the set of cell frequencies) and B is the uniform distribution over the set of 1,820 possible imputed annual hours.

3.5.3 Posterior Distribution

The parameter for the Dirichlet posterior distribution, $N_Z + \alpha$, is computed by taking the sum of the likelihood counts and prior proportions in the last two sections:

$$N_Z + \alpha = [\lambda \cdot L_{SCEF} + (1 - \lambda) \cdot L_{CPS}] + [0.99A + 0.01B]$$

where the variables are as described above. To construct the posterior distribution for PTSI workers, I modified a complete posterior distribution for all workers (full-time and part-time, federal and private) that satisfied the above-

described parametrization.⁷ The modifications were two-fold: 1) I first restricted the number of hours that could be worked annually for all workers in the estimation sample. Since the maximum number of hours that can be worked by a PTSI workers is 1,820, this restriction deleted columns 1,821 to 5,148 from the complete table of multinomial probabilities; and 2) I renormalized the rows of the modified matrix, so that all of the row probabilities summed to one. This modified posterior distribution was then used to impute missing hours/earnings for PTSI workers.

3.5.4 Hours/Earnings Imputation

The imputation procedure for annual hours and earnings for part-time, seasonal/non-intermittent (PTS) workers⁸ is performed in the following steps for each of the 14,400 combinations of the conditioning variables, Z , in each year:

- **Step 1:** Draw once from the Dirichlet posterior to get the multinomial parameter, $\hat{\theta}$, and probabilities for each hours-covariate cell
- **Step 2:** Draw once from a multinomial distribution with parameter $\hat{\theta}$ to get imputed annual hours, $H \in (1, \dots, 1,820)$, for each PTS worker
- **Step 3:** Divide imputed PTS annual hours, H , by the total number of annual hours for full-time federal workers (2,080)⁹ to get the deflation value, δ .

⁷The complete posterior was constructed by Ian Schmutte for the Census Bureau's Human Capital Estimates Project hours imputation.

⁸Since seasonal workers may work full-time, but fewer than 12 months during the year, we can treat them similarly in the imputation procedure to part-time workers. In the data, both sets of workers work less than full time (annually), and both sets have an annual full-time earnings equivalent reported.

⁹An assumption is made that full-time federal workers will work a standard 40-hour work-week.

- **Step 4:** Multiply the deflation value, δ , by the reported annual earnings equivalent in the OPM data to get the imputed annual PTS earnings measure
- **Step 5:** Divide the imputed annual PTS earnings measure by the number of quarters worked to get the imputed quarterly PTS earnings measure.

Since the work schedule for intermittent workers is irregular, by definition, we modify Step 5 of this algorithm by assigning imputed earnings equally across the four quarters, or when a regularly worked set of quarters is observed in the data for a specific worker, we only assign imputed earnings to those observed quarters.¹⁰

3.6 Quality Assessment

We analyze the results of the hours and earnings imputations to determine whether they are internally and externally consistent.

3.6.1 Internal Consistency

We aim for consistent hours distributions for PTSI workers across years/quarters given that the aggregate demographic information on which the conditioning variables are derived do not have significant differences across time periods. The results from the hours imputation were tested using 14 consecutive quarters of the OPM data (2000Q1 - 2003Q4). Annual hours for PTSI workers were imputed in each quarter and frequency distributions were graphed by annual hours.

¹⁰For example, if a worker is observed in the data as working in November and December over a sufficiently long time period, then we assign the imputed earnings to the fourth quarter for that worker.

The imputed hours frequency graphs for all PTSI workers, as well as for part-time non-seasonal workers, are displayed in Figures 3.1-3.16 and Figures 3.17-3.32, respectively (see appendix). Qualitatively, the graphs are almost identical for each of the 16 quarters. The imputed hours frequencies show spikes at regular hours intervals in each year/quarter (400, 600, 1000, 1500, 1800), corresponding approximately to 8-, 12-, 20-, 30-, and 35-hour workweeks. The spikes are also of the same magnitude in each of the graphs, with the modal imputed hours worked for PTSI workers equal to 1800 hours (35 hours/week).

3.6.2 External Consistency

We also aim for imputed hours and earnings for PTSI workers that are consistent with mean hours and earnings information reported in survey data. The imputed OPM data for PTSI workers for the years 2000-2003 were compared with the Merged Outgoing Rotation Groups (MORG) data in the CPS. The MORG data contains detailed information on usual weekly earnings and hours worked for the set of households followed monthly in the CPS.¹¹

In testing for external consistency, we restrict the imputation sample to part-time non-seasonal workers to be consistent with the CPS part-time worker designation, which does not explicitly account for seasonal or intermittent workers. In the MORG data, part-time workers with positive hours and earnings in the public administration industry (2002 NAICS: 9370-9590) were selected for comparison. *Exact values for mean hours and earnings cannot be displayed in this paper due to confidentiality reasons.*

Differences in (weighted) means for samples with unequal variances and

¹¹We do not compare annual earnings between the OPM and MORG data, since the variable asking how many weeks the interviewee has worked is only asked of 12% of workers in the MORG files.

Table 3.2: Imputed OPM Hours/Earnings and MORG Data Comparisons, 2000-2003

	Imputed Weekly Hours	
Year	Test Statistic	Percentage Difference (%)
2000	3.49	4.18
2001	4.24	4.33
2002	3.34	3.49
2003	1.25	1.30
Average	3.08	3.33
	Imputed Weekly Earnings	
Year	Test Statistic	Percentage Difference (%)
2000	1.52	3.17
2001	1.10	2.13
2002	0.52	1.02
2003	0.17	0.34
Average	0.83	1.67

Notes: The test statistic is a t-statistic derived from a difference in means calculation with unequal variances and sample sizes. Sample sizes are as follows: MORG: 2000 (688); 2001 (871); 2002 (805); 2003 (794), OPM: 2000 (39,637); 2001 (32,734); 2002 (33,614); 2003 (32,243). OPM imputed hours restricted to part-time non-seasonal workers. MORG workers are part-time workers in the public administration industry with positive hours worked and weekly earnings. Weekly hours and earnings in OPM data based on a 52-week workyear.

sample sizes were tested between the imputed OPM data and the MORG data for both weekly hours and weekly earnings. Percentage differences in means were also computed. The results are displayed in Table 3.2.

As we can see from the table, the imputed weekly hours differ from weekly hours in the MORG by approximately 3 percent, although the test statistic is statistically significant on average. The imputed weekly earnings differ on average by approximately 2 percent, and the match appears to grow monotonically stronger over time. The differences between mean earnings in the imputed OPM data and the MORG data are not statistically significant for the years that were tested.

3.7 Conclusion

In this paper I have described a Bayesian imputation procedure that imputes annual hours worked and corresponding earnings for part-time, seasonal, and intermittent federal workers in the OPM personnel data files. This procedure is necessary because OPM only reports full-time annual equivalent earnings for these workers, which does not account for their restricted part-time and intermittent work schedules. Using a template probability model composed of an informative Dirichlet prior, the functional form for the posterior distribution and the parameters for this distribution are illustrated.

The empirical Bayes procedure is described and the results of the imputation for hours and earnings are discussed. The empirical procedure produces results that are both internally and externally consistent for 14 consecutive quarters of data tested. In particular, the frequency distribution of imputed hours is qualitatively similar across these quarters. Spikes in the data are found at almost identical intervals, with similar magnitudes. The modal number of hours worked is approximately 35 hours/week in all quarters.

The imputed data are also matched in a consistent way to that found in the CPS Merged Outgoing Rotation Groups (MORG). For part-time workers in the public administration industry with positive hours and earnings in the MORG, we showed that the imputed OPM data differ by approximately 2 percent for weekly earnings and by approximately 3 percent for weekly hours worked. The match for imputed earnings grows monotonically stronger over time for the years examined.

