From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

Doctoral Dissertations, 2011-2026 Economics

2012

Essays in Labor Economics: Empirical Methods and Evaluations of Social and Employment Programs

Pei, Zhuan

Princeton, NJ : Princeton University

http://arks.princeton.edu/ark:/88435/dsp015425k9722

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

ESSAYS IN LABOR ECONOMICS: EMPIRICAL METHODS AND EVALUATIONS OF SOCIAL AND EMPLOYMENT PROGRAMS

Zhuan Pei

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF

ECONOMICS

Adviser: David S. Lee

June 2012

© Copyright by Zhuan Pei, 2012. All Rights Reserved.

Abstract

This dissertation consists of three essays in labor economics, which are motivated by under-appreciated but important policy rules in social and employment programs. The essays attempt to use these rules to evaluate program effects, inform policy design, and study methodological issues that arise in the process.

The first chapter, "Dynamic Opting-in Incentives in Income-tested Social Programs: Evidence from Medicaid/CHIP", examines families' responses to the continuous eligibility provision in Medicaid/CHIP and proposes a framework to evaluate the optimal eligibility recertification frequency. Conventional studies of labor supply in the presence of income-tested social programs implicitly assume that income eligibility for program participation is constantly monitored by the government. However, this is not how most of these programs operate in practice, and the time until the next eligibility recertification can be as long as a year. In particular, the Balanced Budget Act of 1997 gives states the option of insuring children in their Medicaid/CHIP program continuously for up to 12 months regardless of changes in family income. The long recertification period in effect increases the size of the benefit notch, and neoclassical labor supply models predict that agents may lower their labor supply before the application month to gain program eligibility and then increase their labor supply until the next eligibility check. I use the 2001 and 2004 panels of Survey of Income and Program Participation (SIPP) to empirically examine the income and labor supply responses of parents whose children are publicly insured. Comparing theoretical predictions and the empirical evidence points to little labor supply response. Given the absence of strategic behavior, I propose a simple framework to compute the optimal length of the continuous eligibility period relying on the mechanical properties of the income processes observed in SIPP, and derive a mapping from the recertification cost parameters to the optimal monitoring frequencies.

In the second essay, "Regression Discontinuity Design with Measurement Error in the Running Variable", I extend the regression discontinuity (RD) design to allow for the running variable to be mismeasured. The need for the method arises, for example, when a researcher tries to apply an RD design with noisy income measures from survey data to evaluate the effects of income-tested

social programs using discontinuities in eligibility rules. This paper provides sufficient conditions to non-parametrically identify the true running variable distribution and RD treatment effect when the measurement error is independent of the true running variable. A simple estimation procedure is proposed based on a minimum distance formulation, and the resulting estimators are root-N consistent, asymptotically normal and efficient. Simulations show that the procedure is informative for typical sample sizes encountered in relevant empirical studies.

In the third essay, “Quasi-Experimental Identification and Estimation in the Regression Kink Design", co-authored with David Lee and David Card, we consider nonparametric identification of the average marginal effect of a continuous endogenous regressor in a generalized non-separable model when the regressor of interest is a known, deterministic, but kinked function of an observed continuous assignment variable. This design arises in many institutional settings where a policy variable of interest (such as weekly unemployment benefits) is mechanically related to an observed but potentially endogenous variable (like previous earnings). We characterize a broad class of models in which a “Regression Kink Design" (RKD) provides valid inferences for the underlying marginal effects. Importantly, this class includes cases where the assignment variable is endogenously chosen. As in a regression discontinuity design, the required identification assumption implies strong and testable predictions for the pattern of predetermined covariates around the kink point. Standard local linear regression techniques can be adapted to obtain “nonparametric" RKD estimates. We illustrate the RKD approach by examining the effect of unemployment insurance (UI) benefits on the duration of benefit claims, using rich microdata from the state of Washington. We find that a 10 percentage point increase in the UI replacement rate leads to a 1.6 week increase in the duration of insured unemployment, which is in the higher range of magnitudes found in the existing literature.

Acknowledgements

I am deeply indebted to my principal adviser, David S. Lee, for his invaluable guidance and encouragement. I feel extremely fortunate to be his student because of his brilliance, dedication and patience. It is through the interactions with him as a student and a coauthor that I have grown to be a researcher in economics, and it will always be my goal to achieve his standard for scholarship in my academic pursuit.

I would also like to thank Bo Honoré for many helpful discussions. He introduced me to techniques in advanced econometrics and continued to serve as an important adviser throughout my graduate career. Bo is always a joy to talk to, and I am grateful for every opportunity to hear his honest opinions on my research.

Several visitors to the Industrial Relations Section played an important part in shaping this dissertation. Damon Clark, Lars Lefgren, Jordan Matsudaira and Lara Shore-Sheppard could not have been more kind and supportive. I feel fortunate to have had the opportunity to interact with them.

I have also benefitted from discussions with other Princeton faculty members–Orley Ashenfelter, Janet Currie, Kirill Evdokimov, Hank Farber, Alan Krueger, Ilyana Kuziemko, Alex Mas, Ceci Rouse and Andrew Shephard to name a few. Their numerous ideas and suggestions have greatly improved this dissertation.

Linda Belfield, Kathleen DeGennaro, Laura Hedden, Joyce Howell, Jeannie Moore, Annamarie Scarpati and Patti Tracey provided wonderful technical and administrative support and have made my graduate student life much easier. I gratefully acknowledge financial support from Princeton University and the Richard A. Lester Fellowship.

I cannot be where I am today without the help of many outstanding teachers and educators. Laura Chihara, Gao Hong Dice, Bob Dobrow, Yanying Feng, Bryan Garman, Jack Goldfeather, Nathan Grawe, Maria Koziebrodzka, Mark Krusemeyer, Dafa Li, Lucia Pierce, Jenny Wahl, Joel Weisberg, Guohua Xia, Tiansheng Zheng deserve special mention.

My host families–the Fink's, Liotta's and Sherberne/Van Heuvelen's–have provided unflinch-

ing support over the years, for which I am forever grateful. I would also like to thank many of my peers for their friendship and encouragement, and to list a few of them: Ling Ling Ang, Eric Auerbach, Silvia Barcellos, Leandro Carvalho, Dong Beom Choi, Eleanor Choi, Hyun Soo Choi, Rafael Dix-Carneiro, Qi Ge, Qiang Guo, Nik Harmon, Tatiana Homonoff, Jian Kang, Jia Li, Marjolaine Gauthier-Loiselle, Marta Lachowska, Yan Lau, Nicholas Lawson, Jianing Liu, Andrew Marder, Ashley Miller, Stephen Nei, Xiaotong Niu, Uliana Makarov, Yi Shen, Herbert Sun, Tao Wang, Nathan Wozny and Ming Yang.

I thank my parents Shouyong Pei and Anling Wang for always being there for me. Finally, I thank Pauline Leung, whose emotional support has helped me through the ups and downs of graduate school and the job market. Because I have discussed the essays below with her more than anyone else, I owe this dissertation to her.

Contents

Abstract iii

Acknowledgements v

1 Dynamic Opting-in Incentives in Income-tested Social Programs: Evidence from Medicaid/CHIP 1

1.1 Introduction 1

1.2 Institutional Background of Medicaid and CHIP 5

1.3 Theoretical Framework of Eligibility Recertification in Transfer Programs 7

1.3.1 Transfer Program and Labor Supply: Static Models 8

1.3.2 Continuous Eligibility and Labor Supply–Dynamic Models 17

1.4 Data and the Construction of the Analysis Sample 20

1.4.1 Data Sources 20

1.4.2 Adjustment to Potential Measurement Error in Public Insurance Coverage . 22

1.5 Descriptive Analysis of Income and Labor Supply Responses 24

1.6 Calibration 27

1.7 Optimal Length of the Continuous Eligibility Period 31

1.8 Conclusion 37

2 Regression Discontinuity Design with Measurement Error in the Running Variable 58

2.1 Introduction 58

2.2 Identification 61

2.2.1 Baseline Statistical Model 61
2.2.2 Identification of the True running Variable Distribution under Perfect Compliance 64
2.2.3 Running Variable and Measurement Error Have Discrete and Unbounded Support 70
2.2.4 Identification of the RD Treatment Effect 78
2.3 Estimation 81
2.3.1 Theoretical Procedures 81
2.3.2 Potential Issues in Practical Implementation 85
2.3.3 Simulations 86
2.4 Conclusion 88

3 Quasi-Experimental Identification and Estimation in the Regression Kink Design 98
3.1 Introduction 98
3.2 Nonparametric Regression and the Regression Kink Design 101
3.2.1 Background 101
3.2.2 Necessary and Sufficient Conditions for RKD in a Non-Separable Model . 104
3.2.3 Defining Causal Effects in an Ideal Randomized Setting 106
3.2.4 Local Random Assignment From a Regression Kink Design 109
3.3 Unemployment Insurance in Washington State: Background and Data 114
3.3.1 The Washington Reemployment Bonus Experiment 114
3.3.2 UI Institutions in Washington State 115
3.3.3 Data Issues 117
3.4 Empirical Results from the RKD Analysis 119
3.4.1 Outcomes of Interest 119
3.4.2 Graphical Presentation 121
3.4.3 Estimation Results 123
3.5 Conclusion 127

Chapter 1

Dynamic Opting-in Incentives in Income-tested Social Programs: Evidence from Medicaid/CHIP

1.1 Introduction

An implicit assumption in labor supply studies of income-tested social programs is that program eligibility is being constantly monitored. However, this is not how these programs operate in reality and the time between two consecutive eligibility certifications can be as long as a year. Although this fact is recognized in several studies of program participation,[1] a formal investigation has not been carried out to address how program participants adjust their labor supply behavior in response to the dynamic incentives created by the lack of constant income monitoring. In this paper, I attempt to fill this gap by examining families' behavioral responses to the continuous eligibility provision for children participating in Medicaid and the State Children's Health Insurance Program (SCHIP or simply CHIP). In addition, the positive analysis of labor supply responses can inform the normative question of often the government should check program eligibility, which I also address in this study.

The income testing in transfer programs reflects the government's redistributive taste, and

[1] In the context of Food Stamp, for example, Kornfeld (2002), Currie and Grogger (2001) and Kabbani and Wilde (2003) find that shortening the recertification period reduces the participation rate. Currie and Gahvari (2008) and Currie (2004) both recognize long recertification periods in transfer programs in general, and the latter notes that The Special Supplemental Nutrition Program for Women, Infants and Children (WIC) has a fixed recertification time during which families are eligible for benefits irrespective of income changes.

continued eligibility monitoring ensures that the program is targeting the needy. However, if monitoring is costly and incomes of program participants change little over time, it may be sensible for the government to decrease the frequency of eligibility checks and offer a period of "continuous eligibility." From the participants' point of view, there are two channels through which a transfer program becomes more valuable when a period of continuous eligibility is granted. First, transaction costs associated with benefit renewal decrease with fewer renewal periods as pointed out by Currie and Grogger (2001) and Kabbani and Wilde (2003). The second and rarely considered point is that the guarantee of continuous eligibility effectively changes a participant's budget constraint. If the budget constraint non-linearity created by the eligibility requirements distorts a family's labor supply choice, the distortion is eliminated in any period in which eligibility is not checked, allowing the family a more optimal consumption bundle. Increasing the recertification period effectively decreases the number of periods that a family will face the more stringent budget constraint, creating strong incentives for an otherwise ineligible family to opt into the program. That is, families may be induced to temporarily lower their income, gain program eligibility, and revert back to their "optimal" consumption bundle after having acquired the government benefit for the entire continuous eligibility period. As a result, the lengthening of the recertification period may create movements in the income process through labor supply around eligibility checks.

On the one hand, as the length of the continuous eligibility period increases, more families may be expected to participate in the program, resulting in increased expenditure and budget pressure for the government. In addition, these newly participating families of higher income are arguably not the intended beneficiaries of the government transfer. On the other hand, if the continuous eligibility period is significantly shortened, the transaction costs of frequent eligibility recertifications may become insurmountable for a part of the low-income group most in need of the transfer (e.g., single parent families–see Currie and Grogger (2001)). In the case of health insurance, studies (e.g., Olson et al. (2005)) have shown that children who experience interruptions in health insurance coverage are more likely to have unmet health care needs, and

therefore imposing large transaction costs on otherwise eligible families is socially suboptimal.[2] Furthermore, verifying eligibility in short intervals leads to increases in administrative cost for the government as well.

Given the tradeoffs of increasing the continuous eligibility period, understanding the behavioral reaction of economic agents to the lack of monitoring has important policy implications. If extensive strategic dip-and-rebound behavior in income is found, then it suggests that the recertification period may be too long. If no strategic behavior is found, on the other hand, labor supply responses to the continuous eligibility provision can be ruled out, and I can compute the optimal eligibility recertification period based on the mechanical properties of the empirical income processes. The framework I propose or variants thereof may be applied to study the optimal eligibility recertification frequency under the recent Affordable Care Act–for example, the Medicaid expansion to cover families under 133% of the Federal Poverty Line elicits the question of how often family incomes may need to be verified but its answer is not provided with the Act.

In this paper, I carry out an empirical investigation of the effect and optimality of the continuous eligibility provisions in the context of Medicaid/CHIP. Along with creating the SCHIP program, the Balanced Budget Act of 1997 gave states the option of continuously insuring children for up to 12 months in their public insurance programs regardless of changes in family income during that period. As a result, a third of the states implemented the continuous eligibility option in their public insurance program for children. These states present an opportunity to gauge the significance of the strategic behavior, which sheds light on the choice of the optimal continuous eligibility period.

The contributions of this paper are three-fold. First, I recognize the potential dynamic impact of a long continuous eligibility or recertification period on the labor supply decisions of program participants. I derive qualitative and quantitative predictions of the family income process using neo-classical labor supply models that incorporate the budget constraint relevant for continuous

[2]This is in contrast to the argument in Nichols and Zeckhauser (1982), and is further discussed in section (1.7).

eligibility. Second, I empirically examine the model predictions using data from the Survey of Income and Program Participation (SIPP). Third, I compute the length of the optimal continuous eligibility period under a simple social welfare specification.

Empirically, I find no evidence of the short-term dip-and-rebound strategic behavior in average income as predicted by the neoclassical models for families residing in states that provide 12 months of continuous eligibility, and I statistically reject the model in most subsamples. With labor supply responses practically ruled out, I propose a simple framework to compute the length of the optimal continuous eligibility period using the mechanical properties of the income processes observed in SIPP, and derive a mapping from the recertification cost parameters to the optimal monitoring frequencies. Under moderate costs to both the government and the participate, the calculation suggests that the optimal recertification period may be no shorter than 12 months. Therefore, it may be beneficial for states that still provide a 6-month renewal period, namely Georgia and Texas, to consider halving the renewal frequency; it may also be beneficial for those currently offering 12 months of continuous eligibility not to switch back to a 6-month period as in the case of–for example–Connecticut, Indiana, Nebraska, Washington and New Mexico in the early 2000's.

The remainder of the paper is organized as follows. Section (1.2) provides an overview of the Medicaid/CHIP institutions. Section (1.3) presents a series of models to illustrate the tradeoff in a continuous eligibility provision and to theoretically analyze families' responses to such a provision. Section (1.4) describes data used, and empirical results are presented in Section (1.5). Section (1.6) calibrates the labor supply model and compares the quantitative prediction to the empirical results. Section (1.7) conducts a simulation exercise to obtain the optimal choice of the continuous eligibility period length. Section (3.5) concludes.

1.2 Institutional Background of Medicaid and CHIP

The Medicaid program was created by the Social Security Amendments of 1965 and provides health insurance to low-income populations. The program originally targeted those traditionally eligible for welfare–single-parent families, and the aged, blind and disabled. However, eligibility for public insurance through Medicaid, and later, through SCHIP, has expanded substantially over time particularly for the population of dependent children.

Over the 1980s, the link between Medicaid and welfare for children was gradually severed through a series of legislative acts. In 1984, the Deficit Reduction Act required states to cover children less than five years old born after September 30, 1983 living in families income-eligible for Aid to Families with Dependent Children (AFDC), regardless of family structure. Further decoupling occurred with passage of the Omnibus Budget Reconciliation Acts (OBRA) of 1986 and 1987, which allowed states to raise the income limits for Medicaid eligibility above the AFDC thresholds. OBRA 1987 also required states to cover all children less than seven years old born after September 30, 1983 living in families with incomes below the AFDC income threshold. Pregnant women and infants living in families with incomes below 75% of the federal poverty level (FPL) were granted mandatory eligibility through the Medicare Catastrophic Coverage Act of 1988. Also in 1988, the passage of the Family Support Act required states to continue Medicaid coverage for up to one year for families who lost AFDC benefits due to increased earnings.

The two largest federal expansions were included in OBRA 1989 and OBRA 1990, which became effective in April 1990 and July 1991 respectively. OBRA 1989 required states to offer Medicaid coverage to pregnant women and children up to age six with family incomes below 133% of the FPL. OBRA 1990 required states to cover children born after September 30, 1983 with family incomes below 100% of the FPL. The two expansions remain the mandated minimum federal standards for children today: a child under the age 6 is eligible for Medicaid if her family income is below 133% of the FPL, and a child between the age of 6 and 18 is eligible if her family income is below 100% of the FPL (for a detailed account of the major Medicaid legislations by 1997, see Gruber (2003)).

While states are required to adhere to these minimum federal standards from OBRA 1990, the creation of the State Children's Health Insurance Program (SCHIP) in 1997 allowed many states to further expand their public insurance programs above these standards. Unlike Medicaid, SCHIP provided states with block grants to fund coverage for children and left the implementation of program up to the individual states, subject to some rules to prevent crowd out of private insurance and to meet federal benefit standards. Specifically, states could choose to use their funds by expanding their existing Medicaid program, creating a separate program for children who do not qualify for the existing Medicaid program, or a combination of both. In some separate state programs, families who exceed a certain income level were also required to pay a premium for coverage. As a result of the block granting structure of SCHIP, states varied widely in their implementation of public insurance for children. One particular feature of some state programs is a continuous eligibility period as permitted by the Balanced Budget Act of 1997, which provides children with uninterrupted coverage for up to 12 months after confirming eligibility, regardless of whether their families' incomes rise above the eligibility requirement during this period. The focus of this paper is the effect of this continuous eligibility period on public insurance coverage for children and on the labor supply of their families.

In practice, Medicaid and CHIP eligibility is established based on the most recent monthly income, and official income proof[3] needs to be submitted with the application in most cases. However, with the implementation of a continuous eligibility period, some states explicitly mandate that once a family is qualified for coverage, they are eligible for coverage for a fixed, continuous period after that point. After this period, the family is once again required to confirm their eligibility, either by reporting any changes in income, or by actually sending proof of income with a renewal application. Families typically receive a package containing renewal materials 60 to 90 days prior to the expiration of their benefits.

[3]On the current application form for New York State, for example, it states that an applicant must provide a letter,written statement, or copy of check or stubs, from the employer, person or agency providing the income... [the applicant should] [p]rovide the most recent proof of income before taxes and any other deductions. The proof must be dated, include the employee's name and show gross income for the pay period. The proof must be for the last four weeks, whether you get paid weekly, bi-weekly, or monthly. It is important that these be current."

Finally, along with the continuous eligibility provision, the Balanced Budget Act of 1997 also gave states the option of allowing presumptive eligibility for children. That is, states may allow children who appear eligible to obtain temporary Medicaid/CHIP eligibility (so that they may immediately access health care services) while their eligibility based on income is being confirmed.[4] Since children who are covered under presumptive eligibility do not always need to meet the usual income requirements, I analyze the effect of the continuous eligibility provision on family income both including and excluding families in states that provide presumptive eligibility. Twelve states provided presumptive eligibility to children in my sample and are listed in the Appendix.

1.3 Theoretical Framework of Eligibility Recertification in Transfer Programs

The income or means testing in transfer programs reflects the government's redistributive taste and its intention in targeting the needy. If income does not change over time and those in need remain in need, then there is no point in monitoring program eligibility once families are allowed in the program. Therefore, the necessity of eligibility recertification stems from the possibility that a family having entered the program with a low income is no longer in need following a large positive income change. From the perspective of the government, this family should be taken off the program roster based on an income eligibility recertification so as to alleviate the pressure on the government's budget. If monitoring eligibility is costless for both the government and family, then the government should perform an eligibility check every period to ensure that the transfer targets the needy as is shown formally below. When there is a cost to eligibility monitoring, the choice of the length of the recertification period in part reflects the compromise between incurring this cost and transferring benefits to those who are not the neediest.

It may be tempting to allow a long continuous eligibility period in the case of high

[4]Presumptive eligibility to infants and pregnant women was granted a decade earlier by OBRA '86.

administrative cost or low income volatility, but policy makers should be wary of the potentially large labor supply distortions that result. In the extreme scenario mentioned above where the continuous eligibility period is infinite–once eligible due to a low monthly income, families can claim program benefits for a life-time–the vast majority of the families may decide to temporarily lower their income and participate in the program, which will render the system unsustainable. Therefore, understanding families' income and labor response to the dynamic opt-in incentive is key in the consideration of the optimal continuous eligibility period. In subsection (1.3.1) I first review the prediction of a class of standard static (i.e., assuming constant eligibility monitoring) neo-classical labor supply models in the presence of an in-kind transfer, and I show in subsection (1.3.2) that the dynamic problem with continuous eligibility provisions can be reduced to two static problems with different budget constraints. The solution to these problems predicts a dip and rebound in average income at each eligibility check.

1.3.1 Transfer Program and Labor Supply: Static Models

In this subsection, I analyze the labor supply decisions when eligibility for Medicaid/CHIP is recertified every period. That is, families are eligible for benefits only if their income is below a cutoff as is assumed in the conventional labor supply framework in the literature. The analysis is standard, and implications–at least the qualitative ones–have been explored in other studies (e.g., Blank (1989) and Yelowitz (1995)). I review it here using the utility functional form from Saez (2010) because results derived below will be relevant for the dynamic problem with continuous eligibility provision in section (1.3.2). Subsection (1.3.1) discusses several extensions of the baseline model: it explores the implication of allowing agents only discrete labor supply choices and the consequences of introducing welfare stigma, income effects and incorporating heterogeneity in the elasticity of labor supply.

Baseline Labor Supply Model

The utility functional form of the baseline model is taken from Saez (2010), who studies the bunching behavior of economic agents in response to non-linearities in the tax schedule. The particular utility functional form has also been used in other recent papers, e.g., Chetty et al. (2011), that study the response to nonlinearities in the budget constraint.

Agents choose continuous pre-tax income Z and post-tax income (consumption) C to maximize utility that increases in C and decreases in Z. The implicit assumption is that Z is an increasing function of labor supply, and it is equivalent to formulating the utility function in terms of C and hours worked H as noted below. Specifically, the utility function is of quasi-linear form

$$u(C, Z) = C - \frac{n}{1 + 1/e} \left(\frac{Z}{n} \right)^{1+1/e} \tag{1.1}$$

where n and e are parameters indicating taste for work and the responsiveness of pre-tax income to a change in tax rate. Solving the optimization problem implies that an agent chooses optimal pre-tax income $Z^* = n(1-t)^e$ when facing the budget constraint $C = (1-t)Z$. As pointed out by Saez (2010), $Z^* = n$ when $t = 0$, and n can be interpreted as the choice of potential income in the absence of tax and transfer programs. An agent with a larger n both works and consumes more, and n is assumed to be smoothly distributed according to density f_n across the population.

The parameter e is the elasticity of pre-tax income with respect to (one minus) the marginal tax rate because of the following identity $\frac{(1-t)}{(Z^*)} \frac{d(Z^*)}{d(1-t)} = e$. Note that if the utility function is written in terms of consumption and hours worked $u(C, H) = C - \frac{n}{1+1/e} (\frac{wH}{n})^{1+1/e}$ for an agent with taste n and wage rate w, it is also true that $\frac{(1-t)}{(H^*)} \frac{d(H^*)}{d(1-t)} = e$. Therefore I will refer to e interchangeably as the income elasticity or the labor supply elasticity in the subsequent sections of this paper. As is well known, the quasi-linear utility functional form implies no income effects, so e is both the compensated and uncompensated elasticity (in fact, e is also the Frisch elasticity of income/labor supply). I assume e to be constant in this section but the consequence of allowing heterogeneity in e, as well as that of allowing income effects, will be discussed in section (1.3.1).

The presence of Medicaid/CHIP induces at least one discontinuity in the relationship between consumption and income. For simplicity of exposition, however, I include only one such notch and a single marginal tax rate in the presentation in this section.[5] When eligibility is checked every month, this budget constraint can be thought of as being static. Following the "notch" specification adopted by Blank (1989) and Yelowitz (1995) with no saving or borrowing, the budget constraint is

$$C = [Z(1-t)+g]1_{[Z\leqslant\gamma]} + Z(1-t)1_{\{Z>\gamma\}} \tag{1.2}$$

where γ is the Medicaid/CHIP eligibility cutoff, g the monthly value of public insurance and t the marginal tax rate. As pointed out by, for example Blinder and Rosen (1985), Blank (1989) and Kleven and Waseem (2011), no family will choose income to be just above the threshold. This is intuitive because a family consumes more and works less by choosing its income to be at the eligibility cutoff rather than just above it. Certain families who would have chosen $Z > \gamma$ in the absence of Medicaid/CHIP would now switch to γ. Solving the optimization problem predicts the choice of Z^* for a family of type n:

$$Z^* = \begin{cases} n(1-t)^e & \text{if } n \leqslant n_\gamma \text{ or } n > \bar{n} \\ \gamma & \text{if } n \in (n_\gamma, \bar{n}) \end{cases}$$

where $n_\gamma = \frac{\gamma}{(1-t)^e}$ is the type of agent who would choose income at γ in the absence of the notch and $\bar{n}$ is the highest type of agent who would bunch at γ in the presence of the notch. Figure 1.1 provides an illustration: an agent with $\bar{n}$ is indifferent between the consumption-income bundle at the notch $(\gamma(1-t)+g, \gamma)$ and her optimal choice in the absence of the notch $(\bar{n}(1-t)^{1+e}, \bar{n}(1-t)^e)$. Therefore, $\bar{n}$ is the solution to the equation

$$\gamma(1-t)+g-\frac{\bar{n}}{1+1/e}(\frac{\gamma}{\bar{n}})^{1+1/e} = \bar{n}(1-t)^{1+e} - \frac{\bar{n}}{1+1/e}(1-t)^{1+e} \tag{1.3}$$

[5] As mentioned in section (1.2), families enrolled in CHIP with income above the 150% FPL may be subject to moderate premiums and co-payments, which implies a lower CHIP notch than that of Medicaid. The empirical impacts of presence of other notches (induced by Medicaid or other transfer programs) and differential tax rates will be discussed in section (1.6).

It follows that the distribution of pre-tax income Z is given by:

$$f_Z(z) = \begin{cases} \frac{1}{(1-t)^e} f_n(\frac{z}{(1-t)^e}) & \text{if z<}\gamma \text{ and } z \geqslant \bar{n}(1-t)^e \\ 0 & \text{if } z \in (\gamma, \bar{n}(1-t)^e) \end{cases}$$

and

$$\Pr(Z = \gamma) = F_n(\bar{n}) - F_n(n_\gamma)$$

In summary, the standard static model makes the prediction that, when a benefit notch is introduced, agents originally choosing income just above the eligibility cutoff will lower their labor supply to locate at the cutoff and become just eligible for benefit. Thus, there is income bunching at the eligibility cutoff and a drop in density to 0 right above the cutoff.

Extensions of the Baseline Model: Discrete Labor Supply Choices, Heterogeneous Elasticity, Welfare Stigma and Income Effects

In the previous subsection (1.3.1), the income/labor supply choice is assumed to be continuous, the income/labor supply elasticity e is held constant across agents, perfect compliance (i.e. those eligible will participate in Medicaid/CHIP) is assumed, and there are no income effects in the utility function (1.1). This subsection investigates the implication of relaxing these assumptions and shows that the qualitative predictions in the previous subsections still hold true.

Discrete Labor Supply Choices In the preceding subsection, agents' pre-tax income choice is assumed to be continuous which implies that agents are free to choose their hours and hence perfectly control their income. Obviously, this may not be a realistic restriction per Ashenfelter (1980), Ham (1982), Kahn and Lang (1991), Altonji and Paxson (1992), Dickens and Lundberg (1993) and Chetty et al. (2011). In this subsection, I will first derive the theoretical prediction only allowing an agent finitely many hours choices.[6] The main implication is still that certain agents will lower their labor supply in order to claim benefit when a notch is introduced. But

[6]The working paper versions of Saez (2010), Saez (1999) and Saez (2002), address this extension in their simulation section but do not discuss the predictions from a theoretical perspective.

rather than bunching at the eligibility cutoff, there is only a discontinuous drop in the density of income at the cutoff.

Because of the discrete labor supply restriction, Z in this section is written explicitly as wH where w is considered to be distributed smoothly among agents. For exposition purposes, I discuss only the case when H can only vary along the extensive margin; that is, an agent can only work full time or not work at all. The general case where H is allowed more than two choices is explored in the Appendix. Let $H = 0$, $H = 1$ denote the labor supply choice of not working and working full time respectively. If workers are constrained to only these two labor supply options, then the maximization problem becomes $\max_{H\in\{0,1\}} u(C, wH)$ subject to the budget constraint (1.2) where Z is replaced by wH, and we solve the maximization problem by considering the following two scenarios.

1. $w \leqslant \gamma$. An agent with potential monthly wage below the cutoff can claim benefits whether she works or not. In other words, the budget constraint she faces is only the segment to the left of γ: $C = (1-t)wH + g$. Consequently, maximizing utility involves the comparison of $u(g, 0)$ and $u((1-t)w + g, 1)$. To characterize the solutions, consider the agent of type $\bar{n}^l$ who is indifferent between choosing $H = 0$ and $H = 1$ at wage w.[7] Therefore, $\bar{n}^l$ solves

$$u(g,0) \equiv g = (1-t)w + g - \frac{n}{1+1/e}(\frac{w}{n})^{1+1/e} \equiv u((1-t)w + g, 1)$$

which implies that $\bar{n}^l(w) = (\frac{w^{1+1/e}}{[(1-t)w](1+1/e)})^e$. Since $\frac{n}{1+1/e}(\frac{w}{n})^{1+1/e}$ is decreasing in n (i.e. the disutility of working is less for an agent with high n), agents with wage w and of type $n \geqslant \bar{n}^l(w)$ choose $H = 1$ and those with $n < \bar{n}^l(w)$ choose $H = 0$.

2. $w > \gamma$. An agent with potential monthly wage above the cutoff is eligible for benefits only if she chooses not to work. The type of agent who is indifferent between working and not working at wage w equates $u(g, 0)$ and $u((1-t)w, 1)$. Because her type $\bar{n}^r$ solves

$$u(g,0) \equiv g = (1-t)w - \frac{n}{1+1/e}(\frac{w}{n})^{1+1/e} \equiv u((1-t)w, 1)$$

[7]The superscript l here stands for left as w lies to the left of γ. The superscript r will be used in the next case.

$\bar{n}^r(w) = (\frac{w^{1+1/e}}{[(1-t)w-g](1+1/e)})^e$.[8] Analogous to the case above, agents with $n \geqslant \bar{n}^r(w)$ choose to work full time while those with $n < \bar{n}^r(w)$ choose not to work.

To summarize, if $\bar{n}_{0,1}$ denotes the type of agents who are indifferent between working and not working, then

$$\bar{n}_{0,1}(w) = \begin{cases} \bar{n}^l(w) & \text{if } w \leqslant \gamma \\ \bar{n}^r(w) & \text{if } w > \gamma \end{cases}$$

$\bar{n}_{0,1}$ varies smoothly with w within each case, but there is a discontinuous increase in $\bar{n}_{0,1}$ as w crosses γ. When w and n follow a smooth joint distribution $f_{n,w}$ over the first quadrant of $\mathbb{R}^2$, this discontinuous drop in threshold agent type implies no bunching but a discontinuity in the density of pre-tax income $Z = wH$ at γ. To see this, notice that the c.d.f of Z evaluated at $z > 0$ is

$$F_Z(z) = \Pr(Z = 0) + \Pr(0 < Z \leqslant z) = \Pr(H = 0) + \Pr(H = 1, w \leqslant z) \tag{1.4}$$

On two sides of the eligibility cutoff γ, the values of F_Z are

$$\begin{aligned} F_Z(z) &= \begin{cases} \Pr(H = 0) + \Pr(w \leqslant z, n \geqslant \bar{n}^l(w)) & \text{if } z \leqslant \gamma \\ \Pr(H = 0) + \Pr(w \leqslant \gamma, n \geqslant \bar{n}^l(w)) + \Pr(\gamma < w \leqslant z, n \geqslant \bar{n}^r(w)) & \text{if } z > \gamma \end{cases} \\ &= \begin{cases} \Pr(H = 0) + \int_0^z \int_{\bar{n}^l(w')}^{\infty} f_{n,w}(n', w') dn' dw' & \text{if } z \leqslant \gamma \\ \Pr(H = 0) + \int_0^{\gamma} \int_{\bar{n}^l(w')}^{\infty} f_{n,w}(n', w') dn' dw' + \int_{\gamma}^z \int_{\bar{n}^r(w')}^{\infty} f_{n,w}(n', w') dn' dw' & \text{if } z > \gamma \end{cases} \end{aligned}$$

Since $\int_0^z \int_{\bar{n}_l(w')}^{\infty} f_{n,w}(n', w') dn' dw'$ is continuous in z and $\lim_{z \downarrow \gamma} \int_{\gamma}^z \int_{\bar{n}^r(w')}^{\infty} f_{n,w}(n', w') dn' dw' = 0$, $F_Z(z)$ is continuous at γ. Hence, there is no bunching at the eligibility cutoff unlike in section (1.3.1) where agents can choose along the intensive margin of labor supply.

However, the p.d.f of Z, $f_Z(z)$, is not continuous at γ. By continuity of $f_{n,w}$, $\bar{n}^l$ and $\bar{n}^r$ along

[8] Note that a positive n^r exists–n^r has to be positive for the marginal utility of work to be negative–when $(1-t)w > g$, which means that the post-tax income of working full time at wage w is larger than the value of benefit g. This is most likely satisfied in reality for families with a wage above the CHIP cutoff.

with an application of the Fundamental Theorem of Calculus,

$$\begin{aligned}\lim_{z\uparrow\gamma} f_Z(z) &= \int_{\bar{n}^l(\gamma)}^{\infty} f_{n,w}(n',\gamma)dn' \\ \lim_{z\downarrow\gamma} f_Z(z) &= \int_{\bar{n}^r(\gamma)}^{\infty} f_{n,w}(n',\gamma)dn'\end{aligned}$$

Since $\bar{n}^r(\gamma) > \bar{n}^l(\gamma)$, $\lim_{z\uparrow\gamma} f_Z(z) > \lim_{z\downarrow\gamma} f_Z(z)$ which implies a discontinuous drop in the income density at γ.

The constraint that workers can only work full time or not work at all is too restrictive. In reality, workers may and do work part time. It is plausible that employers offer several hours-of-work choices to their employees. As shown in the Appendix, the result of no bunching but a density discontinuity at the cutoff holds true when H takes on a finite number of values.

Heterogeneous labor supply elasticities Instead of requiring agents to share the same labor supply elasticity, the first part of this subsection studies the pre-tax income distribution when elasticities are heterogeneous across families. In section (1.3.1), the threshold taste parameter–$\bar{n}$–is a function of e, and all statements are true for each $e > 0$. Now suppose that e is heterogeneous and distributed smoothly across agents. In the case where there is no constraint on labor supply, the discontinuous drop in the income density at γ is

$$\lim_{z\uparrow\gamma} f_Z(z) - \lim_{z\downarrow\gamma} f_Z(z) = \int_0^{\infty} \frac{1}{(1-t)^e} f_{n|e}(n_\gamma(e)|e) f_e(e) de \tag{1.5}$$

since $\lim_{z\uparrow\gamma} f_Z(z) = \int_0^{\infty} \frac{1}{(1-t)^e} f_{n|e}(n_\gamma|e) f_e(e) de$ and $\lim_{z\downarrow\gamma} f_Z(z) = 0$, and the fraction of agents bunching at γ is

$$\Pr(Z = \gamma) = \int_0^{\infty} (F_{n|e}(\bar{n}(e)|e) - F_{n|e}(n_\gamma(e))) f_e(e) de \tag{1.6}$$

where $\bar{n}(e)$ and $n_\gamma(e)$ are as defined in subsection (1.3.1): $\bar{n}(e)$ is the solution to (1.3) and $n_\gamma(e) = \frac{\gamma}{(1-t)^e}$. Since the integrands in both (1.5) and (1.6) are positive, there is still bunching and a discontinuous drop in the income density at γ. Analogously, the result of no bunching but a density discontinuity at γ presented earlier in this subsection also holds when labor supply is

constrained to several choices–since the result holds for all e, it also holds when integrating over the density of e.

Non-participation To account for non-participation among eligible agents, I follow a conventional approach by Moffitt (1983) and introduce welfare stigma. As pointed out in Moffitt (1983), however, the stigma term can also encapsulate more than the simple psychological cost of being perceived as a beneficiary of government programs. For example, it may incorporate the cost of applying for benefit such as filling out the required forms and learning about program rules. The simplest formulation of welfare stigma is a flat cost to participating in welfare programs. The maximization problem becomes:

$$\max_{C,Z,P} u(C,Z) - \phi P$$

where an agent's welfare participation decision $P \in \{0,1\}$ depends on the stigma parameter $\phi > 0$.

In effect, introducing welfare stigma shifts down the program segment of the budget constraint $[Z(1-t)+g]1_{[Z\leqslant\gamma]}$ by ϕ and therefore reduces the public insurance notch to $\max\{g-\phi,0\}$. If ϕ is constant across agents and $\phi < g$, then all the analyses in (1.3.1) carry through by replacing g with $\tilde{g} = g - \phi$. When ϕ is heterogeneous, the income distribution is smooth for the sub-population with $\tilde{g} = g - \phi \leqslant 0$, and analyses from previous subsections only hold true for those with $\tilde{g} > 0$. In the entire population, the qualitative predictions from (1.3.1) are still valid if (n, e, ϕ) follows a smooth distribution supported on $\mathbb{R}^4_{++}$, although the bunching and density discontinuities are less pronounced due to the existence of non-participants.[9]

Income effects As mentioned in section (1.3.1), the quasi-linear functional form of (1.1) eliminates income effects. This may be reasonable in the context of a tax rate change (i.e. a kink in the budget constraint), since as Chetty et al. (2011) note, tax rate changes have little effect on

[9]Note that allowing heterogeneity in ϕ is equivalent to allowing heterogeneity in g, but a more general interpretation of the heterogeneity in the notch size is permitted. For example, families with healthier children arguably value health insurance less than those with sicker children and would hence face a larger notch.

average tax rates. In the case of a notch, however, the absence of income effects in modeling may no longer be appropriate. Here I explore the implication of using a functional form that allows non-zero income effects.

Consider the utility function

$$u(C, Z) = \frac{C^{1-\rho}}{1-\rho} - \frac{n}{1+1/e}(\frac{Z}{n})^{1+1/e} \tag{1.7}$$

which displays constant relative risk aversion in consumption, and which encapsulates the quasi-linear utility (1.1) as a special case when $\rho = 0$. When facing a budget constraint $C = (1-t)Z$, the optimal interior pre-tax income choice is $Z^* = (1-t)^{\frac{e}{\rho e+1}} n^{\frac{1}{1+\rho e}}$. Therefore, $n^{\frac{1}{1+\rho e}}$ is the agent's desired income choice when $t = 0$. Instead of the Marshallian, Hicksian and the Frisch elasticity all being e as in the quasi-linear case, e for the utility functional form (1.7) is only interpreted as the Frisch elasticity of income/labor supply, i.e. the elasticity holding marginal utility of wealth constant. This interpretation of e will be convenient for the dynamic problem we consider below. Note that the Marshallian elasticity of labor supply with respect to the marginal tax rate reduces to $\frac{\partial Z^*}{\partial(1-t)}\frac{1-t}{Z^*} = \frac{e}{\rho e+1} < e$ when $\rho > 0$ whereas the Hicksian elasticity varies across agents.[10]

The analyses undertaken in sections (1.3.1) carry through with the more general utility function (1.7) although the expressions for the various $\bar{n}$'s will change. Therefore, the introduction of non-zero income effects does not change the qualitative predictions. That is, there is income bunching at the eligibility cutoff when agents have prefect control over their income and a discontinuous drop in income density at the cutoff when agents face a menu of finitely many labor supply choices. The intuition is that these predictions hinge on the convexity of the indifference curves, which is not altered when curvature in consumption utility is introduced.

[10] See MaCurdy (1981) and Browning et al. (1985) for discussions on the magnitudes of the three elasticities.

1.3.2 Continuous Eligibility and Labor Supply–Dynamic Models

This section extends the static framework in the previous section to incorporate continuous eligibility provisions. In essence, the provisions allow a more generous budget constraint over time than (1.2). More specifically, families that are just approved for public insurance can have income above γ and remain covered until the eligibility recertification a year later. To characterize a family's consumption and labor supply decisions in the presence of continuous eligibility provisions, I cast the family's utility maximization problem in a dynamic programming framework.

Formally, the state variable s is the number of months until recertification (s is defined to be 0 for those not claiming benefits since they will face the eligibility check when they apply), and let τ be the number of months of provided continuous eligibility. In each period, an agent chooses whether or not to participate in the program:

$$V_s = \max_{P_s} P_s V_s^1 + (1 - P_s) V_s^0$$

where $P_s = 0, 1$ denotes participation choice, and V_s^1 and V_s^0 are utilities associated with participating and not participating in the program when agents are s months away from an eligibility check. Formally, the expressions for V_s^1 and V_s^0 are

$$V_s^1 = \max_{C,Z} \{u(C, Z) + \beta V_{s'}\}$$
$$\text{s.t. } Z < \gamma \text{ if } s = 0;\ C = (1 - t)Z + g$$
$$s' = \begin{cases} s - 1 & \text{if } s > 0 \\ \tau - 1 & \text{if } s = 0 \end{cases}$$

$$V_s^0 = \max_{c,z} \{u(C, Z) + \beta V_{s'}\}$$
$$\text{s.t. } C = (1 - t)Z$$
$$s' = \begin{cases} s - 1 & \text{if } s > 0 \\ 0 & \text{if } s = 0 \end{cases}$$

For illustration purposes, first consider the simple case when $\tau = 2$, in which case s takes on the value 0 or 1. Let $\{C_s^p, Z_s^p\} = \text{argmax} V_s^p$ for $p = 0, 1$. The dynamic problem is thus simplified

to

$$\begin{aligned} V_0 &= \max_{P_0} P_0\{u(C_0^1, Z_0^1) + \beta V_1\} + (1 - P_0)\{u(C_0^0, Z_0^0) + \beta V_0\} \\ V_1 &= \max_{P_1} P_1\{u(C_1^1, Z_1^1) + \beta V_0\} + (1 - P_1)\{u(C_1^0, Z_1^0) + \beta V_0\} \end{aligned}$$

and I will characterize the optimal P_s, C_s^p and Z_s^p's below.

First note that choosing $P_1 = 1$ strictly dominates $P_1 = 0$ because (C_1^0, Z_1^0) lies in the interior of the budget set for an agent with $s = 1$. In other words, when benefit can be claimed at no cost (i.e. no restrictions on income), a rational family will do so. This reasoning simplifies the expression for V_1: $V_1 = u(C_1^1, Z_1^1) + \beta V_0$. Plugging in this expression of V_1 into that of V_0 leads to

$$V_0 = \max_{P_0} P_0\{u(C_0^1, Z_0^1) + \beta u(C_1^1, Z_1^1) + \beta^2 V_0\} + (1 - P_0)\{u(C_0^0, Z_0^0) + \beta V_0\}$$

For the agents indifferent between choosing $P_0 = 0$ and $P_0 = 1$,

$$V_0 = u(C_0^1, Z_0^1) + \beta u(C_1^1, Z_1^1) + \beta^2 V_0 = u(C_0^0, Z_0^0) + \beta V_0$$

and therefore $V_0 = \frac{u(C_0^0, Z_0^0)}{1-\beta}$. It follows that

$$u(C_0^1, Z_0^1) + \beta u(C_1^1, Z_1^1) = u(C_0^0, Z_0^0) + \beta u(C_0^0, Z_0^0) \tag{1.8}$$

If u has the functional form in (1.1), then $C_1^1 = C_0^0 + g$ and $Z_1^1 = Z_0^0$ because of quasilinearity. Consequently, $u(C_1^1, Z_1^1) = u(C_0^0, Z_0^0) + g$, and (1.8) leads to

$$u(C_0^1, Z_0^1) + \beta g = u(C_0^0, Z_0^0) \tag{1.9}$$

Suppose (C_0^1, Z_0^1) satisfying (1.9) is an interior solution. Then the convex indifference curve passing through the bundle (C_0^1, Z_0^1) is tangent to the program segment of the budget constraint and therefore lies above the non-program budget constraint $C = (1 - t)Z$. Consequently,

$u(C_0^1, Z_0^1) > u(C_0^0, Z_0^0)$ implying that $u(C_0^1, Z_0^1) + \beta g > u(C_0^0, Z_0^0)$, contradicting (1.9). Therefore, the (C_0^1, Z_0^1) that satisfies (1.9) has to be a corner solution with $Z_0^1 = \gamma$. Denote the indifferent agent's type by $\bar{n}^{dynamic}$ and expanding (1.9) using the quasi-linear functional form leads to

$$\gamma(1-t)+(1+\beta)g-\frac{\bar{n}^{dynamic}}{1+1/e}(\frac{\gamma}{\bar{n}^{dynamic}})^{1+1/e}=\bar{n}^{dynamic}(1-t)^{1+e}-\frac{\bar{n}^{dynamic}}{1+1/e}(1-t)^{1+e} \quad (1.10)$$

Equation (1.10) states that an agent of type $\bar{n}^{dynamic}$ is indifferent between choosing her interior solution on the budget constraint segment $C = (1-t)Z1_{\{Z>\gamma\}}$ and the post-tax/pre-tax income bundle $(\gamma(1-t)+(1+\beta)g,\, Z)$. Analogous to the analyses in section (1.3.1) and illustrated in the right panel of Figure 1.1, agents with $n \leqslant \bar{n}^{dynamic}$ choose to participate in the program and those with $n > \bar{n}^{dynamic}$ do not. While those with $n \in (0, n_\gamma]$ participate without altering their labor supply, those with $n \in (n_\gamma, \bar{n}^{dynamic}]$ will lower their labor supply and income when $s = 0$ to gain eligibility but revert back to their desired interior solution with an income above γ when their eligibility is not checked (i.e. $s > 0$). Comparing (1.10) to (1.3) reveals that doubling the length of the recertification period in effect doubles the benefit notch if $\beta \approx 1$. It is easy to show that for a general recertification period τ, the size of the benefit notch an agent faces is effectively $\sum_{i=0}^{\tau-1} \beta^\tau g \approx \tau g$ when making her participation decision when $s = 0$.

There are two extensions to consider. First if a cost is associated with applying for benefit, then the marginal agents at period $s = 0$ will compare choosing a notch of size $\tau g - \phi$ to their interior pre-tax income choice on $C = (1-t)Z$ where ϕ entails the application cost. The results above hold for agents whose $\phi < \tau g$. Second, when income effects are take into account, the solution may no longer be obtained analytically. Qualitatively, once a new applicant family is approved for benefit, the transfer may reduce labor supply through the income effect channel. This will imply that the rebound in income after starting a public insurance spell will not be as large as when income effects are absent. In section (1.6), I will calibrate the model that allows income effects and compare the predicted effect to that observed empirically.

To summarize, the dynamic labor supply models make the prediction that average income drops at the income eligibility check and rebounds afterward though the magnitudes of the dip and rebound depend on factors such as the size of income effects. In the following sections, I will examine whether agents empirically behave as predicted by the labor supply model.

1.4 Data and the Construction of the Analysis Sample

1.4.1 Data Sources

To examine the income and labor supply responses to the continuous eligibility provisions in Medicaid/CHIP, I use data from the 2001 and 2004 panels of Survey of Income and Program Participation. SIPP is a representative household survey designed to provide detailed information on incomes and labor force and government program participation. Each of the panel files contains four rotation groups, and they span the period from October 2000 to December 2003 and from October 2003 to December 2007 respectively. All of the rotation groups in the 2001 panel provide information for 36 consecutive months, and those in the 2004 panel for 48 months. Each adult member of the participating household was interviewed every four months about his or her experiences since the last interview (i.e. four-month reference period).

The chief advantage of SIPP over other candidate datasets (CPS, PSID, HIS, etc.) is its panel structure at the monthly frequency and the rich array of variables including detailed information on income, program participation, and family structures. Since the focus of the study is to examine families' income and labor supply behavior before and during their children's Medicaid/CHIP spells, SIPP is the best choice among public use survey data sets for the purpose of this study.

There are, however, several limitations of the data used. First is the existence of the well-known seam bias, which refers to the fact that changes in status are under-reported within a four-month reference period while over-reported between two reference periods (see, for example, Pischke (1995) and Ham et al. (2009) for details). As noted above, the interviews are

not conducted every month but every four months, and children's reported public insurance coverage spells are therefore much more likely to start on the first month of the reference period than the second, third or fourth. In fact, about 80% of the fresh spells in the 2001 panel and 90% of the fresh spells in the 2004 panel start on the first month of the reference period. Complications created by the seam bias will be addressed in section (1.6).

Second, Medicaid and CHIP coverages cannot be reliably distinguished, and this is true with all survey data.[11] Therefore, I will use public insurance coverage which encapsulates both Medicaid and CHIP, and the phrase "public insurance" will be used interchangeably with Medicaid/CHIP or simply Medicaid. When computing the value of the benefit notch, I will use the CHIP government spending per enrollee, which is lower than that of Medicaid. The implication of ignoring the higher benefit notch Medicaid applicants face will be discussed in section (1.6).

Third, identifiers of families in less populated states are missing from the 2001 panel. Families in Maine and Vermont share the same state identifier as well as those in North Dakota, South Dakota and Wyoming. Because these states have different Medicaid/CHIP policy parameters, they are excluded from analyses. Due to the larger sample size of the 2004 panel, all fifty states plus the District of Columbia have their own identifier, and therefore all states are included.

The main analysis sample consists of children who started a public insurance spell during the SIPP panel. The restriction to those with "fresh" spells (as opposed to the left-truncated spells that start with the child's first appearance in the panel) comes from the necessity of identifying when the family applied for benefit, which is not possible with the left-truncated spells. In addition, children younger than the age of 1, children whose families moved to another state during the spell and children who were on the Supplemental Security Income (SSI) program at the beginning of a public insurance spell are excluded from the analysis sample. Infants are excluded because most states have been extending presumptive eligibility to infants since the 1990's; children whose families moved across states pose a challenge in assigning

[11] An expert at the U.S. Census Bureau noted in a correspondence that "respondents rarely know with certainty whether their child is in Medicaid or CHIP... We found this out with the 2004 SIPP instrument, where question order happened to be revised so that CHIP was asked about before Medicaid. Here we observed that respondents were most likely to answer the question asked first, resulting in higher reported levels of CHIP than of Medicaid for Panel 2004)."

Medicaid/CHIP parameters; and children who are on the SSI program are conferred automatic eligibility for public insurance. As shown in Table (1.2), the analysis sample consists of–for the 2001 and 2004 panels respectively–7158 and 8321 fresh spells in total and 3096 and 3312 fresh spells from children in the states that offer continuous eligibility.

Nuclear families for each child are constructed using information on the relationship to household and family reference person (head). In cases where a child and his or her parent(s) live with other adults, however, families include only the children and parent(s) of the appropriate subfamily. This definition corresponds to the family assistance unit that would be potentially eligible for Medicaid/CHIP. Family level variables are then calculated by aggregating over individual family members, and family income includes earned and unearned incomes excluding welfare receipts and children's incomes.[12]

The state level Medicaid/CHIP data are extracted from reports issued and databases maintained by various organizations. The policy parameters (e.g., continuous eligibility, presumptive eligibility, income eligibility cutoffs, etc.) come from NGA (2000-2008), Kaiser (2002-2011) and CMS (Various Years). Medicaid/CHIP spending and enrollment data are extracted from the Kaiser Foundation State Health Facts database and the CMS Medicaid Statistical Information Statistics System.

1.4.2 Adjustment to Potential Measurement Error in Public Insurance Coverage

The accurate identification of a public insurance spell start is important for the empirical analysis in the following sections. In the data, many fresh spells are preceded by a short gap in public insurance coverage. Short coverage gaps are not inconsistent with what is found using administrative data. For example, using administrative data from Ohio, Fairbrother et al. (2011) show that 40% of the children whose coverage was not renewed at month 12 re-enroll within a

[12]The exclusion of children's income is due to the fact that student income is disregarded for the purpose of Medicaid/CHIP eligibility determination. Whether or not other adult family members' incomes are included in the computation of family income in addition to those of the parents makes little difference empirically.

short period. However, it does cause concern regarding the reliability of identifying the start of a spell. Although it is rare for families to report coverage while they are not on public insurance (Card et al. (2004)), under-reporting coverage at a particular month during a long spell will lead to the false identification of starting a fresh spell.

To address this potential measurement error problem, I construct a subsample consisting of children who report no public insurance coverage for 12 consecutive months before the start of a spell. This subsample will be henceforth referred to as the "long gap" sample. As shown in Table (1.2), the 2001 panel long gap sample includes 501 spells from 501 children in 302 sample units and the 2004 panel long gap sample includes 815 spells from 804 children in 494 sample units. Note that the long gap sample contains mostly single spells by virtue of its construction.

Table 1.3 presents summary statistics for both the full sample and the long gap sample in the 2001 and 2004 SIPP panels. In the full sample, a child switches onto public insurance during a SIPP panel when she is between 8 and 9 years old on average. About half of the children are female; a quarter are black in the 2001 panel and only 17% in the 2004 panel. The average child lives in a four-person nuclear family, and 40% to 50% come from a single-parent family. The vast majority of the families are working families–only around 10% do not report earnings around the time their child starts a public insurance spell and less than 5% of the parents claim unemployment benefits. Less than 7% of the families in all samples receive welfare cash transfers from TANF programs. Many children report private insurance coverage while on Medicaid/CHIP, and this dual or overlap coverage phenomenon has been observed by other studies. Gruber and Simon (2008) note that there is no clear evidence on the "correct" interpretation of the overlap but state two hypotheses: 1. It could be individuals moving from private to public insurance or 2. CHIP claimants report being on both public and private insurance because it is often delivered by HMOs. Under both hypotheses, I can interpret the switch from no public insurance to public insurance as the start of a public insurance spell, and therefore dual coverage does not pose a threat to the empirical studies that ensue. Finally, only a small fraction of parents switch onto Medicaid with their children.

In comparison, children in the long gap sample come from families in better economic conditions. Fewer children (10 percentage points) come from single-parent families, and family incomes are more than 30% higher as compared to the full sample. In the next section, I will present results for both the full sample and the long gap sample. In either sample, I find little evidence of the predicted strategic dip-and-rebound behavior.

1.5 Descriptive Analysis of Income and Labor Supply Responses

In the section, I present descriptive evidence on families' income response over their childrens' Medicaid/CHIP spell. I follow a flexible specification adopted by Jacobson et al. (1993). Specifically I estimate

$$y_{it} = \omega_i + \nu_t + \sum_{|k| \geqslant m} D_{it}^k \delta_k + \epsilon_{it} \tag{1.11}$$

where ω_i and ν_t are individual and calendar month fixed effects respectively and D_{it}^k is a set of dummy variables indicating months after the start of a public insurance spell.[13] $D_{it}^k = 1$ if child i started her public insurance spell at month $t - k + 1$. As a special case, a child with $D_{it}^0 = 1$ started her public insurance spell in month $t + 1$, and another child with $D_{it}^1 = 1$ started her spell in month t. Month 0 is the omitted category and the δ_k's measure the difference in the average outcome k months after the start of a spell relative to the value at the beginning of the spell. I examine the income and labor supply responses 24 months before and after the beginning of a spell. Since families were interviewed for 36 months in the 2001 panel and 48 months in the 2004 panel, it is possible to allow $m = 35$ for the 2001 panel and $m = 47$ for the 2004 panel. However, there are few families who start a spell at the second month or the last month of the panel which render the estimation of δ_m and δ_{-m} imprecise for large m. Even though the sample period for $m = 24$ covers two years of data surrounding an additional recertification period ($m = 12$), I will

[13] Inclusion of time-varying covariates such as state monthly unemployment rate changes the empirical results little.

only focus on the initial entry of program participants. The reason is mentioned in Section (1.2): families receive a package containing renewal materials 60-90 days prior to the end of the 12-month period, which creates ambiguity in the timing of potential strategic behavior at eligibility recertification.

According to the theoretical models in Section (1.3), the δ_k's are expected to be positive for k not being a multiple of 12 (eligibility check points) if there is strategic behavior. However, even in the absence of any strategic behavior, one might expect a mechanical dip as pointed out by Ashenfelter (1978) and Ashenfelter and Card (1985). For example, if selecting into public insurance is based on $y_{i(t_0-k)} < \bar{y}$ where t_0 is the month of starting the spell and that the ϵ_{it}'s are serially correlated, then a dip and rebound may be expected based purely on mean reversion, though not as abrupt as predicted by the strategic behavior. Given the existence of the mean-reversion mechanism, the dip and rebound due to the strategic behavior should be more pronounced.

Note that many families in states that provide 12-months of continuous eligibility do not report coverage for all 12 months after beginning the spell. Likewise, many families only experience short gaps leading up to the beginning of a public insurance spell. This makes it difficult to interpret the δ_k's in an event-study framework, which usually calls for single status transitions. Therefore, I will truncate the analysis sample for each child and focus on the single transitions. Specifically, let k^+ denote the first month after month 0 the child switches off public insurance, and let k^- denote the last month before month 0 the child was covered by public insurance. I will discard all observations after k^+ and those before k^-.

Figure (1.2) and Figure (1.3) plot the movement of (unweighted) average family income over the 24 periods around the beginning of a public insurance spell. Both point estimates and standard errors are shown, where the standard errors are clustered at the sample unit level. None of the figures show a pronounced dip-and-rebound in the six months before and after the spell start. For the 2001 Panel, the income trend leading up to the beginning of spell is flat in both samples, however, the 95% confidence interval does not rule out a downward trend. For the full sample

specification, the income increases gradually especially after 12 months, but the period immediately following the spell start shows no rebound. Even the upward income trend between 12 months and 24 months after the spell is not statistically different from zero, and it disappears altogether when restricting to the long gap sample in Figure (1.3). In the 2004 panel, the income process shows a persistent downward trend in both samples, and the estimates of δ_k are significantly positive for many $k < 0$ and negative for many $k > 0$.

Unfortunately, SIPP does not collect information on hours worked on a monthly level but usual hours worked are reported for the entire wave. If the labor supply models presented above are true and imply only supply movements immediately before and after a public insurance spell, then using the usual hours worked variable may cloud the dip-and-rebound pattern. Instead, I describe the movement of the less refined labor supply variables at the monthly level before and after the start of a public insurance spell. In particular, I construct a dummy variable indicating whether or not the head of the family–defined to be father in a two-parent family and all single-parent family heads–worked more than 35 hours for all weeks during a month, and plot its movement in Figure (1.4) and Figure (1.3) for the full sample and long gap sample respectively. In all figures, there is a slight downward trend for the two years before the spell start, and it continues in the 2001 panel for at least 6 months in both samples. The trend flattens after the insurance spell in the full sample of the 2004 panel, but the downward trend continues as with the 2001 panel when I restrict to the long gap sample.

Even though the strategic behavior is not visible for the full and long gap sample, the question remains whether income responses are pronounced for a particular subgroup, which may not be detected when averaging with others. I present income responses for four subgroups that may be more likely to respond to the dynamic opt-in incentive. The four subgroups are: (1) children in two-parent families during the sample period, (2) children whose parents worked in the construction industry during the sample period, (3) children whose mother is college educated and (4) children in states offering no presumptive eligibility to non-infants. Two-parent families may be less credit constrained and have more flexibility in adjusting labor supply; many jobs in

the construction industry are of seasonal nature and parents may time their government benefit application when they do not work; highly educated parent(s) may be more likely to understand program rules; and parents in states that do not offer presumptive eligibility (see section (1.2) for definition) will need to have their income low enough so that their children can acquire public health insurance. Figures (1.6), (1.7), (1.8) and (1.9) plot the income responses for the four subgroups respectively. Contrary to expectation, they do not show the apparent dip-and-rebound pattern. In fact, the trends are fairly similar to those observed in the full sample.

In summary, both the labor supply theory and mechanical mean reversion predict a dip-and-rebound in family income and labor supply around the start of a public insurance spell. However, I do not find such patterns using the various subsamples of the SIPP 2001 and 2004 panel. In some cases, the average incomes are significantly below that of month 0, but small sample size leads to a 95% confidence interval whose upper bound is positive. Therefore, strategic behavior cannot be strictly ruled out. In the next section, I present results from calibrating the labor supply models using various elasticity measures and test whether the empirical estimates are consistent with the quantitative theoretical predictions.

1.6 Calibration

In this section, I provide quantitative predictions by calibrating the simple models presented in section (1.3). Specifically, I focus on the dynamic model where the labor supply choice is continuous and the flow utility is of the form (1.3.1) and (1.7) and attempt to find the labor supply elasticity parameter that is consistent with the empirical evidence. The case where labor supply choices are discrete is currently being investigated and will be included in a future version of the paper.

For the calibration exercise, I generate 100,000 families for whom I assign the taste parameter n, income eligibility cutoff γ, size of the benefit notch as well as the tax rate t. The distribution of n is non-parametrically estimated using the family income distribution from SIPP data. In

particular, recall that the optimal pretax income choice for a family with quasi-linear utility and taste parameter n facing the budget constraint $C = (1-t)Z$ is $Z^* = (1-t)^e n$. For each family in SIPP, I assign the marginal federal income tax rate based on their income and family composition.[14] Using families with children residing in states that offer continuous eligibility, I impute their value of n as $Z^*/(1-t)^e$ where Z^* is family income and estimate the distribution of n using a kernel density estimator for each e as in Brewer et al. (2010). In the case where the flow utility is of the form (1.7), $n = \frac{Z^*}{(1-t)^e}[(1-t)Z^*]^{\rho e}$.

In this section, I consider only the CHIP benefit notch, which is the notch at the highest public insurance eligibility cutoff. The CHIP notch is smaller than that of Medicaid as most states demand no premium payment for Medicaid and lower or no copay.[15] Ignoring the larger Medicaid notch will bias downward the theoretical prediction of the size of dip and rebound as discussed below. An estimate of the benefit notch value g comes from the CHIP spending data collected by the Kaiser Foundation and the Center for Medicare and Medicaid Services. The spending variable excludes beneficiary and third-party payment and should reflect expected government subsidy. Unfortunately, state-by-state spending per child enrollee figures are not readily available from either source for years earlier than 2004. Therefore, I use the average per-enrollee-spending for the entire U.S. as a measure of the notch, which had increased from \$835 in 2001 to \$1217 in 2007 in nominal terms (approximately 5% annual growth rate). The monthly benefit amount per child g is $\frac{1}{12}$ of the annual spending per-enrollee spending, and the size of the notch a family faces is g times the number of children they have. The marginal distribution of the number of children in each family in the simulation sample mimics that in SIPP–approximately 32%, 37% and 21% of families have 1, 2 and 3 children in both the 2001 and 2004 panels of SIPP. Finally, the CHIP income eligibility cutoff γ is the average of the cutoffs families face in continuous eligibility states in SIPP sample.

I show calibration results by using two models–1. the quasi-linear model, i.e. $\rho = 0$ and 2. the

[14] Specifically, I assume that parents in a dual-headed family file jointly and claim the deduction accordingly and that all families claim standard deductions.

[15] According to the Kaiser Family Foundation, the annual per child government spending in Medicaid and CHIP are \$2171 and \$1363, respectively, for the 2008 fiscal year.

model with $\rho = 0.74$, which is the average value of ρ from empirical literature as surveyed by Chetty (2006). As expected, when income effects are taken into account, the magnitudes of the dip and the rebound are smaller as the increase/decrease in income render agents demand more/less leisure. Also, the rebound is smaller in magnitude than the dip, reflecting the demand for more leisure as children in families acquire public insurance.

Calibration results from the SIPP 2001 and 2004 panels are presented in Table 1.4 and Table 1.5 respectively as well as in Figure1.10. I compare model predicted income responses based on e=0.15[16], e=0.05 and e=0 to those observed empirically, and model predictions are compared to both the empirical point estimates and the upper end points of the 95% confidence intervals to account for sampling errors. Even the smallest model predicted rebound responses, which are based on $e = 0$, are larger than the upper end points of the 95% confidence interval, except for month 10 in the long gap sample of the 2004 panel.[17] The comparison therefore points to little labor supply response.

Note that the theoretical prediction of dip and rebound magnitudes may still be an underestimate for two reasons. First, as pointed out previously, the existence of the Ashenfelter dip should accentuate the dip and rebound. Second, because of the high income cutoff of CHIP, there are many programs agents eligible for CHIP may qualify for if they reduce their income. For example, even though a five-year-old child in a family with income at 140% of the FPL is eligible for CHIP in practically all states, the family will face a more generous transfer by reducing their income to 133% of the FPL. In this case, the child will qualify for Medicaid in every state and the family will incur no premium payments and the lowest co-pay. If the family is willing to reduce their income further to 130% of the FPL, they will also gain eligibility for the Supplemental Nutrition Assistance Program (formerly Food Stamps).

Formal statistical tests are carried out to examine whether the average empirical rebound magnitude is consistent with the model predictions. Let $\bar{\delta}_S$ denote the average δ_s's for s between 1

[16] $e = 0.15$ is the average Hicksian elasticities in the empirical micro literature for non-top income population as surveyed by Chetty (Forthcoming).

[17] Note that even $e = 0$, which corresponds to the case where the indifference curve in the flow utility is Leontief, still implies strategic behavior. This is the result of having a notch–as opposed to a kink–in the budget constraint.

and S: $\frac{1}{S}\sum_{s=1}^{S}\delta_s$, where the δ_s's are the coefficients in (1.11), and let $\delta^{e=0}$ be the rebound magnitude predicted by the model with $e=0$. The baseline test is whether the average empirical rebound magnitude for the eleven months after starting a public insurance spell is as large as predicted: (1) H_0: $\bar{\delta}_{11}=\delta^{e=0}$ vs. H_1: $\bar{\delta}_{11}<\delta^{e=0}$. In light of the ambiguity introduced by receiving the renewal package 60 to 90 days prior to the end of the 12-month period, the potential strategic behavior for recertification may occur as early as during month 10. Therefore, I drop month 10 and 11 from the baseline in an alternative test: (2) H_0: $\bar{\delta}_9=\delta^{e=0}$ vs. H_1: $\bar{\delta}_9<\delta^{e=0}$. Finally, other tests are called for in the presence of seam bias. As noted in Pischke (1995) and Ham et al. (2009), seam bias is typically interpreted as the result of the "telescoping behavior"–survey respondents answer retrospective questions using their most recent status. In the most extreme scenario where every respondent telescopes in reporting public insurance coverage, which will imply the largest measurement error in the transitions into public insurance, the true month-0 income is observed only for those who actually start a public insurance spell in month 1. Assuming that there is equal probability of starting a spell in month 1, 2, 3 and 4 of a particular wave, then the expected dip-and-rebound magnitude is only a quarter of what is predicted by the neo-classical model. Hence, the final hypothesis I test is (3) H_0: $\bar{\delta}_8=\frac{1}{4}\delta^{e=0}$ vs. H_1: $\bar{\delta}_8<\frac{1}{4}\delta^{e=0}$ where I only include eight months of data because month 9 may be contaminated by the telescoping behavior in wave 3 after the start of the public insurance spell.

Table 1.6 presents the p-values associated with the three statistical tests above. The null hypothesis for all three tests are rejected based on all four samples (full and long gap sample for the 2001 and 2004 panels respectively) in favor of the alternative at the 5% level, except for test (3) in the 2001 panel full sample, which is rejected at the 10% level. To summarize, the empirical evidence points to little labor supply response and is in general inconsistent with the labor supply model. The findings likely imply frictions in income adjustments or that the perceived value of CHIP for families above the eligibility cutoff is lower than the expected government subsidy. Identifying the precise reason for the lack of strategic behavior is beyond the scope of this paper. With the labor supply responses practically ruled out, however, I will use a mechanical model in

the subsequent analysis in the sense that incomes are drawn from a stochastic process as opposed to being controlled by the families. I compute the optimal continuous eligibility period length based on observed income processes from SIPP under various recertification cost parameters.

1.7 Optimal Length of the Continuous Eligibility Period

Because of the lack of income/labor supply responses to the continuous eligibility provision shown in section (1.6), I compute the optimal length of the recertification period for Medicaid/CHIP, τ, based on a mechanical model for individual behavior. In this model, labor supply considerations are absent from families' optimizing decisions and incomes simply follow a stochastic process. After the realization of income Z, consumption is determined by

$$C = [(1-t)Z + g]P + (1-t)Z(1-P)$$

where P is an indicator variable denoting program participation status, and agents' utility u only depends on the consumption level C.

In order to determine the optimal transfer policy, I specify next the social welfare function, which contains two components. The first component is a standard Bergson-Samuelson functional of weighted individual utilities, W, and the second component is surplus in the government's budget, S, which can be used to finance a public good (Salanie (2003)). I assume that the two components are additive and that the welfare resulting from the public good is linear in its spending. As an illustration, when eligibility check is performed every month and take-up rate is 100% (i.e. $P = 1_{[Z \leqslant \gamma]}$) and when eligibility monitoring is free, the per-period social welfare for each value of γ and t is given by:

$$\underbrace{\int \Psi(u(C(z)))f_Z(z)dz}_{W} + \omega \underbrace{[R - \Pr(Z \leqslant \gamma)g]}_{S} \tag{1.12}$$
$$\text{s.t.} \quad C(z) = [(1-t)z + g]1_{[z \leqslant \gamma]} + (1-t)z1_{[z>\gamma]}$$

Ψ is an increasing and concave function that weights the utilities of individual agents according to the social planner's redistributive taste, and ω reflects the contribution of S to overall social welfare relative to that of agents' utilities. t is the pre-determined marginal tax rates on income, and f_Z and F_Z specify the p.d.f and c.d.f. of pre-tax income Z respectively (assuming the stationarity of the income process which will be relaxed later). The government collects per-agent revenue R, which may contain income tax revenue $\int tzf_Z(z)dz$ as well as sources not explicitly modeled here,[18] and it is assumed that R is sufficiently large to cover program spending: $R \gg \Pr(Z \leqslant \gamma)g$.

The formulation of the social welfare function (1.12) differs slightly from a textbook approach (e.g., Salanie (2003)) in the follow respects. First, government surplus does not usually enter the social welfare function directly but through a balanced budget constraint. As mentioned above, however, the direct welfare impact of the S term can be attributed to its interpretation as spending on a public good (p. 81 Salanie (2003)). Salanie (2003) notes that the dependence of utility on S is neglected in his model because the spending on the public good is held constant. The specification (1.12) simply extends that of Salanie (2003) by allowing the production of the public good to be variable, and the additional advantage of this specification will be clear in the remainder of this subsection.

Second, having a "notched" lump sum transfer schedule with the associated cutoff γ as the policy instrument is not prevalent in the optimal design literature. In fact, if the income tax schedule is completely flexible and that $\Psi \circ u$ is strictly concave, then the government should choose a transfer function that equalizes consumption across agents when labor supply decisions are not considered in the model (a special case is studied as early as in Edgeworth (1897)). When labor supply incentives are considered, the seminal paper of Mirrlees (1971) shows that the marginal tax rate always lies between zero and one which precludes a discrete drop in consumption as pre-tax income increases if the optimal tax schedule is completely flexible. However, Blinder and Rosen (1985) and Slemrod (2010) argue that it is possible to institute a

[18]For example, part of the federal CHIP funding comes from tobacco taxes.

notch as part of an optimal schedule when the set of income tax instruments is limited, e.g., linear.[19] Given their theoretical argument and the practical relevance of a notch-based transfer schedule, I continue with the specification of (1.12).

A dynamic extension of the baseline formulation (1.12) is called for when evaluating the optimal recertification period. I consider a T-period problem, where the public insurance program becomes available in period 1, and in every period, families' eligibility depends on their income and program participation history through the continuous eligibility provision. For example, when the continuous eligibility period is 3, a family is assumed to automatically participate in the program in period 2 and 3 if it was eligible and participated in the program in period 1. In period 4 when the family's eligibility is recertified, the participation status will depend on whether their period-4 income falls below the threshold. Formally, the social welfare function becomes

$$\begin{aligned} & \sum_{m=1}^{T} \beta^{m-1} E[\Psi(u(C_m - \phi R_m))] - \omega E[R - (gP_m + \kappa R_m)] \\ \text{s.t.} \quad & C_m = [(1-t)Z_m + g]P_m + (1-t)Z_m(1-P_m) \end{aligned} \tag{1.13}$$

The expectation is taken over the joint distribution of $\{Z_1, ...Z_T\}$; P_m and R_m are dummy variables indicating whether a family participates in the program and whether program eligibility is certified in month m, and they are determined by the family income histories and the recertification period as illustrated above. In addition to spending on public insurance benefits, each eligibility check costs the government and the participating family κ and ϕ to perform, respectively.

In this paper, I consider the problem where the government chooses the continuous eligibility period τ to maximize social welfare taking tax rate t and eligibility cutoff γ as given (subsequently denoted by t^* and γ^*). Although it is tempting to cast the model in continuous time

[19]The theoretical properties of means-tested in-kind transfers in an optimal-design context have also been studied in Nichols and Zeckhauser (1982), Blackorby and Donaldson (1988), Gahvari (1995), Cremer and Gahvari (1997), Singh and Thomas (2000), etc., These studies typically consider the problem with two types of agents and a transfer scheme that ensures second-best allocation, i.e. the high type does not pretend to be the low type and claims benefit transfer. See Currie and Gahvari (2008) for a survey.

instead of (1.13) and analytically solve for τ, this approach is not feasible for the problem at hand due to its discrete nature. The dynamic budget constraints resulting from the continuous eligibility provision can only be specified discretely for each month, and it is not obvious how to formulate its continuous counterpart or even if it is at all possible. Therefore, I proceed to determine the optimal continuous eligibility period length by comparing numerically calculated social welfare corresponding to different recertification periods.

Obtaining the empirical income processes is crucial to the computation of social welfare. For the numerical exercise, I again rely on the 2001 and 2004 panels of SIPP but only keep the families that appear in all months of the panels for the observability of income processes over a long period of time (three years for the 2001 panel and four years for the 2004 panel, which translate to $T = 36$ and $T = 48$ for the two panels respectively). In my simulations, once I specify an eligibility recertification period, I can impute each family's monthly program participation decision based on its income history (assuming full take-up) and calculate its consumption accordingly.

The remaining missing piece for the numerical exercise of determining the optimal τ is specifying the value of ω and the functional form of Ψ. In order to obtain ω, I assume that the observed policy parameter γ^* is the solution to the frictionless optimization problem (1.12) where the cost of eligibility certification is zero. That is, the government abstracts away from the monitoring problem when determining the eligibility cutoff.[20] I can then solve for ω following the first order condition of γ as

$$\omega = \{\Psi([(1-t^*)\gamma^* + g]) - \Psi([(1-t^*)\gamma^*])\}/g \tag{1.14}$$

where individual utility u is assumed to be linear in consumption. For the weighting function Ψ, I follow Brewer et al. (2010) and specify $\Psi(u) = \frac{u^{1-\nu}}{1-\nu}$ with the redistributive taste parameter $\nu = 1$, and it follows that $\omega = \frac{1}{g}\log\frac{(1-t^*)\gamma^*+g}{(1-t^*)\gamma^*}$. Because each family faces different eligibility

[20] ω will be smaller if monitoring cost is taken into consideration: $\omega = \{\Psi([(1-t^*)\gamma^* + g - \phi]) - \Psi([(1-t^*)\gamma^*])\}/(g+\kappa)$, and the resulting continuous eligibility period will be longer.

cutoffs, benefit notches and tax rates depending on its composition and income, I calculate ω using the average g, γ^* and t^* families face in SIPP similar to section (1.6), and the resulting value of ω is approximately 0.00026 in the both the 2001 and 2004 panels. Finally, β, the monthly discount rate is taken to be $(0.95)^{(1/12)}$.

In my simulations, I assume that families have no participation history prior to the beginning of the panel as mentioned above. This assumption precludes the possibility that a family with income above the eligibility cutoff can participate in the program at the very beginning of the panel. Consequently, families will incur a cost for eligibility check, ϕ, when they first become eligible during the panel (i.e. when I first observe their income dip below the cutoff). After identifying the first time a family participates in the program, participation status in the ensuing months can be determined under each specified continuous eligibility period length, τ, and the observed income process–in each month, families claim program benefits if they are still within the eligibility period or if their income is below the cutoff. With the ingredients g, t^*, γ^* and ω known for each family, I will be able to compute the welfare in (1.13) under combinations of τ, κ and ϕ.

Specifically, I choose $\tau = 1, 2, ..., 36$ months, and κ and ϕ take on the values of \$0, \$9.5, \$19, \$28.5 and \$38 in 2010 dollars respectively. Note that 36 is the number of months in the SIPP 2001 panel, and so the data will not be informative for $\tau > 36$. The basis for the choices of κ is that the median hourly wage rate for government program interviewers is around \$19 in May 2010 according to the Occupational Employment Statistics database of the Bureau of Labor Statistics, and the cost parameters correspond to 0, 0.5, 1, 1.5 and 2 hours of work for eligibility recertification respectively (similar estimates of recertification costs are carried out in Irvin et al. (2001)[21]). There is no formal reason to choose the same cost parameters for program participants as those for κ, and I do so here simply for convenience. In general, it may be more costly for a family to have their eligibility certified because it involves finding out information, gathering

[21]Irvin et al. (2001) simulates the impact of implementing the 12-month continuous eligibility provision on Medicaid coverage, payment and administrative costs using program data for four states between 1994 and 1995. The study serves as a good benchmark, but it does not consider the potential labor supply effects of the implementation. Moreover, it does not model the impact under alternative lengths of the continuous eligibility period.

proof of incomes, filling out the application forms and sometimes traveling to meet face-to-face with their case worker on a work day. While $38 (2 hours of work for a case worker) is probably too high a cost on the government, it may not be unreasonable for a family trying to continue benefits.

Table 1.7 presents the optimal length of the continuous eligibility period from my simulations under the 25 combinations of κ and ϕ (5 values for κ and 5 for ϕ).[22] Not surprisingly, the government should certify eligibility every period when it is costless to do so and should check less frequently as costs increase. An increase in the cost on families, ϕ, is more likely to lengthen the recertification period than an equal-sized increase in κ. The two SIPP panels give similar results on the optimal level τ. In particular, the estimates from both panels point to an optimal τ of 12 months for all values of κ when $\phi = \$19$, which may be reasonable and possibly an underestimate for the value of ϕ.

There are two caveats in interpreting the results of Table 1.7. First, given that the simulation sample consists of families that responded to interviews for all months during the SIPP panels, their income process may be different from those that are not included in the sample. Therefore, the resulting optimal length of the continuous eligibility period is sample-specific and may not apply for the general population. Second, I have so far abstracted away from the consideration of imperfect take-up and self-selection. Incorporating imperfect take-up may lengthen the optimal continuous eligibility period. In fact, simulations that allow random non-participation among eligible families point to longer optimal recertification periods, reflecting the cost in social welfare of having eligible families dropping out of insurance coverage due to frequent eligibility checks. When setting the monthly enrollment probability to be 12.5%, which amounts to an 80% annual participation rate as per Kaiser (2002-2011) assuming independent enrollment decision from month to month, the implied average optimal recertification period is 30 months or more under all cost parameter combinations.

[22]Note that the prevalence of the optimal continuous eligibility periods that are multiples of 4 in Table 1.7 is a result of the 4-month seam structure of SIPP. Since most of the changes occur between waves, a τ which is not a multiple of 4 may lead to much more income change during the continuous eligibility period and is therefore less likely to be the maximizer of the social welfare function.

A related point is self-selection; that is, those who take up benefits place a particularly high valuation on program benefits, in which case there is variation in g across agents and possibly over time as well. Nichols and Zeckhauser (1982) argue that the government may consider imposing an "ordeal" mechanism so that only those most in need will select into the program, and one such "ordeal" is frequent eligibility checks. However, Currie and Grogger (2001) show that single-parent families, who were arguably among the most needy, disproportionately dropped out of the Food Stamp program when the frequency of recertifications had increased. This empirical example demonstrates that an ordeal mechanism may have the opposite effect than intended by Nichols and Zeckhauser (1982) if valuations of benefit and the severity of "ordeal" due to frequent recertifications are positively correlated for a given family. This correlation is key in a formal analysis of the effect of recertification frequency, and will be investigated in a future version of the paper.

1.8 Conclusion

This paper presents both a positive and normative analysis regarding the continuous eligibility provisions in a means-tested program. For the positive analysis, it investigates both theoretically and empirically the impact of continuous eligibility on the income and labor supply responses of program participants. Neo-classical labor supply models predict that a long eligibility recertification period provides strong dynamic opt-in incentives wherein families lower their income to gain program eligibility, acquire government-provided benefits for the continuous eligibility period and revert back to their "optimal" interior consumption bundle. Using the 2001 and 2004 panels from SIPP, I follow a flexible specification as in Jacobson et al. (1993) to describe the income and labor supply processes for families participating in Medicaid/CHIP and the point estimates do not indicate the dip-and-rebound strategic behavior around the time a child initially gains public insurance coverage. In addition, neo-classical dynamic labor supply models based on Saez (2010) are calibrated using family income and composition information, Medicaid/CHIP

policy parameters and income tax rates. Comparing the magnitudes of the predicted strategic behavior to those observed empirically rejects the simple labor supply model in most subsamples.

With the positive analysis practically ruling out labor supply responses, I propose a framework utilizing the mechanical properties of the income processes in SIPP to answer the normative question: what the length of the continuous eligibility period should be. I derive a mapping from various combinations of cost parameters associated with eligibility recertification to the optimal length of the continuous eligibility period. Under moderate cost parameter values ($19 to the program participants for going through the application procedure), the evidence is suggestive that the optimal recertification period is no shorter than 12 months. Consequently, it may be beneficial for states currently allowing a 6-month renewal period to extend it to 12 months.

Appendix

States Providing Continuous Eligibility and Presumptive Eligibility

Continuous Eligibility

- The states that provided 12 months of continuous coverage in the sample period (2001-2007) are: Alabama, California, Washington D.C., Idaho, Illinois, Iowa, Kansas, Louisiana, Mississippi, Maine, Michigan (after Jan 2003), New York, North Carolina, Ohio, Pennsylvania, South Carolina, Washington, and West Virginia (after Oct 2002).
- The states that did not provide 12 months of continuous eligibility are: Colorado, Georgia, Hawaii, Kentucky, Missouri, Montana, Nevada, New Hampshire, North Dakota, Oklahoma, Rhode Island, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Wisconsin, and Wyoming.[23]
- In the states that remain–Alaska, Arizona, Arkansas, Connecticut, Delaware, Florida, Indiana, Maryland, Massachusetts, Minnesota, Nebraska, New Jersey, New Mexico, and Oregon–the rules for continuous eligibility are complicated with changes or different implementations for different programs during my sample period, and are thus dropped from my analysis sample.

Presumptive Eligibility

- The states that provided presumptive eligibility are California, Connecticut, Florida, Illinois, Massachusetts, Michigan, Missouri, Nebraska, New Hampshire, New Jersey, New York and Oklahoma.

Discrete Labor Supply Choices

In this section, I show that there is no income bunching but a discontinuity in the income density at the eligibility cutoff in the case where finitely many choices of hours are allowed. Formally, let

[23]Note that many of the states that did not provide continuous eligibility allowed a 12-month renewal period.

the menu of labor supply choices be $\{h_1, h_2, ..., h_d\}$ where $0 = h_1 < h_2 < ... < h_d = 1$. For each w, let $\bar{n}_{h_i,h_{i+1}}(w)$, $i = 1, 2, ...d$, be the type of agent indifferent between choosing $H = h_i$ and $H = h_{i+1}$, and it follows that agents facing wage w of type $n \in (\bar{n}_{h_{i-1},h_i}(w), \bar{n}_{h_i,h_{i+1}}(w)]$ choose $H = h_i$.[24] $\bar{n}_{h_i,h_{i+1}}(w)$ varies smoothly with w except when $wh_i = \gamma$ and $wh_{i+1} = \gamma$. There is a discontinuous increase when $w = \gamma/h_{i+1}$ and a discontinuous decrease when $w = \gamma/h_i$. The c.d.f. of Z at $z > 0$ is

$$F_Z(Z < z) = \Pr(H = 0) + \sum_{i=2}^{d} \int_0^{z/h_i} \int_{\bar{n}_{h_{i-1},h_i}(w')}^{\bar{n}_{h_i,h_{i+1}}(w')} f_{n,w}(n', w')dn'dw'$$

F_Z is continuous at γ because $\bar{n}_{h_i,h_{i+1}}(w)$ is right continuous for all i and that $f_{n,w}$ is continuous. However,

$$\begin{aligned}
\lim_{z\uparrow\gamma} f_Z(z) &= \textstyle\sum_{i=2}^{d} \quad \frac{1}{h_i} \int_{\bar{n}^l_{h_{i-1},h_i}(\gamma/h_i)}^{\bar{n}^l_{h_i,h_{i+1}}(\gamma/h_i)} f_{n,w}(n', \frac{\gamma}{h_i})dn' \\
\lim_{z\downarrow\gamma} f_Z(z) &= \textstyle\sum_{i=2}^{d} \frac{1}{h_i} \quad \int_{\bar{n}^r_{h_{i-1},h_i}(\gamma/h_i)}^{\bar{n}^r_{h_i,h_{i+1}}(\gamma/h_i)} f_{n,w}(n', \frac{\gamma}{h_i})dn'
\end{aligned}$$

where

$$\begin{aligned}
\bar{n}^l_{h_i,h_{i+1}}(\frac{\gamma}{h_i}) &\equiv \lim_{w\uparrow(\frac{\gamma}{h_i})} \bar{n}_{h_i,h_{i+1}}(w) > \lim_{w\downarrow(\frac{\gamma}{h_i})} \bar{n}_{h_i,h_{i+1}}(w) \equiv \bar{n}^r_{h_i,h_{i+1}}(\frac{\gamma}{h_i}) \\
\bar{n}^l_{h_{i-1},h_i}(\frac{\gamma}{h_i}) &\equiv \lim_{w\uparrow(\frac{\gamma}{h_i})} \bar{n}_{h_{i-1},h_i}(w) < \lim_{w\downarrow(\frac{\gamma}{h_i})} \bar{n}_{h_{i-1},h_i}(w) \equiv \bar{n}^r_{h_{i-1},h_i}(\frac{\gamma}{h_i})
\end{aligned}$$

It follows that

$$\int_{\bar{n}^r_{h_{i-1},h_i}(\gamma/h_i)}^{\bar{n}^r_{h_i,h_{i+1}}(\gamma/h_i)} f_{n,w}(n', \frac{\gamma}{h_i})dn' < \frac{1}{h_i} \int_{\bar{n}^l_{h_{i-1},h_i}(\gamma/h_i)}^{\bar{n}^l_{h_i,h_{i+1}}(\gamma/h_i)} f_{n,w}(n', \frac{\gamma}{h_i})dn'$$

for all i, and therefore $\lim_{z\uparrow\gamma} f_Z(z) > \lim_{z\downarrow\gamma} f_Z(z)$.

[24] $\bar{n}_{h_d,h_{d+1}}$ is defined to be ∞.

Table 1.1: Total Counts of Individuals and Sample Units and Those Covered by Public Insurance During Panel

(a) Total Individual and Sample Unit Counts

	No. of Individuals		No. of SU's	
	2001	2004	2001	2004
Total	104053	131549	35106	43540
Children Living with Parent(s)	29549	37333	14323	17780

(b) Individual and Sample Unit Counts: on Public Insurance During Panel

	No. of Individuals on Public Insurance		No. of SU's on Public Insurance	
	2001	2004	2001	2004
Total	23152	33986	10001	14454
Children Living with Parent(s)	11109	17458	5272	8235

Notes: Panel (a) shows the total number of children, individuals and sample units in the 2001 and 2004 SIPP panels. The bottom row of the right column shows number of sample units that have children living with parents. Panel (b) shows the total number of children, individuals and sample units ever on public insurance. The top row of the right column shows the number of sample units that had at least a member on public insurance during the panel, and the bottom row shows the number of sample units that had at least a child on public insurance during the panel.

Table 1.2: Public Insurance Spell, Child and Sample Unit Counts by Spell Types, Continuous Eligibility States and Analysis Sample

(a) Public Insurance Spell, Child and Sample Unit Counts by Spell Types and Continuous Eligibility States

	Total No. of Public Insurance Spells		No. of Kids with Public Insurance Spells		No. of SU with Kids on Public Insurance	
	2001	2004	2001	2004	2001	2004
Pub Insurance spells	16109	23109	10656	17190	5024	8149
Left-Truncated Spells	8402	13996	7407	11929	3683	5848
Fresh Spells	7707	9113	5759	7850	3033	4390
Fresh Spells Ex. Infants & State Movers & SSI Kids	7158	8321	5360	7188	2866	4099
12-Month Continuous Elig.	3096	3312	2310	2843	1257	1642
No Continuous Elig.	2270	2535	1742	2217	934	1255
Other States	1792	2474	1319	2147	685	1223

(b) Public Insurance Spell, Child and Sample Unit Counts in Analysis Sample

	Subsample Public Insurance Spells		No. of Kids with Public Insurance Spells		No. of SU with Kids on Public Insurance	
	2001	2004	2001	2004	2001	2004
Fresh Spells: States with 12 Months of Continuous Elig.	3096	3312	2310	2843	1257	1642
Long Gap Sample	501	815	501	804	302	494
Long Gap and No Presumptive Elig.	232	389	232	382	134	232

Notes: This table shows the number of observations excluded in each subsample. The full analysis sample consists of 3096 spells from 2310 children in 1257 sample units in the 2001 panel and 3312 spells from 2843 children in 1642 sample units in the 2004 panel. For the 2001 panel, the analysis sample is reached by excluding 8402 left-truncated Spells, 549 spells from infants and children in families that moved during the panel, and 4062 spells from children not in the states that provide continuous eligibility. For the 2004 panel, the analysis sample is reached by excluding 13996 left-truncated Spells, 792 spells from infants and children in families that moved during the panel, and 5009 spells from children not in the states that provide continuous eligibility.

Table 1.3: Variable Means for Children in Analysis Samples

	2001 Panel				2004 Panel			
	Full Sample		Long Gap Sample		Full Sample		Long Gap Sample	
	Month 0	Month 1	Month 0	Month 1	Month 0	Month 1	Month 0	Month 1
Age	8.17	8.26	9.15	9.14	8.76	8.85	9.27	9.36
Female	0.5	0.5	0.5	0.5	0.48	0.48	0.5	0.5
Black	0.25	0.25	0.2	0.2	0.17	0.17	0.12	0.11
Family Size	4.1	4.1	4.12	4.13	4	4	4.2	4.2
Single Parent Family	0.49	0.49	0.38	0.38	0.41	0.41	0.27	0.27
Family Income (nominal)	2139	2048	2914	2847	3585	3615	4860	4828
Family Income (in 2010 $)	2582	2468	3481	3396	3990	4014	5312	5268
Fraction without Earnings	0.11	0.12	0.05	0.07	0.08	0.07	0.04	0.05
On Welfare	0.05	0.07	0.01	0.02	0.03	0.04	0.01	0.01
On Medicaid	0	1	0	1	0	1	0	1
On Private Insurance	0.35	0.26	0.42	0.31	0.5	0.46	0.63	0.57
Mom on Medicaid	0.19	0.36	0.07	0.24	0.18	0.27	0.01	0.15
Dad on Medicaid	0.08	0.16	0.03	0.1	0.07	0.12	0.04	0.07
Mom on Private Insurance	0.36	0.3	0.47	0.41	0.54	0.52	0.69	0.68
Dad on Private Insurance	0.43	0.38	0.52	0.47	0.66	0.65	0.77	0.77
Mom on UI	0.02	0.02	0.03	0.04	0.01	0.02	0.02	0.02
Dad on UI	0.05	0.05	0.07	0.08	0.03	0.02	0.02	0.03

Notes: Variable means for children and their families in the various analysis samples right before (Month 0) and during the first month (Month 1) of public insurance spell.

Table 1.4: Empirical vs. Predicted Income Dip and Rebound Measures for Various Labor Supply Elasticities: SIPP 2001 Panel

	Empirical Responses		Empirical Responses		Predicted Responses			Predicted Responses		
	Full Sample		Long Gap Sample		No Inc. Eff. ($\rho = 0$)			W. Inc. Eff. ($\rho = 0.74$)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Month in Spell	Point Estimates	Upper 95% CI	Point Estimates	Upper 95% CI	e=0.15	e=0.05	e=0	e=0.15	e=0.05	e=0
-11	-124	217	49	763	661	494	415	480	428	415
-10	46	280	148	815	661	494	415	480	428	415
-9	-99	222	36	735	661	494	415	480	428	415
-8	-14	294	97	751	661	494	415	480	428	415
-7	69	318	170	838	661	494	415	480	428	415
-6	10	247	17	661	661	494	415	480	428	415
-5	64	302	33	677	661	494	415	501	435	415
-4	34	239	29	565	661	494	415	501	435	415
-3	-54	64	-229	305	661	494	415	501	435	415
-2	-57	31	-200	216	661	494	415	501	435	415
-1	-23	62	-132	285	661	494	415	501	435	415
0	0	0	0	0	0	0	0	0	0	0
1	-77	39	-230	214	661	494	415	480	428	415
2	-50	77	-191	294	661	494	415	480	428	415
3	-28	113	-143	351	661	494	415	480	428	415
4	-16	143	-259	269	661	494	415	480	428	415
5	-68	141	-512	183	661	494	415	480	428	415
6	-40	186	-515	221	661	494	415	480	428	415
7	-77	243	-557	220	661	494	415	501	435	415
8	-50	292	-624	184	661	494	415	501	435	415
9	-28	393	-731	192	661	494	415	501	435	415
10	-16	337	-736	202	661	494	415	501	435	415
11	-69	381	-582	398	661	494	415	501	435	415

Notes: Empirical estimates are based on specification (1.11) where the point estimates trace out average family income in 2010 dollars around the start of a child's public insurance spell. Predicted responses are obtained from models described in section (1.3) with parameters calibrated using 2001 SIPP panel income data, federal income tax rates, published Medicaid/CHIP eligibility cutoffs and CHIP spending data.