3.8 Appendix. Imputed Hours Frequency Graphs

Figure 3.1: Imputed Hours Frequency, all PTSI Workers, 2000Q1

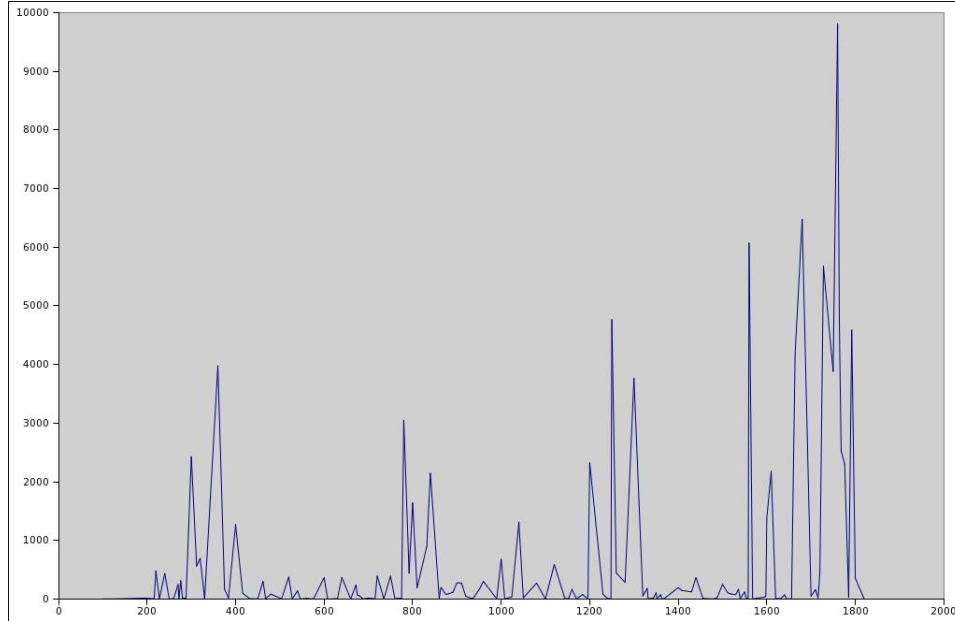


Figure 3.2: Imputed Hours Frequency, all PTSI Workers, 2000Q2

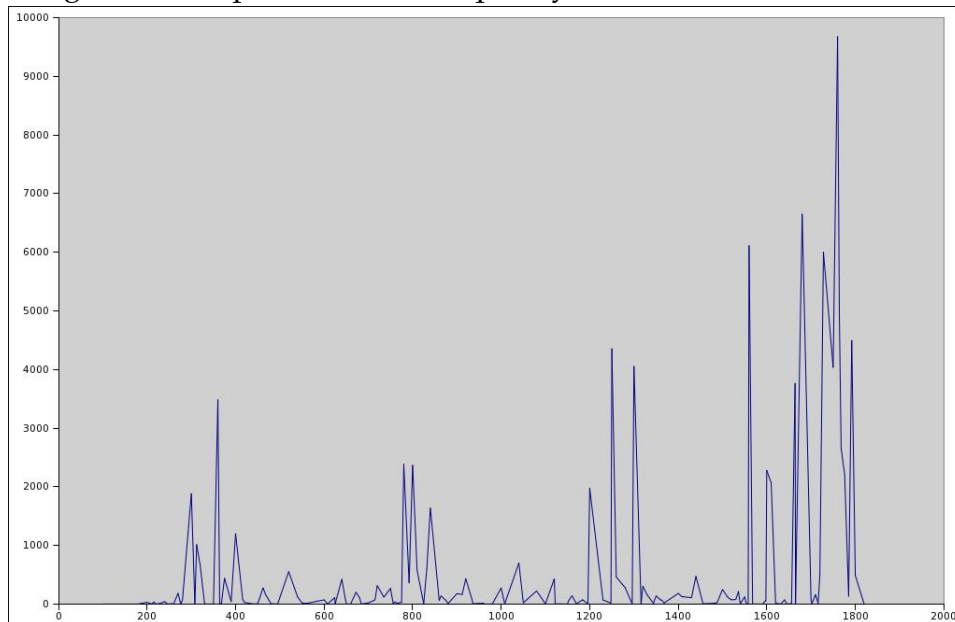


Figure 3.3: Imputed Hours Frequency, all PTSI Workers, 2000Q3

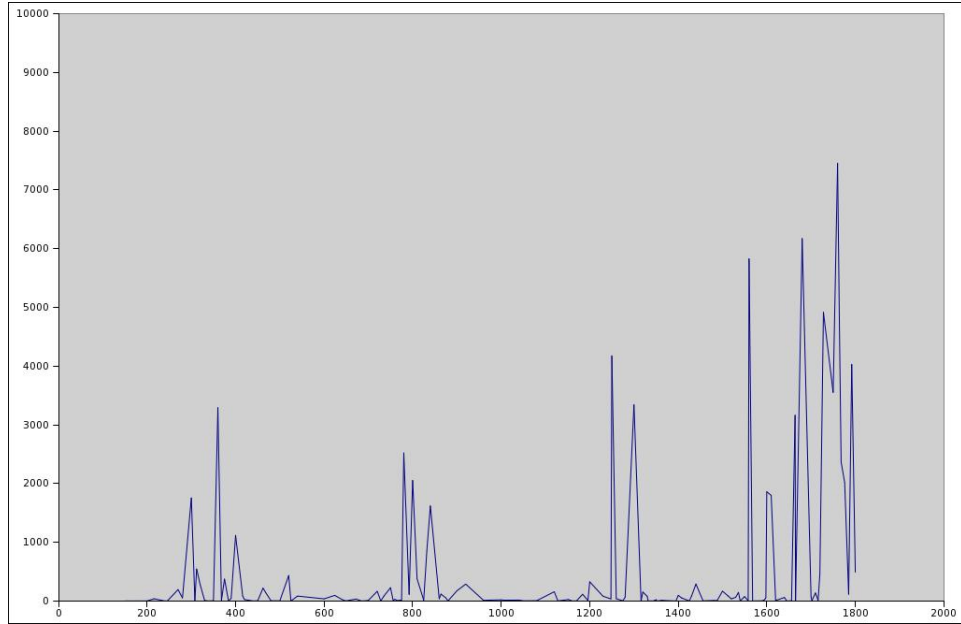


Figure 3.4: Imputed Hours Frequency, all PTSI Workers, 2000Q4

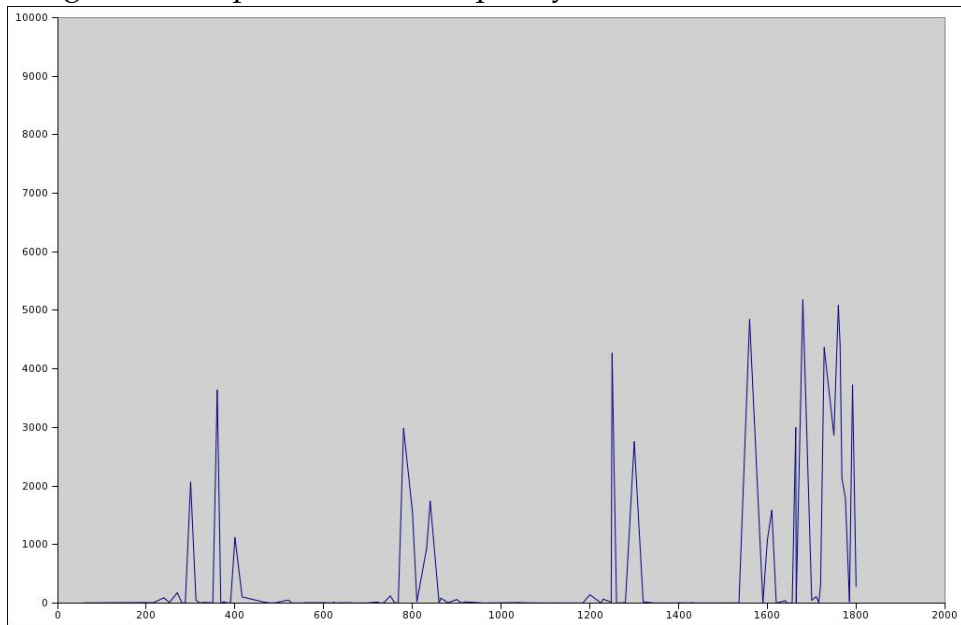


Figure 3.5: Imputed Hours Frequency, all PTSI Workers, 2001Q1

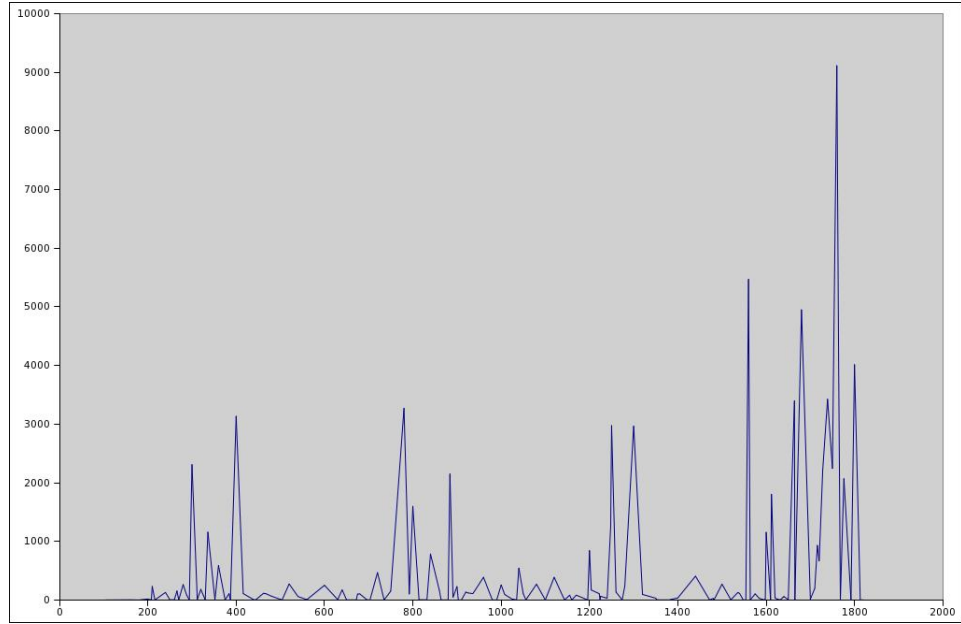


Figure 3.6: Imputed Hours Frequency, all PTSI Workers, 2001Q2