Table 1.5: Empirical vs. Predicted Income Dip and Rebound Measures for Various Labor Supply Elasticities: SIPP 2004 Panel

	Empirical Responses		Empirical Responses		Predicted Responses			Predicted Responses		
	Full Sample		Long Gap Sample		No Inc. Eff. (ρ=0)			W. Inc. Eff. (ρ=0.74)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Month in Spell	Point Estimates	Upper 95% CI	Point Estimates	Upper 95% CI	e=0.15	e=0.05	e=0	e=0.15	e=0.05	e=0
-11	407	944	517	1355	687	524	388	508	412	388
-10	462	995	536	1391	687	524	388	508	412	388
-9	302	786	348	1158	687	524	388	508	412	388
-8	388	914	361	1172	687	524	388	508	412	388
-7	101	403	243	908	687	524	388	508	412	388
-6	119	423	267	932	687	524	388	508	412	388
-5	128	458	259	939	687	524	388	531	420	388
-4	75	382	147	852	687	524	388	531	420	388
-3	-28	130	27	415	687	524	388	531	420	388
-2	-27	122	51	427	687	524	388	531	420	388
-1	-77	41	1	160	687	524	388	531	420	388
0	0	0	0	0	0	0	0	0	0	0
1	23	242	-245	35	687	524	388	508	412	388
2	-3	261	-356	-40	687	524	388	508	412	388
3	-40	197	-510	-134	687	524	388	508	412	388
4	-43	201	-607	-196	687	524	388	508	412	388
5	-250	-23	-686	-140	687	524	388	508	412	388
6	-342	-112	-753	-182	687	524	388	508	412	388
7	-363	-127	-884	-284	687	524	388	531	420	388
8	-346	-96	-908	-262	687	524	388	531	420	388
9	-353	5	-1035	44	687	524	388	531	420	388
10	-374	44	-684	849	687	524	388	531	420	388
11	-452	-43	-1350	-166	687	524	388	531	420	388

Notes: Empirical estimates are based on specification (1.11) where the point estimates trace out average family income in 2010 dollars around the start of a child's public insurance spell. Predicted responses are obtained from models described in section (1.3) with parameters calibrated using 2004 SIPP panel income data, federal income tax rates, published Medicaid/CHIP eligibility cutoffs and CHIP spending data.

Table 1.6: Statistical Tests of Theoretical Predictions

	p-values from SIPP 2001 Panel		p-values from SIPP 2004 Panel	
	Full Sample	Long Gap Sample	Full Sample	Long Gap Sample
(1) H_0: $\bar{\delta}_{11} = \delta^{e=0}$ vs. H_1: $\bar{\delta}_{11} < \delta^{e=0}$	0.000***	0.003***	0.000***	0.000***
(2) H_0: $\bar{\delta}_{9} = \delta^{e=0}$ vs. H_1: $\bar{\delta}_{9} < \delta^{e=0}$	0.000***	0.003***	0.000***	0.000***
(3) H_0: $\bar{\delta}_{8} = \frac{1}{4}\delta^{e=0}$ vs. H_1: $\bar{\delta}_{8} < \frac{1}{4}\delta^{e=0}$	0.059*	0.047**	0.004***	0.000***

Notes: Presented are p-values from statistical tests of theoretical predictions based on the neo-classical model in subsection (1.3.2) with $e = 0$. Test (1) is the baseline, (2) incorporates the timing of the receipt of the renewal package and (3) incorporates seam bias as discussed in section (1.6).

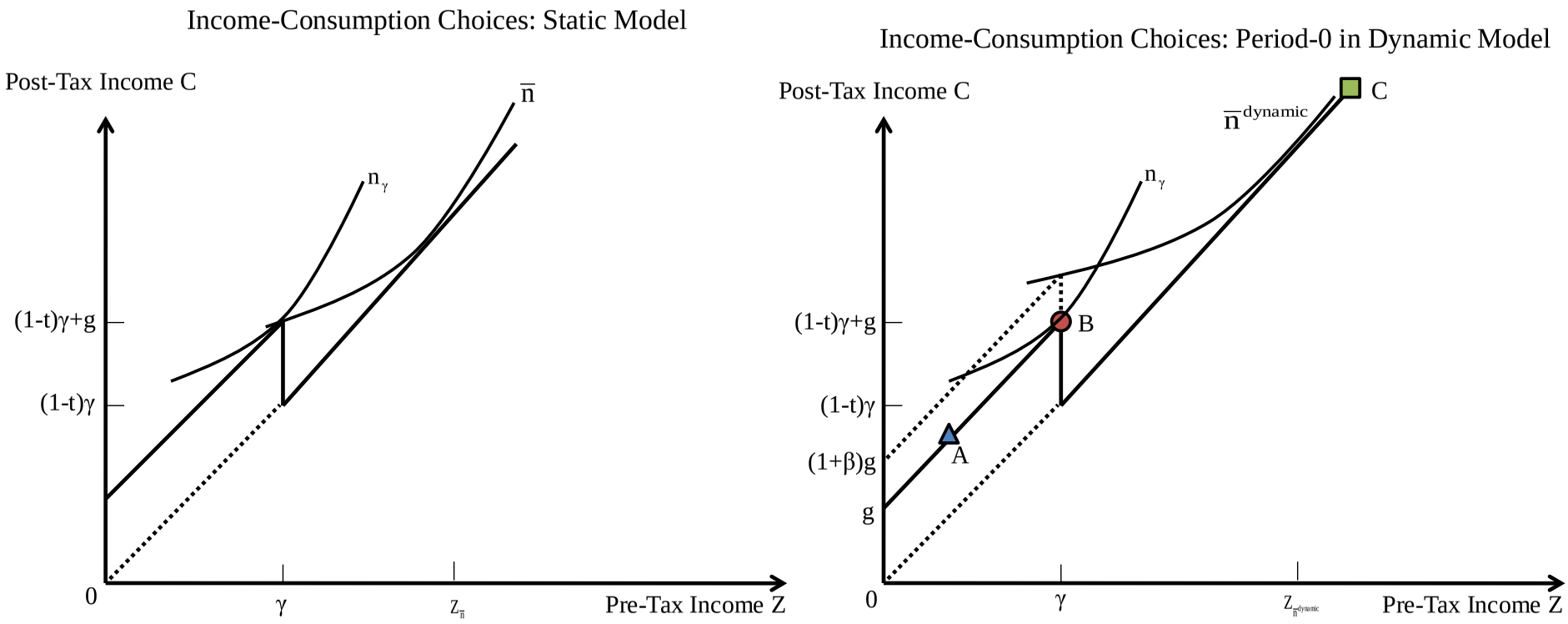
* p<0.1; ** p<0.05; *** p<0.001.

Table 1.7: Optimal Length of the Continuous Eligibility Period from Simulations

Cost of Recertification		Optimal Cont. Elig. Period Length (τ)	
ϕ (in 2010 dollars)	κ (in 2010 dollars)	SIPP 2001 Panel	SIPP 2004 Panel
\$0	\$0	1	1
\$0	\$9.5	4	4
\$0	\$19	4	4
\$0	\$28.5	4	4
\$0	\$38	6	8
\$9.5	\$0	6	4
\$9.5	\$9.5	8	8
\$9.5	\$19	9	8
\$9.5	\$28.5	12	8
\$9.5	\$38	12	12
\$19	\$0	12	12
\$19	\$9.5	12	12
\$19	\$19	12	12
\$19	\$28.5	12	12
\$19	\$38	12	12
\$28.5	\$0	12	12
\$28.5	\$9.5	12	12
\$28.5	\$19	18	12
\$28.5	\$28.5	18	16
\$28.5	\$38	18	16
\$38	\$0	18	16
\$38	\$9.5	18	16
\$38	\$19	18	16
\$38	\$28.5	18	24
\$38	\$38	18	24

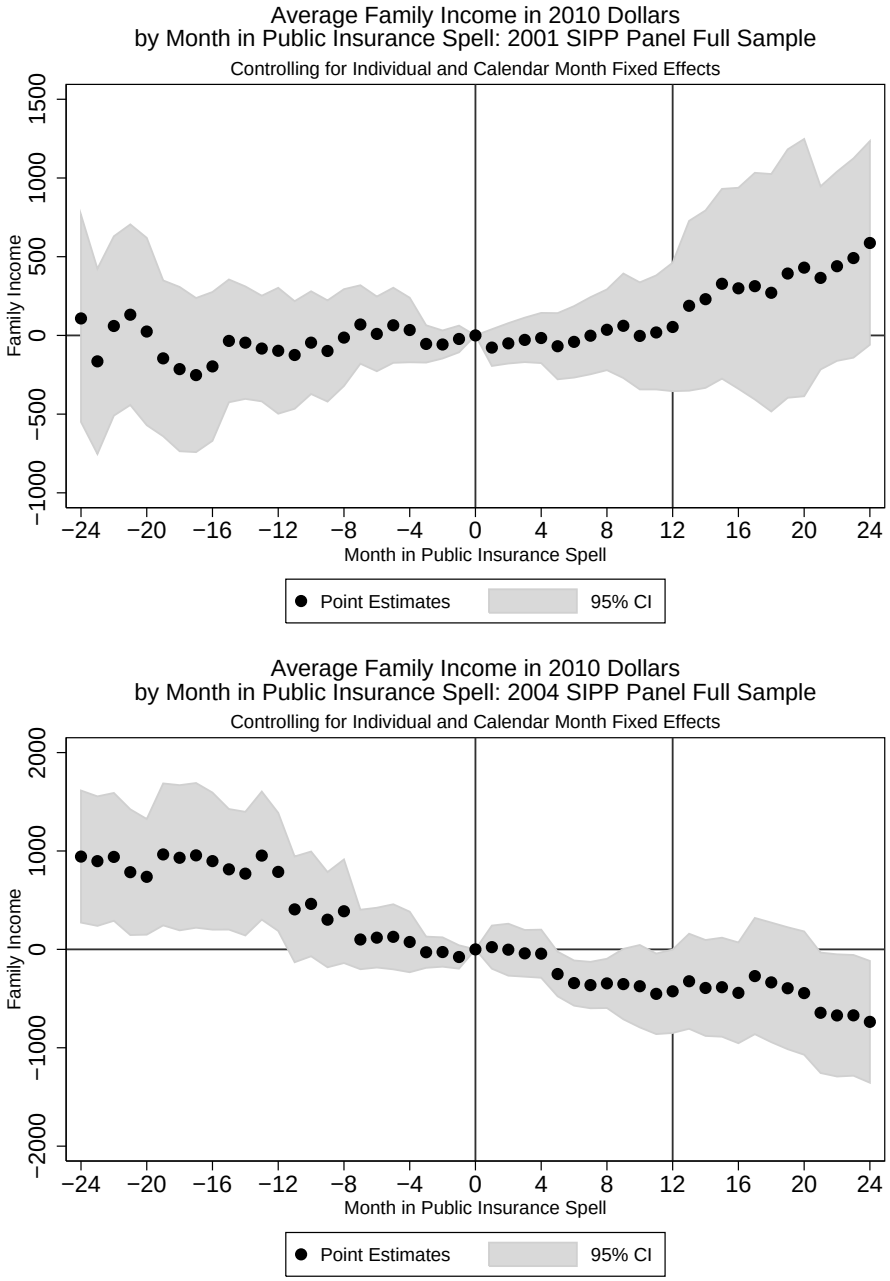
Notes: The optimal lengths of the continuous eligibility period, τ, are calculated based on the model in section (1.7). The choice set of τ is $\{1, 2, ..., 36\}$, and details regarding the model parameters and specification are presented in section (1.7). The simulation sample consists of families who responded to SIPP interviews for every month during the 2001 and 2004 panels.

Figure 1.1: Income-Consumption Choices in the Static and Dynamic Models



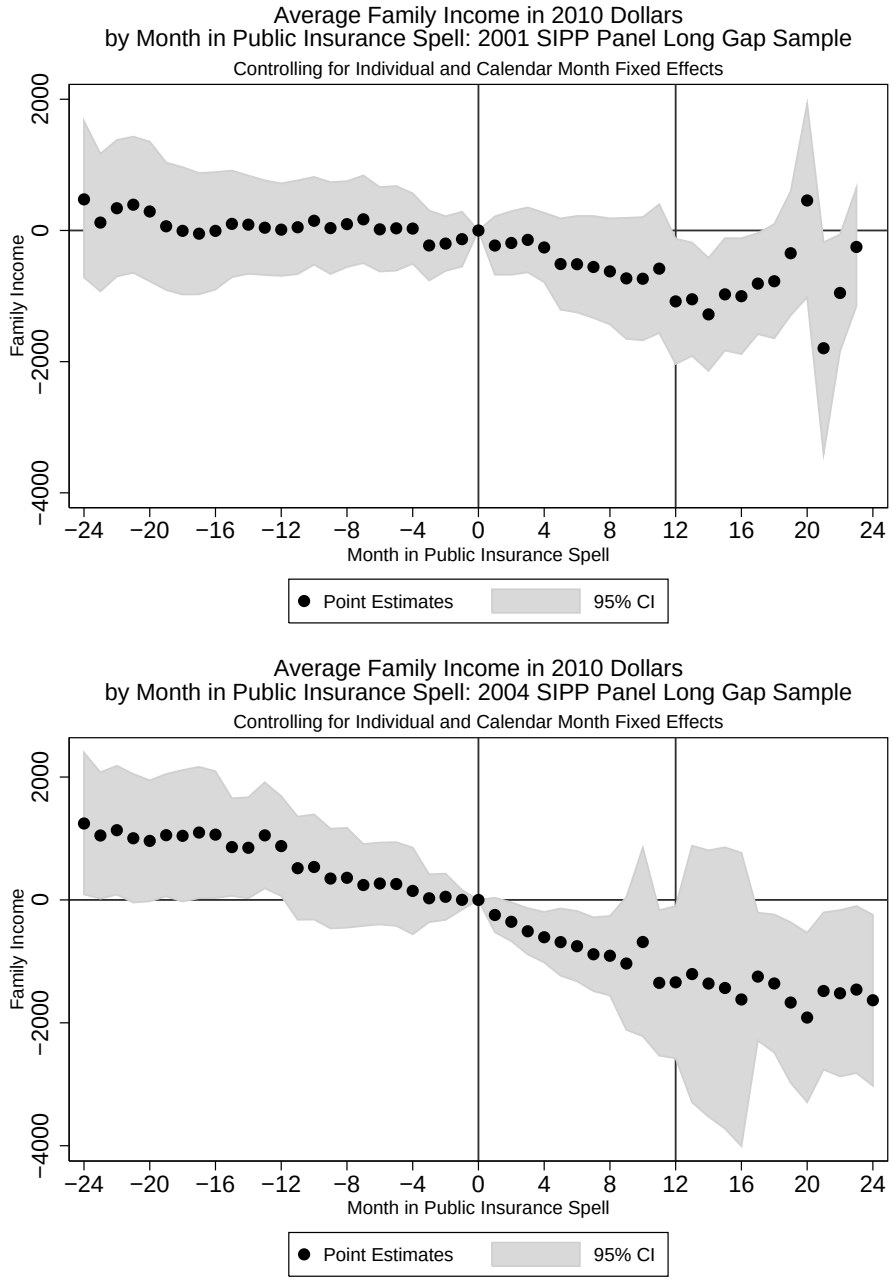
Notes: The left and right panels illustrate agents' income-consumption choices in the static (subsection 1.3.1) and dynamic model (subsection 1.3.2) respectively. The static model assumes that income eligibility is checked every period whereas it is checked every two periods in the dynamic model. In the case of the static model which is presented in the left panel, agents whose type is in the interval $[n_\gamma, \bar{n}]$ choose pre-tax income $Z = \gamma$. For the 2-period dynamic model which is presented in the right panel, agents whose type is in the interval $[n_\gamma, \bar{n}^{dynamic}]$ choose pre-tax income γ in period 0. The fact that eligibility is only checked once every two periods in effect increases the size of the benefit notch from g to $(1+\beta)g$.

Figure 1.2: Average Family Income by Month in Medicaid Spell: Full Sample



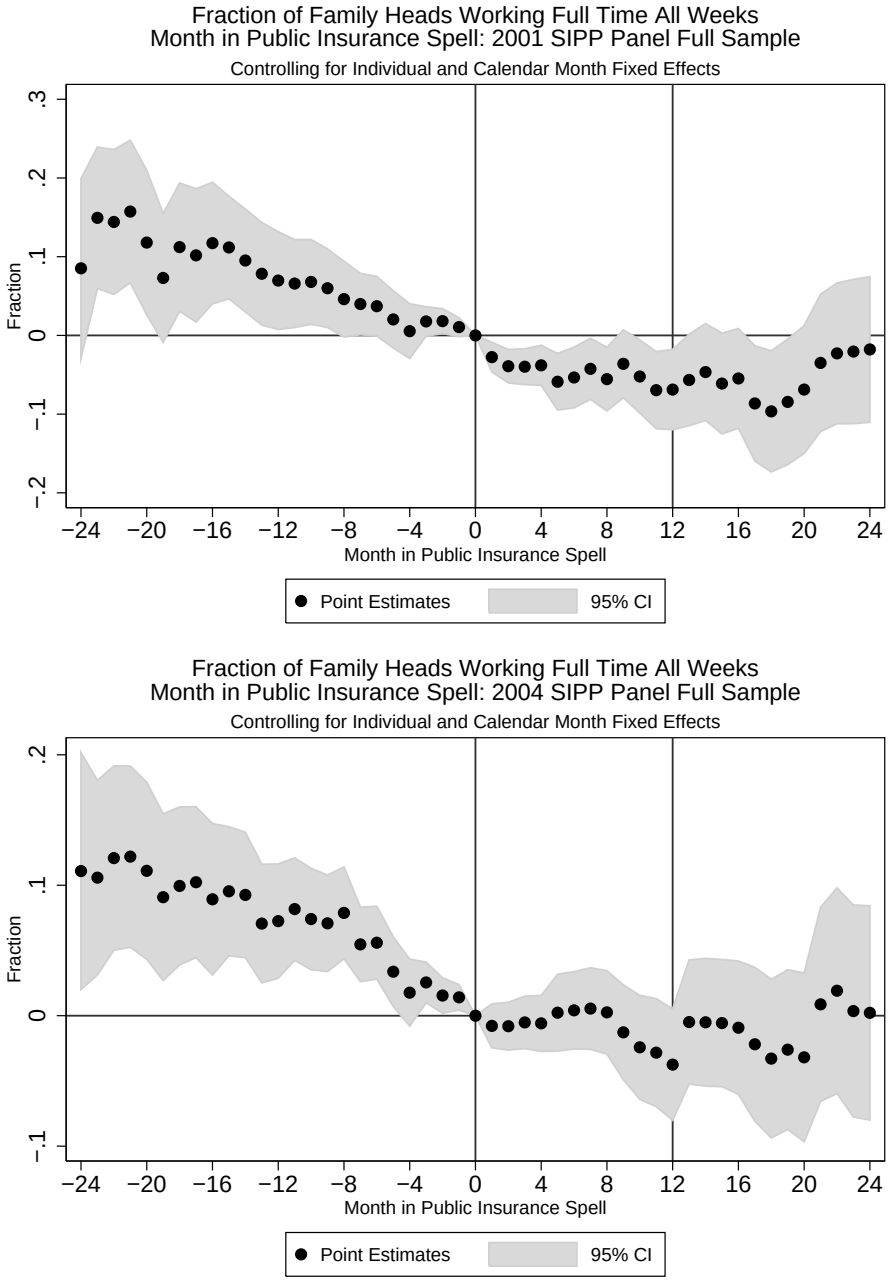
Notes: Plotted are coefficients from regressions of family monthly income in 2010 dollars on a set of indicator variables for months before or since the start of a fresh Medicaid/CHIP spell. The top panel plots the coefficients from the SIPP 2001 panel and the bottom plots those from the 2004 panel. The regression includes individual and calendar month fixed effects. The standard error is clustered at the SIPP sampling unit level. Month 0 (the month right before the start of a fresh Medicaid/CHIP spell) is the omitted category, and the corresponding coefficient is 0 by construction. The sample in the 2001 panel includes 3096 fresh spells for 2310 children from 1257 sample units, and the 2004 panel sample includes 3312 fresh spells for 2843 children from 1642 sample units. The sample is truncated by discarding the observations with Medicaid coverage before the start of a spell and those without Medicaid coverage after.

Figure 1.3: Average Family Income by Month in Medicaid Spell: Long Gap Sample



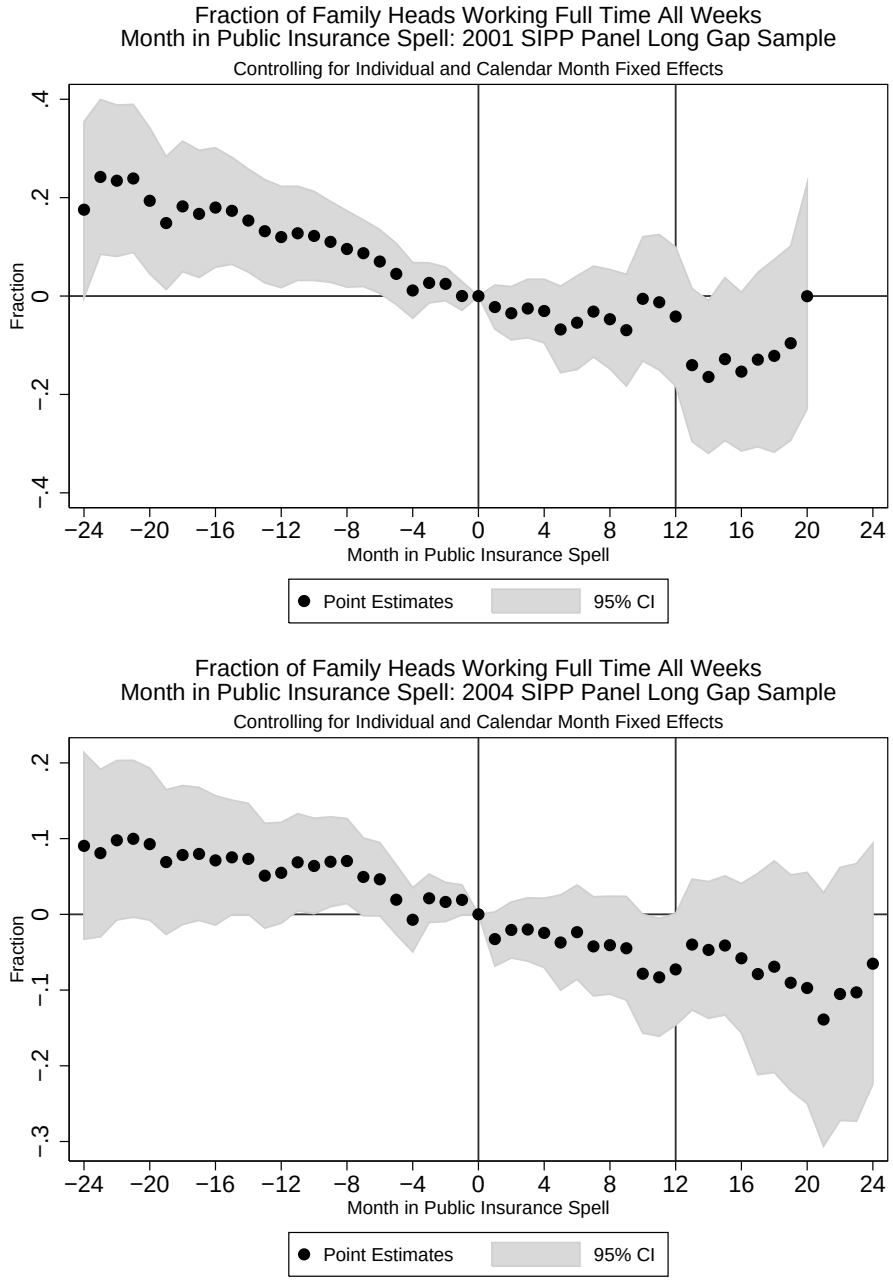
Notes: Plotted are coefficients from regressions of family monthly income in 2010 dollars on a set of indicator variables for months before or since the start of a fresh Medicaid/CHIP spell. The top panel plots the coefficients from the SIPP 2001 panel and the bottom plots those from the 2004 panel. The regression includes individual and calendar month fixed effects. The standard error is clustered at the SIPP sampling unit level. Month 0 (the month right before the start of a fresh Medicaid/CHIP spell) is the omitted category, and the corresponding coefficient is 0 by construction. The sample only includes children who were not covered by public insurance for 12 months before the start of a spell. The 2001 panel sample includes 501 fresh spells for 501 children from 302 sample units, and the 2004 panel sample includes 815 fresh spells for 804 children from 494 sample units. The sample is truncated by discarding the observations with Medicaid coverage before the start of a spell and those without Medicaid coverage after.

Figure 1.4: Fraction of Family Head Working Full Time in Medicaid Spell: Full Sample



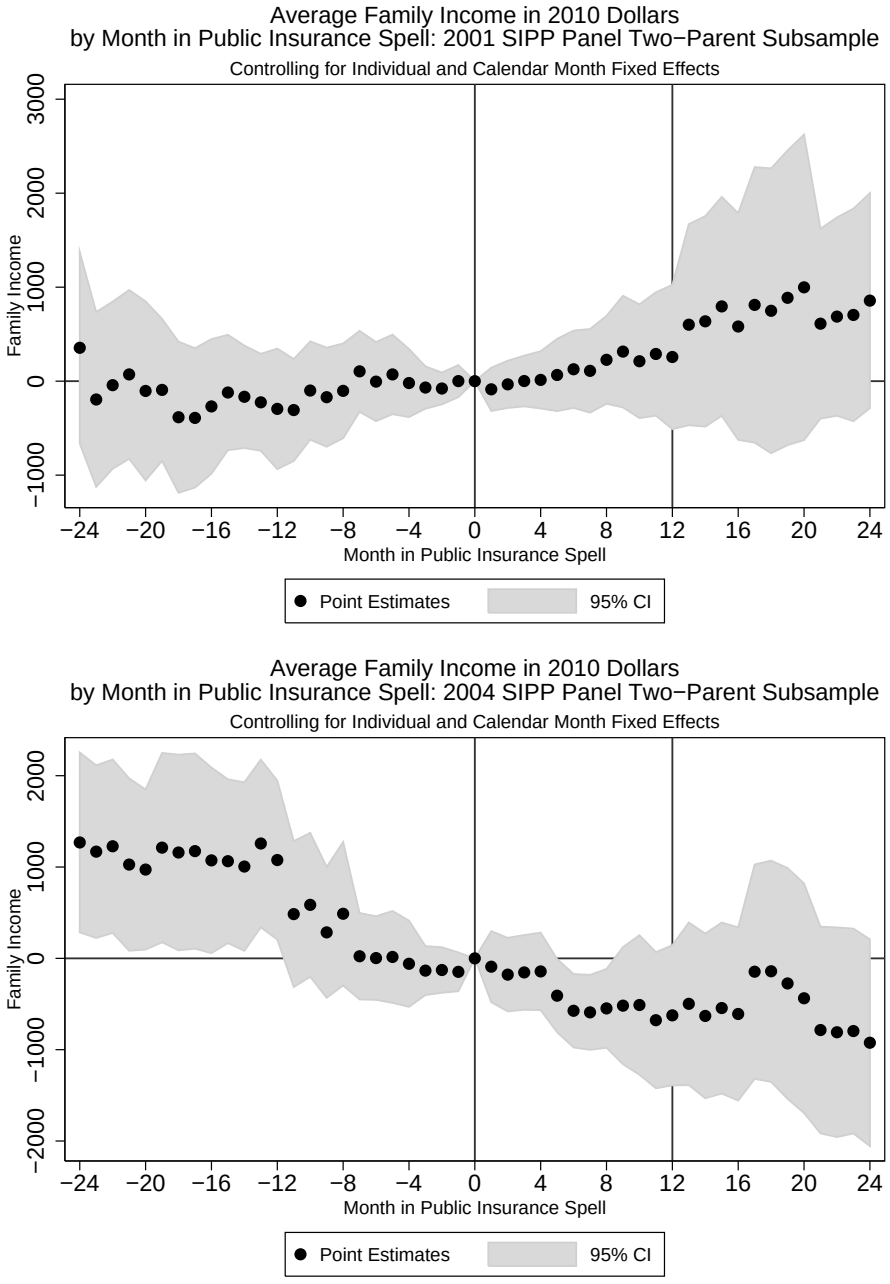
Notes: Plotted are coefficients from regressions of the indicator of whether the family head worked full time on a set of indicator variables for months before or since the start of a fresh Medicaid/CHIP spell. The top panel plots the coefficients from the SIPP 2001 panel and the bottom plots those from the 2004 panel. The regression includes individual and calendar month fixed effects. The standard error is clustered at the SIPP sampling unit level. Month 0 (the month right before the start of a fresh Medicaid/CHIP spell) is the omitted category, and the corresponding coefficient is 0 by construction. The sample in the 2001 panel includes 3096 fresh spells for 2310 children from 1257 sample units, and the 2004 panel sample includes 3312 fresh spells for 2843 children from 1642 sample units. The sample is truncated by discarding the observations with Medicaid coverage before the start of a spell and those without Medicaid coverage after.

Figure 1.5: Fraction of Family Head Working Full Time in Medicaid Spell: Long Gap Sample



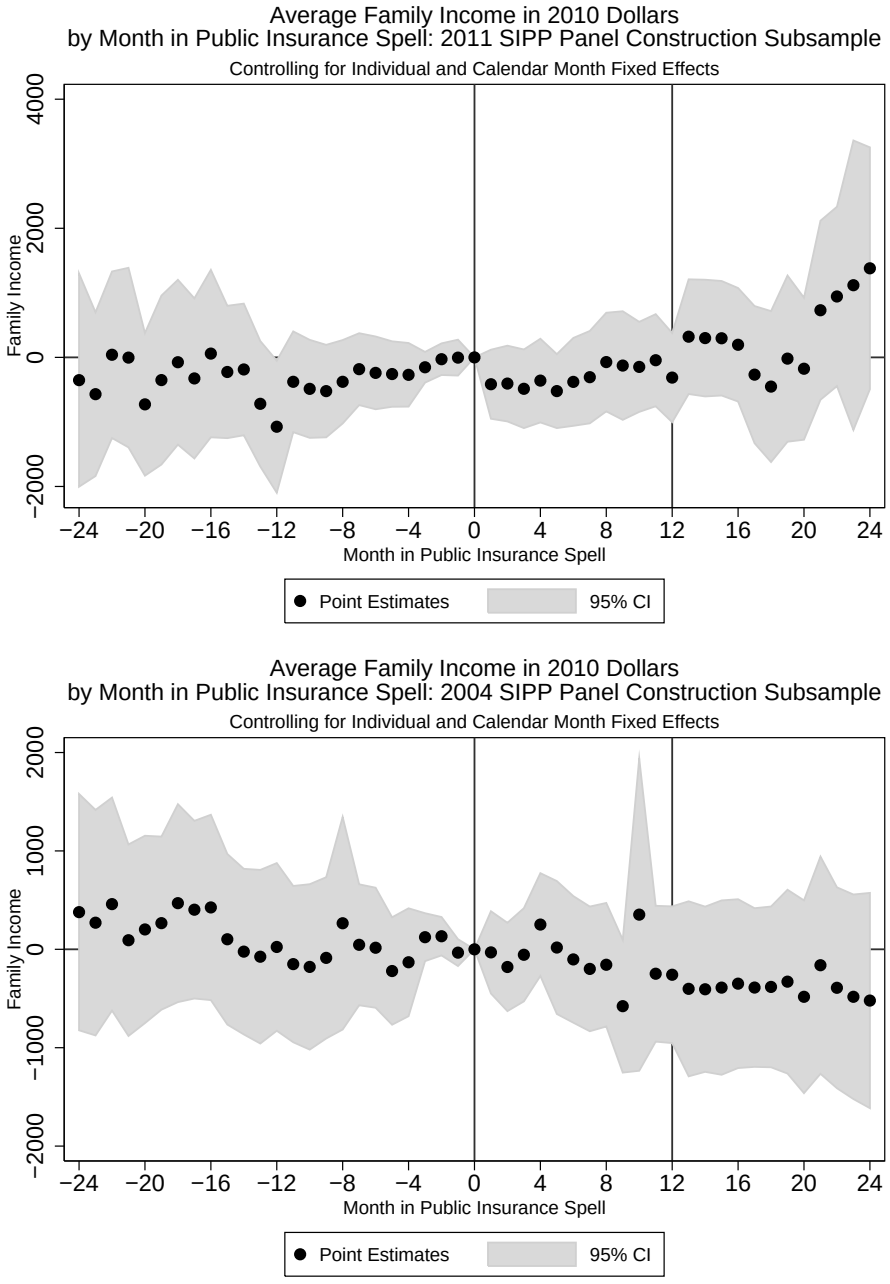
Notes: Plotted are coefficients from regressions of the indicator of whether the family head worked full time on a set of indicator variables for months before or since the start of a fresh Medicaid/CHIP spell. The top panel plots the coefficients from the SIPP 2001 panel and the bottom plots those from the 2004 panel. The regression includes individual and calendar month fixed effects. The standard error is clustered at the SIPP sampling unit level. Month 0 (the month right before the start of a fresh Medicaid/CHIP spell) is the omitted category, and the corresponding coefficient is 0 by construction. For the long gap sample, the 2001 panel includes 501 fresh spells for 501 children from 302 sample units, and the 2004 panel includes 815 fresh spells for 804 children from 494 sample units. The sample is truncated by discarding the observations with Medicaid coverage before the start of a spell and those without Medicaid coverage after.

Figure 1.6: Average Family Income by Month in Medicaid Spell: Subsample with Two-Parent Families



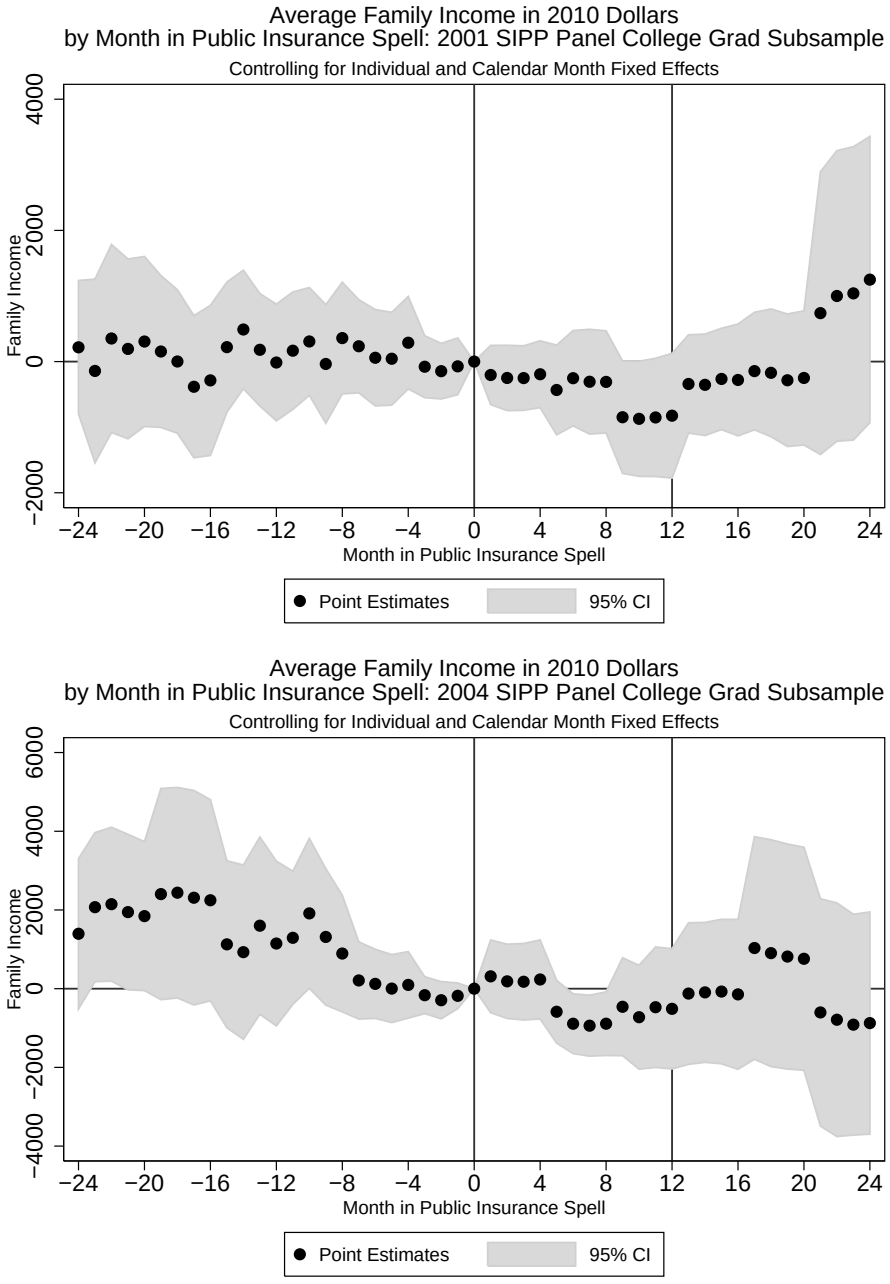
Notes: Plotted are coefficients from regressions of family monthly income in 2010 dollars on a set of indicator variables for months before or since the start of a fresh Medicaid/CHIP spell for children living with both parents in the sample period. The top panel plots the coefficients from the SIPP 2001 panel and the bottom plots those from the 2004 panel. The regression includes individual and calendar month fixed effects. The standard error is clustered at the SIPP sampling unit level. Month 0 (the month right before the start of a fresh Medicaid/CHIP spell) is the omitted category, and the corresponding coefficient is 0 by construction. The 2001 panel sample includes 1472 fresh spells for 1106 children from 578 sample units, and the 2004 panel sample includes 1860 fresh spells for 1630 children from 903 sample units. The sample is truncated by discarding the observations with Medicaid coverage before the start of a spell and those without Medicaid coverage after.

Figure 1.7: Average Family Income by Month in Medicaid Spell: Subsample with Construction-Worker Parents



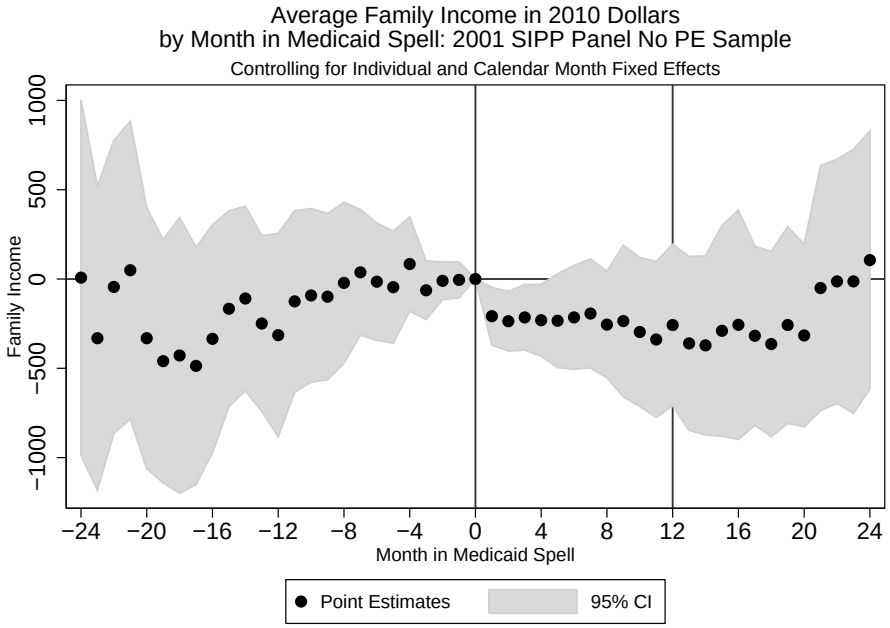
Notes: Plotted are coefficients from regressions of family monthly income in 2010 dollars on a set of indicator variables for months before or since the start of a fresh Medicaid/CHIP spell for children whose parent worked in the construction industry during the sample period. The top panel plots the coefficients from the SIPP 2001 panel and the bottom plots those from the 2004 panel. The regression includes individual and calendar month fixed effects. The standard error is clustered at the SIPP sampling unit level. Month 0 (the month right before the start of a fresh Medicaid/CHIP spell) is the omitted category, and the corresponding coefficient is 0 by construction. The 2001 panel sample includes 321 fresh spells for 254 children from 120 sample units, and the 2004 panel sample includes 395 fresh spells for 359 children from 196 sample units. The sample is truncated by discarding the observations with Medicaid coverage before the start of a spell and those without Medicaid coverage after.

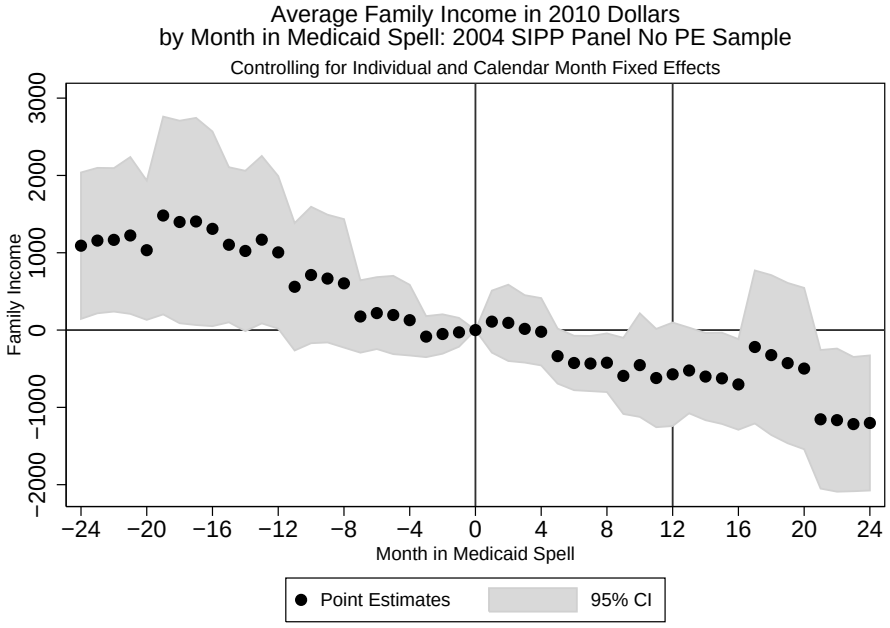
Figure 1.8: Average Family Income by Month in Medicaid Spell: Subsample with College-Grad Mother



Notes: Plotted are coefficients from regressions of family monthly income in 2010 dollars on a set of indicator variables for months before or since the start of a fresh Medicaid/CHIP spell for children whose mother was a college graduate. The top panel plots the coefficients from the SIPP 2001 panel and the bottom plots those from the 2004 panel. The regression includes individual and calendar month fixed effects. The standard error is clustered at the SIPP sampling unit level. Month 0 (the month right before the start of a fresh Medicaid/CHIP spell) is the omitted category, and the corresponding coefficient is 0 by construction. The 2001 panel sample includes 472 fresh spells for 369 children from 237 sample units, and the 2004 panel sample includes 671 fresh spells for 614 children from 386 sample units. The sample is truncated by discarding the observations with Medicaid coverage before the start of a spell and those without Medicaid coverage after.

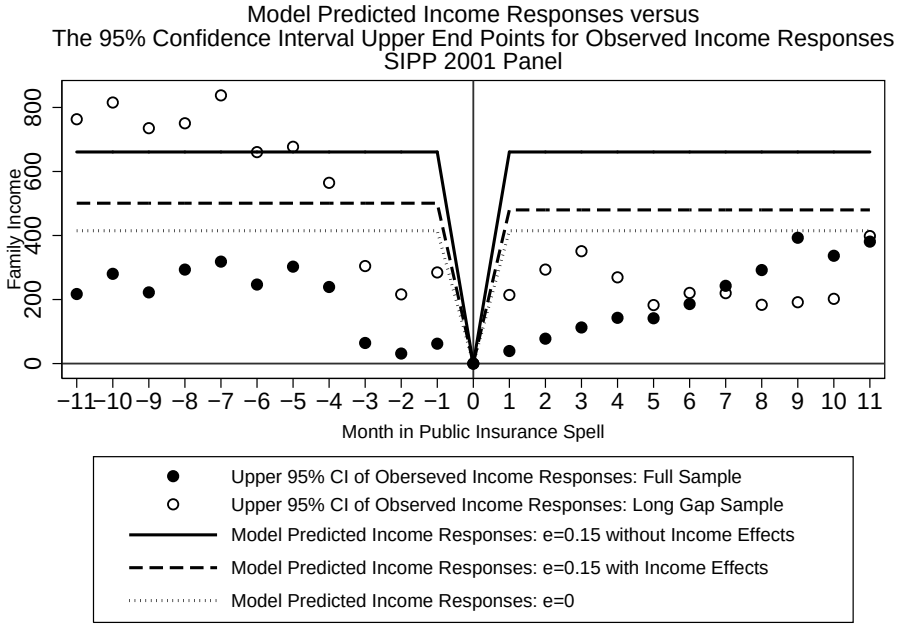
Figure 1.9: Average Family Income by Month in Medicaid Spell: Subsample with No-Presumptive-Eligibility States

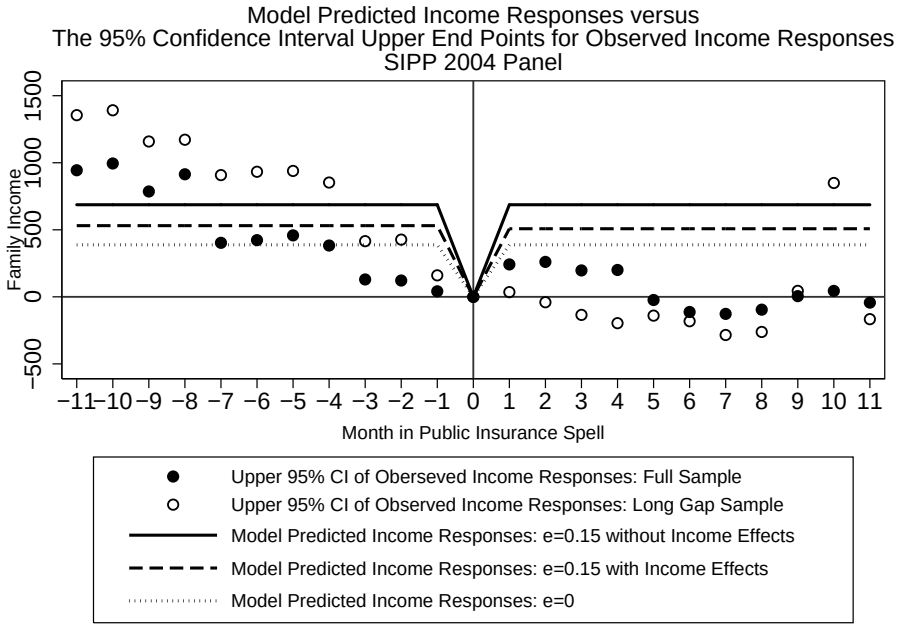




Notes: Plotted are coefficients from regressions of family monthly income in 2010 dollars on a set of indicator variables for months before or since the start of a fresh Medicaid/CHIP spell for children living in states that do not offer presumptive eligibility to non-infants. The top panel plots the coefficients from the SIPP 2001 panel and the bottom plots those from the 2004 panel. The regression includes individual and calendar month fixed effects. The standard error is clustered at the SIPP sampling unit level. Month 0 (the month right before the start of a fresh Medicaid/CHIP spell) is the omitted category, and the corresponding coefficient is 0 by construction. The 2001 panel sample includes 1294 fresh spells for 993 children from 536 sample units, and the 2004 panel sample includes 1578 fresh spells for 1365 children from 787 sample units. The sample is truncated by discarding the observations with Medicaid coverage before the start of a spell and those without Medicaid coverage after.

Figure 1.10: 95% Confidence Interval Upper End Points for Observed Income Responses versus Model Predicted Income Responses





Notes: The top and bottom panels are graphical representations of Table 1.4 and Table 1.5 respectively. The solid dot, hollow dot, solid line and dashed line respectively represent columns (2), (4), (5) and (8), and the dotted line represents both column (7) and (10) in each of the two tables.

Chapter 2

Regression Discontinuity Design with Measurement Error in the Running Variable

2.1 Introduction

Over the past decade, many studies in economics have relied on the Regression Discontinuity (RD) Design to evaluate the effects of a wide range of programs. In a classic sharp RD design, agents are eligible to participate in a program if and only if the value of her running variable (sometimes called the “assignment variable" or “forcing variable") exceeds or falls below a known threshold. However, researchers have encountered departures from the classical RD designs. One of these departures entails the situation in which the running variable is not available but a noisy measure of it is observed. The occurrence of such situation is common when survey data are used, in which the value of the running variable is self-reported as opposed to being extracted from an administrative database.

A typical example is the application of RD that uses income as a running variable to study the effect of means-tested transfer programs where the eligibility depends on whether income falls below a certain threshold. However, most administrative data cannot be used for an RD because they only include the treatment population, namely those who enroll in the program, and contain little information on the various outcomes for applicants who are denied benefits. Therefore, practitioners may be forced to rely on survey data in order to apply an RD design. For instance,

Schanzenbach (2009) uses the Early Childhood Longitudinal Study to study the effect of school lunch on obesity and compares obesity rates for children below and above the reduced-price lunch cutoff for the National School Lunch Program. Hullegie and Klein (2010) study the effect of private insurance on health care utilization in the German Socio Economic Panel by using a policy rule that obliges workers with income below a threshold to participate in the public health insurance system. Koch (2010) uses the Medical Expenditure Panel Survey to study health insurance crowd-out by focusing on income cutoffs in the Children's Health Insurance Program (CHIP). de la Mata (2011) estimates the effect of Medicaid/CHIP coverage on children's health care utilization and health outcomes with the Panel Study of Income Dynamics (PSID) and its Child Development Study (CDS) supplement.

The studies above all use income data gathered from surveys as their running variable in their RD analyses, but measurement error in survey data has been widely documented and studied (see Bound et al. (2001) for a review). Yet, the presence of measurement error in the running variable directly threatens the source of identification in an RD design, which hinges on the discontinuous relationship between the treatment and running variable. Even if there is perfect compliance with the discontinuous rule, there may not be a discontinuity in the probability of treatment conditional on the observed noisy running variable. Instead of a step function, the first-stage relationship–the probability of treatment conditional on the noisy running variable–will likely be smoothly S-shaped. This S-shaped pattern is seen, for example, in Figure 2 of de la Mata (2011), which plots the fraction of children covered by Medicaid against reported family income from PSID. About 10% of the families with annual income $20,000 above the eligibility cutoff report Medicaid coverage–a likely indication of measurement error[1]–and it is not convincing that there is (or should be) a discontinuity at the Medicaid eligibility cutoff in this first-stage relationship. This lack of discontinuity casts serious doubt on the identification and estimation of the program effect.

The measurement error problem has not been addressed in the literature with the exception of

[1]There are several sources of measurement error. For example, the author uses annual income measures and the annual equivalent of monthly Medicaid cutoff, whereas Medicaid eligibility is assigned based on monthly income in practice. In addition, reported income is subject to recall error and certain deductions when determining eligibility, such as child care and work related expenses, may not be observed in the data (e.g. Card and Shore-Sheppard (2004)).

Hullegie and Klein (2010).[2] The authors adopt a parametric Berkson measurement error specification, in which the true running variable is the sum of the observed running variable and an independent normally distributed measurement error. This specification implies, however, that the distribution of the true running variable is smooth and precludes the testing of density discontinuity, which has been a central element in assessing the RD design validity as per McCrary (2008). Testing is particularly important when income is the running variable (as is the case for the studies listed above) because neo-classical models in labor and public economics predict income sorting at the discontinuity in the budget constraint. In this paper, I adopt the more conventional classical measurement error model (Bound et al. (2001)), which allows non-smoothness in the running variable density. I provide sufficient conditions for non-parametrically identifying the underlying true running variable distribution, using only the mis-measured running variable and program participation status. This identification result can certainly be used to test the validity of RD design, but it can also be applied to assess the degree of sorting when an RD design is invalid and to estimate labor supply elasticities in the presence of a benefit notch.[3] Following the identification of the true running variable distribution, I will show that the RD treatment effect is identified under the non-differential measurement assumption.

A simple procedure is proposed for estimating the true running variable distribution as well as the RD treatment effect parameter. The procedure fits into a minimum distance framework, and I will use standard techniques to show that the estimators are efficient, $\sqrt{N}$-consistent and asymptotically normal. Monte Carlo simulations verify that the true running variable distribution and the RD treatment effect parameter can indeed be recovered using the proposed method. The

[2]It should be pointed out that the problem of heaping in the running variable raised in recent literature (Almond et al. (2010), Almond et al. (2011), Barreca et al. (2011b), and Barreca et al. (2011a)) is not the focus of this paper. In the heaping setup, treatment assignment is based on the observed value of the running variable. The problem at hand is one where I do not observe the variable determining treatment.

[3]As stated in Saez (1999), neo-classical labor supply theory predicts that discontinuities in the budget constraint, such as those at the Medicaid and School Lunch Program eligibility cutoff, will generate bunching or discontinuity in the income distribution and that the degree of the non-smoothness will be informative of the labor supply elasticity. In fact, a recent study Kleven and Waseem (2011) has found bunching behavior in the presence of income tax notches in Pakistan using administrative data. However, the existence of income tax notches is rare at best for other countries, and income-tested transfer programs are the main sources that generate discontinuities in the budget constraint. As mentioned above, survey data are most likely needed for analyzing agents' responses to the incentives created by these programs, and measurement error will need to be dealt with.

procedure produces informative RD treatment effect estimates and is capable of detecting discontinuity in the true running variable distribution for typical sample sizes in relevant empirical studies.

The remainder of the paper is organized as follows. Section 2.2 introduces the statistical model and discusses identification of the true running variable distribution as well as that of the RD treatment effect parameter for both the sharp and fuzzy design. An estimation procedure is developed in section 2.3, and simulations are performed to evaluate the proposed procedure. Section 3.5 concludes and charts out future research directions.

2.2 Identification

2.2.1 Baseline Statistical Model

In a conventional sharp regression discontinuity design, the econometrician observes running variable X^*, eligibility/treatment D^* and outcome Y where

$$\begin{aligned} Y &= h(D^*, X^*, \epsilon) \\ D^* &= 1_{[X^*<0]} \end{aligned} \tag{2.1}$$

h is a function continuous in its second argument, and the eligibility cutoff is normalized to 0. Note that it may be common for $D^* = 1_{[X^* \geqslant 0]}$ to be the treatment determining mechanism in many applications, most of the motivating examples in the Introduction follow $D^* = 1_{[X^*<0]}$, i.e. program eligibility depends on whether the running variable (income) falls below a known cutoff. It is a standard result (e.g. Hahn et al. (2001)) that the treatment effect $\delta_{sharp} = E[y(1,0,\epsilon) - y(0,0,\epsilon)|X^* = 0]$ is identified as

$$\delta_{sharp} = \lim_{x^* \uparrow 0} E[Y|X^* = x^*] - \lim_{x^* \downarrow 0} E[Y|X^* = x^*]$$

when conditional expectation of the response function $E[h(D^*, X^*, \epsilon)|X^* = x^*]$ is continuous at $x^* = 0$ for $D^* = 0, 1$. In this paper, I consider the extension where X^* is not observed but a noisy measure of it is. Let X be the observed running variable, and $u \equiv X - X^*$ the measurement error. As mentioned in section 2.1, a key assumption that distinguishes this study from Hullegie and Klein (2010) is that the measurement error is independent of the true running variable as opposed to the observed running variable. Formally,

Assumption 1 (Independence). $X^* \perp\!\!\!\perp u$

The first step will be the identification of the distribution of X^*. As mentioned in section 2.1, not only is the identification of the X^* distribution crucial for the recovery of the RD treatment effect parameter δ_{sharp}, it is also central for testing the validity of the RD design and can even be applicable to a range of economic problems beyond the scope of RD.[4] However, it is not possible to identify the distribution of X^* from the observed distribution of X, and economists have traditionally proposed strategies using a repeated and possibly noisy measure of the true explanatory variable (e.g. Ashenfelter and Krueger (1994), Black et al. (2000), Hausman et al. (1991), Li and Vuong (1998), Schennach (2004)). Nevertheless, such a measure may not be available in the data–for example, it is not evident that there exists an alternative measure of monthly family income in the data used to evaluate the effect of public insurance or school lunch programs. What is helpful in the RD context is that the observed program eligibility $D^* = 1_{[X^*<0]}$ (the case of imperfect compliance will be discussed below), which is a deterministic function of X^*, is informative of the value of X^*. Therefore, it becomes an interesting question as to whether and under what additional assumptions is the distribution of X^* identifiable from the joint distribution of (X, D^*).

This question is addressed in subsection 2.2.2. In particular, I focus on the non-parametric identification of the running variable distribution and the RD treatment effect parameter under the assumption that X^* and u are discrete. A discrete running variable setup may appear odd given

[4] An example is the estimation of the labor supply elasticity using the Medicaid notch proposed by Saez (1999) that will be biased in the presence of measurement error.

the continuity assumption of $E[h(D^*, X^*, \epsilon)|X^* = x^*]$ in identifying the treatment effect in a sharp RD design, but it is necessary in many policy contexts where an RD design appears compelling (Lee and Card (2008)). Typical running variables that are intrinsically discrete or reported in coarse intervals include age/birthdate (Card and Shore-Sheppard (2004), Snyder and Evans (2006), Dobkin and Ferreira (2010), etc), student enrollment (Angrist and Lavy (1999), Asadullah (2005), Urquiola (2006), etc) and test score (e.g. Jacob and Lefgren (2004), Matsudaira (2008), etc). Even if the running variable is truly continuous (as is the case of income), the discretization of X^* can be thought of as a binned-up approximation (this is common practice in graphical presentations of most RD applications). Sufficient conditions for identification will be provided in subsection 2.2.2 for the case in which X^* and u have bounded support, and an example is constructed in subsection 2.2.3 to show that the model is in general not identified when the bounded support assumption of X^* and u is relaxed. Identification in the case of X^* and u having continuous distributions is being investigated, and preliminary results are presented in the Appendix.

The observability of program eligibility, or equivalently perfect compliance, is assumed in the sharp RD model (2.1). In most applied contexts (such as those in the studies cited above), however, this assumption is rarely satisfied. In all social programs, for example, the take-up rate of entitlement programs among eligible individuals and families is not close to 100%[5]. For these programs, only take-up D is observed in the data, which is no longer a deterministic function of true running variable X^*. As a consequence, additional assumptions on the measurement error distribution may be needed for the identification of the distribution of X^*, and I will explore them in subsection 2.2.3. Finally in subsection 2.2.4, I will discuss how the treatment effect from the regression discontinuity design can be identified by assuming non-differential measurement error.

[5]See Currie (2004) for a survey on benefit take-up in social programs.

2.2.2 Identification of the True running Variable Distribution under Perfect Compliance

Running Variable and Measurement Error Have Discrete and Bounded Support

In this section, I investigate the identifiability under Assumption 1 of the distributions of $X^*$and u from the joint distribution of X and D^* in model (2.1) where X^* and u are discrete and bounded. Formally, what I observe are

$$\begin{aligned} D^* &= 1_{[X^*<0]} \\ X &= X^* + u \end{aligned} \tag{2.2}$$

The identification follows two steps: 1) the identification of the support of X^* and u and 2) the identification of the probability mass at each point in the support of X^* and u. In additional to independence between X^* and u, the identification result relies on the assumption of positive mass around the threshold 0 in the X^* distribution and a technical rank conditional to be discussed in detail later.

Denote the support of any random variable Z by support_Z, and let $\min\{\text{support}_Z\} = L_Z$ and $\max\{\text{support}_Z\} = U_Z$ for a discrete and bounded Z. Without loss of generality, I consider the case where $\text{support}_{X^*}, \text{support}_u \subseteq \mathbb{Z}$, the set of integers. Formally, the discrete and bounded support assumption is written as

Assumption DB (Discrete and Bounded Support).
$\text{support}_{X^*} \subseteq \{L_{X^*}, L_{X^*}+1, ..., U_{X^*}-1, U_{X^*}\}$ and $\text{support}_u \subseteq \{L_u, L_u+1, ..., U_u-1, U_u\}$ where $L_{X^*}, U_{X^*} L_u, U_u \in \mathbb{Z}$.

Abstracting from sampling error, the joint distribution of (X,D^*) observed by the econometrician is fully characterized by the distribution of X conditional on D^* and the marginal distribution of D^*. The assumption of independence between X^* and u gives strong implications

relating their respective supports to the observed support of X, conditional on D^*. Specifically,

$$\begin{aligned} \min\{\text{support}_{X|D^*=d}\} &= \min\{\text{support}_{X^*|D^*=d}\} + \min\{\text{support}_u\} \\ \max\{\text{support}_{X|D^*=d}\} &= \max\{\text{support}_{X^*|D^*=d}\} + \max\{\text{support}_u\} \text{ for } d = 0, 1 \end{aligned} \tag{2.3}$$

which entail four restrictions (four equations in the equation array (2.3)) on six unknowns ($\min\{\text{support}_{X^*|D^*=0}\}$, $\max\{\text{support}_{X^*|D^*=0}\}$, $\min\{\text{support}_{X^*|D^*=1}\}$, $\max\{\text{support}_{X^*|D^*=1}\}$, $\min\{\text{support}_u\}$ and $\max\{\text{support}_u\}$). In order to identify the supports of X^* and u, I impose the additional assumption

Assumption 2 (Threshold Support). $-1, 0 \in \text{support}_{X^*}$

Assumption 2 states that there exist agents with X^* right at and below the eligibility threshold 0. This is not a strong assumption and has to be satisfied in all valid RD designs because the quasi-experimental variation of an RDD comes from the agents around the threshold. It is straightforward to show that the addition of this week assumption is sufficient for identifying the supports.

Lemma 1 (**Support Identification in a Sharp Design**) Under Assumptions DB, 1 and 2, L_{X^*}, U_{X^*}, L_u and U_u are identified.

Proof. The relationship $D^* = 1_{[X^*<0]}$ implies

1) $\min\{\text{support}_{X^*|D^*=1}\} = \min\{\text{support}_{X^*}\} = L_{X^*}$,

2) $\max\{\text{support}_{X^*|D^*=0}\} = \max\{\text{support}_{X^*}\} = U_{X^*}$

3) $\max\{\text{support}_{X^*|D^*=1}\} < 0$

4) $\min\{\text{support}_{X^*|D^*=0}\} \geqslant 0$.

By Assumption 2, $\Pr(X^* = -1|D^* = 1) = \frac{\Pr(X^*=-1)}{\Pr(D^*=1)} > 0$ and $\Pr(X^* = 0|D^* = 0) = \frac{\Pr(X^*=0)}{\Pr(D^*=0)} > 0$. Consequently, Assumption 2 translates into statements

about the support of $X^*|D^* = 0$ and $X^*|D^* = 1$:

$$\begin{aligned} \max\{\text{support}_{X^*|D^*=1}\} &= -1 \\ \min\{\text{support}_{X^*|D^*=0}\} &= 0 \end{aligned}$$

i.e. it pins down two of the six unknowns in (2.3). It follows that the remaining four unknowns in (2.3), L_{X^*}, U_{X^*}, L_u and $U_u$are now exactly identified:

$$\begin{aligned} L_u &= L_{X|D^*=0} \\ U_u &= U_{X|D^*=1} + 1 \\ L_{X^*} &= L_{X|D^*=1} - L_{X|D^*=0} \\ U_{X^*} &= U_{X|D^*=0} - U_{X|D^*=1} - 1 \end{aligned}$$

□

Intuitively, those in the program $D^* = 1$ but appear ineligible $X \geqslant 0$ have a positive measurement error $u > 0$, and analogously those with $D^* = 0$ but $X < 0$ have a negative measurement error $u < 0$. This is essentially the insight behind Lemma 1.

With the support of X^* identified, I next derive the identification of the probability mass of X^* at every point in its support. Denote the probability mass of X^* by p_i at each integer i, and denote that of u by m_i. Let the conditional probability masses of the observed running variable X be $q_i^1 \equiv \Pr(X = i|D^* = 1)$ for $i \in \{L_{X|D^*=1}, L_{X|D^*=1} + 1, ..., U_{X|D^*=1} - 1, U_{X|D^*=1}\}$, $q_j^0 \equiv \Pr(X = j|D^* = 0)$ for $j \in \{L_{X|D^*=0}, L_{X|D^*=0} + 1, ..., U_{X|D=0} - 1, U_{X|D=0}\}$, and the marginal probabilities $r^1 \equiv \Pr(D^* = 1)$ and $r^0 \equiv \Pr(D^* = 0)$.

Under the independence assumption of X^* and u, the distribution of $X|D^*$ is the convolution

of the distribution of $X^*|D^*$ and that of u. In particular,

$$\begin{aligned} q_i^1 &= \frac{\sum_{k<0} p_k m_{i-k}}{\sum_{k<0} p_k} \\ q_j^0 &= \frac{\sum_{k\geqslant 0} p_k m_{j-k}}{\sum_{k\geqslant 0} p_k} \end{aligned} \tag{2.4}$$

Additionally, the marginal probabilities of D give rise to two more restrictions on the parameters of interest, namely

$$\begin{aligned} r^1 &= \sum_{k<0} p_k \\ r^0 &= \sum_{k\geqslant 0} p_k \end{aligned} \tag{2.5}$$

Note that $r^1, r^0 > 0$ under Assumption 2, and the q_i^1 and q_j^0's are thus well-defined. Note also that $\sum_k p_k = 1$ follows from $r^1 + r^0 = 1$ and (2.5), and $\sum_k m_k = 1$ follows from $\sum_i (q_i^1 r^1 + q_i^0 r^0) = 1$, and they are therefore redundant constraints.

Together, (2.4) and (2.5) represent $2K_u + K_{X^*}$ restrictions on $K_u + K_{X^*}$ parameters, where $K_{X^*} = |\{L_{X^*}, L_{X^*}+1, ..., U_{X^*}-1, U_{X^*}\}|$ and $K_u = |\{L_u, L_u+1, ..., U_u-1, U_u\}|$ denote the number of probability mass points to be identified in the X^* and u distributions. Even though there are more constraints than the number parameters, it is not clear that the X^* distribution is always identified because of the nonlinearity in (2.4). To formally investigate the identifiability of the parameters, I introduce the following notations: Let $p_k^1 = \frac{p_k}{r^1}$ for $k \leqslant 0$ and $p_k^0 = \frac{p_k}{r^0}$ for $k > 0$. Define $Q^1(t) = \sum_i q_i^1 e^{ti}$, $Q^0(t) = \sum_j q_j^0 e^{ti}$, $P^1(t) = \sum_k p_k^1 e^{tk}$, $P^0(t) = \sum_k p_k^0 e^{tk}$ and $M(t) = \sum_l m_l e^{tl}$, which are the moment generating functions (MGF's) of the random variables, $X|D=1$, $X|D=0$, $X^*|D=1$, $X^*|D=0$ and u.[6] It is a well-known result that the moment generating function of the sum of two independent random variables is the product of the moment generating functions of the two variables (see for example Chapter 10 of Grinstead and Snell

[6]Because of the bounded support assumption, the defined moment generating functions always exist and are positive for all t.

(1997)). Consequently, equations (2.4) and (2.5) can be compactly represented as

$$\begin{aligned} Q^1(t) &= P^1(t)M(t) \text{ for all } t \neq 0 \\ Q^0(t) &= P^0(t)M(t) \text{ for all } t \neq 0 \\ P^1(0) &= 1 \\ P^0(0) &= 1 \end{aligned} \tag{2.6}$$

For the first two equations above, the coefficients on the e^{ti} term in $Q^1(t)$ and $Q^0(t)$ are q_i^1 and q_i^0 respectively for each i, and those on the e^{ti} term in $P^1(t)M(t)$ and $P^0(t)M(t)$ are $\sum_k p_k^1 m_{i-k}$ and $\sum_k p_k^0 m_{i-k}$ respectively. The last two equations are simply another way of writing (2.5).

Because $P^1(t)$ and $P^0(t)$ are everywhere positive, (2.6) implies that

$$M(t) = \frac{Q^1(t)}{P^1(t)} = \frac{Q^0(t)}{P^0(t)}$$

and it follows that,

$$P^0(t)Q^1(t) = P^1(t)Q^0(t)$$

which eliminates the nuisance parameters associated the measurement error distribution. Matching the coefficient for each of the e^{ti} terms in $P^0(t)Q^1(t)$ to that in $P^1(t)Q^0(t)$ along with the constraint $P^1(0) = P^0(0) = 1$ results in the following linear system of equations in terms of

the p_k^1's and p_k^0's:

$$\underbrace{\begin{bmatrix} q^1_{U_u-1} & 0 & \cdots & 0 & -q^0_{U_{X^*}+U_u} & 0 & \cdots & 0 \\ q^1_{U_u-2} & q^1_{U_u-1} & \cdots & 0 & -q^0_{U_{X^*}+U_u-1} & -q^0_{U_{X^*}+U_u} & \cdots & 0 \\ \vdots & q^1_{U_u-2} & \cdots & \vdots & \vdots & -q^0_{U_{X^*}+U_u-1} & \cdots & \vdots \\ q^1_{L_u+L_{X^*}} & \vdots & \cdots & 0 & -q^0_{L_u} & \vdots & \cdots & 0 \\ 0 & q^1_{L_u+L_{X^*}} & \cdots & q^1_{U_u-1} & 0 & -q^0_{L_u} & \cdots & -q^0_{U_{X^*}+U_u} \\ \vdots & 0 & \cdots & q^1_{U_u-2} & \vdots & 0 & \cdots & -q^0_{U_{X^*}+U_u-1} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & q^1_{L_u+L_{X^*}} & 0 & 0 & \cdots & -q^0_{L_u} \\ 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \end{bmatrix}}_{\mathbf{Q}:\,(K_{X^*}+K_u)\times K_{X^*}} \underbrace{\begin{bmatrix} p^0_{U_{X^*}} \\ p^0_{U_{X^*}-1} \\ \vdots \\ p^0_0 \\ p^1_{-1} \\ p^1_{-2} \\ \vdots \\ p^1_{L_{X^*}} \end{bmatrix}}_{\mathbf{p}:\,K_{X^*}\times 1} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{b}:\,(K_{X^*}+K_u)\times 1} \tag{2.7}$$

Standard results in linear algebra can be invoked to provide identification of the probability masses. Denote system (2.7) with the compact notation $\mathbf{Qp} = \mathbf{b}$, where $\mathbf{Q}$ is the $(K_{X^*} + K_u) \times K_{X^*}$ data matrix, $\mathbf{p}$ is the $K_{X^*} \times 1$ parameter vector and $\mathbf{b}$ is the $(K_{X^*} + K_u) \times 1$ vector of 0's and 1's. The parameter vector $\mathbf{p}$ is identified if and only if $\mathbf{Q}$ is of full rank. Note that there are more rows than columns in $\mathbf{Q}$, i.e. $K_{X^*} + K_u > K_{X^*}$, and therefore I introduce the following assumption

Assumption 3 (Full Rank). $\mathbf{Q}$ in equation (2.7) has full column rank.

Note that Assumption 3 is not always satisfied, and an example is provided in the appendix. At the same time, Assumption 3 is directly testable because $\text{rank}(\mathbf{Q}) = \text{rank}(\mathbf{Q^T Q})$, and $\mathbf{Q}$ is of full rank if and only if $\det(\mathbf{Q^T Q}) = 0$. The distribution of the determinant estimator can be obtained by a simple delta method because it is a polynomial in the q_i^0 and q_j^1's, the observed probability masses. Characterizing the set of X^* and u distributions that guarantee full column rank of $\mathbf{Q}$ is a task left for future research.

With Assumption 3, the $\mathbf{p}$ vector is identified. Because $p_k = r^1 p_k^1$ for $k < 0$ and $p_k = p_k^0 r^0$ for $k \geqslant 0$ and because r^1 and r^0 are observed, uniqueness of the p_k^1's and the p_k^0's under Assumption 3 implies the uniqueness of the p_k's. Although parameters of the the measurement error

distribution are eliminated in (2.7), they are identified after the identification of the p_k's as shown in the Appendix. Formally, the identification of probability masses is summarized in the following Lemma.

Lemma 2 (**Probability Mass Identification in a Sharp Design**) Suppose $(\{p_k\}, \{m_l\})$ ($k \in \{L_{X^*}, L_{X^*}+1, ..., U_{X^*}-1, U_{X^*}\}$, $l \in \{L_u, L_u+1, ..., U_u-1, U_u\}$) solves the system of equations consisting of (2.4) and (2.5). Then $(\{p_k\}, \{m_l\})$ is the unique solution if and only if Assumption 3 holds.

Combining Lemma 1 and 2 implies the identification of model (2.2):

Proposition 1 Under Assumption DB, 1, 2, and 3, the distributions of X^* and u are identified.

In the next section, I discuss the implications of alternatives to Assumption DB by considering the case where the supports of X^* and u are unbounded. I construct an example to show that in general model (2.2) is not identified.

2.2.3 Running Variable and Measurement Error Have Discrete and Unbounded Support

Following the identification result in 2.2.2, a natural question arises: how does the result extend to the case where the support of the discrete running variable is unbounded? While a sufficient condition for identification is left for future research, I will show in this section that the model is not always identified by constructing two sets of observationally equivalent distributions. While the general non-identifiability result may not be surprising given the absence of an infinite-support-counterpart to Assumption 3, the construction of the example is not straightforward as the technique used in the construction of a not-full-rank $\mathbf{Q}$ in the Appendix no longer applies when supports of X^* and u are infinite.

I construct two sets of infinitely supported distributions of X^* and u that are observationally equivalent, i.e. they give rise to the same joint distribution (X, D^*). In particular, I specify

discrete probability mass functions, $\{p_a^1, p_a^0, m_a\}$ and $\{p_b^1, p_b^0, m_b\}$ (where p_j^1 and p_j^0 $(j = a, b)$ denote the conditional probability mass functions of $X^*|D^* = 1$ and $X^*|D^* = 0$ respectively) such that

1. the support of p_a^1, p_b^1 is the set of negative integers $\{-1, -2, -3, ...\}$;

2. the support of p_a^0, p_b^0 is the set of non-negative integers $\{0, 1, 2...\}$;

3. $q^1 \equiv p_a^1 * m_a = p_b^1 * m_b$ and $q^0 \equiv p_a^0 * m_a = p_b^0 * m_b$ where $*$ denotes convolution.

As in the previous section, the probability mass functions q^1 and q^0 are the observed distribution of the noisy running variable X conditional on $D = 1$ and $D = 0$ respectively. Note that Assumption 1 and 2 still hold in the construction of the example.

It is useful to consider yet again the moment generating functions of the distributions, which I denote by $\{P_a^1(t), P_a^0(t), M_a(t)\}$ and $\{P_b^1(t), P_b^0(t), M_b(t)\}$.[7] Again, I can translate the convolutions of the distributions $p_a^1 * m_a = p_b^1 * m_b$ and $p_a^0 * m_a = p_b^0 * m_b$ into products of MGF's $P_a^1(t)M_a(t) = P_b^1(t)M_b(t)$ and $P_a^0(t)M_a(t) = P_b^0(t)M_b(t)$. It follows then that

$$\begin{aligned} P_b^1(t) &= P_a^1(t)\frac{M_a(t)}{M_b(t)} \\ P_b^0(t) &= P_a^0(t)\frac{M_a(t)}{M_b(t)} \end{aligned} \tag{2.8}$$

Loosely speaking, the supports of p_a^1 and p_b^1 are preserved under convolution with the "distribution" represented by $\frac{M_a}{M_b}(t)$.

To construction the two sets of distributions, I will first specify $\frac{M_a}{M_b}(t)$, $P_a^1(t)$, $P_a^0(t)$ and $M_b(t)$, and then show that $P_b^1(t)$ and $P_b^0(t)$ obtained following (2.8) are moment generating functions for valid probability distributions that are supported on the negative and non-negative integers respectively. Finally, I will check that $M_a(t)$ constructed by $\frac{M_a}{M_b}(t)M_b(t)$ represents a valid probability distribution.

[7] When the support is unbounded, the question of convergence naturally arises regarding the moment generating functions. I am not concerned with the convergence issue and will use the MGF's in the formal sense as I am only interested in the coefficients of the e^{ti} terms.

Let

$$\begin{aligned}
\frac{M_a}{M_b}(t) &= c_{a/b}(x + \sum_{n\neq 0}(-x)^{|n|-1}e^{tn}) \\
P_a^1(t) &= c_a^1(\frac{x^2}{1+x^2}e^{-t} + \sum_{n\leqslant -2} x^{|n|-1}e^{tn}) \\
P_a^0(t) &= c_a^0(\frac{x^2}{1+x^2} + \sum_{n\geqslant 1} x^{|n|-1}e^{tn}) \\
M_b(t) &= \frac{1}{2}(P_a^1(t) + P_a^0(t))
\end{aligned}$$

where x is any constant in the interval $(0,1)$, and $c_{a/b} = \frac{x+1}{x^2+x+2}$, $c_a^1 = c_a^0 = \frac{1-x+x^2-x^3}{x+x^2}$ are normalizing constants so that $\frac{M_a}{M_b}(0) = P_a^1(0) = P_a^0(0) = 1$ (and consequently $M_b(0) = 1$). Using (2.8), I obtain

$$\begin{aligned}
P_b^1(t) &= c_{a/b}c_a^1\left(xe^{-t} + \sum_{n\leqslant -2 \text{ and } n \text{ even}} \frac{x^{|n|}(x^2+3)}{x^2+1}e^{tn} + \sum_{n\leqslant -3 \text{ and } n \text{ odd}} (x^{|n|} + x^{|n|-2})e^{tn}\right) \\
P_b^0(t) &= c_{a/b}c_a^0\left(x + \sum_{n\geqslant 1 \text{ and } n \text{ odd}} \frac{x^{|n|+1}(x^2+3)}{x^2+1}e^{tn} + \sum_{n\geqslant 2 \text{ and } n \text{ even}} (x^{|n|+1} + x^{|n|-1})e^{tn}\right) \\
M_a(t) &= \frac{1}{2}[P_b^1(t) + P_b^0(t)]
\end{aligned}$$

Note that $P_b^1(t)$ only contains negative powers of e^t and that $P_b^0(t)$ only contains non-negative powers of e^t. Also, all coefficients of powers of e^t in P_b^1, P_b^0 and M_a are strictly positive with $P_b^1(0) = P_b^0(0) = M_a(0) = 1$. Thus, P_b^1, P_b^0 and M_a represent valid probability distributions that satisfy the support requirement mentioned above. Hence, (2.1) is not always identified when the supports of X^* and u are infinite.

Extending the support of X^* and u to infinity only represents one departure from Assumption DB. Another natural alternative is the case where X^* and u are continuously distributed and is currently investigated. Preliminary results are presented in the Appendix.

Imperfect Compliance

As mentioned in section 2.2.1, the assumption of perfect compliance or equivalently the observability of eligibility ($D^* = 1_{[X^*<0]}$) is not often satisfied, as is the case in almost all means-tested social programs. Instead, only a measure of program participation D may be available. In this subsection, I consider the more realistic case of imperfect compliance for discrete and bounded running variable X^* and measurement error u. The task becomes the identification of the X^* distribution from the observed joint distribution (X, D).

Rather than having benefit receipt D as a deterministic step function of X^*, $\Pr(D = 1|X^*)$ is potentially an unrestricted function in X^* even though eligibility is still given by $D^* = 1_{[X^*<0]}$. In the extreme, program participation D could be independent from X^* and therefore would not provide any information for X^*. If this were the case, deconvolving $X^*$and u from the observed joint distribution (X, D) would not be possible. In many programs (typically means-tested transfer programs), however, it is the case that if an agent's true running variable is above the eligibility threshold, she is forbidden from participating in the program, that is

Assumption 4 (One-sided Fuzzy). $\Pr(D = 1|X^* = x^*) = 0$ for $x^* \geqslant 0$.

Assumption 4 informs the researcher that any agent with $D = 1$ has true running variable $X^* < 0$. It follows that the upper end point in the $X|D = 1$ distribution will identify U_u (the upper end point of the u distribution as defined in 2.2.2) provided that Assumption 1 and 2 hold and for the $D = 1$ population. Unlike in the perfect compliance scenario where $X^* \perp\!\!\!\perp u$ conditional on D^* directly follows from Assumption 1, an additional assumption is needed to ensure the independence between X and u conditional on $D = 1$:

Assumption 1F (Strong Independence). $u \perp\!\!\!\perp X^*, D$.

and the required extension of Assumption 2 is

Assumption 2F (Threshold Support: Fuzzy). $-1 \in \text{support}_{X^*|D=1}$ and $0 \in \text{support}_{X^*}$.

I make two remarks regarding Assumption 1F and 2F. First, note that a weaker version of Assumption 1F, $u \perp\!\!\!\perp X^*$ conditional on D, suffices for the results below. However, it may be difficult to justify this weaker condition economically. In fact, this assumption does not even imply that $X^* \perp\!\!\!\perp u$ unconditionally, which is the baseline model in the measurement error literature. Therefore, I propose the stronger Assumption 1F. Second, even though Assumption 2F is stronger Assumption 2, it needs to be satisfied in a valid fuzzy RD design. It states that there is non-zero take-up just below the eligibility cutoff, without which there may not be a first-stage discontinuity.

The difficulty still remains in distinguishing between non-compliance and ineligibility after the introduction of Assumption 1F and 2F. An agent with $D = 0$ and $X = -1$ could have true income $X^* = 1$ (with an implied measurement error $u = -2$) and is not program eligible; or she could be eligible with income $X^* = -1$ (with an implied measurement error $u = 0$) but chooses not to participate in the program. On the one hand, if every observation with $D = 0$ is treated as ineligible, then the lower end point in the support of u, L_u, is that in the $X|D = 0$ distribution. On the other hand, if every observation with $D = 0$ is treated as an eligible non-take-up, then L_u is 0. Clearly, the two treatments imply different distributions. However, if the researcher believes that the identified length of the right tail in the u distribution sheds light on the length of its left tail, it may be reasonable to assume

Assumption 5 (Symmetry in Support). $L_u = -U_u$

which is weaker than imposing symmetry in the measurement error distribution as is conventional in the literature.

With the additional Assumptions 4, 1F, 2F and 5, the supports of the X^* and u are identified:

Lemma 1F (**Support Identification in a Fuzzy Design**) Under Assumption 1F, 2F, 4 and 5 the

upper and lower end points of the u distribution are given by

$$\begin{aligned} U_u &= U_{X|D=1} + 1 \\ L_u &= -(U_{X|D=1} + 1) \end{aligned} \tag{2.9}$$

and those of the $X^*|D = d$ $(d = 0, 1)$ distributions are given by:

$$\begin{aligned} L_{X^*|D=1} &= L_{X|D=1} - L_u \\ L_{X^*|D=0} &= L_{X|D=0} - L_u \\ U_{X^*|D=1} &= -1 \\ U_{X^*|D=0} &= U_{X|D=0} - U_u \end{aligned} \tag{2.10}$$

As in subsection 2.2.2, the right tail of $X^* \geqslant 0$ and $D = 1$ population provides information on the length of the right tail of the measurement error distribution thanks to Assumption 4. The length of the left tail of the measurement error distribution is then obtained following Assumption 5. As it turns out, the identification of probability masses can proceed analogously as in subsection 2.2.2.

Because of the existence of non-participants, however, the distribution of X^* conditional $D = 0$ is also supported on negative integers. The number of parameters therefore is larger than that in the perfect compliance case even if the support of the unconditional X^* distribution does not change. It is straightforward to show that the convolution relationships under Assumption 1F again lead to a system of equations $\mathbf{Q_F p_F} = \mathbf{b_F}$:

$$\underbrace{\begin{bmatrix} q^1_{U_u-1} & 0 & \cdots & 0 & -q^0_{U^0_{X^*}+U_u} & 0 & \cdots & 0 \\ q^1_{U_u-2} & q^1_{U_u-1} & \cdots & 0 & -q^0_{U^0_{X^*}+U_u-1} & -q^0_{U_{X^*}+U_u} & \cdots & 0 \\ \vdots & q^1_{U_u-2} & \cdots & \vdots & \vdots & -q^0_{U_{X^*}+U_u-1} & \cdots & \vdots \\ q^1_{L_u+L^1_{X^*}} & \vdots & \cdots & 0 & -q^0_{L^0_u+L_u} & \vdots & \cdots & 0 \\ 0 & q^1_{L_u+L^1_{X^*}} & \cdots & q^1_{U_u-1} & 0 & -q^0_{L^0_u+L_u} & \cdots & -q^0_{U_{X^*}+U_u} \\ \vdots & 0 & \cdots & q^1_{U_u-2} & \vdots & 0 & \cdots & -q^0_{U_{X^*}+U_u-1} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & q^1_{L_u+L^1_{X^*}} & 0 & 0 & \cdots & -q^0_{L_u} \\ 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \end{bmatrix}}_{\mathbf{Q_F}:\,(K^F_{X^*}+K_u)\times K^F_{X^*}} \underbrace{\begin{bmatrix} p^0_{U^0_{X^*}} \\ p^0_{U^0_{X^*}-1} \\ \vdots \\ p^0_{L^0_{X^*}+1} \\ p^0_{L^0_{X^*}} \\ p^1_{-1} \\ p^1_{-2} \\ \vdots \\ p^1_{L^1_{X^*}} \end{bmatrix}}_{\mathbf{p_F}:K^F_{X^*}\times 1} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{b_F}:\,(K^F_{X^*}+K_u)\times 1} \tag{2.11}$$

Note that in (2.11), I

1) adopt the notation $L_{X^*|D=d} = L^d_{X^*}$, $U_{X^*|D=d} = U^d_{X^*}$ for $d = 0, 1$;

2) define $K^F_{X^*} = U^0_{X^*} - L^0_{X^*} - L^1_{X^*} + 1$ and

3) use the superscript 1 and 0 to indicate conditioning on $D = 1$ and $D = 0$ (as opposed to $D^* = 1$ and $D^* = 0$ as in (2.7)).

Analogous to (2.7), the number of rows in $\mathbf{Q_F}$ is K_u more than the number of columns. Full column rank in $\mathbf{Q_F}$ will again be a necessary and sufficient condition for identification:

Assumption 3F (Full Rank: Fuzzy). $\mathbf{Q_F}$ in equation (2.11) has full column rank.

Thus, I arrive at the counterpart of Proposition 1 for the fuzzy RD case:

Proposition 1F Under Assumption DB, 1F, 2F, 3F, 4 and 5 the distributions of X^* conditional on D and u are identified.

It is worth noting as a theoretical point that identification is possible in the absence of Assumption 2F, 4F and 5F provided that the researcher has exact knowledge of what U_u and L_u are. In this case, $L_{X^*|D=d}$ and $U_{X^*|D=d}$ can be recovered using this knowledge, and the probability masses are identified analogously if the full rank condition is satisfied. In practice, this

observation has little practical value since researchers rarely–if at all–know the true value of U_u and L_u. Thus, results may depend crucially on the imposed support, and the act of imposing support should be carried out with caution in empirical studies.

A related point is that identification can be obtained with only the independence assumption (Assumption 1 in the sharp case and Assumption 1F in the fuzzy case) if the econometrician has explicit knowledge of the marginal distribution of X^*, say from a validation sample. This is because, as it is easy to show, the marginal distribution of u is identified from the marginal distribution of X and X^* by an overidentified linear system of equations. It follows that the distribution of X^* conditional on D^* in the sharp case or conditional on D in the fuzzy case is identified from the observed $X|D^*$ distribution in the sharp case or $X|D$ distribution in the fuzzy case together with the identified u distribution. In practice, however, it is unlikely that an econometrician can obtain the marginal distribution of X^* in the case of a transfer program even if s/he has access to administrative earnings data. First of all, commonly used administrative earnings records are of quarterly frequency, but program eligibility is usually based on monthly income. Second, the income used for determining program eligibility is typically earnings after certain deductions (child care or work related expenses, for example) plus unearned income. In that sense, the administrative earnings records are also a noisy version of the income for program eligibility determination, not to mention the fact that they may not perfectly measure true earnings either (e.g. Abowd and Stinson (2007)). That said, the possibility of obtaining the marginal distribution of X^* for other applications should not be overlooked.

Finally, one might question the implicit assumption that benefit receipt D is accurately measured when discussing the measurement error in X. For example, errors in reporting program participation status in means-tested transfer programs has been documented in validation studies of survey data. Marquis and Moore (1990) reports that the AFDC under-reporting rate (i.e. those who did not report AFDC receipt among all who received the benefit) in the 1984 SIPP panel could be as high as 50%. The problem with under-reporting Medicaid coverage is also present but appears to be less severe–Card et al. (2004) estimates that the overall error rate in the 1990-93

SIPP Medicaid status is 15% for the state of California.

Under-reporting, however, does not pose a threat to the identification of the X^* distribution as long as those with $D = 1$ indeed received benefits and were therefore eligible. It will still follow that the support of X^* conditional on D and u are identified correctly, and probability masses can be recovered as long as Assumption 1F holds. It will be problematic, however, if those who do not take part in the program report participation. Fortunately, the rate of false-positive reporting associated with transfer programs at least is very small empirically–around 0.2% in the Marquis and Moore (1990) study and 1.5% in Card et al. (2004). This suggests that the reporting error in D will not pose a big threat to the procedure above when applying an RD design using benefit discontinuities at the eligibility cutoff. Further, trimming procedures can be undertaken to correct for the false-positive reporting problem, which will be discussed in detail in subsection 2.3.2.

2.2.4 Identification of the RD Treatment Effect

I this section, I show that the RD treatment effect parameter in both a sharp and one-sided fuzzy design can be identified under conditional versions of Assumption 1, 2 and 3. In essence, these assumptions allow the performance of the deconvolution exercise detailed in section 2.2.2 for each value of Y, the outcome variable. Once I obtain the distribution of X^* conditional on each value of Y, I apply Bayes' Theorem to recover $E[Y|X^* = x^*]$ in the sharp RD case and both $E[D|X^* = x^*]$ and $E[Y|X^* = x^*]$ in the fuzzy case, which are sufficient for identifying the RD treatment effect. For this study, I focus on the case in which Y is binary, but it can be easily extended to the framework where Y is multinomial.

In the sharp RD model (2.1), the treatment effect is $\delta_{sharp} = \lim_{x^* \uparrow 0} E[Y|X^* = x^*] - E[Y|X^* = 0]$ for discrete X^* (note the slight difference between δ_{sharp} presented in subsection 2.2.1 and here: discrete X^* implies $\lim_{x^* \downarrow 0} E[Y|X^* = x^*] = E[Y|X^* = 0]$). Clearly, δ_{sharp} is identified if the entire conditional expectation function $E[Y|X^*]$ is identified. In order to identify $E[Y|X^*]$, I propose the following assumptions which imply that Assumption 1, 2 and 3 hold conditional on Y

Assumption 1Y (Non-Differential Measurement Error). $u \perp\!\!\!\perp X^*, Y$.[8]

Assumption 2Y (Threshold Support). $-1, 0 \in \text{support}_{X^*|Y=y}$ for each $y = 0, 1$.

In order to to state Assumption 3 conditional on Y, note that Assumption 1Y and Assumption 2Y allow the formulation of the conditional counterparts of (2.7), $\mathbf{Q}_Y \mathbf{p}_Y = \mathbf{b}_Y$ for $Y = 0, 1$, where $\mathbf{Q}_Y$ and $\mathbf{p}_Y$ consist of probability masses of q_j^1,q_j^0, p_k^1 and p_k^0 (for the conditional distributions of X and X^*on $D^* = 1$ and 0 respectively) conditional on Y.

Assumption 3Y (Full Rank). The matrix $\mathbf{Q}_Y$ is of full rank for $Y = 0, 1$.

Proposition 2 Under Assumption DB, 1Y, 2Y and 3Y, the RD treatment effect parameter δ_{sharp} is identified for model (2.1).

Proof. Assumption 1Y implies that $X^* \perp\!\!\!\perp u$ conditional on Y. Therefore, the distribution of X^* is identified from the observed joint distribution of (X, D^*) conditional on each value of Y following Proposition 1. That is, I can obtain the X^* distribution conditional on Y, $\Pr(X^* = x^*|Y = y)$ for all x^* and $y = 0, 1$. Consequently, $E[Y|X^* = x^*]$ is recovered by Bayes' Theorem since the marginal distribution of Y is observed in the data. In the binary case,

$$E[Y|X^* = x^*] = \frac{\Pr(X^* = x^*|Y = 1)\Pr(Y = 1)}{\sum_y \Pr(X^* = x^*|Y = y)\Pr(Y = y)} \tag{2.12}$$

δ_{sharp} is consequently identified because it is a function of $E[Y|X^* = x^*]$. □

Identification of the RD treatment effect parameter is obtained analogously in a fuzzy RD setting except that the first stage relationship $E[D|X^*]$ also needs to be recovered. Consider

[8]Non-differential measurement error is a commonly made assumption in the literature (see Carroll et al. (2006) for reference). As with Assumption 1F, a weaker version of Assumption 1FY, $X^* \perp\!\!\!\perp u$ conditional on Y, also delivers the following identification results. However, I adopt Assumption 1Y for its simplicity in economic interpretation. This is also the case for Assumption 1FY for exactly the same reason.

formally the fuzzy RD model where the running variable is measured with error:

$$\begin{aligned} Y &= h(D, X^*, \epsilon) \\ X &= X^* + u \end{aligned} \tag{2.13}$$

where the outcome Y depends on program participation D. Under the assumption that $E[h(d, X^*, \epsilon)|X^* = x^*]$ is continuous at the threshold $x^* = 0$, the ratio

$$\delta_{fuzzy} = \frac{\lim_{c \uparrow 0} E[Y|X^* = c] - E[Y|X^* = 0]}{\lim_{c \uparrow 0} E[D|X^* = c] - E[D|X^* = 0]}$$

for X^* discrete is the average treatment effect of D on Y for the "complier" population that takes up benefit when eligible (e.g. Hahn et al. (2001), Lee and Lemieux (2010)). δ_{fuzzy} is the RD treatment effect to be identified in Model (2.13), and the identification strategy is analogous to that in the sharp case: First identify the X^* distribution conditional on D and Y, and then apply Bayes' Theorem to recover the conditional expectation of Y and D on X^*. Again, assumptions underpinning Proposition 1F are extended to hold conditional on Y:

Assumption 1FY (Strong Independence and Non-Differential Measurement Error: Fuzzy). $u \perp\!\!\!\perp X^*, D, Y$.