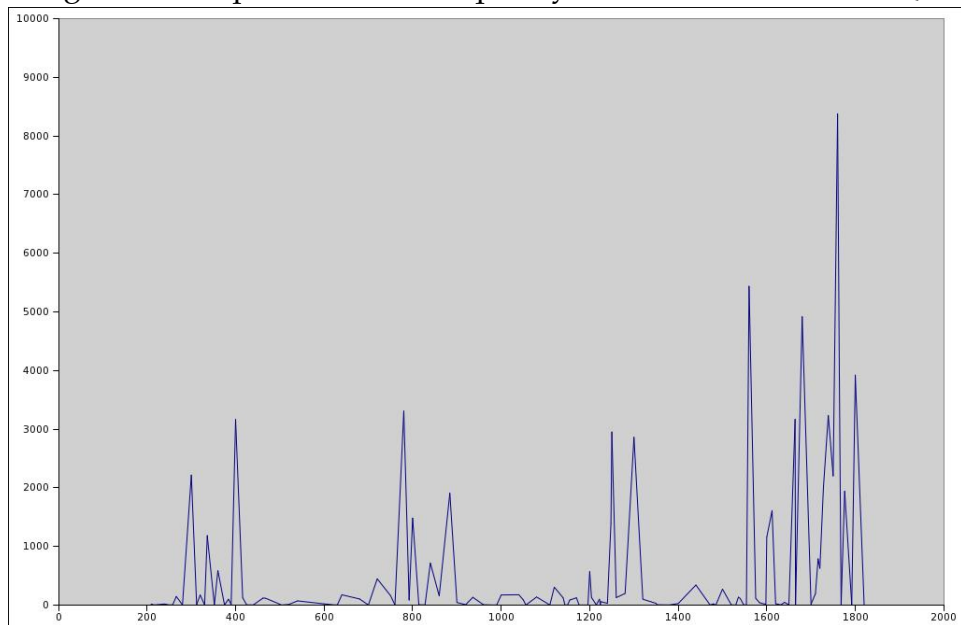


Figure 3.7: Imputed Hours Frequency, all PTSI Workers, 2001Q3

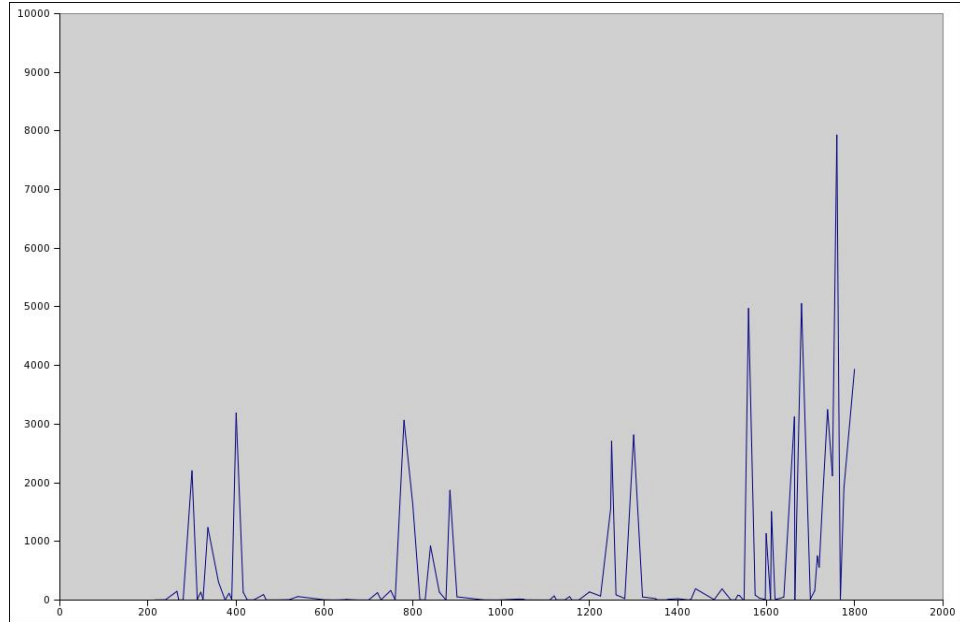


Figure 3.8: Imputed Hours Frequency, all PTSI Workers, 2001Q4

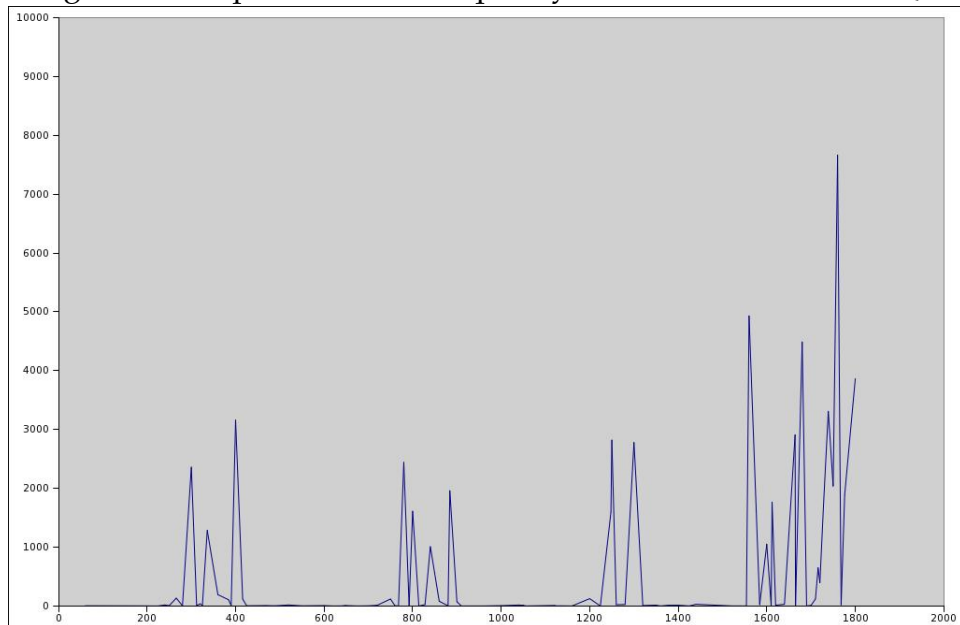


Figure 3.9: Imputed Hours Frequency, all PTSI Workers, 2002Q1

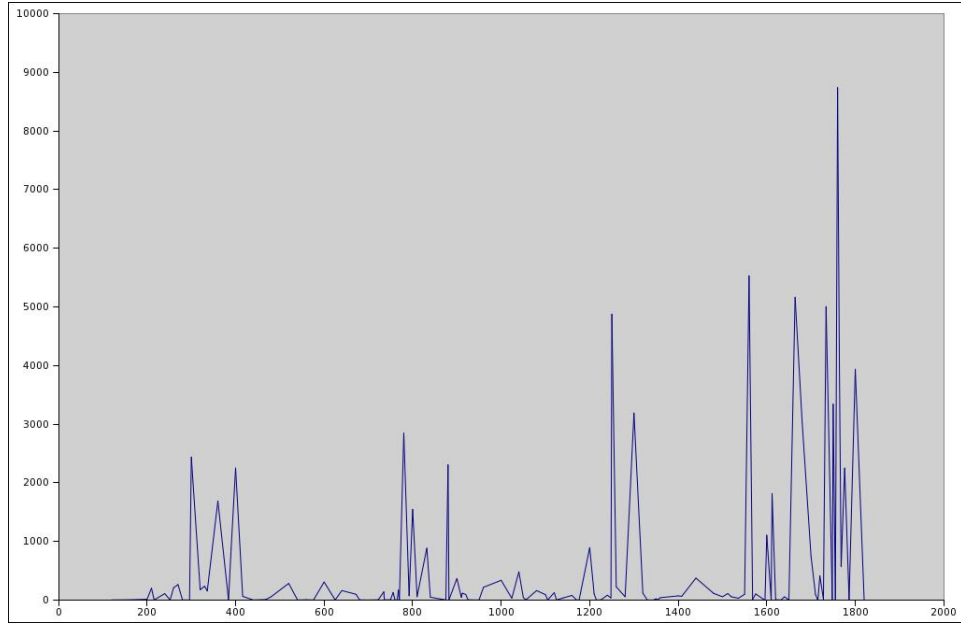


Figure 3.10: Imputed Hours Frequency, all PTSI Workers, 2002Q2

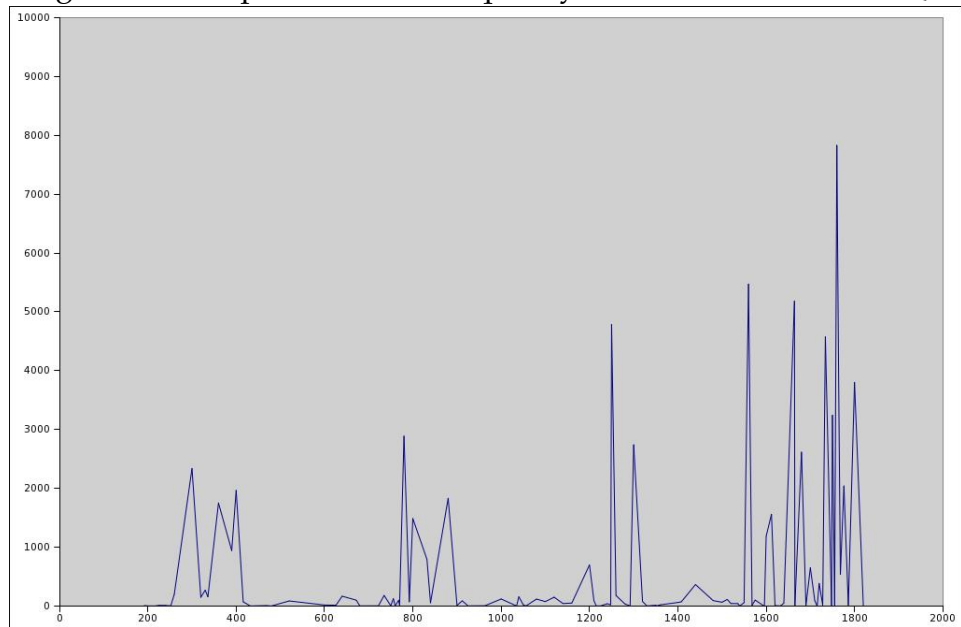


Figure 3.11: Imputed Hours Frequency, all PTSI Workers, 2002Q3

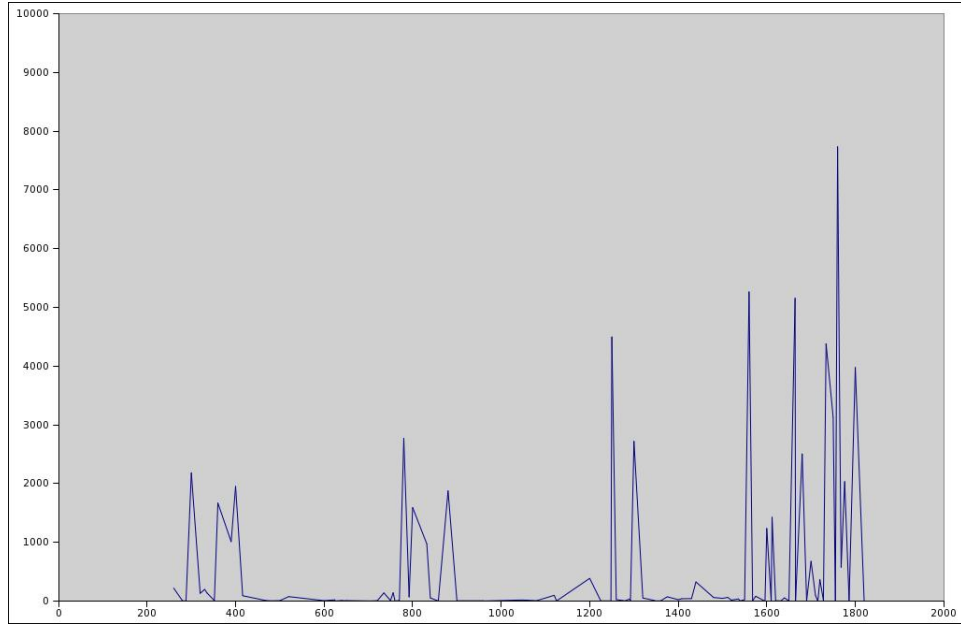


Figure 3.12: Imputed Hours Frequency, all PTSI Workers, 2002Q4