Assumption 2FY (Threshold Support: Fuzzy). $-1 \in \text{support}_{X^*|D=1,Y=y}$ and $0 \in \text{support}_{X^*|Y=y}$ for each $y = 0, 1$.

As in the sharp case, in order to state Assumption 3F conditional on Y, note that Assumption 1FY and Assumption 2FY allow the formulation of the conditional counterparts of (2.11), $\mathbf{Q_{F}}_Y \mathbf{p_{F}}_Y = \mathbf{b_{F}}_Y$ for $Y = 0, 1$, where $\mathbf{Q_{F}}_Y$ and $\mathbf{p_{F}}_Y$ consist of probability masses of q_j^1,q_j^0, p_k^1 and p_k^0 (for the conditional distributions of X and X^*on $D = 1$ and 0 respectively) conditional on Y.

Assumption 3FY (Full Rank: Fuzzy). The matrix $\mathbf{Q_{F}}_Y$ is of full rank for $Y = 0, 1$.

Proposition 2F Under Assumption DB, 1FY, 2FY, 3FY, 4 and 5, the RD treatment effect parameter δ_{fuzzy} is identified for model (2.13).

Proof. Analogous to arguments in the previous subsection, Assumption DB, 1FY, 2FY, 3FY, 4 and 5 are sufficient conditions for identifying the X^* distribution conditional on $D = d$ and $Y = y$ for $d, y = 0, 1$ from that of $(X, D)|Y$. It follows that $\Pr(X^* = x^*|D = d)$ and $\Pr(X^* = x^*|Y = y)$ for $d, y = 0, 1$ are identified because $\Pr(Y = y|D = d)$ and $\Pr(D = d|Y = y)$ are observed in the data:

$$\Pr(X^* = x^*|D = d) = \sum_y \Pr(X^* = x^*|D = d, Y = y) \Pr(Y = y|D = d) \tag{2.14}$$

$$\Pr(X^* = x^*|Y = y) = \sum_d \Pr(X^* = x^*|D = d, Y = y) \Pr(D = d|Y = y) \tag{2.15}$$

Consequently, $E[D|X^* = x^*]$ and $E[Y|X^* = x^*]$ are recovered by an application of the Bayes' Theorem

$$E[D|X^* = x^*] = \frac{\Pr(X^* = x^*|D = 1) \Pr(D = 1)}{\sum_d \Pr(X^* = x^*|D = d) \Pr(D = d)} \tag{2.16}$$

$$E[Y|X^* = x^*] = \frac{\Pr(X^* = x^*|Y = 1) \Pr(Y = 1)}{\sum_y \Pr(X^* = x^*|Y = y) \Pr(Y = y)} \tag{2.17}$$

and δ_{fuzzy} is identified since it is a function of $E[D|X^* = x^*]$ and $E[Y|X^* = x^*]$. □

2.3 Estimation

2.3.1 Theoretical Procedures

As in identification, estimation of the X^* distribution follows two steps: Estimation of its support and estimation of the probability masses at each point in its support. Estimation of support follows Equations (2.10) and (2.9) with the population quantities replaced by sample analogues. I

will abstract away from the sampling error of support and simply assume that the sample is large enough to reveal the true support of the distribution. I present the case in the sharp design setting where I omit the F subscript for notational convenience. Results in the fuzzy case follow by replacing D^* by D.

Given the specification of probability model with independent measurement error, the likelihood function can be explicitly written out by using the p_k^1's, p_k^0's, m_l's and the marginal probability $r = \Pr(D^* = 1)$. Formally, the likelihood for the joint distribution (X,D^*) is

$$\begin{aligned} L(X_i, D_i^*) &= L(X_i|D_i^*)L(D_i^*) \\ &= \{(\sum_k p^1_{X_i-k} m_k)r\}^{D_i^*} \{(\sum_k p^0_{X_i-k} u_k)(1-r)\}^{1-D_i^*} \end{aligned} \tag{2.18}$$

Researchers can directly estimate (2.18) and the resulting estimators are efficient provided that the parameters are in the interior of the parameter space, i.e. strictly between 0 and 1. However, the analytical solutions to maximizing the log likelihood do not appear to exist, and numerically optimizing (2.18) may become computationally intensive as the number of points in the support of X^* and u increases.

An alternative strategy relies on the identification equation (2.7), which fits nicely into a standard minimum distance framework $f(\mathbf{q}, \mathbf{p}) = \mathbf{Qp} - \mathbf{b} = 0$ (Kodde et al. (1990)) from which an estimator of $\mathbf{q}$ can be obtained easily[9]. Because of the linearity in (2.7), the parameter vector of interest $\mathbf{p}$ can be estimated analytically once an estimator of $\mathbf{q}$ are obtained. Estimation follows the following steps:

1. Obtain the estimators $\hat{q}_k^1 = \frac{\sum_i 1_{[X_i=k]} \cdot 1_{[D_i=1]}}{\sum_i 1_{[D_i=1]}}$, $\hat{q}_k^0 = \frac{\sum_i 1_{[X_i=k]} \cdot 1_{[D_i=0]}}{\sum_i 1_{[D_i=0]}}$ and $\hat{r} = \frac{1}{N} \sum 1_{[D_i=1]}$ (N denotes the sample size), as well as $\hat{\Omega}$, which is a consistent estimator for the asymptotic variance-covariance matrix Ω of $\hat{\mathbf{q}}$, (that is, $\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q}) \Rightarrow N(0, \Omega)$). Since $X|D^* = d$ follows a multinomial distribution for each d, $\hat{\Omega}$ is a block-diagonal matrix $\hat{\Omega} = \begin{bmatrix} \hat{\Omega}^{11} & \hat{\Omega}^{10} \\ \hat{\Omega}^{01} & \hat{\Omega}^{00} \end{bmatrix}$ where

[9] $\mathbf{q}$ is the analogue of $\mathbf{p}$ by stacking all the observed conditional probability masses q_k^1 and q_k^0's.

$\hat{\Omega}^{10} = 0$,$\hat{\Omega}^{10} = 0$ and

$$\hat{\Omega}_{ij}^{dd} = \begin{cases} (1-\hat{q}_i^d)\hat{q}_i^d/(d\hat{r}+(1-d)(1-\hat{r})) & \text{if } i=j \\ \hat{q}_i^d\hat{q}_j^d/(d\hat{r}+(1-d)(1-\hat{r})) & \text{if } i\neq j \end{cases}$$

2. Form the estimator of the $\mathbf{Q}$ matrix under perfect compliance in (2.7),$\hat{\mathbf{Q}}$, by replacing the q_k^1 and q_k^0 in $\mathbf{Q}$ with their estimators;

3. Derive a consistent estimator of $\mathbf{p}$: $\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} f'(\hat{\mathbf{q}},\mathbf{p})f(\hat{\mathbf{q}},\mathbf{p}) = (\hat{\mathbf{Q}}'\hat{\mathbf{Q}})^{-1}(\hat{\mathbf{Q}}'\mathbf{b})$;

4. Compute the optimal weighting matrix $\widehat{\mathbf{W}} = (\widehat{\nabla_{\mathbf{q}}f}\hat{\Omega}\widehat{\nabla_{\mathbf{q}}f}')^{-1}$ where $\hat{\Omega}$ is a consistent estimator for the variance-covariance matrix of the $\mathbf{q}$ derived in step 1.[10] $\widehat{\nabla_{\mathbf{q}}f}$ is a consistent estimator for$\nabla_{\mathbf{q}}f$, the Jacobian of f with respect to $\mathbf{q}$. Because $\nabla_{\mathbf{q}}f$ depends on $\mathbf{p}$, step 3 was necessary for first obtaining a consistent estimate of $\mathbf{p}$. It turns out that f is also linear in $\mathbf{q}$, and hence $\widehat{\nabla_{\mathbf{q}}f}$ can be computed analytically;

5. Arrive at the optimal estimator of $\mathbf{p}$: $\hat{\mathbf{p}}_{opt} = (\hat{\mathbf{Q}}'\hat{\mathbf{W}}\hat{\mathbf{Q}})^{-1}(\hat{\mathbf{Q}}'\hat{\mathbf{W}}\mathbf{b})$.

Provided that the true parameter lies in the interior of the parameter space:

Assumption 6 (Interior). $\mathbf{p} \in (0,1)^K$ where K is the length of $\mathbf{p}$

The derivation of the asymptotic distribution of $\hat{\mathbf{p}}_{opt}$ is standard. Specifically,

Proposition 3 Under Assumption DB, 1, 2, 3 and 6 for the perfect compliance case, $\sqrt{N}(\hat{\mathbf{p}}_{opt}-\mathbf{p}) \Rightarrow \mathcal{N}(\mathbf{0},\mathbf{Q}'(\nabla_{\mathbf{q}}f\Omega\nabla_{\mathbf{q}}f')^{-1}\mathbf{Q})$ where Ω is the asymptotic variance-covariance matrix of $\mathbf{q}$, and $\mathbf{Q}$ along with $\nabla_{\mathbf{q}}f = \nabla_{\mathbf{q}}(\mathbf{Q}\mathbf{p}-\mathbf{b})$ are specified in Equation (2.7).

Analogously, for the fuzzy case, I have

Proposition 3F Under Assumption DB, 1F, 2F, 3F, 4, 5 and 6 for the imperfect compliance case, $\sqrt{N}(\hat{\mathbf{p}}_{\mathbf{F}opt}-\mathbf{p}_{\mathbf{F}}) \Rightarrow \mathcal{N}(\mathbf{0},\mathbf{Q}_{\mathbf{F}}'(\nabla_{\mathbf{q}_{\mathbf{F}}}f_{\mathbf{F}}\Omega_{\mathbf{F}}\nabla_{\mathbf{q}_{\mathbf{F}}}f_{\mathbf{F}}')^{-1}\mathbf{Q})$ where $\Omega_{\mathbf{F}}$ is the asymptotic

[10]It turns out that $\nabla_{\mathbf{q}}f\Omega\nabla_{\mathbf{q}}f$ is singular, and is analogous to the case covered in a recent paper by Satchachai and Schmidt (2008) where there are too many restrictions. The study advised against using the generalized inverse, which is confirmed by my own simulations. Instead, they propose dropping one or more restrictions, but stated that the problem of which restrictions to drop has not yet been solved. All the empirical results are based on the last row of the $\mathbf{Q}$ matrix being dropped (dropping other rows had little impact on the empirical results).

variance-covariance matrix of $\mathbf{q_F}$, and $\mathbf{Q_F}$ along with $\nabla_{\mathbf{q_F}} f_{\mathbf{F}} = \nabla_{\mathbf{q_F}}(\mathbf{Q_F p_F} - \mathbf{b_F})$ are specified in Equation (2.11).

The main conclusion of Kodde et al. (1990) shows that $\hat{\mathbf{p}}_{opt}$ is efficient–it has the same asymptotic variance as the maximum likelihood estimator–if $\mathbf{p}$ is exactly or over-identified by $f(\mathbf{q}, \mathbf{p}) = 0$, and $\hat{\mathbf{q}}$ is the maximum likelihood estimator. Since both conditions are satisfied in my setup, I have obtained a computationally inexpensive estimator without sacrificing efficiency. Also note that the assumptions can be jointly tested by the overidentifying restrictions as is standard for minimum distance estimators. In particular, the test statistic $N \cdot (f(\hat{\mathbf{q}}, \hat{\mathbf{p}}_{opt})' \hat{\mathbf{W}} f(\hat{\mathbf{q}}, \hat{\mathbf{p}}_{opt}))$ in the sharp case or $N \cdot (f(\hat{\mathbf{q}}_{\mathbf{F}}, \hat{\mathbf{p}}_{\mathbf{F}opt})' \hat{\mathbf{W}}_{\mathbf{F}} f(\hat{\mathbf{q}}_{\mathbf{F}}, \hat{\mathbf{p}}_{\mathbf{F}opt}))$ in the fuzzy case follows a χ^2-distribution with degrees of freedom equal to K_u, the number of points in the support of the measurement error, when assumptions in Proposition 3 or Proposition 3F are satisfied.

A concern arises because the optimal estimators of p_k^1 and p_k^0 may never sum to 1 by following the procedure above. Therefore, I need to modify the estimation strategy and impose this constraint, as oppose to simply minimizing the distance between the sums and 1. The following modifications will incorporate the matrix constraints $\mathbf{Rp} = \mathbf{c}$, where $\mathbf{c} = \begin{bmatrix} 1 \\ \\ 1 \end{bmatrix}$ and $\mathbf{R} = \begin{bmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ & & & & & \\ 0 & \cdots & 0 & 1 & \cdots & 1 \end{bmatrix}$ that summarize the restriction that the p_k^1 and p_j^0 sum to 1. In step 3, the consistent estimator is instead $\hat{\mathbf{p}} = (\hat{\mathbf{Q}}'\hat{\mathbf{Q}})^{-1}\mathbf{R}'\{\mathbf{R}\hat{\mathbf{Q}}'\hat{\mathbf{Q}})^{-\mathbf{1}}\mathbf{R}'\}^{-1}\mathbf{c}$[11], and in step 5, $\hat{\mathbf{p}}_{opt} = (\hat{\mathbf{Q}}'\hat{\mathbf{W}}\hat{\mathbf{Q}})^{-1}\mathbf{R}'\{\mathbf{R}\hat{\mathbf{Q}}'\hat{\mathbf{W}}\hat{\mathbf{Q}})^{-\mathbf{1}}\mathbf{R}'\}^{-1}\mathbf{c}$, and finally the asymptotic variance of $\hat{\mathbf{p}}_{opt}$ is given by $\mathbf{T}((\mathbf{QT})'\mathbf{W}(\mathbf{QT}))^{-1}\mathbf{T}'$ where $\mathbf{T}$ is a matrix whose columns are the first $K-2$ eigenvectors of the projection matrix $\mathbf{I} - \mathbf{R}'(\mathbf{RR}')^{-1}\mathbf{R}$. The computation for $\hat{\mathbf{W}}$ is unaltered by the imposition of the linear constraints $\mathbf{Rp} = \mathbf{c}$.

Finally, in order to construct the asymptotic distribution of the RD treatment effect estimators, I will need to estimate the variance covariance matrix of $E[Y|X^* = x^*]$ in the sharp RD case and

[11] For a clear exposition of this standard result, see the "makecns" entry in Sta (2010).

$E[D|X^* = x^*]$ and $E[Y|X^* = x^*]$ in the fuzzy RD case for each x^* in the support of X^*. I refer to (2.14), (2.15), (2.16) and (2.17), which show that these two entities are differentiable functions of $\Pr(X^* = x^*|D = d, Y = y)$, $\Pr(D = d|Y = y)$ and $\Pr(Y = 1)$ for $d, y = 0, 1$. The delta method can be directly applied, where the Jacobian of the transformations is derived analytically. A general expression of the RD treatment effect estimator cannot be obtained because it depends on the the functional form of $E[Y|X^*]$ and $E[D|X^*]$ which varies from application to application. Therefore, I am not able to provide the asymptotic distribution for $\hat{\delta}_{sharp}$ and $\hat{\delta}_{fuzzy}$ in the general case, but specific examples are explored in subsection 2.3.3.

2.3.2 Potential Issues in Practical Implementation

There may be several issues in implementing the procedure described in subsection 2.3.1. First of all, in order to have realistic support for the true running variable, the maximum value of the observed running variable needs to be significantly larger for $D = 0$ group than for the $D = 1$ group, since the difference of the two is the upper end point in the true running variable distribution. Also, the left tail of the observed running variable distribution may need to be significantly longer that of the right tail of the $D = 1$ group since the difference in the lengths is the lower bound of the true running variable distribution (following Assumption 5). Since symmetry is a functional form assumption, which may not hold when the running variable is in levels (e.g. in the case of income), a transformation of the observed running variable may be needed. In practice, a Box-Cox type transformation is recommended and practitioners may experiment with various transformation parameters. The over-identification test mentioned in the previous section can be used to help decide which transformation parameters to use.

A related point, as mentioned at the end of subsection 2.2.3, is that someone with very large observed X and not program eligible may actually report program participation ($D = 1$) by mistake. If this is the case, the supports will not be correctly identified, and using a transformation parameter is not sufficient to correct the problem. A trimming procedure should be adopted in practice where outliers in both the left and right tails of the $X|D = 1$ and $X|D = 0$ populations

may be dropped. As with the case of transformation parameters, I recommend trying several trimming percentages and the sensitivity of empirical results should be examined. Finally, a quadratic programming routine with inequality constraints can be used in practice to guarantee non-negativity of the probability masses.

2.3.3 Simulations

In this section, I present results from Monte Carlo simulations that assess the proposed estimation procedure in subsection 2.3.1. I focus on the more complicated fuzzy case and show that the true first and second stage as well as the X^* distribution can indeed be recovered. In the baseline simulation, I generate X^* following a uniform distribution on the set of integers from -10 to 10. u follows a uniform distribution between -3 and 3 and is therefore symmetric in its support (Assumption 5). The true first stage relationship is given by

$$E[D|X^*] = \Pr(D = 1|X^*) = (\alpha_{D^*X^*}X^* + \alpha_{D^*})1_{[X^*<0]} = \alpha_{D^*}D^* + \alpha_{D^*X^*}D^*X^* \tag{2.19}$$

which reflects the one-sided fuzzy assumption (Assumption 4), and the size of the first stage discontinuity is α_{D^*}. The outcome response function is given by the simple constant treatment effect specification

$$E[Y|X^*, D] = \Pr(Y = 1|X^*, D) = \delta_0 + \delta_1 X^* + \delta_{fuzzy}D \tag{2.20}$$

where the treatment effect to be identified is δ_{fuzzy}. Note that (2.19) and (2.20) together imply that the second stage of Y versus X^* is

$$E[Y|X^*] = \Pr(Y = 1|X^*) = \beta_0 + \beta_{D^*}D^* + \beta_1 X^* + \beta_{D^*X^*}D^*X^* \tag{2.21}$$

where $\beta_0 = \delta_0$, $\beta_{D^*} = \alpha_{D^*}\delta_{fuzzy}$, $\beta_1 = \delta_1$ and $\beta_{D^*X^*} = \alpha_{D^*X^*} \cdot \delta_{fuzzy}$.

Figures 2.1 and 2.2 present graphical results based on a sample of 25,000 simulated

observations for the parameter values $\alpha_{D^*X^*} = -0.01$, $\alpha_{D^*} = 0.8$, $\delta_0 = 0.15$, $\delta_1 = -0.01$, and $\delta_{fuzzy} = 0.6$, with the implied coefficients in (2.21) being $\beta_0 = 0.15$, $\beta_{D^*} = 0.48$, $\beta_1 = -0.01$ and $\beta_{D^*X^*} = -0.006$. I choose $N = 25,000$ because it is about the average sample size in the relevant studies–45,722 in Hullegie and Klein (2010), 32,609 in Koch (2010), 11,541 in Schanzenbach (2009) and 2,163 in de la Mata (2011). The top and bottom panels in Figure 2.1 plot the observed first and second stage, i.e. $E[D|X]$ and $E[Y|X]$, respectively. Note that there is no visible discontinuity at the thresholds, and the estimated first and second stage discontinuities based on these observed relationships (as is the case with de la Mata (2011), for example) cannot identify the true discontinuities, which are 0.8 and 0.48 respectively.

Figure 2.2 plots the estimated first and second stage based on procedures developed in subsection 2.3.1 against the actual (2.19) and (2.20) specified with the parameter values above. As is evident from the graphs, the proposed procedures can correctly recover the true first and second stage of the underlying RD design. $\hat{\delta}_{fuzzy}$, the RD treatment effect parameter, is obtained by fitting another linear minimum distance procedure on the estimated $E[D|X^*]$ and $E[Y|X^*]$ (as well as their estimated variance-covariance matrices) with the parametric restrictions (2.19) and (2.21). In my simulated sample, the estimate of the treatment effect is 0.662 with a standard error of 0.061, and its 95% confidence interval $(0.542, 0.782)$ includes $\delta_{fuzzy} = 0.6$. Even when the size of the simulated sample is reduced to that in de la Mata (2011), 2,163, the estimate is 0.466 with a standard error of 0.136 and 95% confidence interval $(0.199, 0.732)$, which is still informative.

As mentioned above, the proposed procedure can also be used to estimate the discontinuity in the density of X^* at the eligibility threshold, which is key in evaluating the validity of an RD design but may in addition shed light on economically interesting quantities (as per Saez (1999)). I perform a simulation exercise to assess the ability of the estimation method to detect non-smoothness in the X^* distribution. In particular, I consider two alternative specifications: 1) I consider the specification above (i.e. that used for Figure 2.1 and 2.2), for which there is no discontinuity in the X^* distribution at the eligibility threshold; 2) X^* is still supported on the set of integers from -10 to 10 but with a discontinuity at the eligibility threshold: $\Pr(X^* = i) = 0.6$

for $i < 0$ and $\Pr(X^* = i) = 0.4$ for $i \geqslant 0$. Figure 2.3 and 2.4 present the observed X and estimated X^* distribution for case 1) and 2) respectively. Note that there is no obvious discontinuity in the observed X distribution at the eligibility threshold in case 2) (top panel of Figure 2.4)–the measurement error has simply smoothed it over. This lack of observed threshold discontinuity illustrates again the problematic nature of using the observed running variable X for RD analyses. In both cases, I test for the threshold discontinuity by fitting a linear minimum distance procedure on the estimated X^* distribution with the restriction

$$\Pr(X^* = x^*) = \gamma_0 + \gamma_{D^*} D^* + \gamma_1 X^* + \gamma_{D^* X^*} D^* X^*$$

Let $\gamma_{D^*}^{1)}$ and $\gamma_{D^*}^{2)}$ be the coefficients of γ_{D^*} in case 1) and 2) respectively, and based on the specifications above, $\gamma_{D^*}^{1)} = 0$ and $\gamma_{D^*}^{2)} = 0.02$. In my simulated samples with 25,000 observations, $\hat{\gamma}_{D^*}^{1)} = 0.006$ with standard error 0.006 and an associated t-statistic of 0.92, and $\hat{\gamma}_{D^*}^{2)} = 0.027$ with standard error 0.008 and an associated t-statistic of 3.47. This exercise demonstrates that the proposed estimation procedure is informative in testing discontinuity in the true running variable distribution for a sample size typical in empirical studies.

2.4 Conclusion

This paper investigates the identification and estimation in the context of a regression discontinuity design where the running variable is measured with error. This is a challenging problem in that presence of the measurement error may smooth out the first stage discontinuity and eliminate the source of identification. The problem has already been encountered in several empirical studies that apply an RD design, but it has not been adequately addressed in the literature. Understanding and solving this problem is important for correctly estimating the program treatment effect and for widening the applicability of the RD design in general.

In this study, I examine the conventional classical measurement error model where the error is assumed to be independent to the true running variable. I focus on the case where the running

variable and the measurement error are discrete and bounded, and propose sufficient conditions to identify the running variable distribution using only its mis-measured counterpart and program eligibility in a sharp RD design. I then extend on the set of assumptions to provide non-parametric identification of both the true running variable distribution and the treatment effect parameter in the more general fuzzy RD design.

Based on derivations in the identification section, a simple estimation procedure is proposed in a minimum distance framework. Following standard arguments, the resulting estimators are $\sqrt{N}$-consistent, asymptotically normal and efficient. Monte Carlo simulations verify that the true running variable distribution and the RD treatment effect parameter can indeed be recovered using the proposed method. The procedure produces informative RD treatment effect estimates and is able to detect discontinuity in the true running variable distribution for sample sizes typical in the relevant literature.

In ongoing research, I am investigating alternatives to the discrete and bounded running variable assumption. As it stands, the identification and estimation depend heavily on the tail behavior of the observed distributions and estimates may be sensitive in empirical applications. One alternative of focus is the case with continuous running variable and normal measurement error. Although identification for this case is charted out in the Appendix, an estimation procedure remains to be developed.[12] There will also be an empirical component in an expanded version of the essay, which will illustrate the methodology by examining the effect of a means-tested transfer program (e.g. crowd-out effect of public health insurance).

[12]The empirical relevance of this setup is questionable due to the slow rate of convergence in similar problems (Butucea and Matias (2005)).

Appendix

Identification of the measurement error distribution in Lemma 2.

The m_l's are identified after the p_k^1's and the p_k^0's are identified because they solve the following linear system:

$$\underbrace{\begin{bmatrix} p^1_{U_{X^*}} & 0 & \cdots & 0 \\ p^1_{U_{X^*}-1} & p^1_{U_{X^*}} & \cdots & 0 \\ \vdots & p^1_{U_{X^*}-1} & \cdots & \vdots \\ p^1_0 & \vdots & \cdots & 0 \\ 0 & p^1_0 & \cdots & p^1_{U_{X^*}} \\ \vdots & 0 & \cdots & p^1_{U_{X^*}-1} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & p^1_0 \\ p^0_{-1} & 0 & \cdots & 0 \\ p^0_{-2} & p^0_{-1} & \cdots & \vdots \\ \vdots & p^0_{-2} & \cdots & \vdots \\ p^0_{L_{X^*}} & \vdots & \cdots & 0 \\ 0 & p^0_{L_{X^*}} & \cdots & p^0_{-1} \\ \vdots & 0 & \cdots & p^0_{-2} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & p^0_{L_{X^*}} \end{bmatrix}}_{(K_{X^*}+2K_u-2)\times K_u} \underbrace{\begin{bmatrix} m_{U_u} \\ m_{U_u-1} \\ \vdots \\ m_0 \\ \vdots \\ m_{L_u+1} \\ m_{L_u} \end{bmatrix}}_{K_u\times 1} = \underbrace{\begin{bmatrix} q^1_{U_{X^*}+U_u} \\ q^1_{U_{X^*}+U_u-1} \\ \vdots \\ q^1_{L_u} \\ q^0_{U_u-1} \\ q^0_{U_u-2} \\ \vdots \\ q^0_{L_u+L_{X^*}} \end{bmatrix}}_{(K_{X^*}+2K_u-2)\times 1} \tag{2.22}$$

Denote system (2.22) with the compact notation $\mathbf{Pm} = \mathbf{q}$, where $\mathbf{P}$ is the $(K_{X^*} + 2K_u - 2) \times K_u$ matrix containing the already known p_k^1 and p_k^0's, $\mathbf{m}$ is the $K_u \times 1$ vector containing the m_l's, and $\mathbf{q}$ is the $(K_{X^*} + 2K_u - 2) \times 1$ vector containing the constant q_i^1's and q_i^0's. The fact $r^1, r^0 > 0$ implies that $K_{X^*} \geq 2$, and $K_u \geq 1$ by construction. Together, they imply

that $K_{X^*} + 2K_u - 2 > K_u$, which means that there are more rows than columns in $\mathbf{P}$. Because $P_k^1 > 0$ for some k, the columns in $\mathbf{P}$ are linearly independent. Therefore, any solution that solves (2.22) is unique, and the parameters m_l's are consequently identified by solving (2.22).

Example Documenting a Non-Identified Case in Subsection 2.2.2

Let $\text{support}_{X^*} = \{-3, -2, -1, 0, 1, 2\}$, the vectors of probability masses $(p^1_{-3}, p^1_{-2}, p^1_{-1}) = (p^0_0, p^0_1, p^0_2) = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$ and $r^1 = \frac{1}{2}$. Let $\text{support}_u = \{-1, 0, 1\}$; and $(m_{-1}, m_0, m_1) = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$. It follows that the observed vectors of probabilities are $(q^1_{-4}, q^1_{-3}, q^1_{-2}, q^1_{-1}, q^1_0) = (q^0_{-1}, q^0_0, q^0_1, q^0_2, q^0_3) = (\frac{1}{8}, \frac{3}{16}, \frac{3}{8}, \frac{3}{16}, \frac{1}{8})$, and the resulting 9×6 matrix

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{8} & 0 & 0 & -\frac{1}{8} & 0 & 0 \\ \frac{3}{16} & \frac{1}{8} & 0 & -\frac{3}{16} & -\frac{1}{8} & 0 \\ \frac{3}{8} & \frac{3}{16} & \frac{1}{8} & -\frac{3}{8} & -\frac{3}{16} & -\frac{1}{8} \\ \frac{3}{16} & \frac{3}{8} & \frac{3}{16} & -\frac{3}{16} & -\frac{3}{8} & -\frac{3}{16} \\ \frac{1}{8} & \frac{3}{16} & \frac{3}{8} & -\frac{1}{8} & -\frac{3}{16} & -\frac{3}{8} \\ 0 & \frac{1}{8} & \frac{3}{16} & 0 & -\frac{1}{8} & -\frac{3}{16} \\ 0 & 0 & \frac{1}{8} & 0 & 0 & -\frac{1}{8} \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

is only of rank 4. The result of non-identification is intuitive because we can "switch" the p and m vectors and the alternative distributions $(\tilde{p}^1_{-3}, \tilde{p}^1_{-2}, \tilde{p}^1_{-1}) = (\tilde{p}^0_0, \tilde{p}^0_1, \tilde{p}^0_2) = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ and $(\tilde{m}_{-1}, \tilde{m}_0, \tilde{m}_1) = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$ give rise to the same distributions of $X|D^* = 1$ and $X|D^* = 0$ as $(p^1_{-3}, p^1_{-2}, p^1_{-1})$, (p^0_0, p^0_1, p^0_2) and (m_{-1}, m_0, m_1).

Identifiability When the Running Variable and Measurement Error Are Continuously Distributed

In this subsection, I first consider the identifiability of Model (2.2) when X^* and u are continuously distributed and focus on the case where they have unbounded support, as is typical in the measurement error literature. While the fully non-parametric identifiability of X^* and u is still being investigated discuss the semiparametric case in this subsection, where the continuous distribution of X^* is not restricted to a particular functional form but u follows a normal distribution with mean 0 and an unknown variance σ^2:

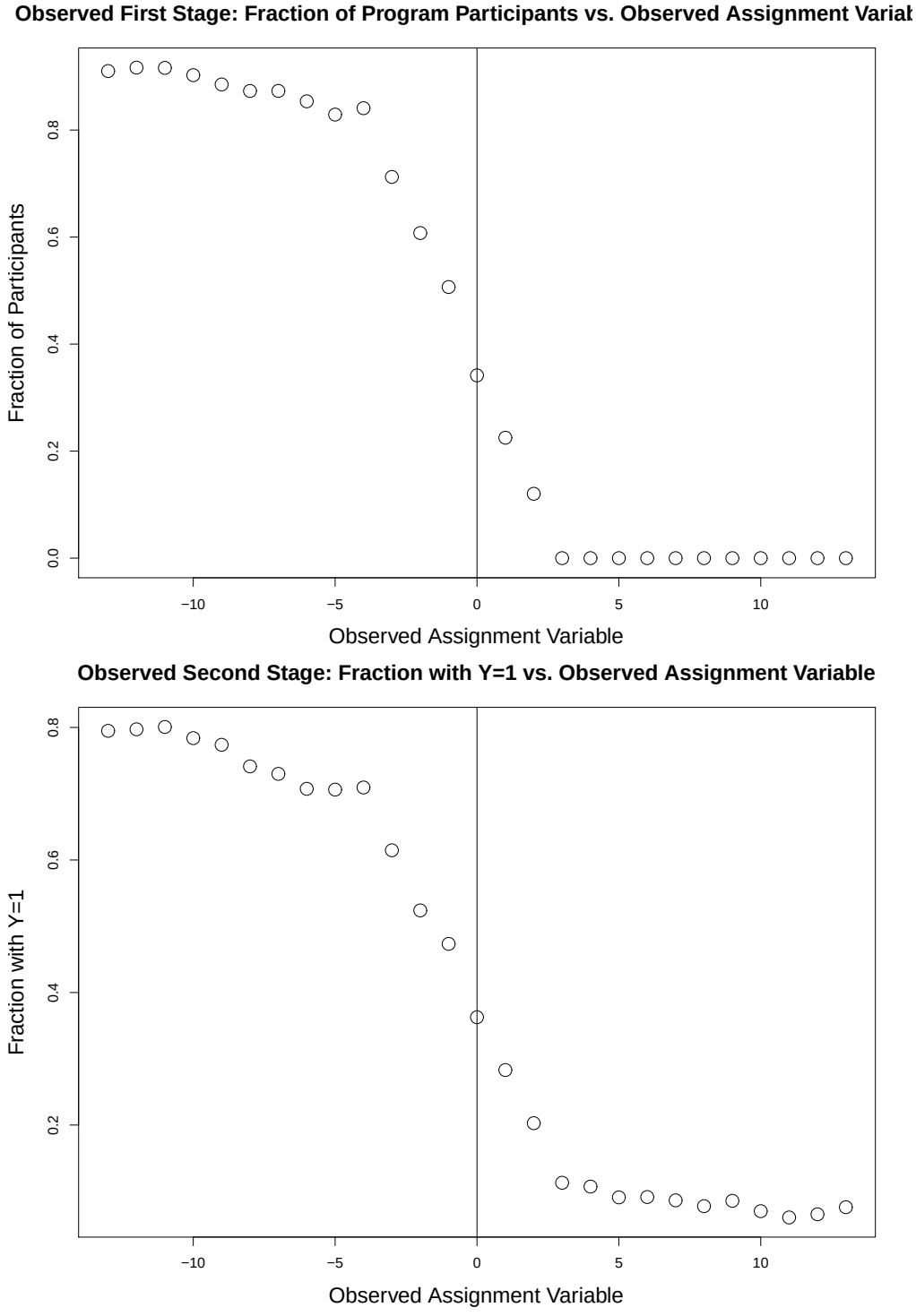
Assumption 7 (Normality). $u \sim \phi(0, \sigma^2)$

The identification of σ and the distribution of X^* from the joint distribution of (X, D^*) follows from a recent study, Schwarz and Bellegem (2010). A rigorous proof can be found in that paper, where the result relies on the fact that the normal distribution is supported on the entire real line. I present an intuitive sketch of the proof. Suppose f_1 and f_2 are the two candidate distributions for $X^*|D^* = 1$ and σ_1 and σ_2 where $\sigma_1 < \sigma_2$ are the two candidates for σ. Let (f_1, σ_1) and (f_2, σ_2) be observationally equivalent, i.e. $f_1 * \phi(0, \sigma_1^2) = f_2 * \phi(0, \sigma_2^2) = g$, where g is the density of $X|D^* = 1$. Then it follows from properties of characteristic functions that $f_1 = f_2 * \phi(0, \sigma_2^2 - \sigma_1^2)$. A contradiction arises because f_1 is only supported on the negative part of the real line but $f_2 * \phi(0, \sigma_2^2 - \sigma_1^2)$ is supported on the entire real line. Hence, $\sigma_1 = \sigma_2$ and the continuous density of $X^*|D^* = 1$ on $x^* \in (-\infty, 0)$ is identified by the one-to-one correspondence between characteristic function and probability density $f_{X^*|D^*=1}(x^*) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{\varphi_{X|D^*=1}(t)}{\varphi_{\phi(0,\sigma^2)}(t)} dt$ (note the characteristic function of $\phi(0, \sigma^2)$ which appears in the denominator of the integrand, $\varphi_{\phi(0,\sigma^2)}(t) = e^{-\frac{1}{2}\sigma^2 t^2}$, is non-zero everywhere). The distribution of $X^*|D^* = 0$ is identified analogously.

Identification under imperfect compliance, i.e. when only benefit receipt D is observed, is similar in spirit to Section 2.2.3, where the result relies on Assumption 1F, 2F, 3 and 4. Here, I appeal to Assumption 1F, 4 and 7 where the strong normality assumption encapsulates symmetry

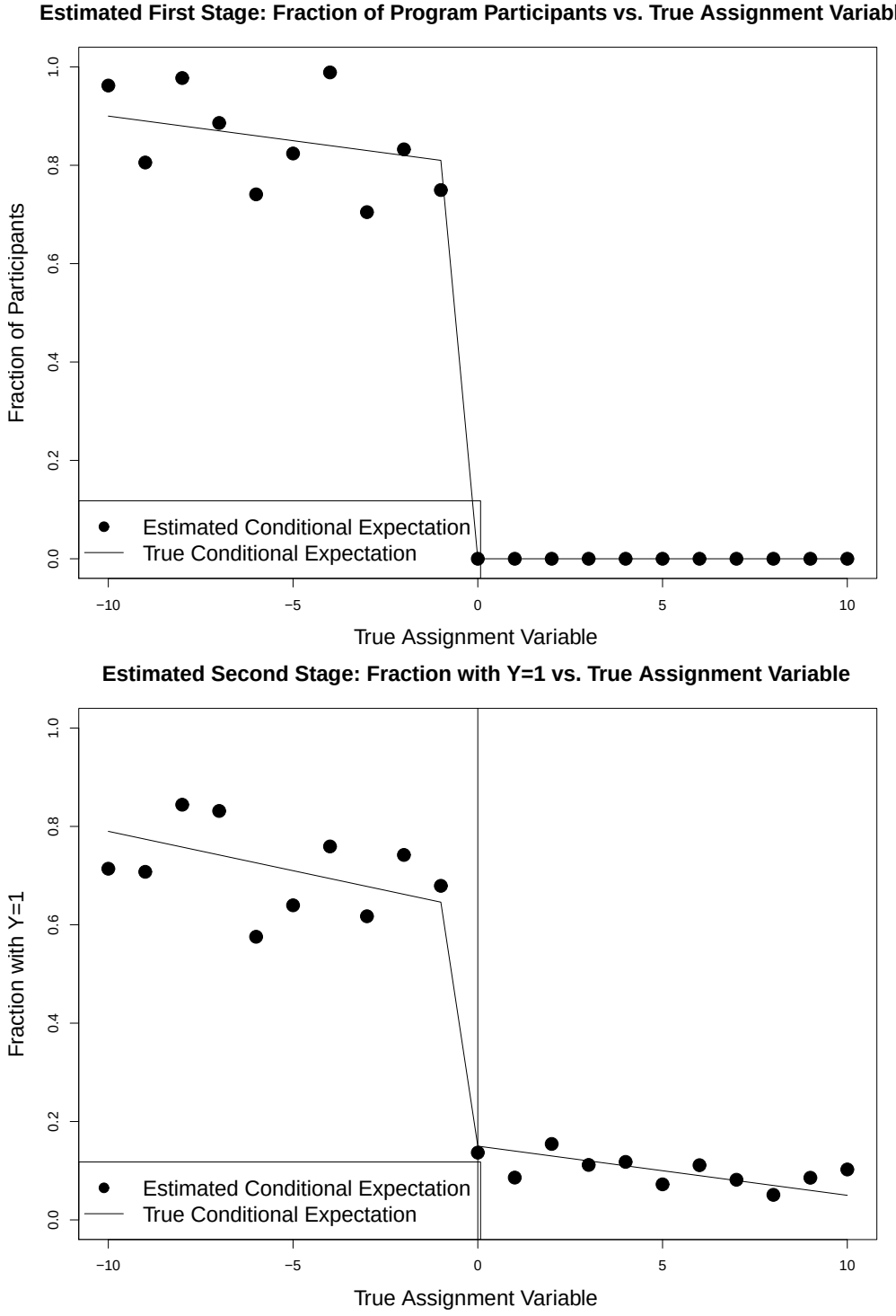
(a stronger version of Assumption 5) and renders the continuous analogue of Assumption 2 unnecessary. $f_{X^*|D=1}$ and σ are identified the same way $f_{X^*|D^*=1}$ and σ are in the previous subsection. Then $f_{X^*|D^*=0}$ is identified from $f_{X^*|D^*=0} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{\varphi_{X|D^*=0}(t)}{\varphi_{\phi(0,\sigma^2)}(t)}dt$ after the identification of σ. It is worth noting that the one-sided fuzzy assumption is no longer needed if the researcher knows what σ is. This is parallel to the discussion in 2.2.3 concerning Assumption 4's redundancy when the support of the measurement error is known. Finally, the identification of the RD treatment effect is obtained with Assumption 1FY.

Figure 2.1: Observed First and Second Stage: Expectation of D and Y Conditional on the Noisy Running Variable X



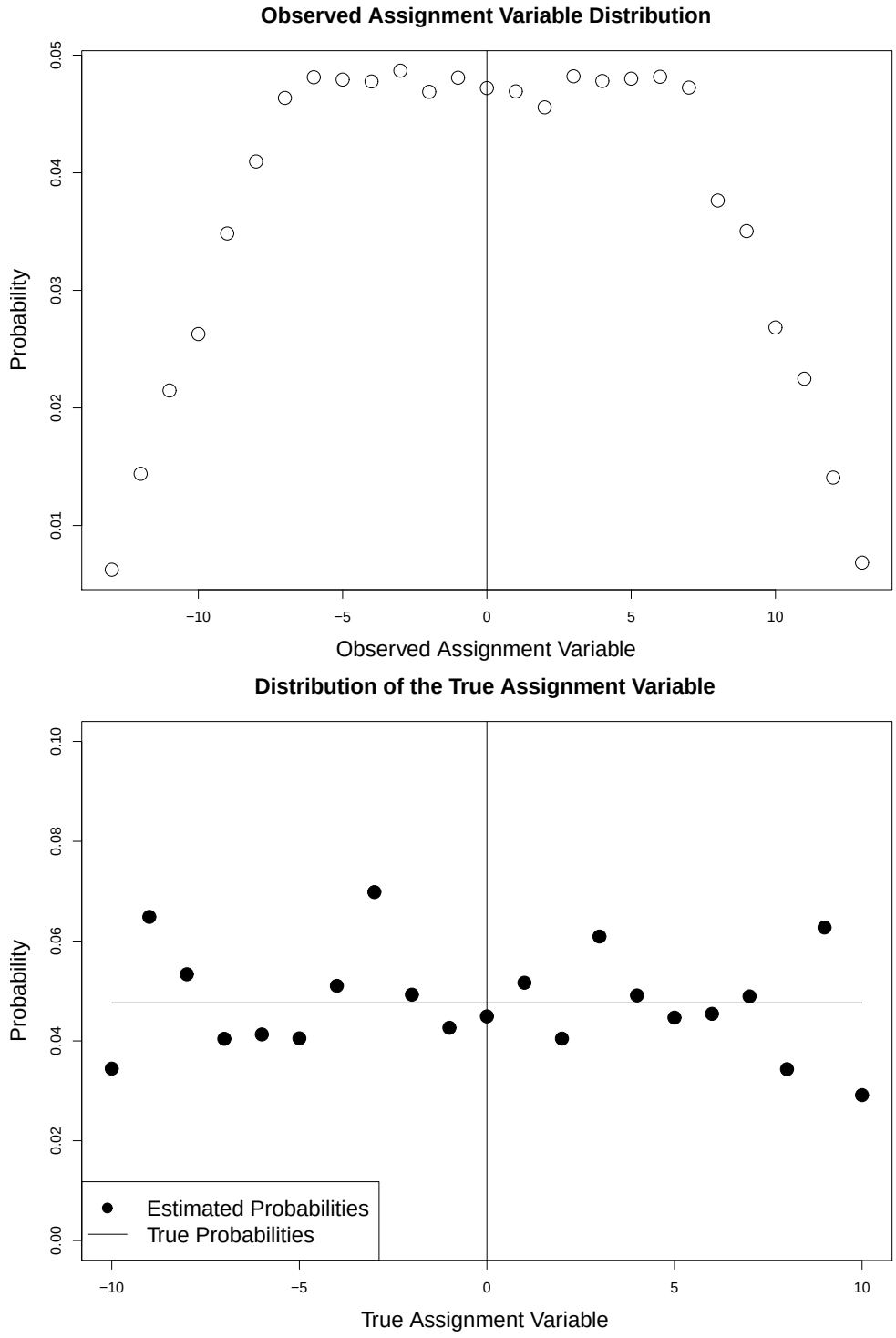
Notes: Simulations are based on a sample of size $N = 25,000$. X^* and u are uniformly distributed on the set of integers in $[-10, 10]$ and $[-3, 3]$, respectively. The true first stage and outcome functions are $E[D|X^*] = (-0.01X^* + 0.8)D^*$ and $E[Y|X^*, D] = 0.15 + 0.6D - 0.01X^*$, which imply a true second stage of $E[Y|X^*] = 0.15 + 0.48D^* - 0.01X^* - 0.006D^*X^*$. Plotted are $E[D|X]$ and $E[Y|X]$ respectively where $X = X^* + u$.

Figure 2.2: Estimated First and Second Stage: Expectation of D and Y Conditional on the True Running Variable X^*



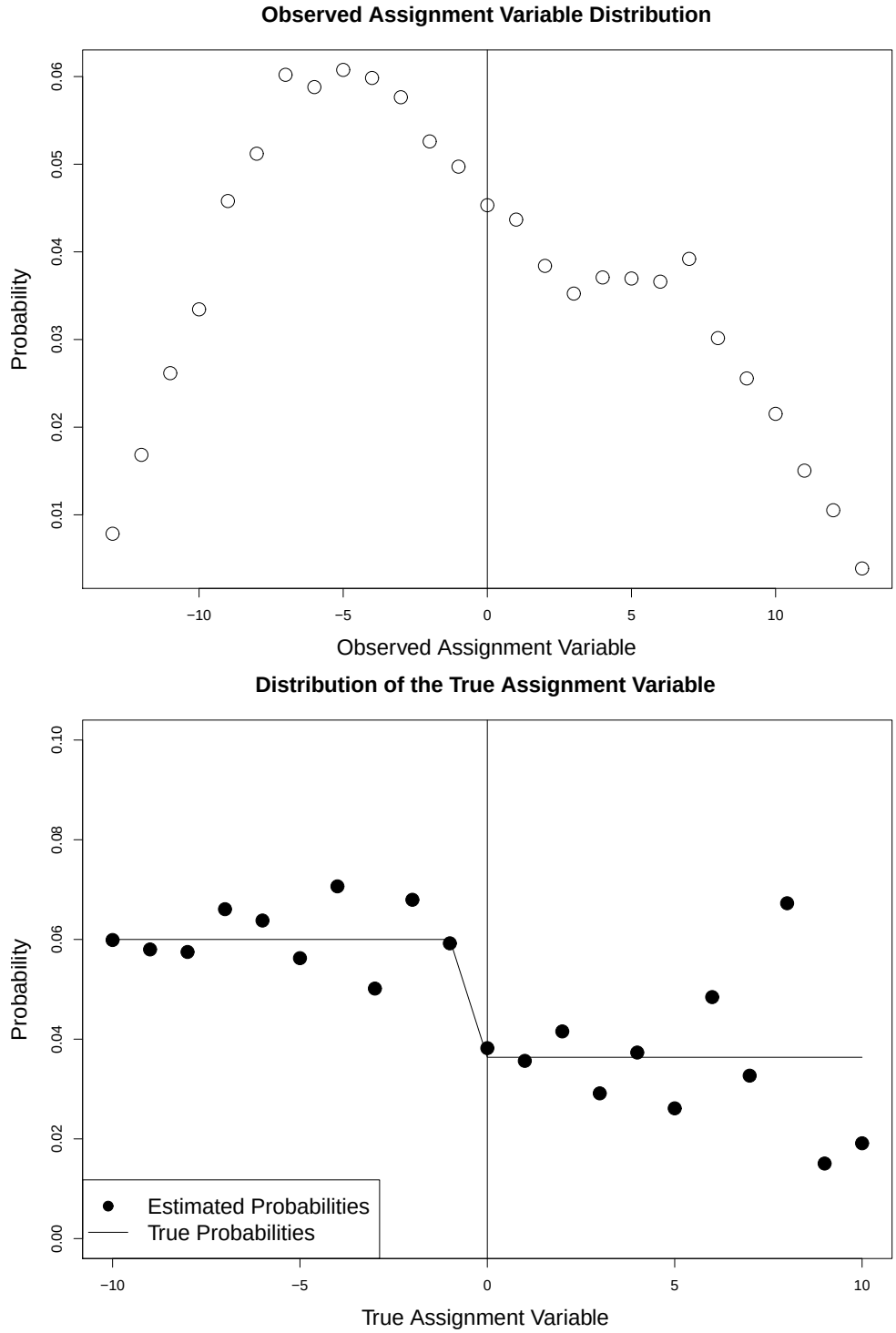
Notes: Simulations are based on a sample of size $N = 25,000$. X^* and u are uniformly distributed on the set of integers in $[-10, 10]$ and $[-3, 3]$. The true first stage and outcome response functions are $E[D|X^*] = (-0.01X^* + 0.8)D^*$ and $E[Y|X^*, D] = 0.15 + 0.6D - 0.01X^*$, respectively, which imply a true second stage of $E[Y|X^*] = 0.15 + 0.48D^* - 0.01X^* - 0.006D^*X^*$. Plotted are the estimated $E[Y|X^*]$ and $E[D|X^*]$ following subsection 2.3.1 with the true conditional expectations specified.