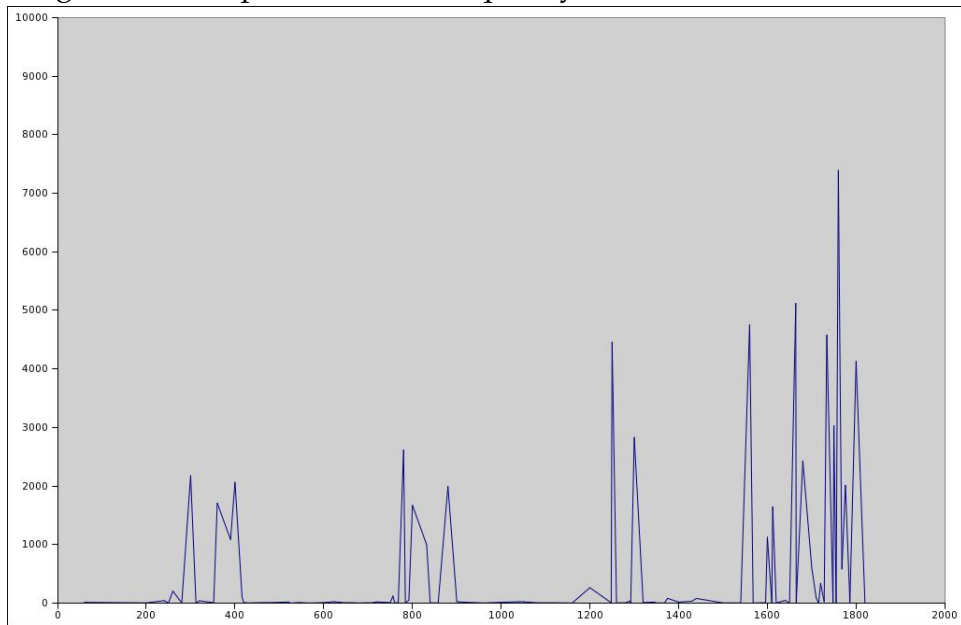


Figure 3.13: Imputed Hours Frequency, all PTSI Workers, 2003Q1

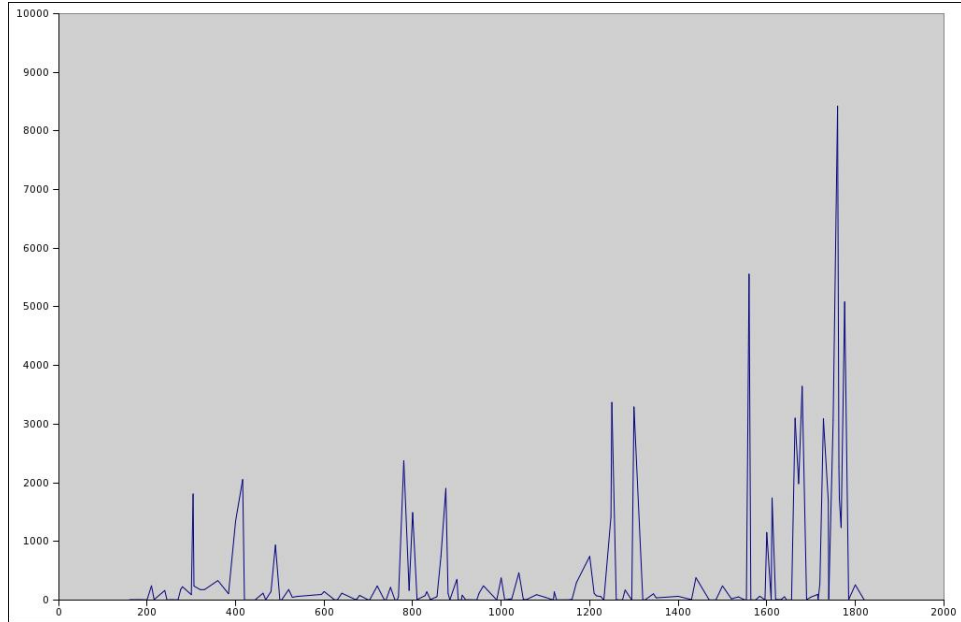


Figure 3.14: Imputed Hours Frequency, all PTSI Workers, 2003Q2

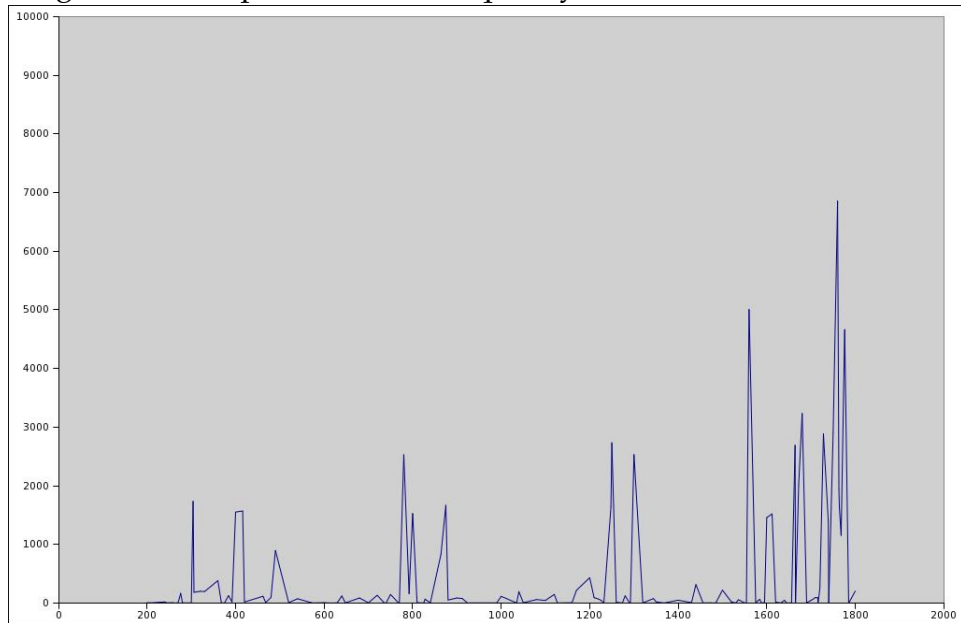


Figure 3.15: Imputed Hours Frequency, all PTSI Workers, 2003Q3

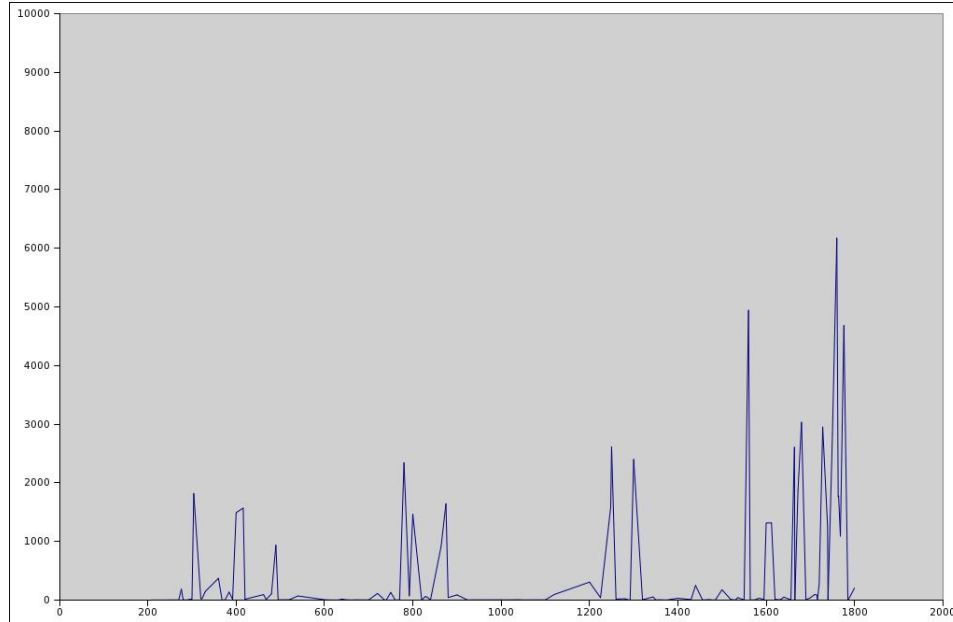


Figure 3.16: Imputed Hours Frequency, all PTSI Workers, 2003Q4

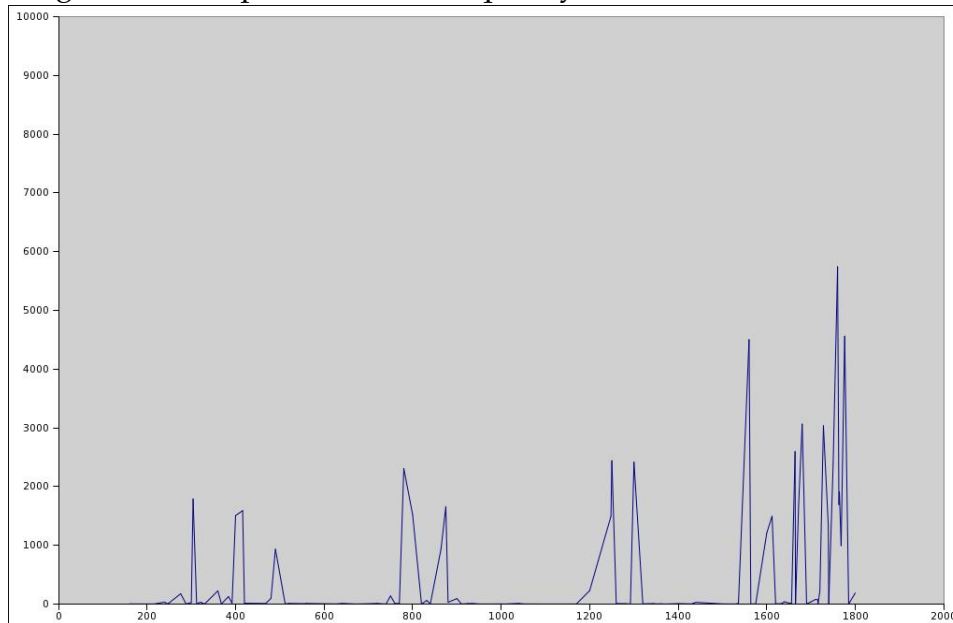


Figure 3.17: Imputed Hours Frequency, all PT Workers, 2000Q1

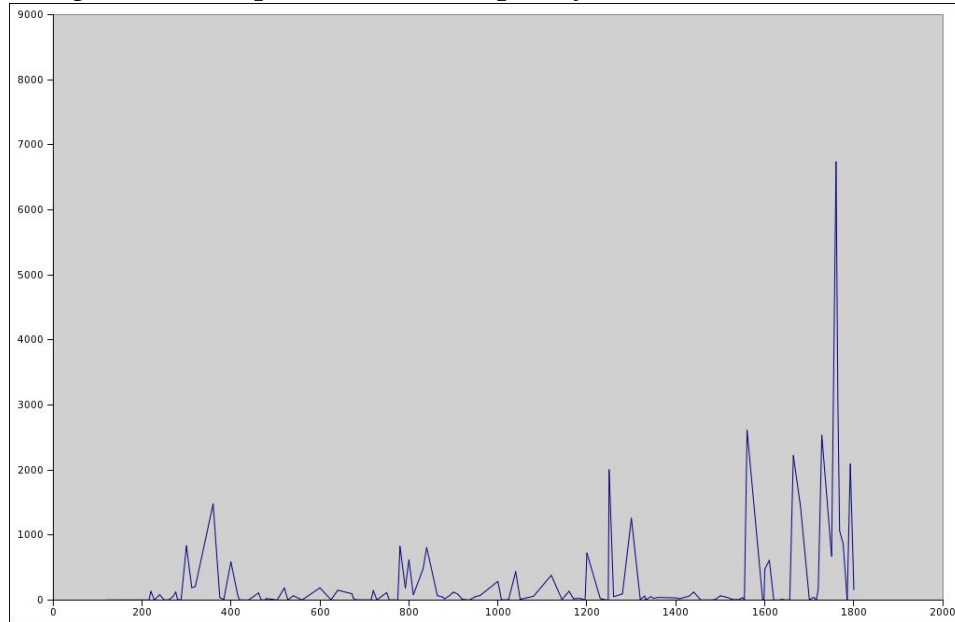


Figure 3.18: Imputed Hours Frequency, all PT Workers, 2000Q2

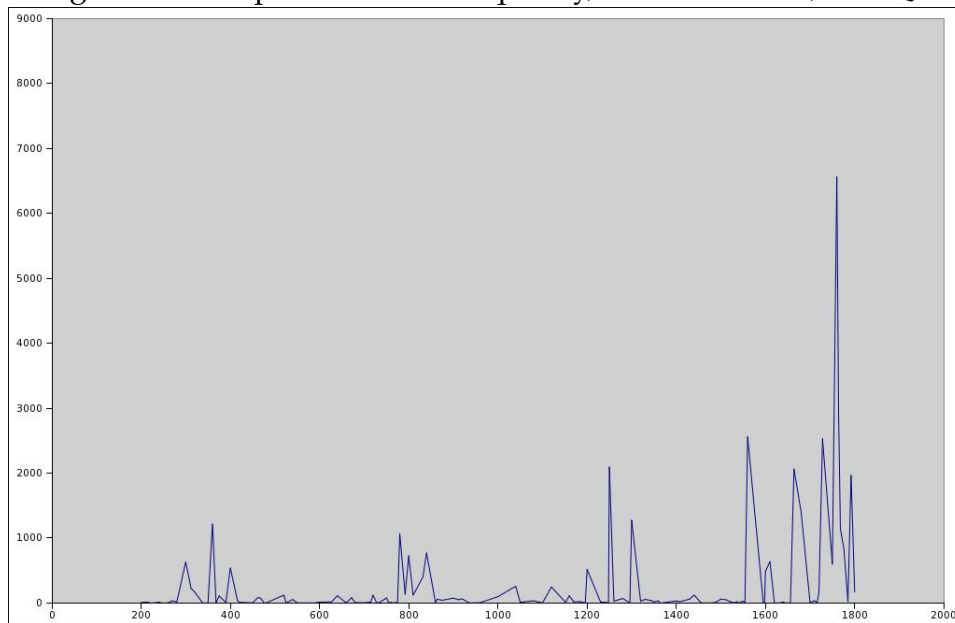


Figure 3.19: Imputed Hours Frequency, all PT Workers, 2000Q3

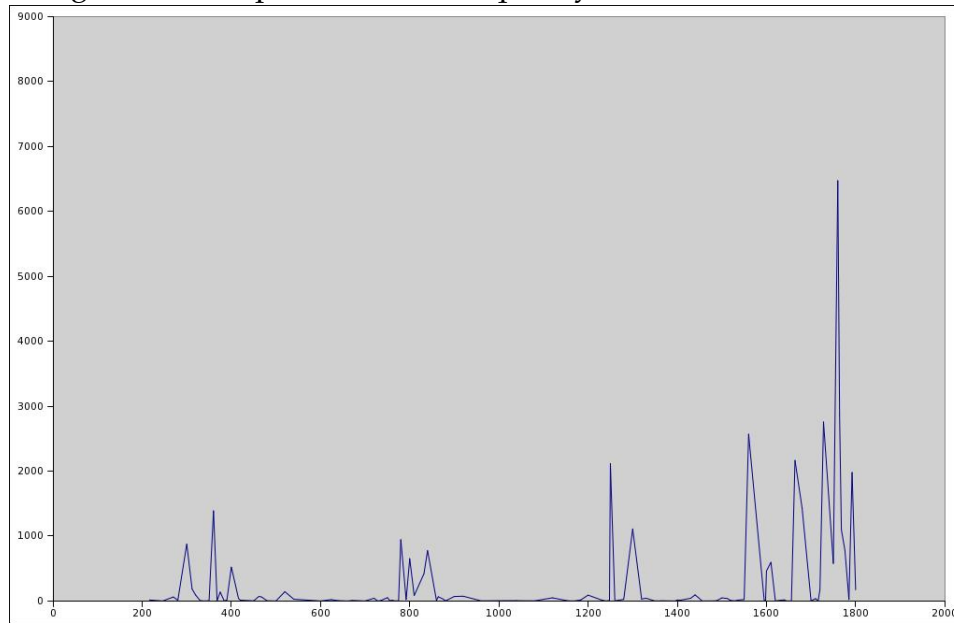


Figure 3.20: Imputed Hours Frequency, all PT Workers, 2000Q4

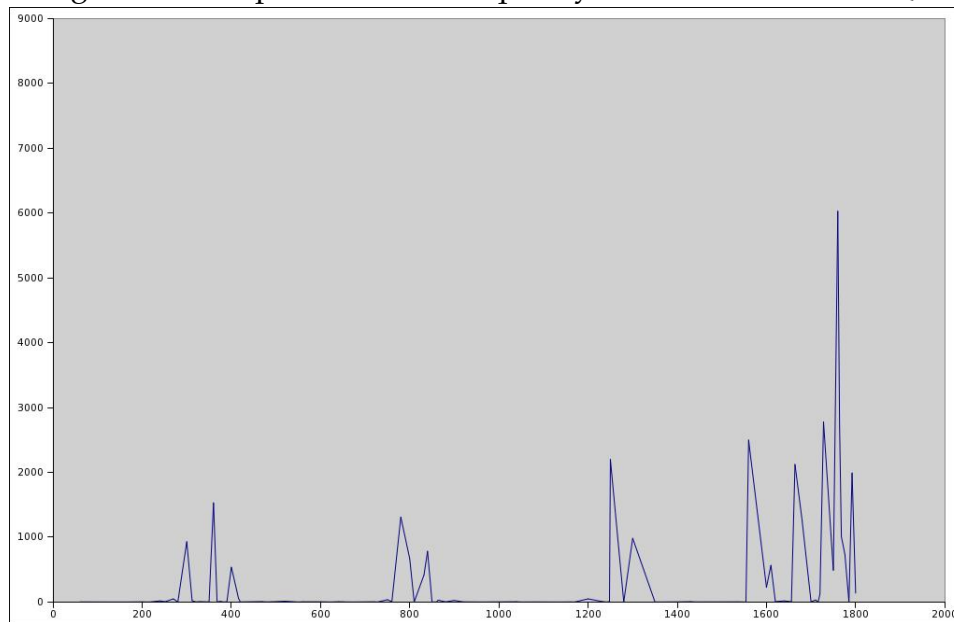


Figure 3.21: Imputed Hours Frequency, all PT Workers, 2001Q1

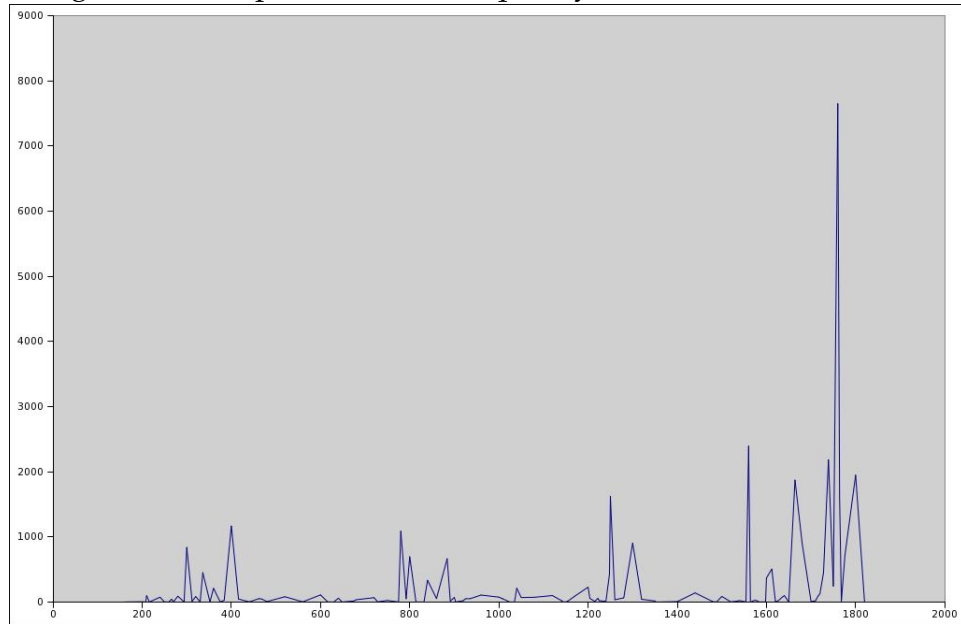


Figure 3.22: Imputed Hours Frequency, all PT Workers, 2001Q2

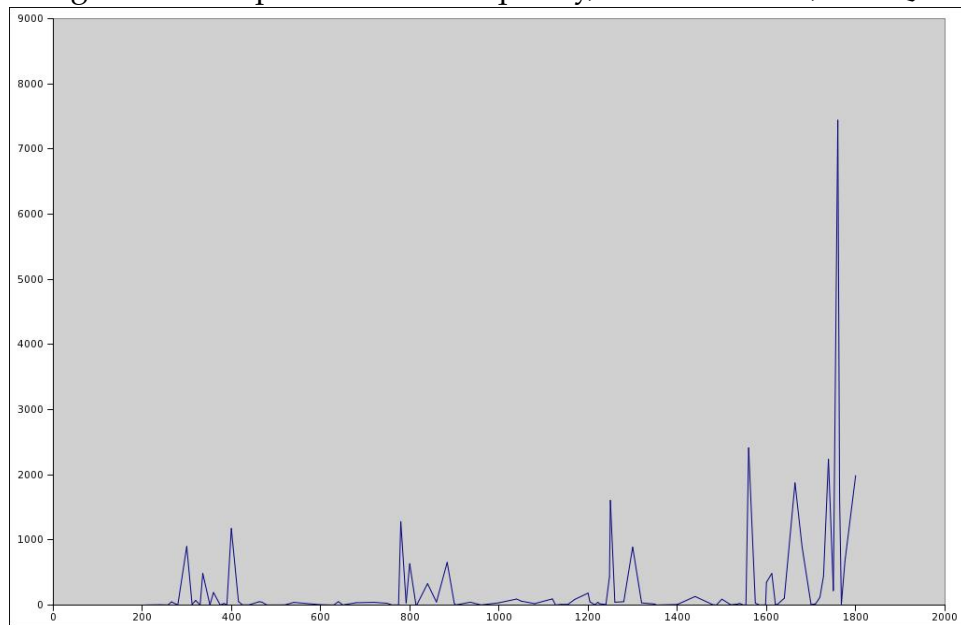


Figure 3.23: Imputed Hours Frequency, all PT Workers, 2001Q3

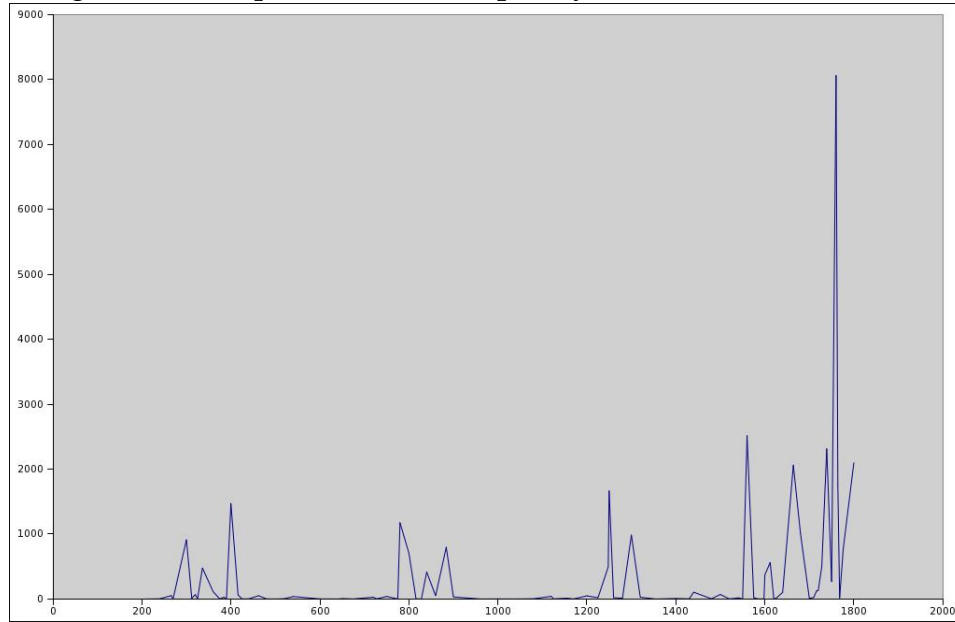


Figure 3.24: Imputed Hours Frequency, all PT Workers, 2001Q4