Figure 2.3: Running Variable Distribution with Uniform X^* Distribution: Observed vs. Estimated



Notes: Simulations are based on a sample of size $N = 25,000$. X^* and u are uniformly distributed on the set of integers in $[-10, 10]$ and $[-3, 3]$, respectively. The true first stage and outcom functions are $E[D|X^*] = (-0.01X^* + 0.8)D^*$ and $E[Y|X^*, D] = 0.15 + 0.6D - 0.01X^*$, respectively, which imply a true second stage of $E[Y|X^*] = 0.15 + 0.48D^* - 0.01X^* - 0.006D^*X^*$. Plotted are the distributions of X and X^*, with the latter against the true uniform distribution specified.

Figure 2.4: Running Variable Distribution when True X^* Distribution is Not Smooth: Observed vs. Estimated



Notes: Simulations are based on a sample of size $N = 25,000$. X^* is supported on the set of integers from -10 to 10 with$\Pr(X^* = i) = 0.6$ for $i < 0$ and $\Pr(X^* = i) = 0.4$ for $i \geqslant 0$. Other specifications are those underlying Figures 2.2, 2.2 and 2.3. Plotted are the distributions of X and X^*, with the latter against the true distribution specified above.

Chapter 3

Quasi-Experimental Identification and Estimation in the Regression Kink Design[1]

3.1 Introduction

Reflecting the widespread concern over the potential fragility of parametric methods, a growing literature considers the identification and estimation of nonparametric regression models with endogenous regressors (e.g., Blundell and Powell (2003); Chesher (2003); Altonji and Matzkin (2005); Florens et al. (2009); Imbens and Newey (2009)). The most general of these models allow both the observed regressors and the unobserved errors to enter an underlying structural function in an arbitrary way. Building on methods for additive models (e.g., Heckman and Robb (1985)) most recent studies follow a control function approach. Conditioning on a suitable control function, the regressor of interest is independent of the unobserved errors, and a variety of non-parametric methods can be used to estimate the associated causal effects. The control function approach relies on the existence of one or more "instruments"– variables that are assumed to be independent of the errors in the regression function. Identification hinges on the validity of the independence assumption, in much the same way that identification in a linear simultaneous equations model depends on orthogonality between the instrumental variables and the additive structural error terms.

[1]This paper has been presented by David S. Lee at the 2010 NBER Summer Institute in Labor Studies.

In some applied contexts, however, it is difficult to find candidate instruments that satisfy the necessary independence assumptions. The problem is particularly acute when the regressor of interest is a policy variable that is mechanically determined by a behaviorally endogenous assignment variable. The level of unemployment benefits, for example, is typically set by a formula that depends on previous earnings. In such settings it is arguably impossible to identify individual characteristics that affect the level of the policy variable and yet are independent of underlying heterogeneity in preferences and/or opportunities. Nevertheless, a common feature of many policy rules is the existence of a "kink", or series of kinks, in the formula that relates the assignment variable to the policy variable. In the case of unemployment benefits, for example, a typical formula provides a fixed fraction of pre-job-loss earnings, subject to a maximum rate. Likewise, the income tax system in most countries is piece-wise linear, with progressively higher tax rates at each kink point. As has been noted in recent studies by Guryan (2003), Nielsen et al. (forthcoming), and Simonsen et al. (2009), the existence of a kinked policy rule holds out the possibility for identification of the effect of the policy variable, even in the absence of traditional instruments. In essence, the idea is to look for an induced kink in the outcome variable that coincides with the kink in the policy rule, and relate the relative magnitudes of the two kinks. While this "regression kink design" (RKD) is potentially attractive, an important concern is the endogeneity of the assignment variable. As noted by Saez (forthcoming), for example, a kink in the marginal tax schedule would be expected to lead to "bunching" of taxpayers at the level of income associated with the kink. Such endogenous sorting could lead to a non-smooth distribution of unobserved heterogeneity around the kink-point, confounding inferences based on a regression kink design.

This paper establishes the conditions under which the behavioral response to a formulaic policy variable like unemployment benefits or the marginal tax rate can be identified within a general class of nonparametric regression models. We show that, in the context of a fully nonparametric regression with non-additive errors, the regression kink design can identify what Altonji and Matzkin (2005) have called the "local average response" to the policy variable, or equivalently the

"treatment on the treated" parameter characterized by Florens et al. (2009). The key condition for identification is that conditional on the unobservables, the density of the assignment variable is smooth – continuously differentiable – at the kink-point in the policy rule. We show that this "smooth density" condition rules out extreme forms of endogenous sorting, which might arise when agents can deterministically manipulate the value of the assignment variable used in the policy formula, while allowing for many other forms of endogeneity in the assignment variable. We also show that the smooth density condition generates strong predictions for the distribution of predetermined covariates among the population of agents located near the kink point. In particular, the conditional distribution functions of observed covariates that are determined prior to the policy variable should have continuous derivatives with respect to the assignment variable at the kink point. Thus, we show that, as in a regression discontinuity design (see Lee and Lemieux (forthcoming)), the validity of the regression kink design can be tested.

Using administrative data from the Unemployment Insurance (UI) system in the state of Washington in 1988, we apply a regression kink approach to estimate the average impact of a marginal increase in weekly UI benefits on total benefits paid, and the duration of UI receipt. We find little evidence that the density of base period earnings (the assignment variable) is discontinuous at the threshold associated with the maximum benefit rate; we also find that the means of key covariates generally have smooth derivatives with respect to the assignment variable, consistent with a valid RK Design. Our estimates suggest that a $1 increase in the benefit amount leads to a 0.04 week increase in the duration of UI receipt and a $18 rise in total benefits paid. These numbers translate into a 1.6 week increase in the duration of insured unemployment in response to a 10 percentage point increase in the UI replacement rate – an estimate that is similar in magnitude to the estimate by Meyer (1990), but somewhat larger in magnitude than estimates reported by Hamermesh (1977) or Moffitt and Nicholson (1982).

The paper is organized as follows. Section 3.2 discusses parameters of interest and identification in the Regression Kink Design. Section 3.3 then describes the institutional details of the UI system in the state of Washington, as background for our empirical analysis, which we

present in Section 3.4. Section 3.5 concludes.

3.2 Nonparametric Regression and the Regression Kink Design

3.2.1 Background

We begin with some background on the existing literature. Consider the model

$$Y = y(B, V, W) \tag{3.1}$$

where Y is an outcome, B is a continuous regressor of interest, V is another covariate that enters the model, and W is an unobservable, non-additive error term. This is a particular case of the model considered by Imbens and Newey (2009); there are two observable covariates and interest centers on the effect of B on Y.[2] As is understood in the literature, this formulation allows for completely unrestricted heterogeneity in the responsiveness of Y to B. In the case where B is binary, the model is equivalent to a potential outcomes framework where the "treatment effect" of B for a particular individual is given by $Y_1 - Y_0 = y(1, V, W) - y(0, V, W)$.

One natural benchmark object of interest in this setting is the "average structural function" (ASF), as discussed in Blundell and Powell (2003):

$$ASF(b, v) = \int y(b, v, w)\, dG(w),$$

where $G(\cdot)$ is the c.d.f. of W. This gives the average value of Y that would occur if the entire population (as represented by the unconditional distribution of W) was assigned to a particular value of the pair (b, v). Florens et al. (2009) call the derivative of the ASF with respect to the

[2] In Imbens and Newey (2009), W is considered to have unknown dimension. We, too, can allow for that generality.

continuous treatment of interest the "average treatment effect" (ATE), which is a natural extension of the average treatment effect familiar in the binary treatment context.

A closely related construct is the "treatment on the treated" (TT) parameter of Florens et al. (2009):

$$TT_b(b,v) = \int \frac{\partial y(b,v,w)}{\partial b} dG(w|b,v).$$

As noted by Florens et al. this is equivalent to the "local average response" (LAR) parameter of Altonji and Matzkin (2005). The TT (or equivalently the LAR) gives the average effect of a marginal increase in b at some specific value of the pair (b,v), holding fixed the distribution of the unobservables at $G(\cdot|b,v)$.

Recent studies, including Florens et al. (2009) and Imbens and Newey (2009) have proposed methods that use an instrumental variable Z to identify causal parameters such as TT or LAR. An appropriate instrument Z is assumed to influence B, but is, at the same time, independent of the non-additive errors in the model. Chesher (2003) observes that such independence assumptions may be "strong and unpalatable", and hence considers using local independence of Z to identify local effects.

As mentioned in the introduction, there are some important contexts – particularly when the regressor of interest is a policy variable that is a deterministic function of a behaviorally endogenous variable – where no instruments can plausibly satisfy the independence assumption, either globally or locally. In the framework of equation (1), consider the case where B represents the level of unemployment benefits available to a newly unemployed worker, Y represents the duration of unemployment, and V represents pre-job-loss earnings. Assume (as in many institutional settings) that unemployment benefits are computed as a fixed fraction of V up to some maximum weekly benefit. Conditional on V there is no variation in the benefit level, so model (1) is not non-parametrically identified. One could try to get around this fundamental non-identification by treating V as an error component that is correlated with B. But in this case, any variable that is independent of V will, by construction, be independent of the regressor of interest B, so it will not be possible to find instruments for B, holding constant the policy regime.

Despite this circumstance, it may still be possible to exploit the kinked benefit rule to identify a causal effect of B on Y, in a similar spirit to the regression discontinuity design of Thistlethwaite and Campbell (1960). The idea is that if B exerts a causal effect on Y, and there is a kink in the deterministic relation between B and V at $v = v_0$ (the lowest level of earnings at which the individual receives the maximum benefit rate) then we should expect to see an induced kink in the relationship between Y and V at $v = v_0$. This identification strategy has been employed in a few empirical studies. Guryan (2001), for example, uses kinks in state education aid formulas as part of an instrumental variables strategy to study the effect of public school spending.[3] More recently, Simonsen et al. (2009) use a kinked relationship between the total expenditure on prescription drugs and the marginal price to study the price sensitivity of demand for prescription drugs. Nielsen et al. (2009), who introduce the term "Regression Kink Design" for this approach, use a kinked student aid scheme to identify the effect of direct costs on college enrollment. Nielsen et al. (2009) make precise the assumptions needed to identify the causal effects in the additive model

$$Y = \tau B + g(V) + \epsilon,$$

where $B = b(V)$ is a deterministic (and continuous) function of V with a kink at $v = 0$. Nielsen et al. (2009) show that if $g(\cdot)$ and $E[\epsilon|V = v]$ have derivatives that are continuous in v at $v = 0$, then

$$\tau = \frac{\lim_{v \to 0^+} \frac{\partial E[Y|V=v]}{\partial v} - \lim_{v \to 0^-} \frac{\partial E[Y|V=v]}{\partial v}}{\lim_{v \to 0^+} \frac{\partial b(v)}{\partial v} - \lim_{v \to 0^-} \frac{\partial b(v)}{\partial v}}.$$

The expression on the right hand side of this equation – the RKD estimand – is simply the change in slope of the conditional expectation function $E[Y|V = v]$ at the kink point ($v = 0$), divided by the change in the slope of the deterministic assignment function $b(\cdot)$ at 0.[4]

[3] Guryan (2003) describes the identification strategy as follows: "In the case of the Overburden Aid formula, the regression includes controls for the valuation ratio, 1989 per-capita income, and the difference between the gross standard and 1993 education expenditures (the standard of effort gap). Because these are the only variables on which Overburden Aid is based, the exclusion restriction only requires that the functional form of the direct relationship between test scores and any of these variables is not the same as the functional form in the Overburden Aid formula."

[4] In an earlier working paper version, Nielsen et al. (2008) provide similar conditions for identification for a less restrictive, additive model, $Y = g(B, V) + \epsilon$.

3.2.2 Necessary and Sufficient Conditions for RKD in a Non-Separable Model

As background for our main results we first specify the necessary and sufficient conditions for an RKD to identify marginal effects in a general nonseparable model, in the same way Hahn et al. (2001) formally established the identifying assumptions for Thistlethwaite and Campbell (1960)'s RD design in the heterogeneous treatment effects, potential outcomes framework.

Proposition 1. *For the nonseparable model (3.1),* $B = b(V)$ *, where* $b(\cdot)$ *is continuous but has discontinuous derivative at* 0*, let* $g(w|v)$ *be the density of* w *conditional on v. If* $\forall w$ *i)* $\frac{\partial y(b,v,w)}{\partial b}$ *is continuous in* v *at* $v = 0$*, ii)* $\frac{\partial y(b,v,w)}{\partial v}$ *is continuous in* v *except possibly at* $v = 0$*, iii)* $\frac{\partial g(w|v)}{\partial v}$ *is continuous in* v *except possibly at* $v = 0$*, then a necessary and sufficient condition for*

$$\frac{\lim_{v\to 0^+} \frac{\partial E[Y|V=v]}{\partial v} - \lim_{v\to 0^-} \frac{\partial E[Y|V=v]}{\partial v}}{\lim_{v\to 0^+} b'(v) - \lim_{v\to 0^-} b'(v)} = \int y_1(b(0),0,w)\, g(w|0)\, dw \equiv TT_b(b(0),0)$$

is that $\left.\frac{\partial E[y(b,V,W)|V=v]}{\partial v}\right|_{b=b(v)}$ *is continuous in* v *at* $v = 0$

To see this, consider the partial derivative of the conditional expectation function $E[Y|V=v] \equiv E[y(B,V,W)|V=v]$ at $v \neq 0$,

$$\begin{aligned}
\frac{\partial E[y(b(V),V,W)|V=v]}{\partial v} &= \int \frac{\partial (y(b(v),v,w)\, g(w|v))}{\partial v} dw \\
&= \int \left(y_1(b(v),v,w)\, b'(v) + y_2(b(v),v,w)\right) g(w|v) + y(b(v),v,w) \frac{\partial g(w|v)}{\partial v} dw \\
&= E[y_1(b(V),V,W)|V=v]\, b'(v) + \int y_2(b(v),v,w)\, g(w|v) + y(b(v),v,w) \frac{\partial g(w|v)}{\partial v} dw \\
&= E[y_1(b(V),V,W)|V=v]\, b'(v) + \left.\frac{\partial E[y(b,V,W)|V=v]}{\partial v}\right|_{b=b(v)}
\end{aligned}$$

By taking the difference between the right and left limits of this expression, we obtain the result that

$$\frac{\lim_{v\to 0^+} \frac{\partial E[y(b(V),V,W)|V=v]}{\partial v} - \lim_{v\to 0^-} \frac{\partial E[y(b(V),V,W)|V=v]}{\partial v}}{\lim_{v\to 0^+} b'(v) - \lim_{v\to 0^-} b'(v)} = E[y_1(b(0),0,W)|V=0]$$

if and only if $\left.\frac{\partial E[y(b,V,W)|V=v]}{\partial v}\right|_{b=b(v)}$ is continuous in v at $v = 0$.[5]

Intuitively, a marginal increase in V induces an effect on Y through b, but also via the functional dependence of $y(B, V, W)$ on V, and via the changing distribution of unobserved heterogeneity (reflected in $g(w|v)$). Only when the latter two effects evolve smoothly as V reaches the kink point will the change in the derivative $\frac{\partial E[y(b(V),V,W)|V=v]}{\partial v}$ isolate the causal effect of b on Y. Formally, condition for the smoothness of $\left.\frac{\partial E[y(b,V,W)|V=v]}{\partial v}\right|_{b=b(v)}$ can be seen in the equation

$$\left.\frac{\partial E\left[y\left(b,V,W\right)|V=v\right]}{\partial v}\right|_{b=b(v)} = \int y_2\left(b\left(v\right),v,w\right)g\left(w|v\right) + y\left(b\left(v\right),v,w\right)\frac{\partial g\left(w|v\right)}{\partial v}dw.$$

It is clear that continuity of $y_2\left(b\left(v\right),v,w\right)$ and $\frac{\partial g(w|v)}{\partial v}$ in v at $v = 0$ will satisfy the condition for identification. That is, it is sufficient that the "direct" marginal impact of v on Y is continuous in v, and that the conditional density of the unobservables W change smoothly with respect to V.

While the assumption that the derivative of the density $g\left(w|v\right)$ is continuous in v may seem like a mild restriction, it is important to emphasize that in many potential applications of the RKD idea, the assignment variable V is an endogenous behavioral outcome. If agents can strategically select a value of V, then we might be concerned that the density of the unobservables may exhibit non-smooth behavior near the kink point.

Arguably the most important development in recent applications of the regression discontinuity design is the recognition that when the assignment variable is endogenously chosen, inferences from an RD design may be rendered invalid (see e.g., the discussion in Lee and Lemieux (forthcoming), and the theoretical treatments in McCrary (2008) and Lee (2008)). Recent RD analyses (for example, Urquiola and Verhoogen (2009)) therefore devote much attention to the possibility of endogenous sorting on the assignment variable. Nevertheless, Lee (2008) shows that when agents have only imprecise control over the assignment variable, a regression discontinuity design can still deliver valid inferences. Our goal is to illuminate the same set of

[5]Note that i) is required for the marginal effect to be well-defined at the point $b\left(0\right)$, ii) and iii) are required in order to evaluate the integrand $\frac{\partial(y(b(v),v,w)g(w|v))}{\partial v}$.

issues for the regression kink design. Thus, our primary contribution is to characterize a broad class of models for which the RK design will isolate quasi-experimental variation in the treatment of interest, and – like a randomized experiment or an RD design – allow for empirical tests of the model's validity. As an illustration of the general principles involved, we sketch two different economic models of behavioral responses to unemployment benefits: one that satisfies the conditions for a valid RKD, and one that does not. Finally, we apply these ideas to evaluate the validity of a RKD analysis of the effect of unemployment benefits on the duration of unemployment among job losers in the state of Washington. Given the results of this analysis we go on to use an RKD to estimate the impact of marginal increases in benefit generosity on employment outcomes and UI claim behavior.

3.2.3 Defining Causal Effects in an Ideal Randomized Setting

Since our goal is to characterize a class of models for which the RKD isolates quasi-experimental or "as good as randomized" variation in a regressor of interest, we first briefly review the statistical implications of running an idealized experiment in which the regressor is randomly assigned. This benchmark will serve as a basis for comparison in the sections to follow when we investigate the RK estimator in the context of non-random treatment assignment.

To envision a randomized scenario that corresponds to a regression kink design setting, a continuous treatment variable is called for. This is different from a classical randomized experiment in which agents are assigned to either a treatment group or a control group.

Let (Y, X, B, V) be observable random variables, where Y denotes the outcome of interest, B is the treatment assignment variable, and X (a vector) and V (a scalar) are variables that are determined prior to B, in that order. The random variable V will play the role of the assignment variable in the RK Design described below. We denote the density of B conditional on V, W as $f_{B|V,W}(\cdot|\cdot,\cdot)$ and the marginal density of B as $f_B(\cdot)$.

Definition. Let W characterize the "type" of the individual, with c.d.f. $G(w)$. All individuals with the same value of W are identical. Conditional on W, however, the distribution of V may be

non-degenerate. Let $Y \equiv y(B, V, W)$, $y_1(b, v, w) \equiv \frac{\partial y(b,v,w)}{\partial b}$, and $X \equiv x(W)$, where $y(\cdot, \cdot, \cdot)$ and $x(\cdot)$ are real-valued functions.

As in Lee (2008), there is no loss of generality in assuming that W is one-dimensional. To give a concrete example, W could represent potential earnings capacity, X could be the level of schooling at the time of applying for UI benefits, and V could represent pre-job-loss earnings. Non-degeneracy of V could arise from pure "randomness" in the determination of pre-job-loss earnings.

Note that there is no loss in generality in excluding X from the structural function $y(\cdot, \cdot, \cdot)$. The variables X could be included as a separate argument, but we are not interested in the marginal effects of X. Thus, we consider the "reduced-form" function $y(\cdot, \cdot, \cdot)$, for which the impact of W is defined to include any indirect effects through X. Note that so far, our setup corresponds to a standard unrestricted non-separable model, except that we have made clear that X is determined before V, which is determined before B. This ordering implies that B cannot enter the function $x(\cdot)$.

Condition 2a $y_1(b, v, w)$ exists for all $(b, v, w) \in R^3$, is integrable with respect to $dG(w)$ for all $b \in R$, and is continuous in b for all w.

Condition 2b (Randomized Treatment Assignment) $f_{B|V,W}(b|v, w) = f_B(b)$ for all $(b, v, w) \in R^3$.

This condition describes a simple assignment mechanism that is the continuous analogue of a classical randomized experiment. In particular, each individual faces the same probability of receiving any level of the treatment variable.

In the context of an unemployment benefit example, we have in mind the following experiment. For each individual, regardless of pre-job-loss earnings, we randomize the weekly benefit amount they can receive and then observe the outcome Y (for example, Y could measure the duration of

insured unemployment). A natural function of interest in this setting is $E\left[y_1\left(B,V,W\right)|B=b,V=v\right]$, which represents the average response of Y to a marginal increase in benefits, for a particular pair (b,v). This function could be averaged over the distribution of V (conditional on B) to obtain $E\left[y_1\left(B,V,W\right)|B=b\right]$, which is the average response of Y to a marginal increase in B at a particular value of b .

The key implications of this assignment mechanism are summarized in the following proposition.

Proposition 2. *If Condition 2a and Condition 2b hold, then*

(a) $Pr(W \leqslant w|B=b) = Pr(W \leqslant w)\ \forall w \in R$ *and* $\forall b$ *in the support of B.*

(b) $\frac{\partial E[Y|B=b,V=v]}{\partial b} = E[y_1(b,V,W)|V=v] \equiv ATE_{b|v} = TT_{b|v} = \int y_1(b,v,w)\frac{f_{V|W}(v|w)}{f_V(v)}dG(w)$ $\forall b$ *in the support of B.*

(c) $Pr(X \leqslant x_0|B=b) = Pr(X \leqslant x_0)\ \forall x_0 \in R$ *and* $\forall b$ *in the support of B.*

The proof of Proposition 2 is in the Appendix. Part (a) states the intuitive consequence of randomized experiment: the distribution of "all other pre-determined factors" is the same, regardless of the level of the treatment B. Part (b) establishes that the partial derivative of $E\left[Y|B=b,V=v\right]$ with respect to b identifies the ATE parameter at b (at $V=v$). $ATE_{b|v}$ is also equivalent to the TT at b, since in this randomized experiment, the distribution of W conditional on B is the same as the unconditional distribution of W, as stated in (a). Part (b) also shows that the average derivative is actually a weighted average across all individuals, where the weights $\frac{f_{V|W}(v|w)}{f_V(v)}$ reflect the relative likelihood that a particular type of individual (identified by a value of w) will have $V=v$. It is also possible to average $ATE_{b|v}$ over the marginal distribution of V to obtain the unconditional ATE of $E[y_1(b,V,W)]$.

Part (c) of the proposition is the most distinctive implication of randomized variation in B. If each agent has the same probability law for B, then the distribution of the pre-determined covariates, X, will be identical for any value of B. Thus, the empirical validity of Conditions 2a and 2b can be tested using baseline covariates. This is analogous to the conventional "test for randomization" that is often employed in randomized controlled trials, whereby the analyst tests

that the distributions of the baseline covariates among the treatment and control groups are statistically similar.

3.2.4 Local Random Assignment From a Regression Kink Design

Identification

We now consider the case where the regressor of interest, B, is mechanically determined as a function of the assignment variable V. We show that if the assignment rule has a kink at $V = 0$ then: (a) RKD can identify the same $TT_{b|v}$ parameter that is identified by the above randomized experiment; and (b) the validity of the RKD can be tested by examining the properties of the conditional distribution of the predetermined covariates $\Pr(X \leq x_0|V = v)$ at $v = 0$.

The random variables are determined in the same sequence as in the experiment: first the pre-determined variables X, then V, and then B. This time, however, B is determined according to the deterministic rule $B = b(V)$. Formally,

Definition. Let (V, W) be a pair of random variables (with W unobservable, V observable), where the distribution of W is given by the c.d.f $G(w)$, and the distribution and density of V conditional on W are given by the c.d.f. $F_{V|W}(v|w)$ and p.d.f. $f_{V|W}(v|w)$. Let $B \equiv b(V)$, let $Y \equiv y(B, V, W)$, and $X \equiv x(W)$. Also, define $y_1(b, v, w) \equiv \frac{\partial y(b,v,w)}{\partial b}$ and $y_2(b, v, w) \equiv \frac{\partial y(b,v,w)}{\partial v}$.

Condition 3a. (Regularity) $y(\cdot, \cdot, \cdot)$ and $x(\cdot)$ are real-valued functions. $y_1(b, v, w)$ is continuous in b and $y_2(b, v, w)$ is continuous in v for all b, v and w.

Relative to the experimental setting, this condition requires that the direct marginal impact of V on Y is smooth.

Condition 3b. (First Stage) $b(\cdot)$ is a known function, everywhere continuous and is continuously differentiable on $(-\infty, 0)$ and $(0, \infty)$, but $\lim_{v\to 0^+} b'(v) \neq \lim_{v\to 0^-} b'(v)$. In addition, $f_{V|W}(0|w)$ is strictly positive for $w \in A$, where $\int_A dG(w) > 0$.

Condition 3c, (Smooth Density) $F_{V|W}(v|w)$ is twice continuously differentiable in v at $v = 0$ for every w. That is,$\frac{\partial f_{V|W}(v|w)}{\partial v}$, the derivative of the conditional probability density function $f_{V|W}(v|w)$ is continuous in v for all w.

This is the key assumption that is required for a valid RK Design. As in Lee (2008), this condition rules out precise manipulation of the assignment variable. But whereas in an RD context continuity of $f_{V|W}(v|w)$ in v is sufficient for valid inferences, in the RK design we need to have a continuous derivative of $f_{V|W}(v|w)$ with respect to v. As we show in the next Section, there are reasonable economic models that predict that the smooth density condition will hold, even though agents have some control over the assignment variable V, and other models where it will not hold. This underscores the need to be able to empirically test the implications of this assumption.

Proposition 3. *If Conditions 3a, 3b and 3c hold, then:*

(a) Pr$(W \leqslant w|V = v)$ *is continuously differentiable in* v *at* $v = 0\ \forall w$.

(b)
$$\frac{\lim\limits_{v\to 0+}\frac{\partial E[Y|V=v]}{\partial v} - \lim\limits_{v\to 0-}\frac{\partial E[Y|V=v]}{\partial v}}{\lim\limits_{v\to 0+}\frac{\partial b(v)}{\partial v} - \lim\limits_{v\to 0-}\frac{\partial b(v)}{\partial v}} = E[y_1(b_0, 0, W)|V = 0] = \int y_1(b_0, 0, w)\frac{f_{V|W}(0|w)}{f_V(0)}dG(w) = TT_{b_0|0}$$
where $b_0 = b(0)$.

(c) Pr$(X \leqslant x_0|V = v)$ *is continuously differentiable in v at* $v = 0\ \forall x_0$.

Proposition 3 is analogous to Proposition 2, and its proof is in the Appendix. Part (a) states that the rate of change in the probability distribution of individual types with respect to the assignment variable V is continuous at $V = 0$.[6] This leads directly to part (b): as a consequence of the smoothness in the underlying distribution of types around the kink, the change in the slope of $E[Y|V = v]$ at $v = 0$ divided by the change in slope in $b(V)$ at the kink point delivers identification of $TT_{b_0|0}$. This is the same parameter identified by the randomized experiment

[6]Note also that (a) implies Proposition 2(a) in Lee (2008), i.e., the continuity of Pr$(W \leqslant w|V = v)$ at $v = 0$ for all w. This is a consequence of the stronger smoothness assumption we have imposed on the conditional distribution of V on W.

above, except that it is evaluated at $B = b_0$ and $V = 0$.[7] The weights in the weighted average interpretation $TT_{b_0|0}$ are also the same as for the experimentally identified $TT_{b|v}$. Note that in contrast to the case of the randomized experiment, $TT_{b|v}$ is in general different from $ATE_{b|v}$, due to the potentially systematic relation between B and W. It should also be clear that Conditions 3a, 3b, and 3c will satisfy the necessary condition of Proposition 1.

Finally, and most importantly, Proposition 3c, which is analogous to Proposition 2c, states that under the required conditions for a valid RKD, any pre-determined variable X should have a c.d.f. that is continuously differentiable (i.e., no kink) with respect to V. It is important to emphasize that this prediction is stronger than the requirement for a valid RD that the distribution of X is continuous with respect to V. In particular, showing that there is a similar distribution of baseline covariates in a neighborhood just to the left and just to the right of $V = 0$ is not enough to verify the prediction of Proposition (3c). Instead, what is needed is evidence on the smoothness of the conditional distribution, based on comparisons of the slope of the conditional expectation function (or the conditional quantile function) of X given v.

Discussion of the Smooth Density Condition–Illustration With a Simple Labor Supply Model

We now use a simple behavioral model of labor supply responses to variation in the unemployment benefit rate to illustrate the substantive content of the smooth density condition (Condition 3c) that is required for valid RK design. We begin with an example where Condition 3c is violated. In this example, sorting is "too extreme" to satisfy the smooth density assumption. Consider a group of workers who are initially employed in a temporary (or seasonal) job. At the beginning of period 1, workers know that the job will end after one period, and that during the

[7] Technically, the TT and LAR parameters do not condition on V. But in the case where there is a one-to-one relationship between B and V, then the trivial integration over the (degenerate) distribution of V conditional on $B = b_0$ will imply that $TT_{b_0|0} = TT_{b_0} \equiv E\left[y_1\left(b_0, V, W\right)|B = b_0\right]$, which is literally the TT and LAR parameters discussed in Florens et al. (2009) and Altonji and Matzkin (2005), respectively. In our application to unemployment benefits, B and V are not one-to-one, since beyond $V = 0$, B is at the maximum benefit level. In this case, TT_b will in general be discontinuous with respect to b at b_0; for $B < b_0$, $TT_b = TT_{b|v}$, but for $B = b_0$, $TT_{b_0} = \int TT_{b_0|v} f_{V|B}\left(v|b_0\right) dv$. In this case, the RKD estimand identifies $\lim_{b \uparrow b_0} TT_b$.

second (and final) period they will receive unemployment insurance benefits that depend on their first-period earnings. In period 1 they earn an exogenous hourly wage w and consume their entire wage income. In period 2, they receive an unemployment benefit b that is based on their first period income (I): $b(I) = \min(\gamma I, \bar{b})$, where $\gamma \in (0, 1)$ represents the replacement rate and $\bar{b}$ represents a maximum benefit rate (both of which are constant). For simplicity we assume that there is no possibility of finding another job in period 2, and that unemployment benefits are the only source of second period income.

Workers can choose how many hours $h \in [0, 1]$ to work in period 1. Workers differ by the relative weight that they assign to consumption versus leisure in their within-period utility functions. We denote the utility function for a worker of type α as u_α(consumption,leisure). We assume that workers have complete knowledge of the mapping $b(\cdot)$ before starting to work in period 1. A worker of type α solves the following problem:

$$\max_h u_\alpha(wh, 1-h) + \beta u_\alpha(b(wh), 1)$$

where $\beta > 0$ is a discount factor.

Consider a Cobb-Douglas utility function:
$u_\alpha(\text{consumption}, \text{leisure}) \equiv \log(\text{consumption}) + \alpha \log(\text{leisure})$. This parametrization leads to the optimal choices:

$$h^* = \begin{cases} \frac{1+\beta}{1+\alpha+\beta} & \text{if } \frac{1+\beta}{1+\alpha+\beta} < k \\ \frac{1}{1+\alpha} & \text{if } \frac{1}{1+\alpha} \geqslant k \\ k & \text{otherwise} \end{cases}$$

where $k = \frac{\bar{b}}{\gamma w}$ is the location of the kink when we plot b (benefits) against h (labor supplied in period 1). The relationship between h^* and α is plotted in Figure 3.1 for the case where $w = 1000$, $\gamma = 0.5$, $\bar{b} = 350$ and $\beta = 0.5$.

As shown in the graph, individuals with $\alpha \in [0.43, 0.64]$ will all choose $h^* = k = 0.7$. Assuming that α is smoothly distributed on (0,1), this means that a discrete mass of workers will

sort to the kink. In general, then, the density of optimal hours, $f_{h^*}(h)$, will have a discontinuous derivative at $h = k$, reflecting the fact that the derivative of the inverse mapping from α to h^* is different from the left and right at $h^* = k$.[8]. Consequently, the derivative of the density of baseline earnings I (the assignment variable) conditional on α (the latent unobservable) will not be the same on the two sides of the kink, violating Condition 3c.

Next we consider a variant of the model in which Condition 3c is satisfied. Suppose that in period 1 workers do not know the precise location of the kink-point k , but instead have a prior represented by the density $f_k(k)$ for $0 < k < 1$, with associated c.d.f. $F_K(k)$. In this case, uncertainty over the location of the kink translates into uncertainty over $\overline{b}$, and the worker only knows that his or her benefit will be $b = \min(\gamma wh, \gamma wk)$ for each potential value of k. Given these prior beliefs the worker maximizes expected utility:

$$\log(wh) + \alpha \log(1-h) + \beta \int (\log(\gamma wh) 1_{[h<k]} + \log(\gamma wk) 1_{[h \geqslant k]}) f_K(k) dk,$$

which can be simplified to:

$$\log(wh) + \alpha \log(1-h) + \beta \{\log(\gamma wh)(1 - F_K(h)) + \int_{-\infty}^{h} \log(\gamma wk) f(k) dk\}.$$

It can be shown that the first order condition for an optimal hours choice in period 1 has the form:

$$1 - h^* - \alpha h^* + \beta(1-h^*)(1 - F_K(h^*)) = 0.$$

The derivative of the optimal hours choice as a function of the worker's type is:

$$\frac{dh^*}{d\alpha} = -\frac{h^*}{(1+\alpha) + \beta(1 - F_K(h^*)) + \beta(1-h^*) f_K(h^*)}.$$

The denominator of this expression is strictly positive, and $\frac{dh^*}{d\alpha}$ is continuous in α at every α and

[8] It is $(1+\beta)\frac{1-c}{c}$ from the left and $\frac{1-c}{c}$ from the right.

h^* that satisfy the first order condition. Assuming that α is smoothly distributed, h^* will then be smoothly distributed, as will baseline earnings $I = wh^*$. Thus, under the assumption that agents have a smooth prior on the location of the kink, Condition 3c will hold. The lack of information among workers rules out the extreme form of sorting in the first example and ensures that there is a smooth mapping from the underlying heterogeneity to the assignment variable.

3.3 Unemployment Insurance in Washington State: Background and Data

3.3.1 The Washington Reemployment Bonus Experiment

In this section we use a regression kink approach to estimate the marginal effect of unemployment benefits on claimant behavior for a sample of individuals in the Washington Reemployment Bonus Experiment (WREB). The WREB was a randomized experiment conducted during 1988 to study the responses of unemployment insurance (UI) claimants to alternative incentive schemes. Specifically, the experiment provided different lump sum bonus amounts to claimants who found a job within alternative time-limits. A total of 15,534 claimants participated in the experiment, of whom 12,452 were assigned to one of the six treatment groups (3 different bonus amounts $\times$ 2 different time-limits) and 3,082 were assigned to the control group. All participants were subject to the same (standard) provisions of the Washington UI system for determining their benefit amounts. To the extent that the bonus provisions may have interacted with the effects of the benefit level, our analysis – which ignores the bonus aspect of the experiment – provides estimates of the marginal impact of higher UI benefit levels averaged across the 7 experimental regimes (6 treatment regimes and 1 control regime).

3.3.2 UI Institutions in Washington State

Benefit Determination Rules

In Washington, as in other US states, UI entitlement is based on a formula that depends on labor market activities in the period before the start of the claim. The Washington formula uses earnings in a "base year", defined as either (i) the first four of the last five completed calendar quarters, or (ii) the last four completed calendar quarters, immediately preceding the start of the claim. Provided that an individual had enough earnings (and, uniquely to the Washington system, worked a minimum of 680 hours) in the base year, he or she is eligible to draw benefits over a year-long period (the so-called "benefit year"). The individual's benefit amount is $1/50$ of total wages earned in the two highest-earning quarters of the base period. Thus, for an individual with constant weekly earnings of $\$I$ over the base year, the benefit amount is $26/50 \times I \approxeq 0.52I$. The benefit amount is subject to a maximum (which was $205/week in the first 6 months of the experimental period, and $209/week in the second half) as well as a minimum (which we ignore by eliminating the small fraction of claimants affected by this provision). Claimants also face a ceiling on the total amount of benefits claimed, which cannot exceed the lesser of $30\times$their weekly benefit amount, or one-third of their base-year earnings. This ceiling determines the maximum number of weeks of UI they can draw.

Formally, the benefit rules can be summarized as follows. Let b denote "weekly benefit amount", $Totalbenefits$ denote "total benefits payable" and $Maxduration$ denote "weeks of unemployment benefit entitlement". Let Q_1, Q_2, Q_3 and Q_4 be the quarterly earnings in the four quarters of the base period, ranked in order of earnings, so Q_1 and Q_2 are the two highest among the four. The rules that determine b, $Totalbenefits$ and $Maxduration$ in the first half of 1988 (the definition is analogous for the last two quarters of 1988) are summarized as follows:

$$b = \begin{cases} \frac{Q_1+Q_2}{50} & \text{if } \frac{Q_1+Q_2}{50} < 205 \\ 205 & \text{if } \frac{Q_1+Q_2}{50} \geqslant 205 \end{cases}$$

$$Totalbenefits = \min(30 \cdot b, \frac{Q_1 + Q_2 + Q_3 + Q_4}{3})$$

$$Maxduration = \frac{Totalbenefits}{b}$$

Define $r \equiv \frac{Q_3+Q_4}{Q_1+Q_2} \in [0, 1]$ and $V \equiv Q_1 + Q_2$. Then with some simplification we can re-write the rules as:

$$b = \begin{cases} \frac{V}{50} & \text{if } V < 10,250 \\ 205 & \text{if } V \geqslant 10,250 \end{cases}$$

$$Totalbenefits = \begin{cases} \frac{3}{5}V & \text{if } V < 10,250 \text{ and } r \geqslant \frac{4}{5} \\ \frac{1}{3}V(1+r) & \text{if } V < 10,250 \text{ and } r < \frac{4}{5} \\ 6,150 & \text{if } V \geqslant 10,250 \text{ and } r \geqslant \frac{4}{5} \\ \frac{1}{3}V(1+r) & \text{if } V \geqslant 10,250 \text{ and } r < \frac{4}{5} \end{cases}$$

$$Maxduration = \begin{cases} 30 & \text{if } V < 10,250 \text{ and } r \geqslant \frac{4}{5} \\ \frac{1}{3}(1+r) \cdot 50 & \text{if } V < 10,250 \text{ and } r < \frac{4}{5} \\ 30 & \text{if } V \geqslant 10,250 \text{ and } r \geqslant \frac{4}{5} \\ \frac{1}{3}(1+r) \cdot \frac{V}{205} & \text{if } V \geqslant 10,250 \text{ and } r < \frac{4}{5} \end{cases}$$

Notice that for claimants with $r \geq \frac{4}{5}$, $Totalbenefits$ "top out" at the same point as b; as a result $Maxduration$ is exactly 30 weeks, regardless of V. For claimants with $r < \frac{4}{5}$, however, $Totalbenefits = \frac{1}{3}V(1+r)$ for all values of V. Since $Maxduration = \frac{Totalbenefits}{b}$, and b flattens out once V reaches $10,250$, there is an "upward" kink in the relation between $Maxduration$ and V at $V = 10,250$ for people with $r < \frac{4}{5}$.

Naturally, having two endogenous regressors that are kinked at the same point will make it impossible to distinguish between the independent effects of b and $Maxduration$. But since the

kink in $Maxduration$ is entirely driven by the kink in b, we can still obtain estimates of the "reduced-form" effect of b. This reduced form effect arises through two channels: the direct effect of b on the outcome of interest, and an indirect effect through $Maxduration$. That is, the identified marginal effect is from manipulating b, while holding constant the other aspects of the rules determining $Totalbenefits$ and $Maxduration$.

3.3.3 Data Issues

Data Source and Variables

Our data are derived from the public use file of the WREB, which combines information from several administrative data sources. Most of the variables, including wage earnings, UI benefits, and demographic characteristics, come from the Benefit Automated System and the WAGE database provided by the Washington State Employment Security Department (WSESD). Also included in the file are variables on local labor market conditions provided by the Labor Market and Economic Analysis Branch of the WSESD. The assignment variables that are crucial to the evaluation of the re-employment bonus experiment come from the Participant Tracking System of the WREB, and will not be used in our analysis. Details on the the construction of the data file can be found in Spiegelman et al. (1992).

The variables relevant for our study are:

- Earnings: quarterly earnings from the first quarter of 1985 to the last quarter of 1989.
- Unemployment Insurance: date of UI claim; b; $Totalbenefits$; $Maxduration$; Net UI payment for every week in the benefit year.
- Baseline covariates: age, gender, race, education, one-digit SIC code of base year employer, Job Service Center where the claim was filed.

Table 3.1 reports summary statistics for these variables for the analysis sample that we describe below.

Adjusting Base Period Earnings for Changes in the Maximum Weekly Benefit

The maximum weekly benefit amount increased from \$205 to \$209 beginning July 1, 1988. To facilitate a pooled analysis using data for the entire year, we adjust V (the sum of earnings in the two highest quarters) for claimants who filed after July 1, 1988 by subtracting \$200 (=\$4 per week $\times$50 weeks) from the sum of their two highest-quarter earnings. With this adjustment, the kink in the benefit rule is at the same point ($V = 10,250$) for all claimants in the sample. The formulaic relationship between b and the adjusted value of V is shown in Figure 3.2 (The Figure also shows the minimum weekly benefit rate, which we ignore in our empirical analysis by eliminating very low-earning claimants). Note that although the relationship has the same kink point for claimants in the two halves of the year, the benefit rates (as a function of V) are slightly higher for claimants from the second half. Estimated treatment effects from the pooled sample therefore represent an average of the marginal effects for these two levels of benefits. If the marginal effects are the same, then the pooled data will yield more efficient estimates.

Measurement Error in Base Period Earnings

In principle, a plot of the actual weekly benefit amounts received by claimants against their normalized base period earnings should replicate Figure 3.2. In practice, the empirical relationship (depicted in Figure 3.3) shows deviations from the UI benefit rules. Out of 15,534 claimants in our overall sample, some 8% (1,249 cases) have benefit amounts that appear to deviate from the formula (this group also includes a small number of claimants with missing data for three or four quarters in the baseline). Figure 3.4(a) plots the histogram of the differences between the actual and predicted values of the weekly benefit amount b. Since 92% of observations have a deviation of precisely 0, the figure is not very informative: Figure 3.4(b) shows the histogram after excluding the $0's$, and suggests that the deviations are slightly left-skewed (i.e, actual benefits tend to be a little lower than predicted, on average). Further investigation revealed that the likely source of the discrepancy between actual and predicted benefits arises because the benefit system data files incorporate an unedited version of quarterly

earnings in the base period, whereas the benefit formula uses a verified measure of earnings.[9]

Under the presumption that the benefit rate b was correctly computed from actual earnings, it is possible to correct the measure of V by inverting the benefit formula. After this correction procedure the only remaining deviations arise from the 90 observations whose actual weekly benefit is at the maximum level ($205 for the claims filed in the first half of 1988, $209 for those filed in the second half of 1988) but whose reported value of V is below the kink point ($10,250). For simplicity we drop these cases from our main analysis sample. The relationship between b and V in the analysis sample is shown in Figure 3.3(b) and (by construction) follows the predicted pattern of Figure 3.2 exactly.[10] A minor complication arises because there is an unusual mass of claimants with filing dates in July 1988 who have values of $b = \$205$, even though though the maximum benefit rate had increased to $209 effective July 1. These claims were most likely processed according to the rules for the first half of the year. Accordingly, we assume that claimants from July 1988 whose actual benefit amount is $205/week, but whose predicted benefit exceeds $205, were processed according to the rules before June 30, 1988.

3.4 Empirical Results from the RKD Analysis

3.4.1 Outcomes of Interest

We focus on three main outcomes associated with the effect of changes in weekly UI benefits: the total amount of UI payments received over the benefit year; the total number of weeks of UI claimed; and the duration of the initial unemployment (claim) spell. Our definition of the initial spell is borrowed directly from Spiegelman et al. (1992): this spell starts with the so-called waiting week (a week during which no payments are received) and ends when there is a gap of at least two weeks in the receipt of benefits.

[9] According to Ken Kline at the Upjohn Institute, if an applicant's earnings do not match the amount in the system database, then the employer is contacted for verification in order to calculate UI benefit. But the system database, which we use, is not subsequently updated with the correct information.

[10] We have performed empirical analyses on both the "raw" data and on the corrected sample, and the results do not differ substantially. We report results from the corrected sample below.

The effect of an increase in the weekly benefit amount b on total UI system costs (per claimant) is of great policy interest in itself. It is important to emphasize, however, that a finding of a significant effect for this outcome does not necessarily imply that higher benefits induce a behavioral response among UI claimants. Even in the absence of any behavioral response, an increase in weekly benefits paid to some group will lead to an increase in total UI benefits that is proportional to the average number of weeks of UI claimed. Thus, a kink in the relationship between total UI payments and base earnings at the point where weekly benefits are capped will in part reflect the purely mechanical relationship between the benefit amount and the cost of payments.