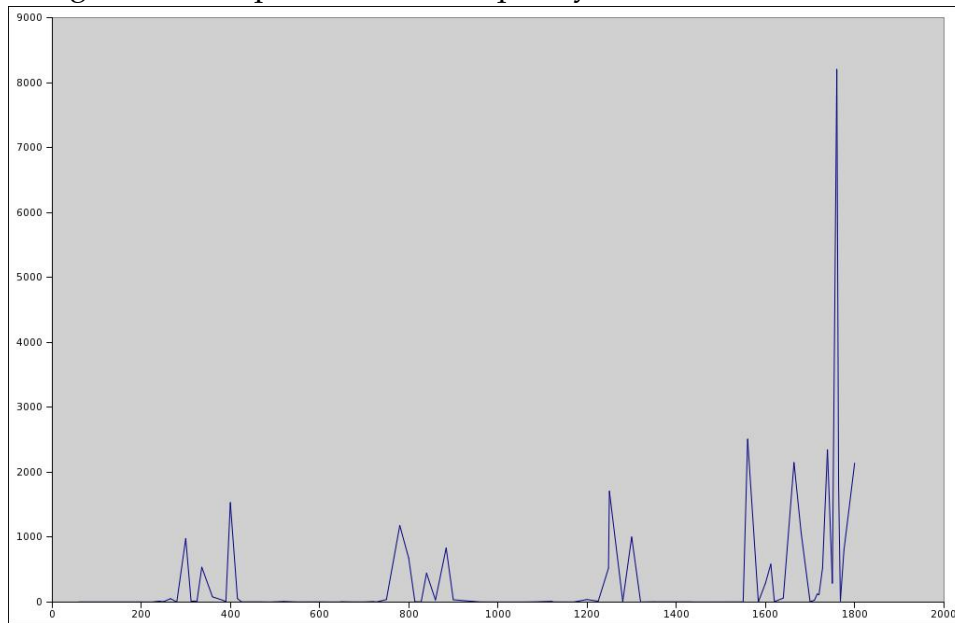


Figure 3.25: Imputed Hours Frequency, all PT Workers, 2002Q1

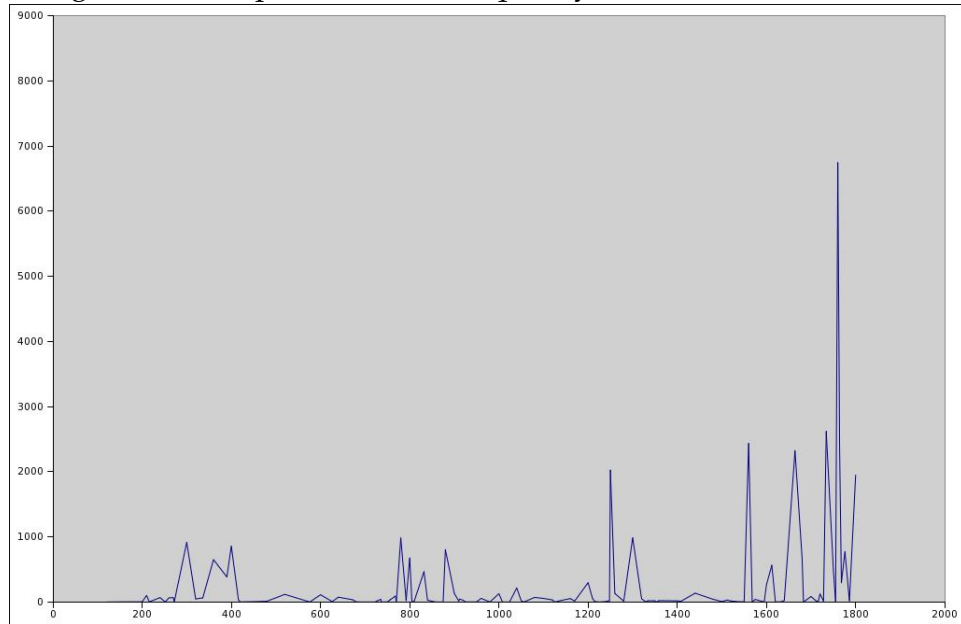


Figure 3.26: Imputed Hours Frequency, all PT Workers, 2002Q2

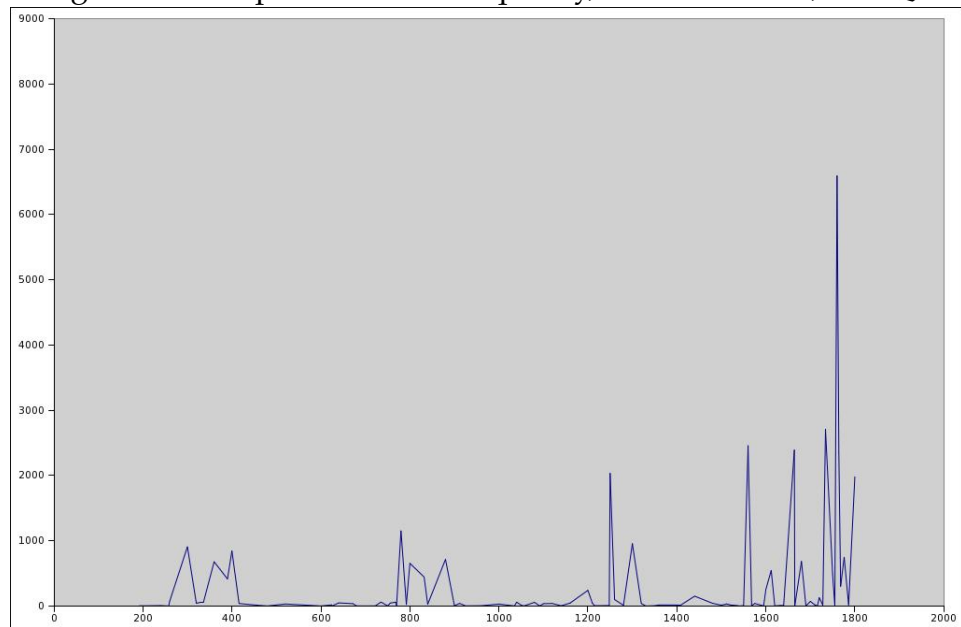


Figure 3.27: Imputed Hours Frequency, all PT Workers, 2002Q3

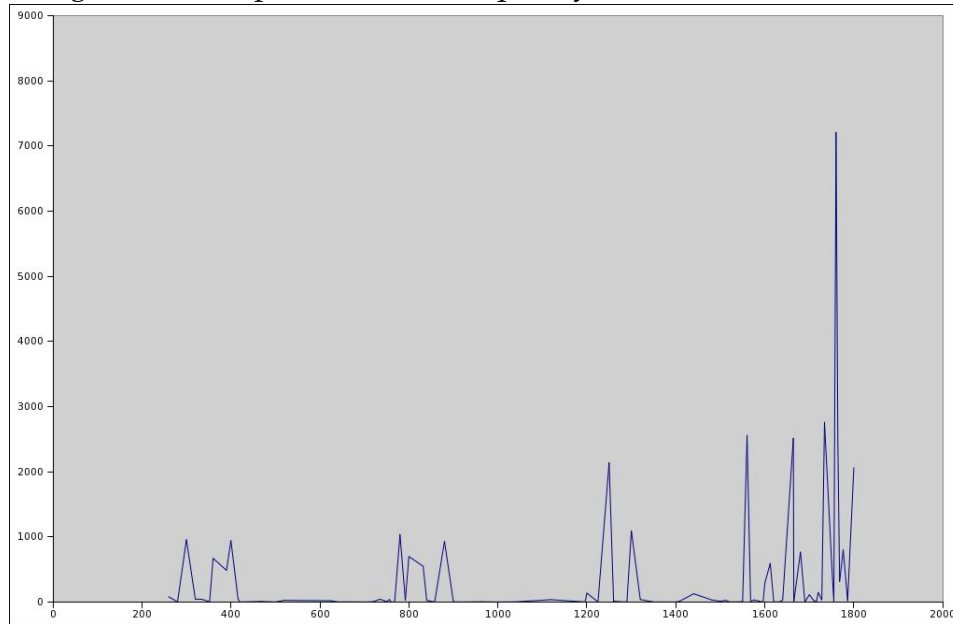


Figure 3.28: Imputed Hours Frequency, all PT Workers, 2002Q4

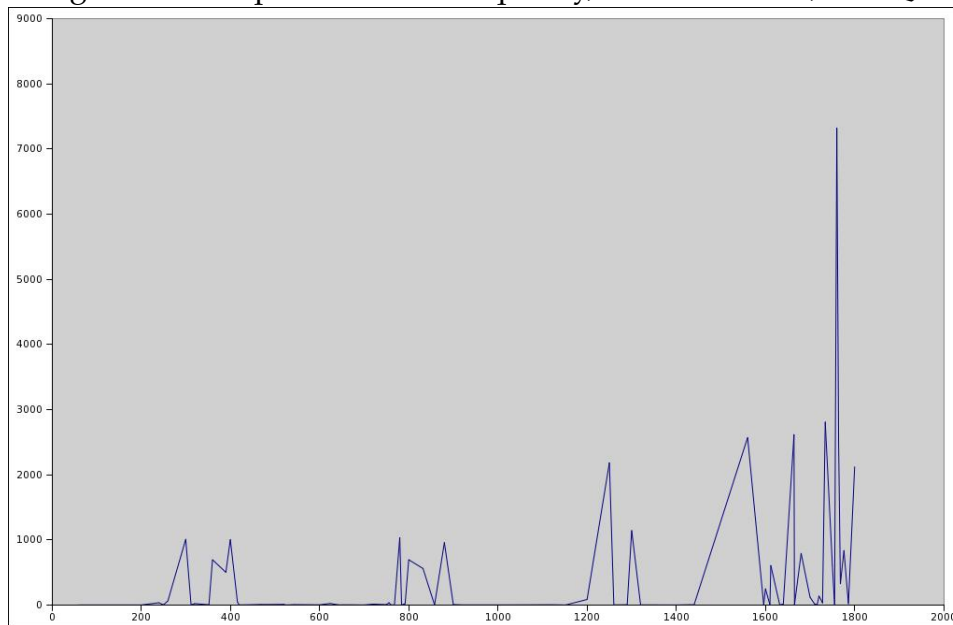


Figure 3.29: Imputed Hours Frequency, all PT Workers, 2003Q1

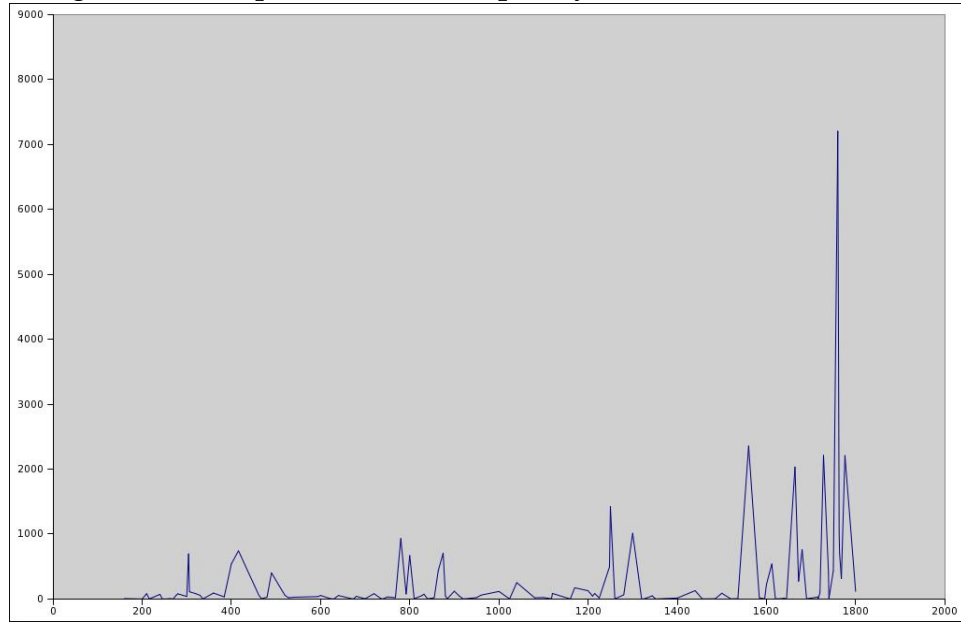


Figure 3.30: Imputed Hours Frequency, all PT Workers, 2003Q2

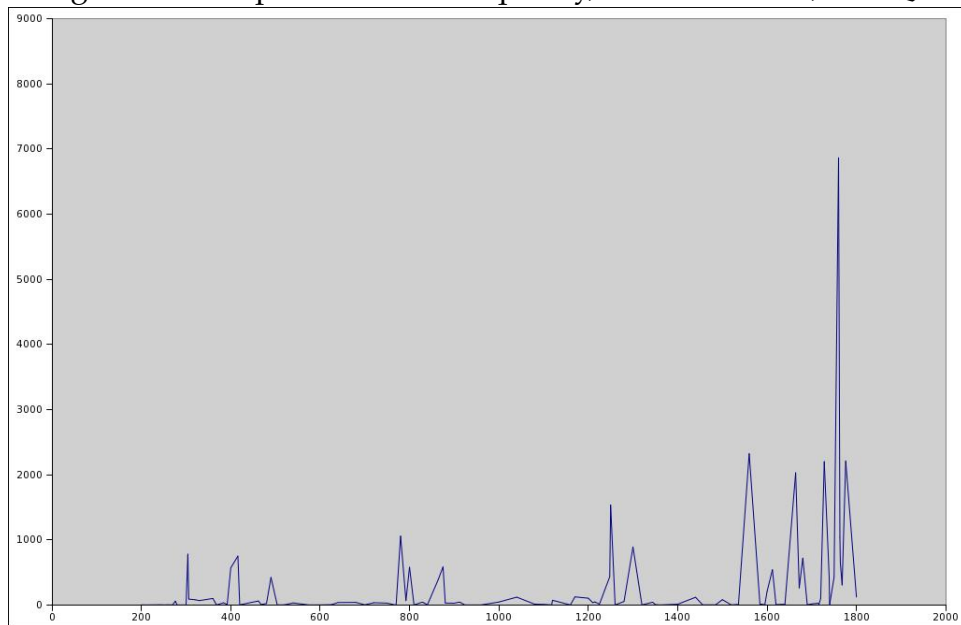


Figure 3.31: Imputed Hours Frequency, all PT Workers, 2003Q3

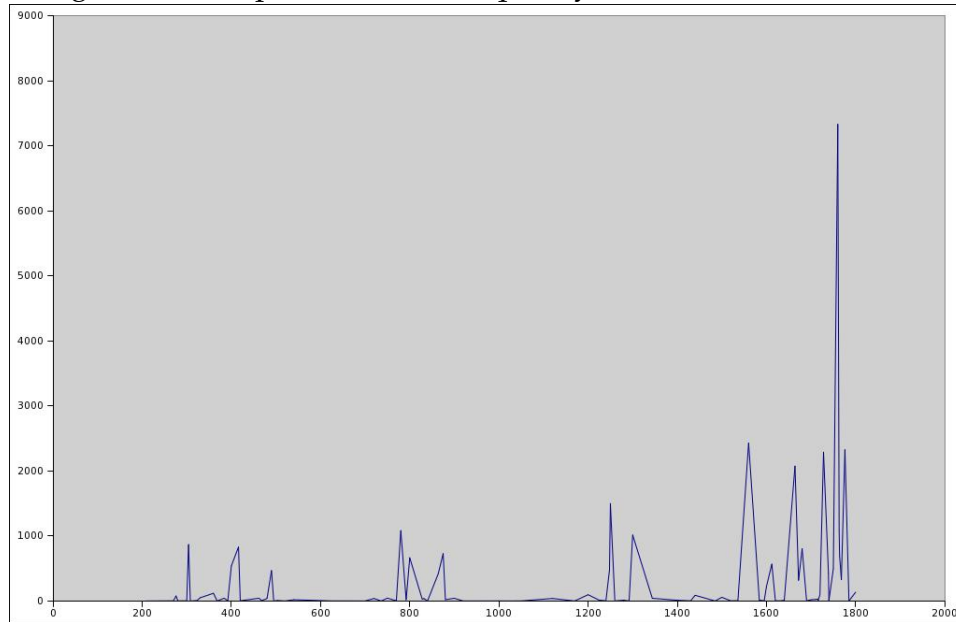
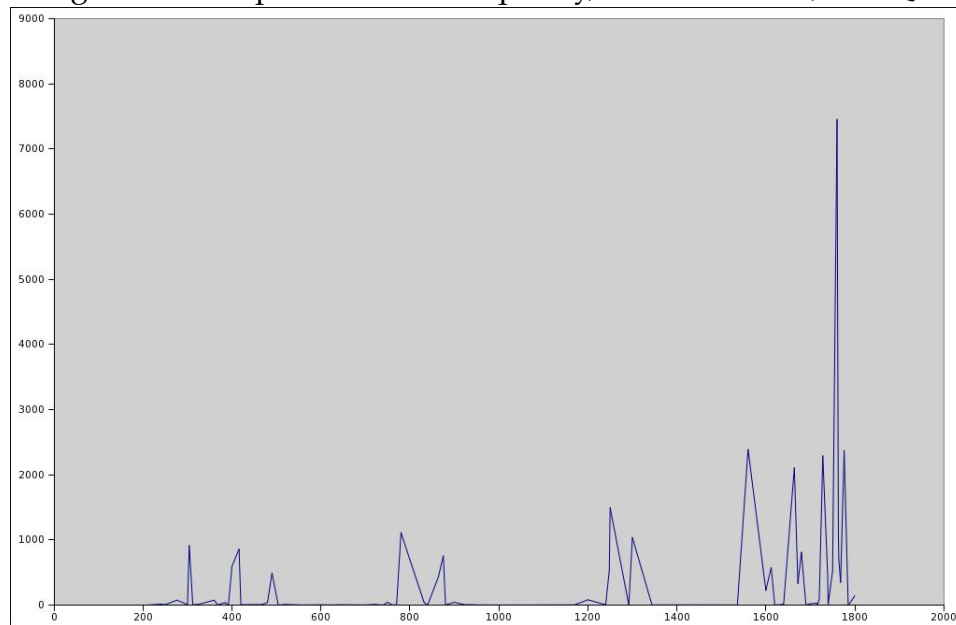


Figure 3.32: Imputed Hours Frequency, all PT Workers, 2003Q4



BIBLIOGRAPHY

- [1] Abowd, John M., Bryce Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, Simon Woodcock (2009). "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators", in: Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts (editors), *Producer Dynamics: New Evidence from Micro Data*, University of Chicago Press, Pages 149-230.
- [2] Current population survey merged outgoing rotation groups (MORG) annual files: 1979-2011 extracts. (machine-readable data file), U.S. Department of Commerce, U.S. Census Bureau. Cambridge, Massachusetts: National Bureau of Economic Research (distributor).
- [3] Office of Personnel Management (OPM): data extract from Central Personnel Data File (CPDF) and Enterprise Human Resources Integration (EHRI), 1990-2012.
- [4] Office of Personnel Management (OPM): Guide to Data Standards, Part A: Human Resources, 2012 (from http://www.opm.gov/feddata/GDS/GDS_A08.pdf).
- [5] U.S. Census Bureau, 2000 Census of Population and Housing, *Sample Census Edited File (SCEF)*, Washington, DC.