It is tempting to interpret the effect of the weekly benefit amount on the number of weeks of UI claimed, or on the length of the initial spell, as a purely behavioral response. In the case of Washington's UI system, however, a straightforward RKD analysis will not isolate a purely behavioral effect on either of these outcomes. The reason is that for claimants with $r = \frac{Q_3+Q_4}{Q_1+Q_2} < \frac{4}{5}$ there is a kink in the relationship between $Maxduration$ (the maximum number of weeks of UI available) and base period earnings V (see Section 3.3.2). For this subgroup (who represent approximately 60% of claimants in the WREB), a small increase in V when V is to the left of the kink-point in the benefit formula leads to a higher benefit rate b, but no increase in $Maxduration$. In contrast, a small increase in V to the right of the kink-point leads to both an increase in b and an increase in $Maxduration$. Thus, a comparison of the slopes of the relationship between V and the duration of UI claims on either side of the kink-point combines a behavioral response and a mechanical entitlement effect. To isolate the behavioral component, we consider artificially censoring the data on the right so that the relationship between the censored maximum duration and V is smooth. Specifically, consider a maximum potential benefit duration measure $Maxduration_{smooth}$ that is constructed by applying the formula for $Maxduration$ that

prevails on the left side of the kink to all claimants:

$$Maxduration_{smooth} = \begin{cases} 30 & r \geqslant \frac{4}{5} \\ \frac{1}{3}(1+r) \cdot 50 & r < \frac{4}{5} \end{cases}.$$

If we conduct an analysis using the number of weeks of benefits claimed (or the duration of the initial UI claim) censored at $Maxduration_{smooth}$, we potentially eliminate the mechanical entitlement effect and isolate the behavioral impact of the change in UI benefits.[11] To evaluate the validity of this simple approach we plotted mean weeks of actual UI entitlement against the value of V for a set of discrete bins (30 different bins of $500 each, equally distributed on the two sides of the kink point). We then compared this to a plot of censored entitlements, censoring actual entitlements at $Maxduration_{Smooth}$. Whereas uncensored mean entitlements show a pronounced kink at $V = 10,250$, the relationship between censored entitlements and base period earnings is smooth (see Figure 3.5), suggesting that the censoring approach will work. We therefore focus on five "outcomes of interest" in our empirical analysis: total UI payment received, weeks of UI claimed, the duration of the initial UI spell, and censoring-adjusted versions of the latter two outcomes.

3.4.2 Graphical Presentation

An attractive feature of a Regression Kink Design is that the results from the analysis can be summarized graphically, in a fashion similar to the way that results from a Regression Discontinuity Design are typically presented. In particular, one can plot the means of the outcomes of interest, as well as the means for the predetermined covariates, against the

[11] In practice we use $Maxduration_{smooth} = \begin{cases} 30 & r > 0.74 \\ \text{ceiling}[\frac{1}{3}(1+r) \cdot 50] & r \leqslant 0.74 \end{cases}$ to incorporate the effect of the way that rounding is implemented in the Washington State UI system. For example, when $Totalbenefits/b = 29.1$, a claimant is entitled to 30 weeks of benefits: he or she can receive full weekly benefits for the first 29 weeks of claim and one tenth of the weekly benefit amount in the 30th week. There are also claimants for whom weeks of UI received are greater than $Maxduration$. This happens when a claimant receives partial benefits during a week when he or she is working part-time while on claim. For simplicity we do not cap the weeks of UI claimed (or initial spell length) at $Maxduration_{smooth}$ for these claimants.

assignment variable V (base period earnings) and look for potential kinks around the kink-point in the formula that maps the assignment variable to the regressor of interest. For such a presentation we need to divide the range of V into suitable "bins." Given the modest sample sizes available, we use \$500 bins.[12] We also limit attention to observations with $V \in [2750, 17750]$, resulting in a graphical analysis with 15 bins on each side of the kink-point ($V = 10250$).[13]

Proposition 3 establishes that there are two key testable implications of a valid RK design. First, the density of the assignment variable V has to be continuously differentiable at the kink-point. Second, the conditional expectations (and conditional quantile functions) of any baseline covariates have to be continuously differentiable at the threshold. As in a RDD, these testable conditions can be visually examined. We proceed by plotting the number of observations in each bin, and the mean values of the covariates for the claimants in each bin, against base period earnings.

Figure 3.6 presents a plot of the "density" of V (i.e., the histogram across the 30 bins).[14] The histogram is somewhat bumpy, with a drop between 15th bin (to the left of the kink) and the 16th bin (to the right), although the drops at other points (e.g., between the 9th and 10th bins, or between the 17th and 18th bins) are similar in magnitude.

Next we examine plots of the conditional means of age, education, gender, race, region, and industry for different values of V. All of these covariates were presumably determined before the claimant's base period earnings. The results are shown in Figures 3.7(a) through (e). As a simple indicator of region we use the fraction of claimants who filed for UI at a Job Service Center in Western Washington.[15]

Inspection of the pattern of the "dots" in Figures 3.7(a)-(e) leads us to conclude that the conditional means of the covariates evolve smoothly across the kink-point in the benefit

[12] Lee and Lemieux (forthcoming) present a formal procedure for choosing bin size, based on goodness-of-fit tests which evaluate the fit of simple models with bin dummies. A bin size of 500 is the largest that passes the two tests suggested by Lee and Lemieux (forthcoming) for all the dependent variables in our analysis.

[13] Outside of the range $[2750, 17750]$ the sample sizes per bin are very small.

[14] Since our interest is in evaluating the smoothness of the density function we do not display the more conventional "smoothed" histogram.

[15] At the time of the WREB the state had 21 Job Service Centers, 14 of which are located in Western Washington.

determination schedule. The evolution of mean education (Figure 3.7(b)) shows some evidence of a kink in the neighborhood of $V = 7250$, but around the critical kink-point ($V = 10250$) it appears to evolve relatively smoothly. On balance there is no strong visual evidence of discontinuities in $\partial E[X|V = v]/\partial v$ at the kink-point. We evaluate the smoothness of the distributions of the covariates more formally in the next section.

The relative smoothness in the conditional means of the covariates around the kink point becomes even more apparent when compared with the patterns for the outcome variables in Figures 3.8(a)-(e). For all five outcomes there is a clearly discernible change in the slope of the relationship with V at the kink-point in the benefit formula. In each case the outcome variable is increasing in V to the left of the threshold and decreasing in V to the right of the threshold. Notice that this is true for total UI benefits (Figure 3.8(a)) which incorporates both a mechanical and behavioral effect of weekly benefits, as well as in the censored versions of weeks of UI claimed (Figure 3.8(d)) and the duration of the initial claim (Figure 3.8(e)), which incorporate only a behavioral component. In the following section, we present the numerical estimates of the change in slopes based on simple parametric specifications.

3.4.3 Estimation Results

Empirical Specification

In empirically implementing the RKD estimator, we follow two complementary approaches. For our first approach we follow Lee and Lemieux (forthcoming) and estimate parametric polynomial models of the form:

$$E\left[Y|V = v\right] = \alpha_0 + \sum_{p=1}^{\overline{p}} [\alpha_p (v-k)^p + \beta_p (v-k)^p \cdot D] \text{ where } |v-k| \leqslant h \tag{3.2}$$

where $k = \$10250$ (the kink point), $D = 1_{[V \geqslant k]}$, an indicator for the event that base period earnings exceeds the kink-point, the α's and the β's are polynomial coefficients, $\overline{p}$ is the maximum polynomial order, and h is the bandwidth that determines the window $[k-h, k+h]$ within which

the sample is selected.[16] In this approach the change in the derivative of the conditional expectation function – the numerator of the RK estimand – is given by the coefficient β_1. Since the slope of the benefit function $b(V)$ changes from $\frac{1}{50}$ to 0 at the kink point, the denominator of the RK estimand is $-\frac{1}{50}$ and so we multiply $\hat{\beta}_1$ by -50 to obtain the TT_b effect of b on Y.

We present a sensitivity analysis, choosing several levels of h ranging from $h = 1000$ to $h = 7500$. For each bandwidth choice we vary $\overline{p}$ from 1 to 3 and report the order of the polynomial preferred by the Akaike Information Criterion (AIC). As suggested in Lee and Lemieux (forthcoming), we also run an additional "unrestricted" model in which we include a set of dummy variables which indicate consecutive intervals (of width 500) in V, and compute a goodness-of-fit test that compare the polynomial model to the dummy variable specification. Within this framework we can also easily probe the sensitivity of the results to inclusion of the baseline covariates, by adding these as additional regressors.

RKD Estimates

Table 3.2 reports RKD estimates using the baseline specification (3.2) and Table 3.3 shows that the results are robust to the inclusion of baseline covariates. For each table, we include point estimates of $\hat{\beta}_1$ and robust standard errors for each of the regressions. We also report the p-values from the Goodness-of-Fit tests including the bin dummies. Results for bandwidths of $h = 7500$, $h = 2500$ and $h = 1000$ are reported. For each bandwidth, we report the coefficient from regressions up to a third order polynomial.

In general, within the bandwidths we consider, the AIC suggests a linear specification, with the exception being that a quadratic is chosen by the AIC for "total weeks claimed". It is also true that the linear specification is not considered too restrictive relative to a model that includes bin dummies: none of the p-values are less than 0.05 in any of the specifications. The most precise estimates of the effect of a dollar increase in benefit on total UI received are around \$17 to \$18,

[16]Note that this specification imposes continuity in the conditional expectation function at the kink-point. We have also estimated all models allowing for a potential discontinuity (by including D as a separate regressor). Estimates of the change in the slope at the kink-point are very similar, and the implied "jumps" in the conditional expectation function are never statistically significant.

while for the point estimates for “Total Weeks Claimed” are around 0.04 of a week. Note that for the bandwidth of 7500, the inclusion of quadratic terms causes the point estimate to fall and become statistically insignificant, although a 0.04 effect could not be ruled out at conventional levels of significance. By comparison, the effects are less sensitive to the inclusion of second order terms for the “Initial Spell Length” variable. As expected, the magnitude of the point estimates increase slightly when we artificially censor the data to isolate the purely behavioral impact of the benefit on the weeks claimed and initial spell variables.

As we might expect, the estimates become less precise as we shrink the bandwidth, and when it becomes 1000, the standard errors are much larger relative to the point estimates, even for “Total UI Received”, which is the most striking kink displayed in our figures. Although we report the point estimates and corresponding p-values for the specification tests for all of these permutations, we focus on the first, second, and fourth rows of Table 3.2 as our preferred specifications, and believe at a minimum that these models are the least likely to be over-fit. The pattern of results in Table 3.3 – where we include baseline covariates – are similar both qualitatively and quantitatively, with the estimated standard errors being slightly smaller, and the point estimates being sometimes higher and sometimes lower, but not by a significant amount, depending on the outcome and specification.

We benchmark our estimates against three studies from the UI literature. Hamermesh (1977) concludes that "the best estimate–if one chooses a single figure–is that a 10-percentage point increase in the gross replacement rate leads to an increase in the duration of insured unemployment of about half a week when labor markets are tight." Moffitt and Nicholson (1982) find that that a 10-percentage point increase in the replacement rate was associated with about a one week increase in the average length of unemployment spells, while the estimate of Meyer (1990) is around one and a half weeks in response to a 10-percentage point increase in replacement rate.

In our setting, a \$1 increase in the weekly benefit amount for the population near $V = 10,250$

corresponds to about a 0.25 percentage point increase in the UI replacement rate[17]. Since our estimates indicate roughly a 0.04 increase in insured spells, this implies that the response to a 10 percentage point increase in replacement rate would be an increase in insured unemployment duration by about $0.04 * (10/0.25) = 1.6$ weeks, which is of a similar magnitude as the estimate in Meyer (1990).

Testing for Kinks in the Density of Baseline Earnings and Conditional Expectation of Covariates

To provide an estimate of a potential kink in the density of V at the threshold, we follow the approach of McCrary (2008) and first collapse the data into equal-sized bins of width 500. The collapsed data set contains 30 observations as we restrict the sample to $V \in [2750, 17750]$. The two key variables in the collapsed data set are: the number of original observations in each bin N_{bin} and the baseline earnings amount each bin is centered around, V_{bin}. We then regress N_{bin} on polynomials of $(V_{bin} - k)$ and the interaction term $1_{[V_{bin} \geqslant k]}(V_{bin} - k)$ where $k = 10,250$ is the kink-point. Because of the small number of observations, we do not interact $1_{[V_{bin} \geqslant k]}$ with higher order polynomial terms of $(V_{bin} - k)$. A fourth order polynomial does a good job fitting the data with an $R^2 = 0.99$. As suggested by Figure 3.6, the coefficient on the interaction term is statistically insignificant (a t-statistic of -0.78).

Table 3.4 is the analogy to Table 3.2, except that the dependent variables are the baseline covariates. If the RKD is valid, and the assumption of a smooth density of V is reasonable, then we expect not to see systematic evidence of kinks. If we again focus on the first, second, and fourth rows, as we did for Tables 3.2 and 3.3, we find that most of the point estimates are statistically insignificant at conventional levels. For the dummy variable indicating the Job Service Center was in western Washington, the point estimates are significant for the linear and quadratic specifications. On the other hand, the specification tests clearly reject those polynomial orders, and the AIC suggests a third order polynomial (third row), and in that row the point

[17]The weekly earnings for the claimants with $V = 10,250$ are about $\$10,250/26 = \$394$. So a \$1 increase is equivalent to an increase in the replacement rate of $\$1/\$394 = 0.25$ percentage point.

estimate is insignificant. In a similar way, the goodness-of-fit statistics for mean Education – for which the point estimates are statistically significant – are rejected at the .10 level. Finally, the point estimates for White are statistically significant in the first and second rows, but not in the third and fourth rows. On balance, if we evaluate the specifications in Tables 3.2, 3.3 and 3.4 with a similar standard of needing to both pass the goodness-of-fit statistic at the .10 level and be chosen by the AIC, we see that almost all of the covariates exhibit no significant kink, while the effects on the outcomes are most striking for "Total UI Received" and "Initial Spell Length".

3.5 Conclusion

This paper considers the identification of marginal effects in nonparametric models of endogenous regressors with nonseparable errors, using the Regression Kink Design. In this context, we establish the necessary condition for identification, which is, loosely speaking, that "all other factors" are evolving smoothly – in the sense of a continuous derivative – with respect to the assignment variable. The problem with such an assumption, is that it involves a statement about the distribution of unobservables, which can be difficult to justify and – given the unobservability of these factors – impossible to test.

Our main contribution is to characterize a class of models that are instead based on an assumption about the distribution of an observable variable, V. In particular, the assumption is that for each agent, the density of V is continuously differentiable at the kink-point. This assumption may follow naturally from models of the underlying behavior. In our context, we have outlined an illustrative model that would suggest the smooth density condition would be violated, and another model in which it would be satisfied. Most importantly, our characterization of a valid RK Design also generates the testable prediction that pre-determined covariates will have a distribution that is "smooth" with respect to V in the kink-point.

Applying these ideas to a study of UI claimant behavior in the state of Washington, we find evidence consistent with a valid RK Design, and estimates of the impact of a marginal increase in

the benefit level on insured unemployment spells that are in the higher range of magnitudes found in the existing literature. In ongoing research, we are investigating the asymptotic properties of non-parametric estimators of the RK Design.

Appendix

Proofs

Proof of Proposition 2:

(a) From Condition 2b, we have that $\frac{f_{B,W,V}(b,w,v)}{f_{W,V}(w,v)} = f_B(b)\ \forall b, w, v$. Therefore, $f_{B,W,V}(b,w,v) = f_B(b) f_{W,V}(w,v)$. Integrating both sides with respect to v gives us $f_{B,W}(b,w) = f_B(b) f_W(w)$, and consequently $f_{B|W}(b,w) = f_B(b)$. Thus,

$$\begin{aligned}
\Pr(W \leqslant w | B = b) &= \int_{-\infty}^{w} f_{W|B}(w'|b) dw' \\
&= \int_{-\infty}^{w} \frac{f_{B|W}(b|w')}{f_B(b)} dG(w') \\
&= \int_{-\infty}^{w} dG(w') \\
&= \Pr(W \leqslant w).
\end{aligned}$$

(b) Condition 2a allows the interchange of differentiation and integration in

$$\frac{\partial E[Y|B=b]}{\partial b} = \int\int \frac{\partial}{\partial b}[y(b,v,w) f_{V,W|B}(v,w|b)] dv dw.$$

Applying the Bayes Rule and invoking Condition 2b, we have

$$\begin{aligned}
\int\int \frac{\partial}{\partial b}[y(b,v,w) f_{V,W|B}(v,w|b)] dv dw &= \int\int \frac{\partial}{\partial b}[y(b,v,w) \frac{f_{B|V,W}(b|v,w)}{f_B(b)}] dF_{V,W}(v,w) \\
&= \int\int \frac{\partial}{\partial b}[y(b,v,w)] dF_{V,W}(v,w) \\
&= E[y_1|B=b]
\end{aligned}$$

(c) The proof is analogous to (a).

Proof of Proposition 3:

(a)

$$\begin{aligned}\frac{\partial}{\partial v}\Pr(W &\leqslant w|V=v) = \frac{\partial}{\partial v}\int_{-\infty}^{w}\frac{f_{V|W(v|w')}}{f_V(v)}dG(w') \\ &= \int_{-\infty}^{w}\frac{\partial}{\partial v}\frac{f_{V|W(v|w')}}{f_V(v)}dG(w')\end{aligned}$$

Since we assume that $f_{V|W}(v|w')$ is continuously differentiable in v at 0 for all w', $\frac{\partial}{\partial v}\Pr(W \leqslant w|V=v)$ is continuous at 0.

(b) On the numerator,

$$\begin{aligned}\lim_{v\to 0^+}\frac{\partial E[Y|V=v]}{\partial v} &= \lim_{v\to 0^+}\frac{\partial}{\partial v}\int y(b(v),v,w)\frac{f_{v|w}(v|w)}{f(v)}dG(w) \\ &= \lim_{v\to 0^+}\int\frac{\partial}{\partial v}y(b(v),v,w)\frac{f_{v|w}(v|w)}{f(v)}dG(w) \\ &= \lim_{v\to 0^+}\int[y_1(b(v),v,w)\frac{\partial b(v)}{\partial v}+y_2(b(v),v,w)]\frac{f_{v|w}(v|w)}{f(v)}dG(w)+ \\ &\quad \lim_{v\to 0^+}\int y(b(v),v,w)\frac{\partial}{\partial v}\frac{f_{v|w}(v|w)}{f(v)}dG(w) \\ &= \lim_{v\to 0^+}\frac{\partial b(v)}{\partial v}\int y_1(b(v),v,w)\frac{f_{v|w}(v|w)}{f(v)}dG(w)+ \\ &\quad \lim_{v\to 0^+}\int y_2(b(v),v,w)\frac{f_{v|w}(v|w)}{f(v)}+y(b(v),v,w)\frac{\partial}{\partial v}\frac{f_{v|w}(v|w)}{f(v)}dG(w)\end{aligned}$$

Similarly,

$$\begin{aligned}\lim_{v\to 0^-}\frac{\partial E[Y|V=v]}{\partial v} &= \lim_{v\to 0^-}\frac{\partial b(v)}{\partial v}\int y_1(b(v),v,w)\frac{f_{v|w}(v|w)}{f(v)}dG(w)+ \\ &= \lim_{v\to 0^-}\int y_2(b(v),v,w)\frac{f_{v|w}(v|w)}{f(v)}+y(b(v),v,w)\frac{\partial}{\partial v}\frac{f_{v|w}(v|w)}{f(v)}dG(w)\end{aligned}$$

By Conditions 3a and 3c, we have

$$\begin{aligned} & \lim_{v\to 0^+} \int y_2(b(v),v,w)\tfrac{f_{v|w}(v|w)}{f(v)} + y(b(v),v,w)\tfrac{\partial}{\partial v}\tfrac{f_{v|w}(v|w)}{f(v)} dG(w) \\ = & \lim_{v\to 0^-} \int y_2(b(v),v,w)\tfrac{f_{v|w}(v|w)}{f(v)} + y(b(v),v,w)\tfrac{\partial}{\partial v}\tfrac{f_{v|w}(v|w)}{f(v)} dG(w) \end{aligned}$$

and therefore,

$$\begin{aligned} & \lim_{v\to 0^+} \frac{\partial E[Y|V=v]}{\partial v} - \lim_{v\to 0^-} \frac{\partial E[Y|V=v]}{\partial v} \\ = & \lim_{v\to 0^+} \frac{\partial b(v)}{\partial v} \int y_1(b(v),v,w)\frac{f_{v|w}(v|w)}{f(v)} dG(w) - \lim_{v\to 0^-} \frac{\partial b(v)}{\partial v} \int y_1(b(v),v,w)\frac{f_{v|w}(v|w)}{f(v)} dG(w) \end{aligned}$$

Conditions 3a, 3b and 3c together guarantee that we can write

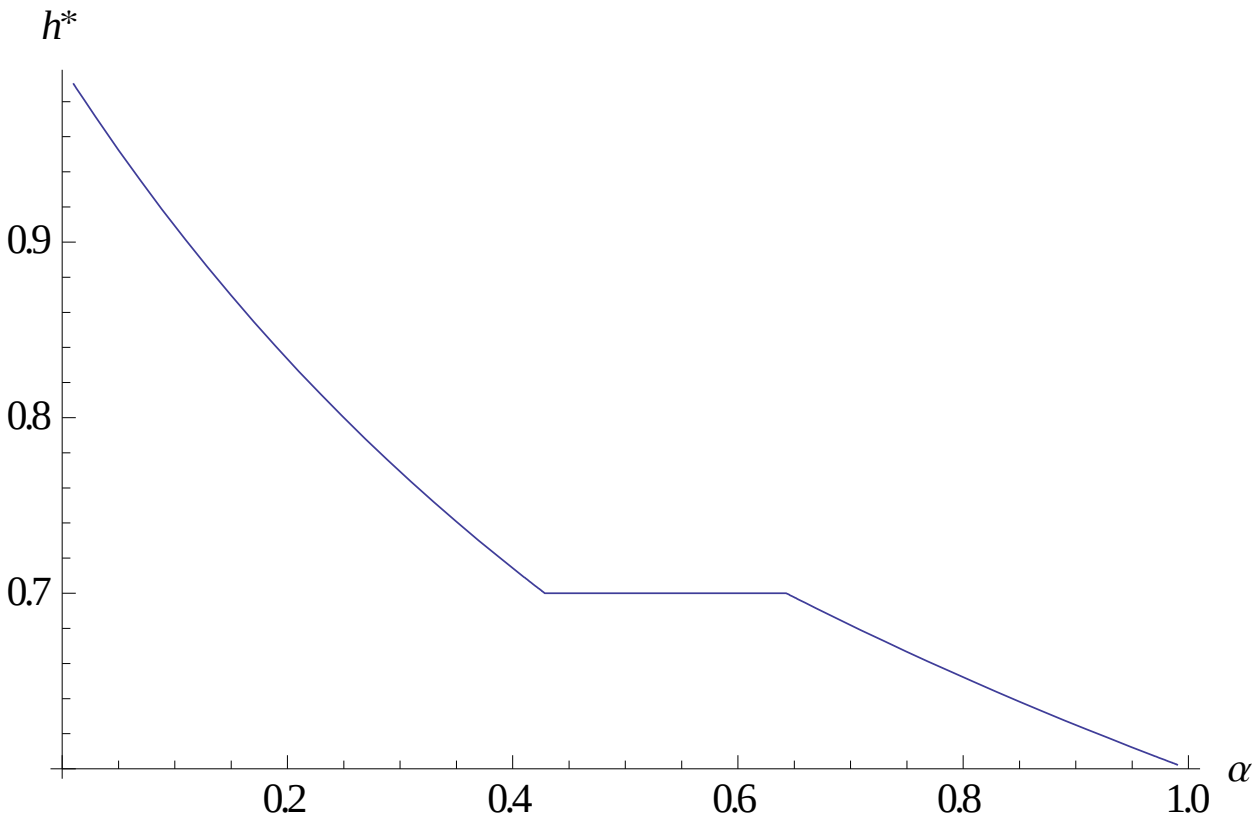
$$\begin{aligned} & \lim_{v\to 0^+} \frac{\partial b(v)}{\partial v} \int y_1(b(v),v,w)\frac{f_{v|w}(v|w)}{f(v)} dG(w) - \lim_{v\to 0^-} \frac{\partial b(v)}{\partial v} \int y_1(b(v),v,w)\frac{f_{v|w}(v|w)}{f(v)} dG(w) \\ = & \left(\lim_{v\to 0^+} \frac{\partial b(v)}{\partial v} - \lim_{v\to 0^-} \frac{\partial b(v)}{\partial v}\right) \int y_1(b(0),0,w)\frac{f_{v|w}(0|w)}{f(0)} dG(w). \end{aligned}$$

Condition 3b guarantees that $\lim_{v\to 0^+} \frac{\partial b(v)}{\partial v} - \lim_{v\to 0^-} \frac{\partial b(v)}{\partial v}$ is nonzero, and hence we have

$$\frac{\lim_{v\to 0^+} \frac{\partial E[Y|V=v]}{\partial v} - \lim_{v\to 0^-} \frac{\partial E[Y|V=v]}{\partial v}}{\lim_{v\to 0^+} \frac{\partial b(v)}{\partial v} - \lim_{v\to 0^-} \frac{\partial b(v)}{\partial v}} = E[y_1(b(0),0,w)|V=0].$$

(c) The proof is analogous to (a).

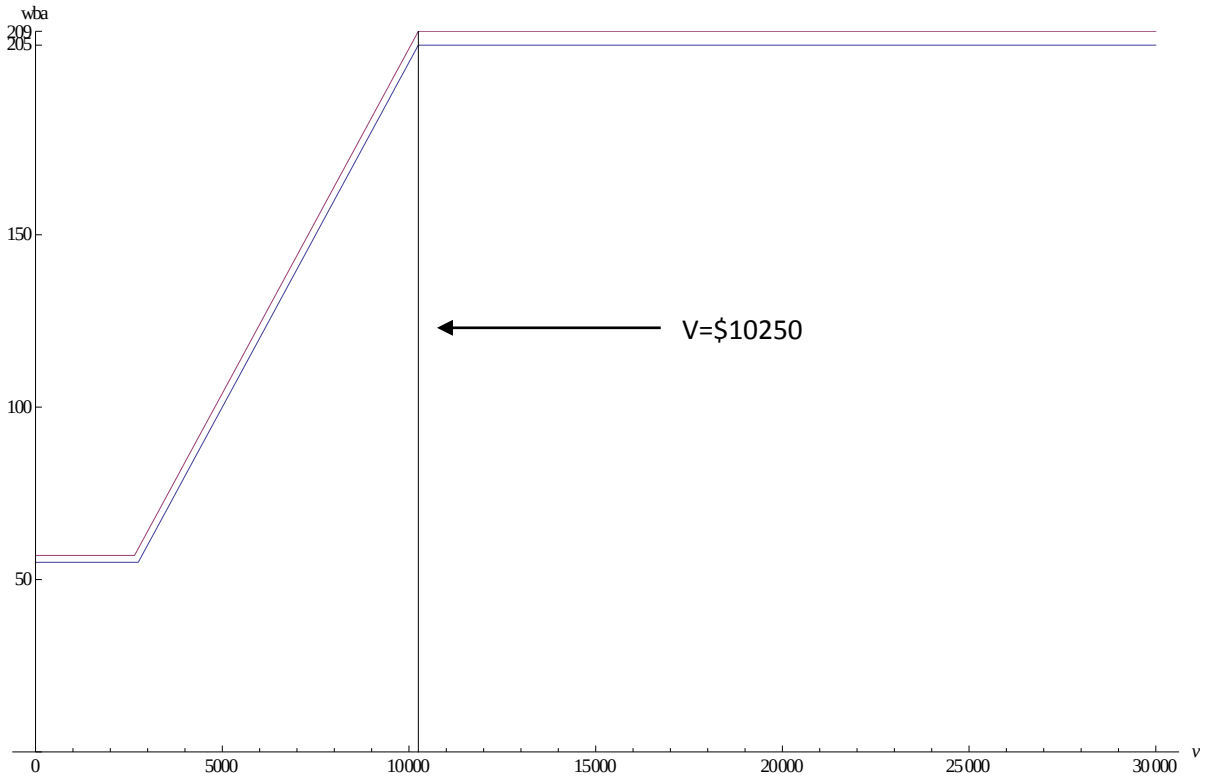
Figure 3.1: The Relationship between h^* and α in the Behavioral Model in Section 2.4.2.



In the behavioral model, h* is the labor supply in the first period and α indicates the relative share of leisure in the Cobb-Douglas utility function. This figure indicates that for a range of α, agents will "bunch" on h*=0.7.

Figure 3.2: Weekly Benefit Amount Determination Rule in Washington State in 1988

WBA as a Function of Normalized Baseline Wage



The rule that determines weekly benefit amount is different in the two halves of 1988. The top line corresponds to the rule after June 30, 1988, and the bottom line corresponds to the rule before June 30, 1988.

The baseline earnings, v, is defined as the sum of the two highest quarterly earnings in the base year. It is normalized here by subtracting $200 from the baseline earnings of claimants who filed in the second half of 1988. According to the normalized rules, there is a kink in the relationship between WBA and V at $10,250 for all UI claimants in 1988.

Figure 3.3(a): Scatter Plot of Weekly Benefit Amount versus Baseline Earnings from Raw Data

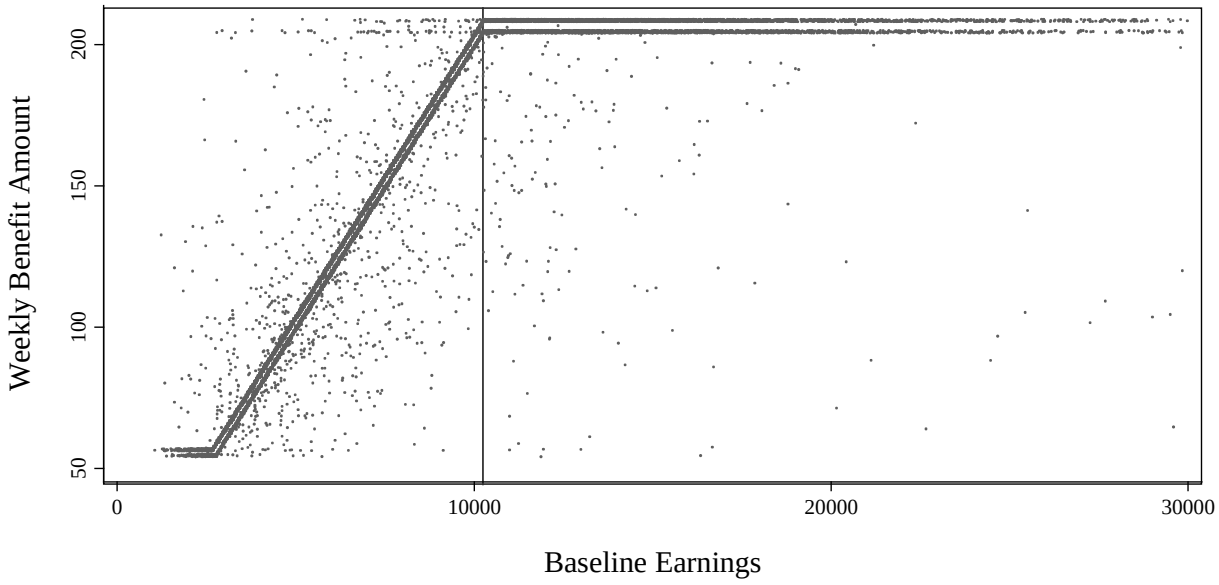
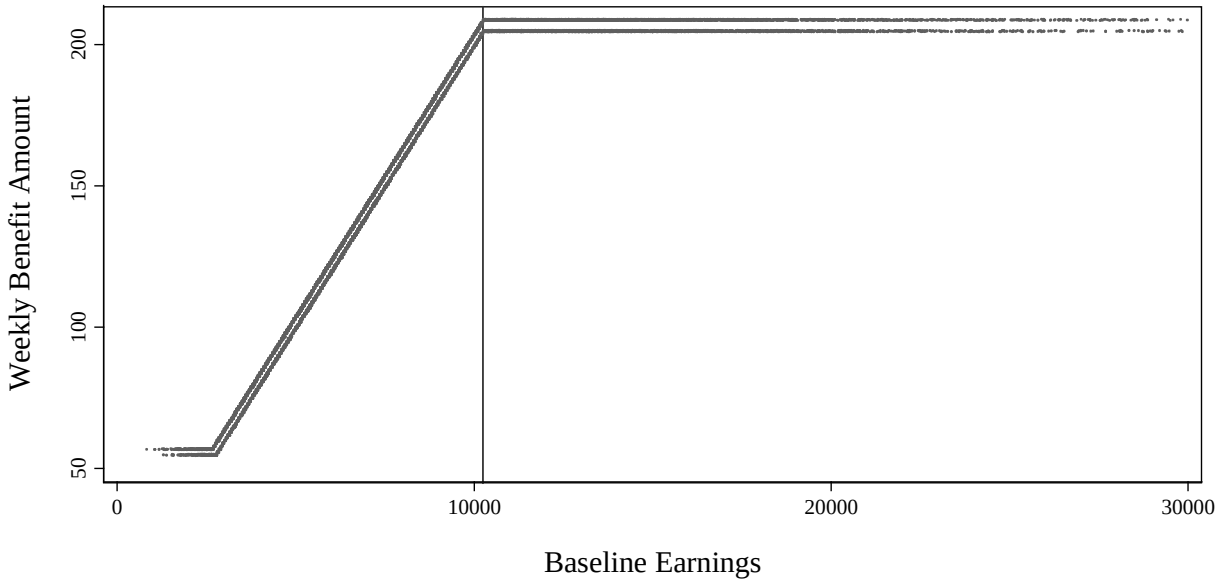


Figure 3.3(b): Scatter Plot of Weekly Benefit Amount versus Baseline Earnings from Imputed Data



Baseline earnings are the sum of the two highest quarterly earnings in the base year. They are normalized here so that the higher kink takes place at $10,250 (the vertical line) for all claimants who filed in 1988. The normalization procedure is described in section 3.3.2 and the caption of Figure 2. The figures above are restricted to the observations whose baseline earnings are less than $30,000 for ease of visual inspection.

Figure 3.4(a): Histogram of Differences between Actual and Predicted WBA in Raw Sample

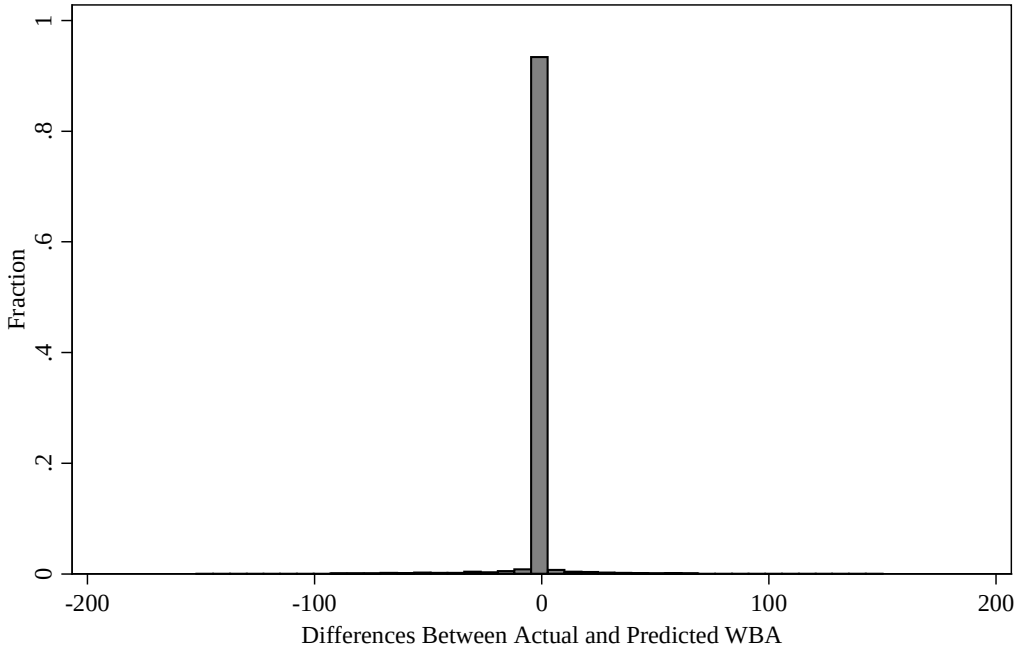
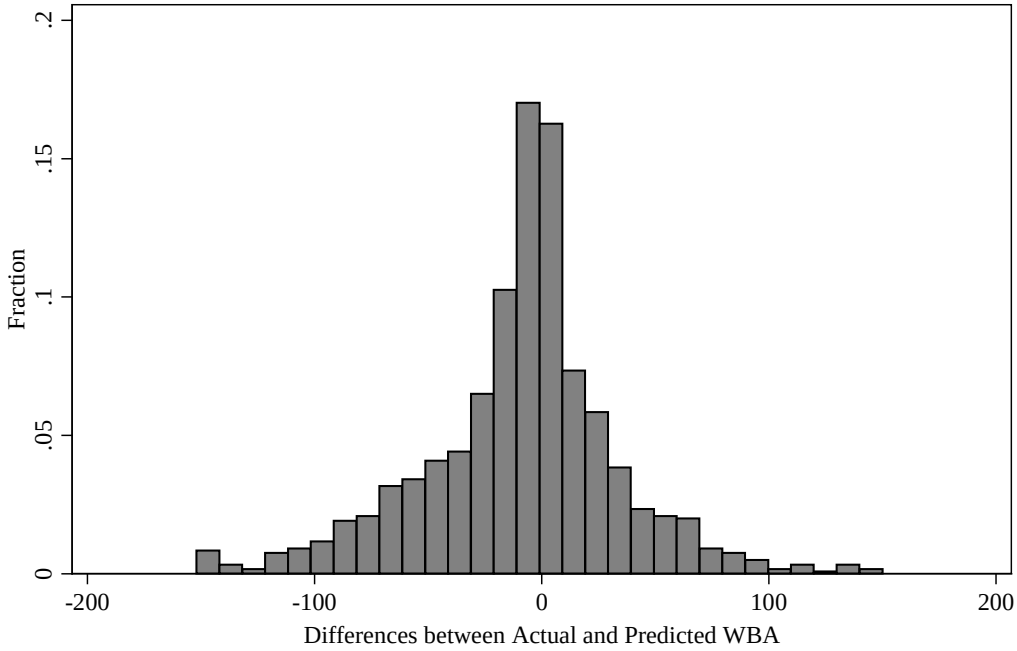
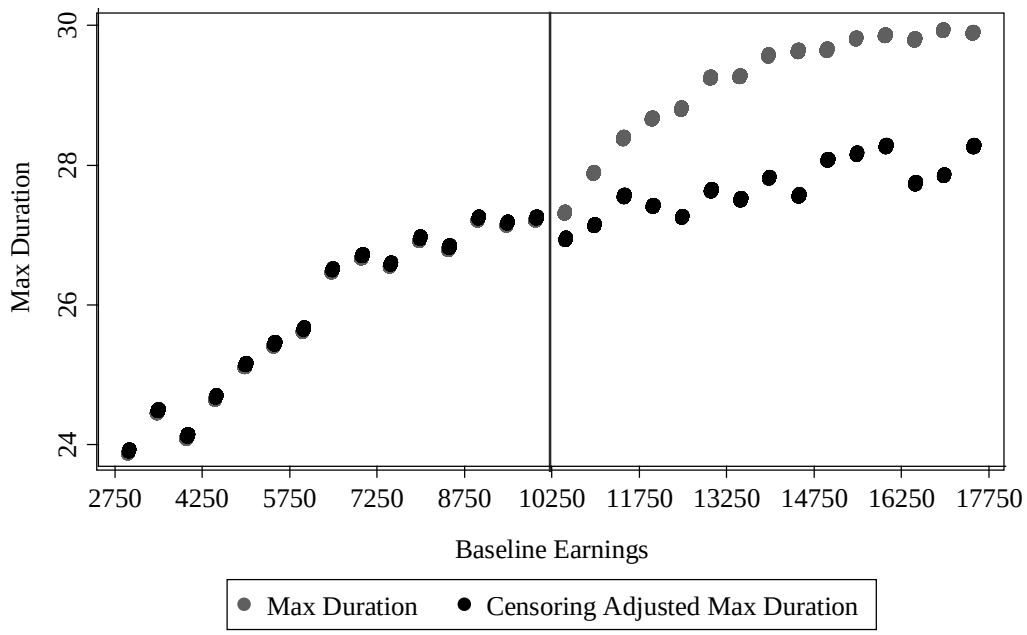


Figure 3.4(b): Histogram of Differences between Actual and Predicted WBA in Raw Sample Conditional on Differences Not Being Zero



Note that the histogram of differences for the imputed sample would simply be a point mass at 0.

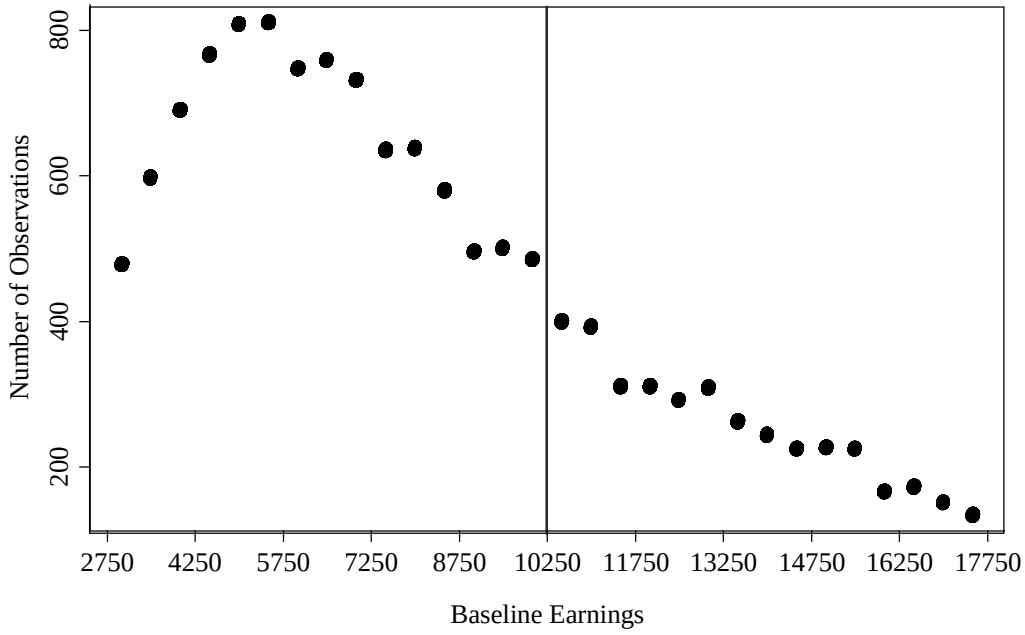
Figure 3.5: Weeks of UI Payments Entitled vs Baseline Earnings



Each black point represents the local average of the actual maximum duration of UI payments over the benefit year in each baseline earnings bin. Each grey point represents the local average of the censoring corrected maximum duration of UI payments over the benefit year in each baseline earnings bin. See Section 4.1 for the construction of the censoring adjusted Max Duration measure.

The baseline earnings bins are of size 500 and are centered at multiples of 500.The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of $10,250.

Figure 3.6: Number of Observations in Each Bin of Baseline Earnings



Each point represents the number of observations in each bin of baseline earnings. The bins are of size 500 and are centered at multiples of 500.The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of $10,250.

Figure 3.7(a): Age vs Baseline Earnings

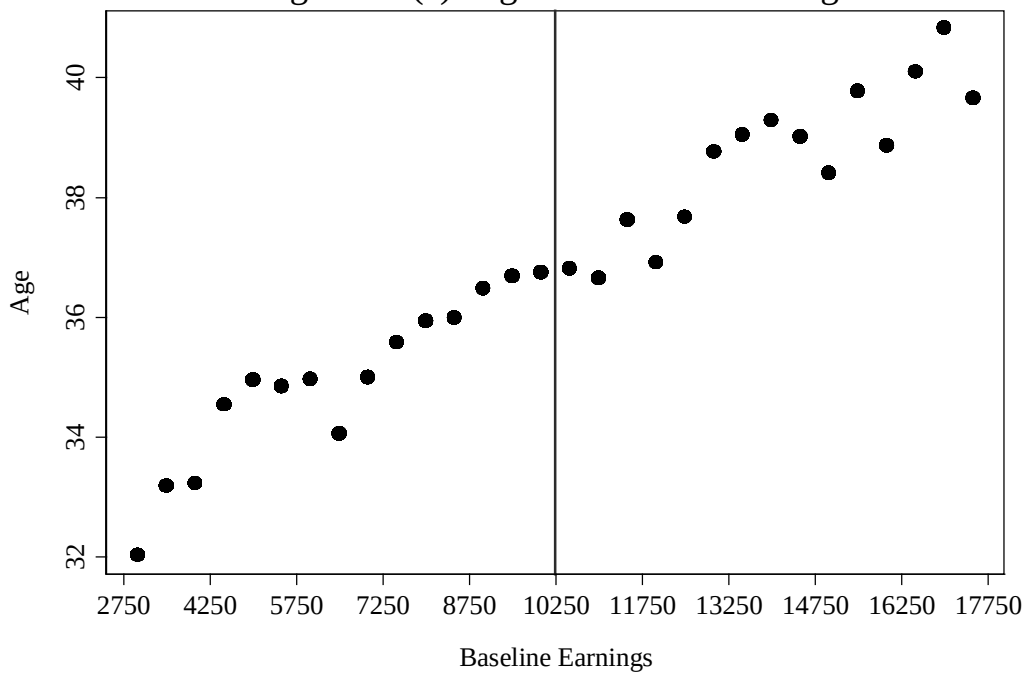
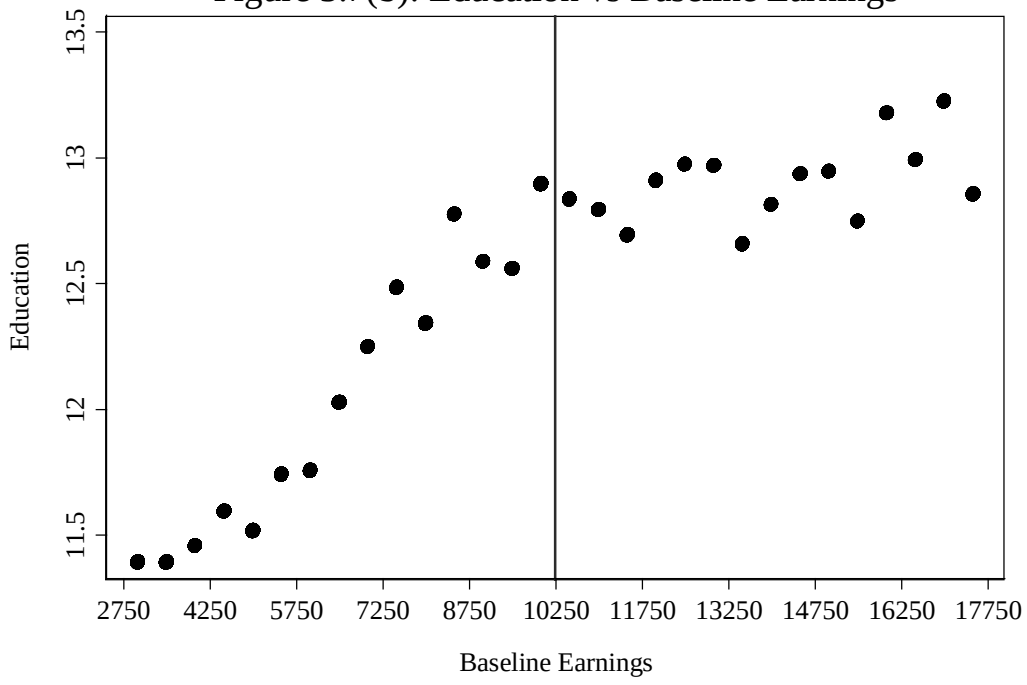


Figure 3.7(b): Education vs Baseline Earnings



Each point in 7(a) represents the local average of the age for claimants in each baseline earnings bin, and likewise in 7(b) the years of educational attainment. The baseline earnings bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of $10,250.

Figure 3.7(c): Fraction of Males vs Baseline Earnings

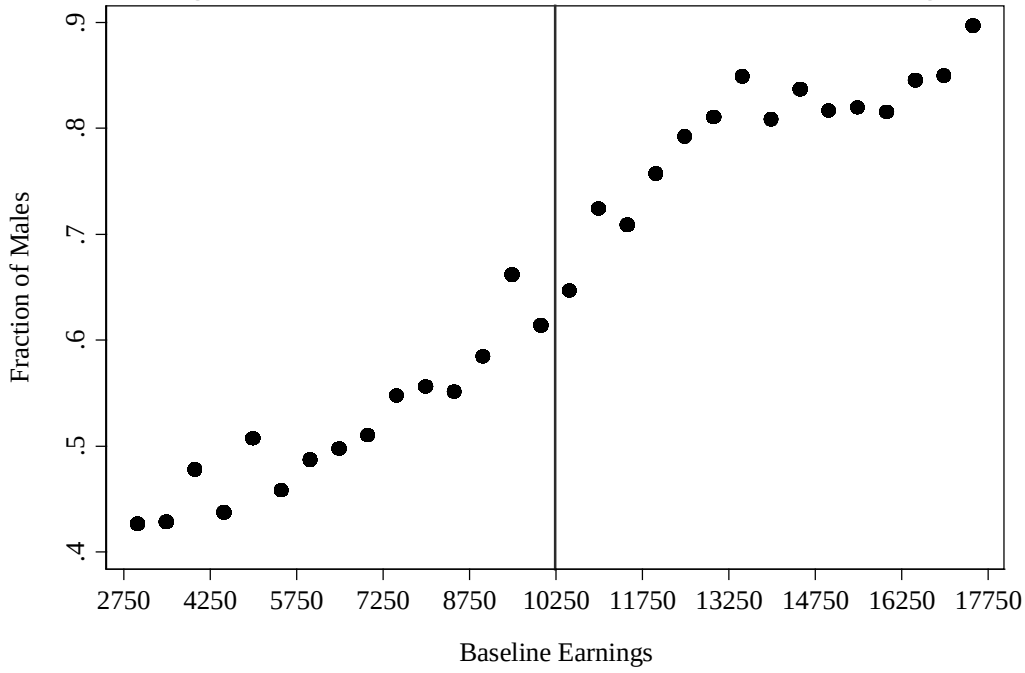
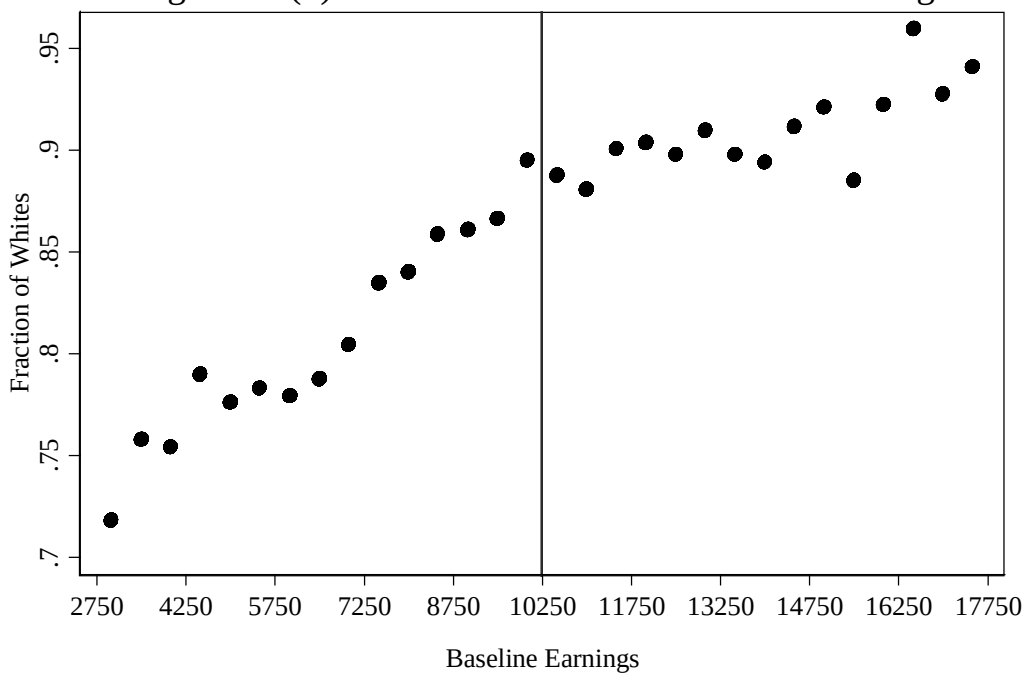
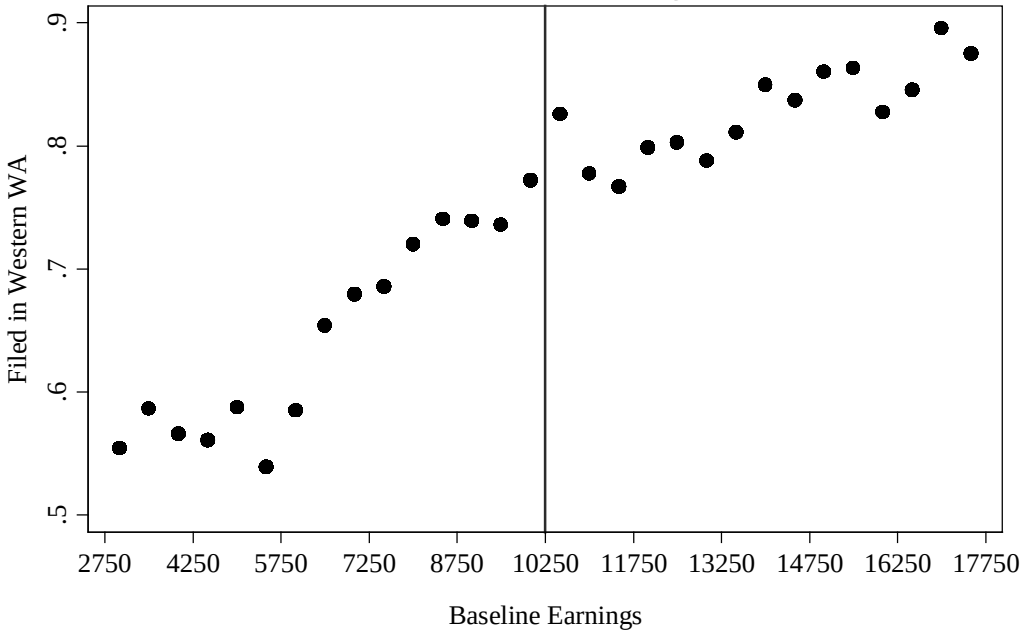


Figure 3.7(d): Fraction of Whites vs Baseline Earnings



Each point in 7(c) represents the fraction of males for claimants in each baseline earnings bin, and likewise in 7(d) the fraction of whites. The bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of $10,250.

Figure 3.7(e): Fraction of Claimants Who Filed Claim in Western WA vs Baseline Earnings



Each point in 7(e) represents the local average of the fraction of claimants who filed at a job service center in Western Washington in each baseline earnings bin. The bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of $10,250.

Figure 3.8(a): Total UI Benefit Payments vs Baseline Earnings

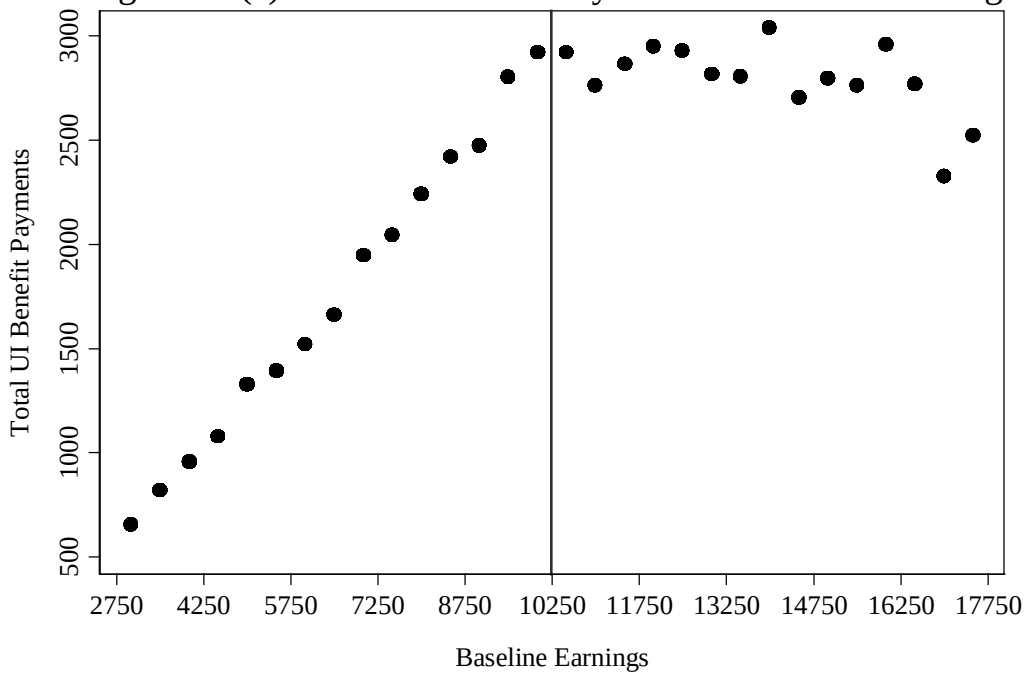
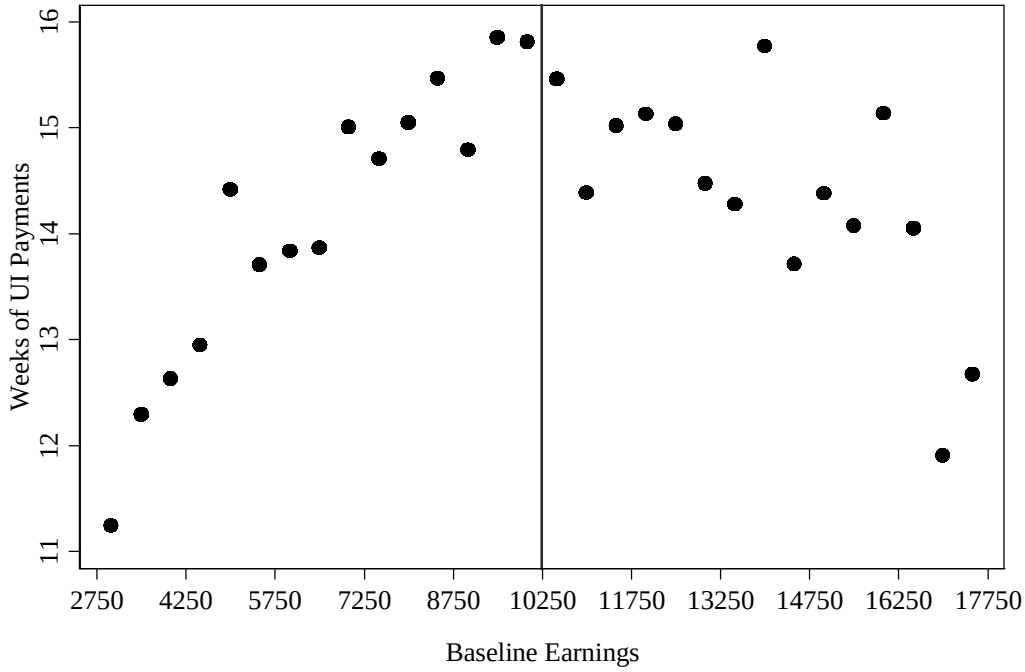


Figure 3.8(b): Weeks of UI Claimed vs Baseline Earnings



Each point in 8(a) represents the local average of the total UI benefit received over the benefit year for claimants in each baseline earnings bin, and likewise in 8(b) the weeks of UI payment claimed. The bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of $10,250.

Figure 3.8(c): Length of Initial Spell vs Baseline Earnings

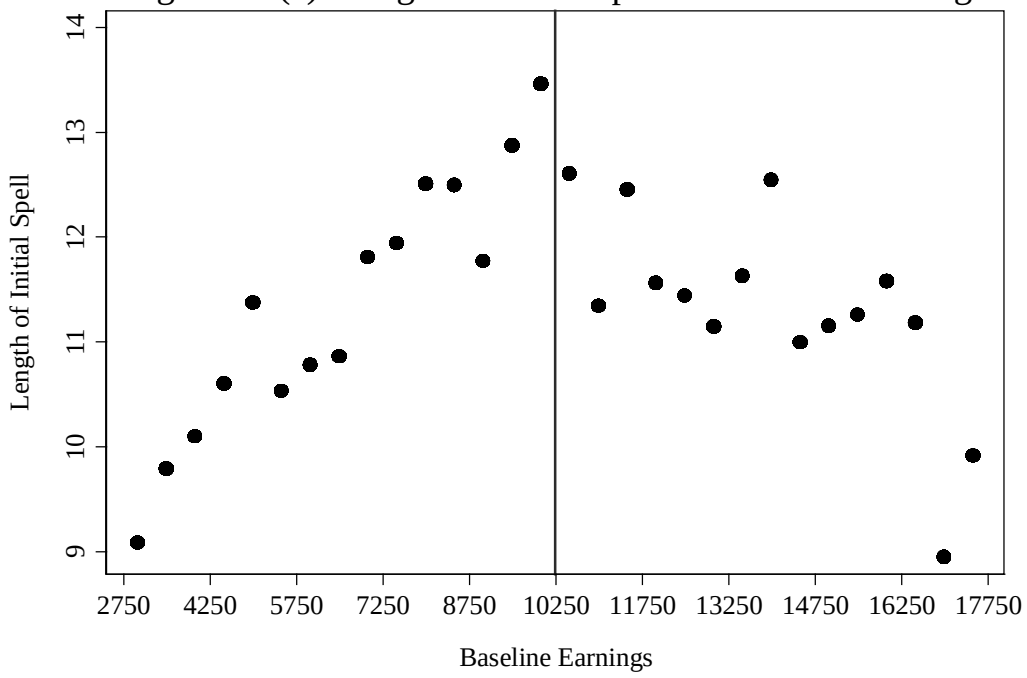
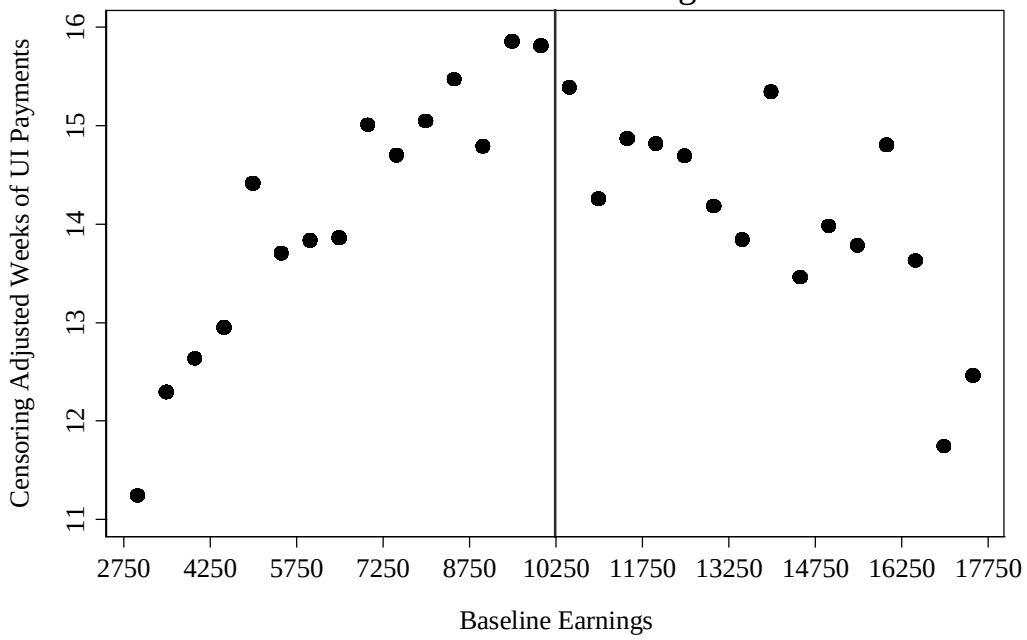
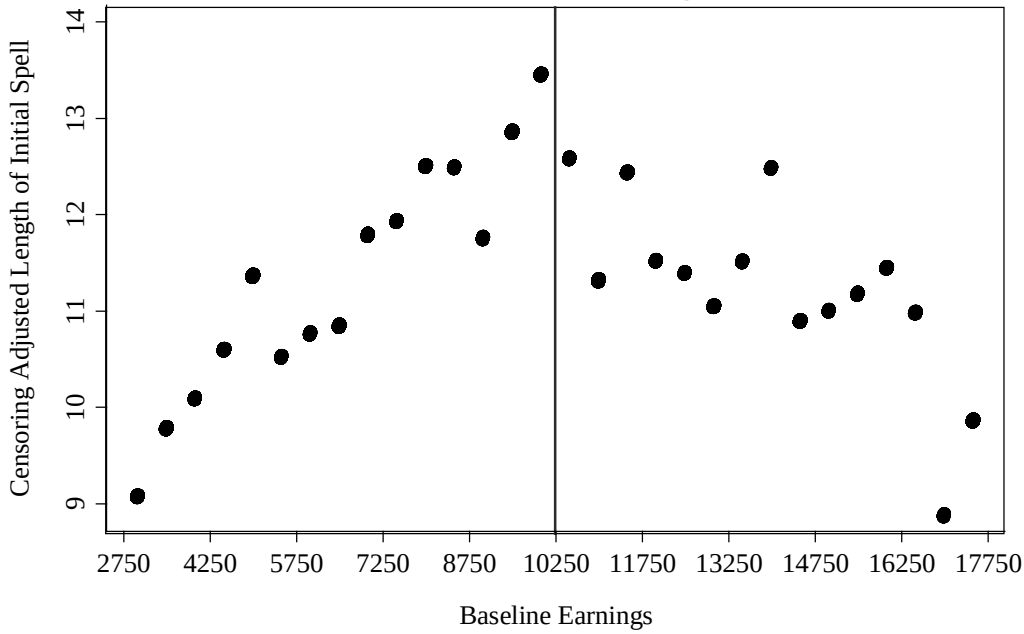


Figure 3.8(d): Censoring Adjusted Weeks of UI Claimed vs Baseline Earnings



Each point in 8(c) represents the local average of the length of the initial spell in the benefit year for claimants in each baseline earnings bin, and likewise in 8(d) the weeks of UI payment claimed, censoring adjusted. See section 4.1 for a discussion of the censoring adjustment procedure. The bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of $10,250.

Figure 3.8(e): Censoring Adjusted Length of Initial Spell vs Baseline Earnings



Each point in 8(e) represents the local average of the length of initial spell, censoring adjusted, for claimants in each baseline earnings bin. See section 4.1 for a discussion of the censoring adjustment procedure. The baseline earnings bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of $10,250.

Figure 3.9(a): Total UI Benefit Payments vs Baseline Earnings Controlling for Covariates

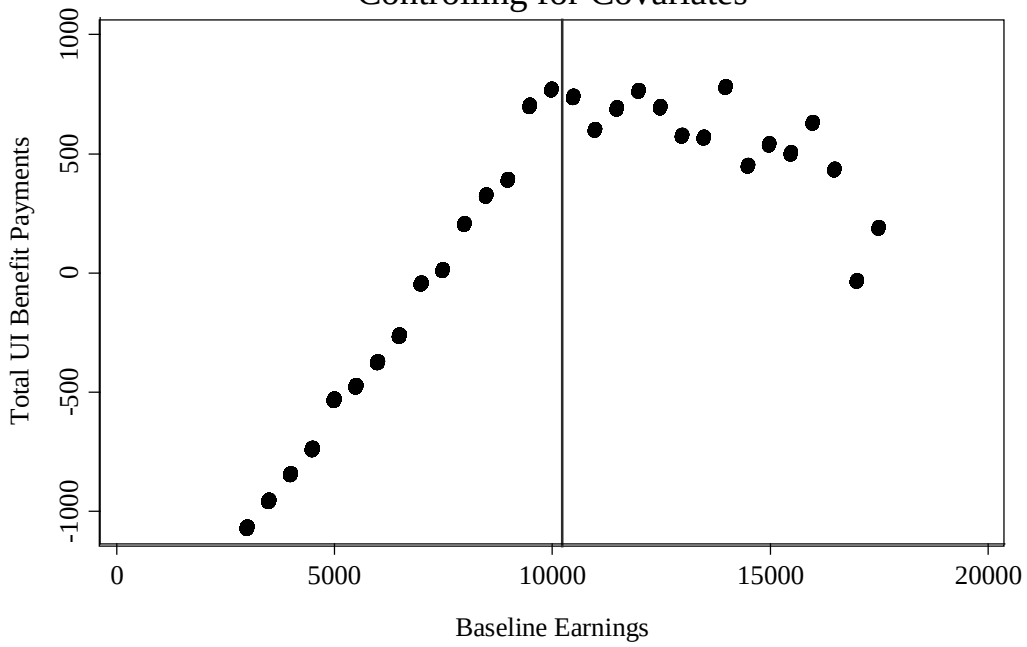
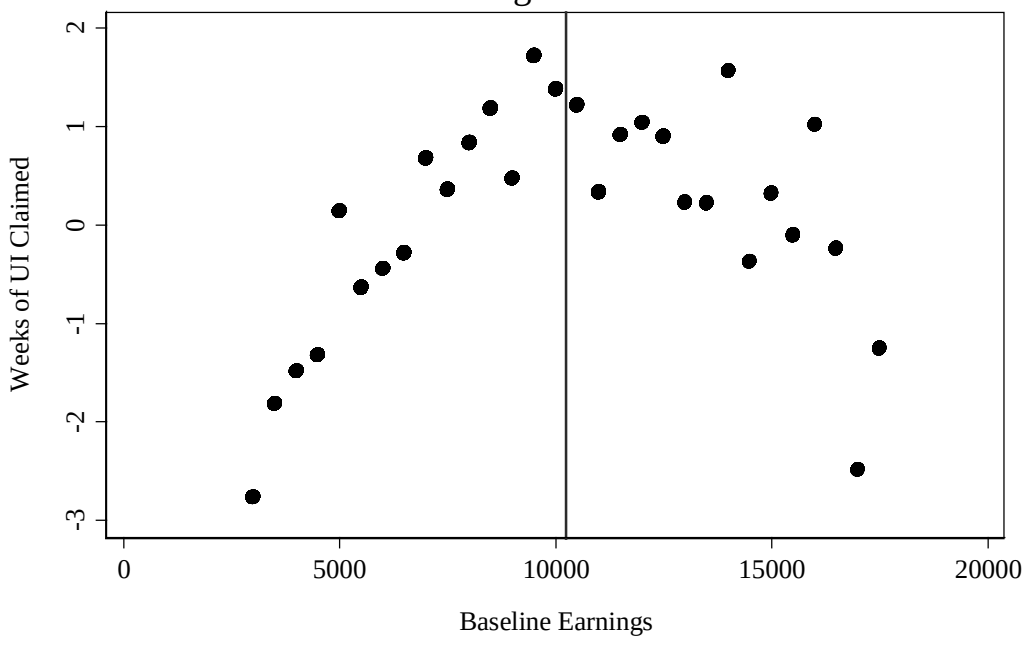


Figure 3.9(b): Weeks of UI Claimed vs Baseline Earnings Controlling for Covariates



Each point in 9(a) represents the local average of the total UI benefit received over the benefit year in each bin of baseline earnings, controlling for covariates, and likewise in 9(b) the weeks of UI claimed. The earnings bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of $10,250.

Figure 3.9(c): Length of Initial Spell vs Baseline Earnings Controlling for Covariates

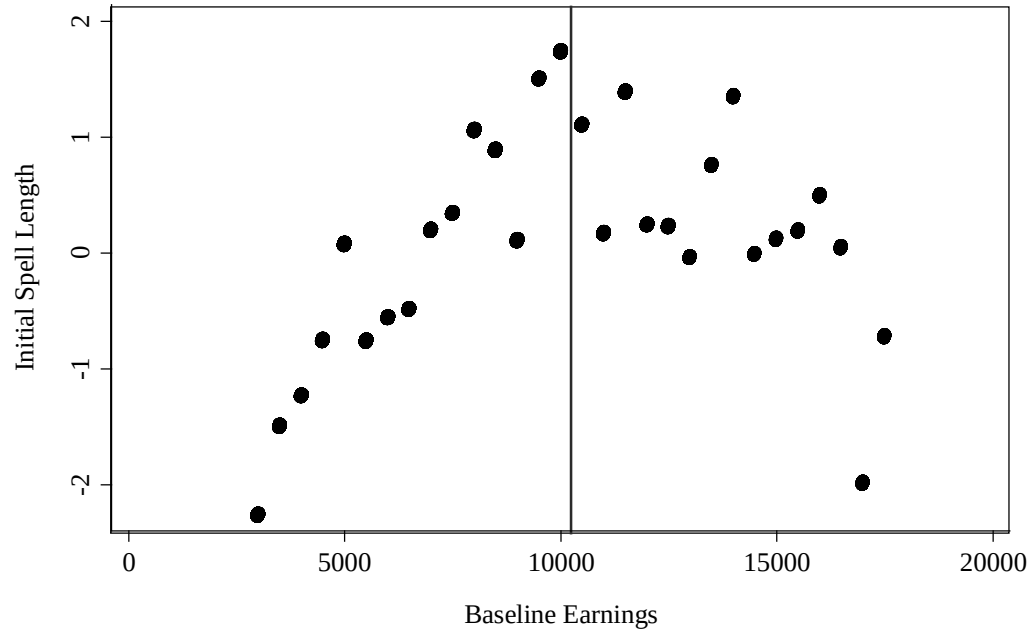
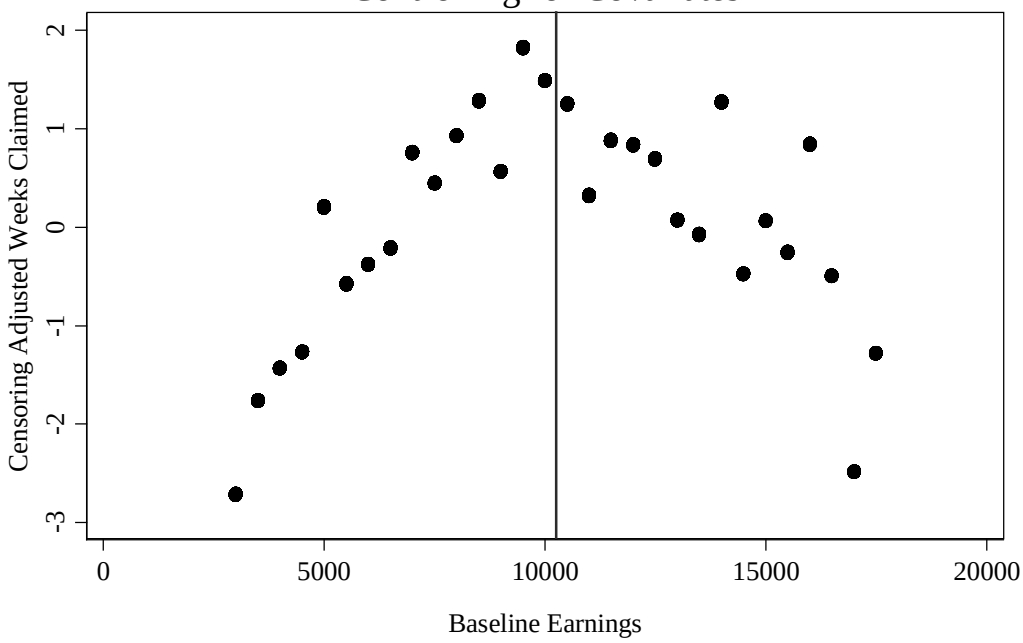
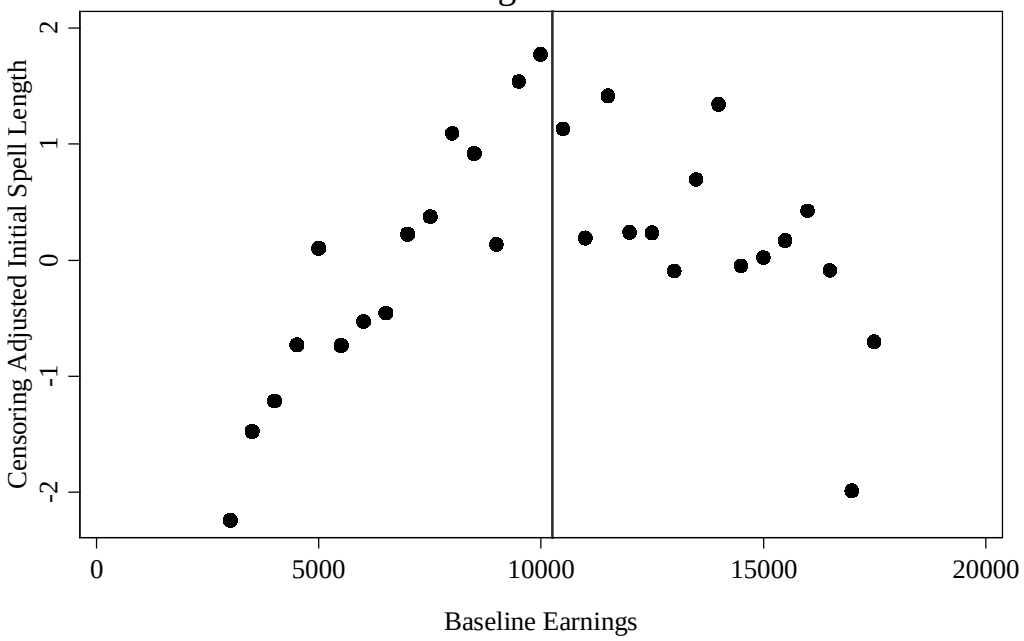


Figure 3.9(d): Censoring Adjusted Weeks Claimed vs Baseline Earnings Controlling for Covariates



Each point in 9(c) represents the local average of the initial spell length controlling for covariates in each bin of baseline earnings, and likewise in 9(d) the length of the weeks of UI claim, censoring adjusted, controlling for covariates. The earnings bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of $10,250.

Figure 3.9(e): Censoring Adjusted Length of Initial Spell vs Baseline Earnings Controlling for Covariates



Each point in 9(e) represents the local average of the censoring adjusted length of the initial spell, and controlling for covariates, in the benefit year in each bin of baseline earnings. See section 4.1 for a discussion of the censoring adjustment procedure. The earnings bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of $10,250.

Table 3.1: Summary Statistics

Variable	Mean	SD	Min	Median	Max	N
Baseline Earnings	9610.58	7176.64	828.53	7900.15	173406.10	15444
Weekly Benefit Amount	152.92	51.90	55	160	209	15444
Maximum Insured Unemployment Duration	26.86	4.17	10	30	30	15444
Total Unemployment Benefit Received During Benefit Year	2044.34	1868.74	0	1512	6488	15444
Number of Weeks claimed for Unemployment Benefit	14.09	10.80	0	13	51	15444
Length of Initial Spell in Weeks	11.28	10.72	0	7	52	15444
No. of Weeks claimed for Unemployment Benefit, Censoring Adj	13.98	10.69	0	13	51	15444
Length of Initial Spell in Weeks, Censoring Adjusted	11.26	10.69	0	7	52	15444
Age	36.24	11.27	15.7	34.2	87.9	15444
Male	0.608	0.488	0	1	1	15444
White	0.839	0.367	0	1	1	15444
Education	12.355	2.674	1	12	19	15444
Claimant filed at a Job Service Center in Western Washington	0.701	0.458	0	1	1	15444

Note that the initial spell includes the waiting week, but the total number of weeks claimed for unemployment benefit does not.

Table 3.2: RK estimates for Outcomes of Interest: Baseline Specification

Bandwidth	Polynomial of Order	Total UI Received	Total Weeks Claimed	Initial Spell Length	Censoring Adjusted Total Weeks Claimed	Censoring Adjusted Intial Spell Length
7500	One	18.1	0.0441	0.0415	0.0461	0.0423
N=13605		(0.9)	(0.0049)	(0.0049)	(0.0048)	(0.0048)
		[0.396]	[0.091]	[0.301]	[0.080]	[0.302]
	Two	16.8	0.0050	0.0343	0.0138	0.0365
		(3.3)	(0.0187)	(0.0188)	(0.0185)	(0.0187)
		[0.514]	[0.394]	[0.378]	[0.417]	[0.385]
	Three	23.9	0.0867	0.1196	0.0968	0.1197
		(8.4)	(0.0462)	(0.0467)	(0.0458)	(0.0466)
		[0.675]	[0.602]	[0.711]	[0.642]	[0.702]
	Optimal Poly Order (AIC)	1	2	1	2	1
2500	One	17.0	0.0296	0.0489	0.0370	0.0496
N=4426		(4.3)	(0.0234)	(0.0238)	(0.0232)	(0.0238)
		[0.732]	[0.700]	[0.490]	[0.690]	[0.489]
	Two	36.1	0.1397	0.2041	0.1398	0.2044
		(17.6)	(0.0947)	(0.0973)	(0.0940)	(0.0972)
		[0.847]	[0.844]	[0.542]	[0.848]	[0.533]
	Three	74.2	0.3718	0.4215	0.3803	0.4180
		(43.2)	(0.2317)	(0.2382)	(0.2303)	(0.2380)
		[0.940]	[0.932]	[0.716]	[0.942]	[0.704]
	Optimal Poly Order (AIC)	1	1	1	1	1
1000	One	28.9	0.1133	0.1603	0.1181	0.1605
N=1786		(17.2)	(0.0909)	(0.0936)	(0.0905)	(0.0935)
		[0.984]	[0.985]	[0.909]	[0.986]	[0.908]
	Two	143.9	0.6556	0.9708	0.6420	0.9676
		(68.3)	(0.3644)	(0.3746)	(0.3631)	(0.3745)
		[0.495]	[0.735]	[0.867]	[0.774]	[0.871]
	Three	53.6	0.1747	0.0046	0.1665	0.0116
		(173.6)	(0.9067)	(0.9301)	(0.9038)	(0.9300)
		[0.207]	[0.401]	[0.425]	[0.422]	[0.426]
	Optimal Poly Order (AIC)	1	1	1	1	1

For each polynomial-bandwidth specification, the first row reports the RKD estimate, the second row the robust standard error, and the third row the p-value from the Goodness of Fit test. For each bandwidth, the optimal order of polynomial is indicated.

Table 3.3: RK estimates for Outcomes of Interest: Specification with Covariates

					Censoring Adjusted	
Bandwidth	Polynomial of Order	Total UI Received	Total Weeks Claimed	Initial Spell Length	Total Weeks Claimed	Intial Spell Length
7500	One	18.1	0.0450	0.0370	0.0469	0.0378
N=13605		(0.8)	(0.0048)	(0.0047)	(0.0048)	(0.0047)
		[0.325]	[0.139]	[0.332]	[0.139]	[0.334]
	Two	16.7	0.0082	0.0255	0.0170	0.0277
		(3.3)	(0.0182)	(0.0181)	(0.0180)	(0.0180)
		[0.570]	[0.374]	[0.382]	[0.397]	[0.383]
	Three	22.2	0.0709	0.1050	0.0807	0.1052
		(8.1)	(0.0450)	(0.0447)	(0.0445)	(0.0446)
		[0.698]	[0.515]	[0.627]	[0.553]	[0.612]
	Optimal Poly Order (AIC)	1	2	1	2	1
2500	One	16.0	0.0234	0.0390	0.0307	0.0397
N=4426		(4.2)	(0.0227)	(0.0226)	(0.0225)	(0.0226)
		[0.818]	[0.745]	[0.421]	[0.729]	[0.417]
	Two	31.1	0.1167	0.1740	0.1169	0.1743
		(17.1)	(0.0921)	(0.0923)	(0.0913)	(0.0922)
		[0.786]	[0.728]	[0.396]	[0.737]	[0.384]
	Three	50.8	0.2472	0.2098	0.2563	0.2061
		(42.1)	(0.2257)	(0.2263)	(0.2242)	(0.2261)
		[0.847]	[0.779]	[0.479]	[0.798]	[0.464]
	Optimal Poly Order (AIC)	1	1	1	1	1
1000	One	21.0	0.0696	0.0872	0.0741	0.0872
N=1786		(16.9)	(0.0890)	(0.0891)	(0.0886)	(0.0890)
		[0.864]	[0.785]	[0.957]	[0.801]	[0.956]
	Two	130.1	0.5439	0.7429	0.5290	0.7397
		(66.8)	(0.3544)	(0.3544)	(0.3527)	(0.3543)
		[0.264]	[0.401]	[0.841]	[0.442]	[0.845]
	Three	45.1	0.1718	-0.0188	0.1608	-0.0107
		(168.7)	(0.8740)	(0.8740)	(0.8704)	(0.8738)
		[0.077]	[0.147]	[0.280]	[0.160]	[0.281]
	Optimal Poly Order (AIC)	1	1	1	1	1

For each polynomial-bandwidth specification, the first row reports the RKD estimate, the second row the robust standard error, and the third row the p-value from the Goodness of Fit test. For each bandwidth, the optimal order of polynomial is indicated.

Table 3.4: RK Estimates for Covariates

Bandwidth	Polynomial of Order	Age	Education	Male	White	West
7500	One	0.0017	0.0108	-0.0002	0.0007	0.0011
N=13605		(0.0047)	(0.0011)	(0.0002)	(0.0001)	(0.0002)
		[0.145]	[0.074]	[0.000]	[0.891]	[0.003]
	Two	-0.0249	0.0146	-0.0009	0.0014	0.0029
		(0.0177)	(0.0041)	(0.0008)	(0.0006)	(0.0007)
		[0.196]	[0.099]	[0.202]	[0.916]	[0.011]
	Three	0.0320	-0.0035	-0.0025	0.0014	0.0002
		(0.0429)	(0.0100)	(0.0019)	(0.0014)	(0.0017)
		[0.203]	[0.372]	[0.320]	[0.904]	[0.074]
	Optimal Poly Order (AIC)	1	1	2	1	3
2500	One	-0.0021	0.0068	-0.0009	0.0008	0.0012
N=4426		(0.0215)	(0.0049)	(0.0010)	(0.0007)	(0.0009)
		[0.733]	[0.115]	[0.281]	[0.967]	[0.372]
	Two	-0.0513	0.0114	-0.0009	-0.0003	0.0049
		(0.0860)	(0.0198)	(0.0040)	(0.0027)	(0.0035)
		[0.491]	[0.084]	[0.186]	[0.701]	[0.330]
	Three	-0.1272	0.0708	-0.0212	0.0023	0.0117
		(0.2052)	(0.0477)	(0.0097)	(0.0066)	(0.0085)
		[0.378]	[0.381]	[0.620]	[0.437]	[0.581]
	Optimal Poly Order (AIC)	1	1	1	1	1
1000	One	-0.0511	0.0280	-0.0073	0.0008	0.0064
N=1786		(0.0808)	(0.0190)	(0.0039)	(0.0026)	(0.0034)
		[0.122]	[0.399]	[0.066]	[0.156]	[0.085]
	Two	-0.1687	0.0236	-0.0273	0.0063	0.0011
		(0.3199)	(0.0766)	(0.0152)	(0.0104)	(0.0132)
		[0.122]	[0.355]	[0.055]	[0.154]	[0.605]
	Three	-0.3036	-0.1795	0.0476	-0.0222	-0.0018
		(0.8119)	(0.1892)	(0.0392)	(0.0251)	(0.0333)
		[0.078]	[0.764]	[0.298]	[0.392]	[0.521]
	Optimal Poly Order (AIC)	1	1	1	1	2

For each polynomial-bandwidth specification, the first row reports the RKD estimate, the second row the robust standard error, and the third row the p-value from the Goodness of Fit test. For each bandwidth, the optimal order of polynomial is indicated. "West" indicates that the Job Service Center at which the claimant files a claim is in western Washington.

Bibliography

Abowd, John M. and Martha Stinson, "Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Survey and SSA Administrative Data:," Technical Report 2007.

Almond, Douglas, Joseph J. Doyle, Amanda Ellen Kowalski, and Heidi L. Williams, "Estimating Marginal Returns to Medical Care: Evidence from At-Risk Newborns," *The Quarterly Journal of Economics*, 2010, *125* (2), 591–634.

_ **, Jr. Joseph J. Doyle, Amanda E. Kowalski, and Heidi Williams**, "The Role of Hospital Heterogeneity in Measuring Marginal Returns to Medical Care: A Reply to Barreca, Guldi, Lindo, and Waddell," *Quarterly Journal of Economics*, November 2011, *126* (4).

Altonji, J.G. and R.L. Matzkin, "Cross section and panel data estimators for nonseparable models with endogenous regressors," *Econometrica*, 2005, *73:4*, 1053–1102.

Altonji, Joseph G. and Christina H. Paxson, "Labor Supply, Hours Constraints, and Job Mobility," *The Journal of Human Resources*, 1992, *27* (2), pp. 256–278.

Angrist, Joshua D. and Victor Lavy, "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, May 1999, *114* (2), 533–575.

Asadullah, M. Niaz, "The Effect of Class Size on Student Achievement: Evidence from Bangladesh," *Applied Economic Letters*, March 2005, *12* (4), 217–221.

Ashenfelter, Orley, "Estimating the Effect of Training Programs on Earnings," *The Review of Economics and Statistics*, 1978, *60* (1), pp. 47–57.

_ , "Unemployment as Disequilibrium in a Model of Aggregate Labor Supply," *Econometrica*, 1980, *48* (3), pp. 547–564.

_ **and Alan Krueger**, "Estimates of the Economic Return to Schooling from a New Sample of Twins," *The American Economic Review*, 1994, *84* (5), pp. 1157–1173.

_ **and David Card**, "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs," *The Review of Economics and Statistics*, 1985, *67* (4), pp. 648–660.

Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell, "Heaping-Induced Bias in Regression-Discontinuity Designs," Working Paper 17408, National Bureau of Economic Research September 2011.

_ **, Melanie Guldi, Jason M. Lindo, and Glen R. Waddell**, "Saving Babies? Revisiting the effect of very low birth weight classification," *The Quarterly Journal of Economics*, 2011.

Black, Dan A., Mark C. Berger, and Frank A. Scott, "Bounding Parameter Estimates with Nonclassical Measurement Error," *Journal of the American Statistical Association*, 2000, *95* (451), pp. 739–748.

Blackorby, Charles and David Donaldson, "Cash versus Kind, Self-Selection, and Efficient Transfers," *The American Economic Review*, 1988, *78* (4), pp. 691–700.

Blank, Rebecca M., "The Effect of Medical Need and Medicaid on AFDC Participation," *The Journal of Human Resources*, 1989, *24* (1), pp. 54–87.

Blinder, Alan S. and Harvey S. Rosen, "Notches," *The American Economic Review*, 1985, *75* (4), pp. 736–747.

Blundell, Richard and James L. Powell, "Endogeneity in Nonparametric and Semiparametric Regression Models," in Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, eds., *Advances in economics and econometrics theory and applications : Eighth World Congress.*, Vol. II of *Econometric Society monographs no. 36*, Cambridge: Cambridge University Press, 2003, pp. 312–357.

Bound, John, Charles Brown, and Nancy Mathiowetz, "Measurement error in survey data," in J.J. Heckman and E. Leamer, eds., *Handbook of Econometrics*, Vol. 5, Elsevier, 2001, chapter Chapter 59, pp. 3705 – 3843.

Brewer, Mike, Emmanuel Saez, and Andrew Shephard, "Means-testing and tax rate on earnings," in J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chot, M. Gammie, P. Johnson, G. Myles, and J. Poterba, eds., *Dimensions of Tax Design: The Mirrlees Review*, Oxford: Oxford University Press for the Institute for Fiscal Studies, 2010, chapter 2.

Browning, Martin, Angus Deaton, and Margaret Irish, "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle," *Econometrica*, 1985, *53* (3), pp. 503–544.

Butucea, Cristina and Catherine Matias, "Minimax Estimation of the Noise Level and of the Deconvolution Density in a Semiparametric Convolution Model," *Bernoulli*, 2005, *11* (2), pp. 309–340.

Card, David and Lara D. Shore-Sheppard, "Using Discontinuous Eligibility Rules to Identify the Effects of the Federal Medicaid Expansions on Low-Income Children," *The Review of Economics and Statistics*, 2004, *86* (3), 752–766.

_ , Andrew K. G. Hildreth, and Lara D. Shore-Sheppard, "The Measurement of Medicaid Coverage in the SIPP: Evidence from a Comparison of Matched Records," *Journal of Business & Economic Statistics*, 2004, *22* (4), 410–420.

Carroll, R.J., D. Ruppert, and L.A. Stefanski, *Measurement error in nonlinear models : a modern perspective*, 2 ed., Chapman & Hall/CRC, 2006.

Chesher, A., "Identification in Nonseparable Models," *Econometrica*, 2003, *71(5)*, 1405–1441.

Chetty, Raj, "A New Method of Estimating Risk Aversion," *The American Economic Review*, 2006, *96* (5), 1821–1834.

_ , "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, Forthcoming.

_ **, John N. Friedman, Tore Olsen, and Luigi Pistaferri**, "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *The Quarterly Journal of Economics*, 2011, *126* (2), 749–804.

CMS, "Title XXI Program Fact Sheets," Technical Report, Center for Medicare and Medicaid Services Various Years.

Cremer, Helmuth and Firouz Gahvari, "In-kind transfers, self-selection and optimal tax policy," *European Economic Review*, 1997, *41* (1), 97 – 114.

Currie, Janet, "The Take Up of Social Benefits," Working Paper 10488, National Bureau of Economic Research May 2004.

_ **and Firouz Gahvari**, "Transfers in Cash and In-Kind: Theory Meets the Data," *Journal of Economic Literature*, 2008, *46* (2), pp. 333–383.

_ **and Jeffrey Grogger**, "Explaining Recent Declines in Food Stamp Program Participation," *Brookings-Wharton Papers on Urban Affairs*, 2001, pp. pp. 203–244.

de la Mata, Dolores, "The effect of Medicaid on Children's Health: a Regression Discontinuity Approach," Working Paper 11/16, HEDG, The University of York July 2011.

Dickens, William T. and Shelly J. Lundberg, "Hours Restrictions and Labor Supply," *International Economic Review*, 1993, *34* (1), pp. 169–192.

Dobkin, Carlos and Fernando Ferreira, "Do school entry laws affect educational attainment and labor market outcomes?," *Economics of Education Review*, 2010, *29* (1), 40 – 54.

Edgeworth, F. Y., "The Pure Theory of Taxation," *The Economic Journal*, 1897, *7* (25), pp. 46–70.

Fairbrother, Gerry, Gowri Madhavan, Anthony Goudie, Joshua Watring, Rachel A. Sebastian, Lorin Ranbom, and Lisa A. Simpson, "Reporting on Continuity of Coverage for Children in Medicaid and CHIP: What States Can Learn from Monitoring Continuity and Duration of Coverage," *Academic Pediatrics*, 2011, *11* (4), 318 – 325.

Florens, J. P., J. J. Heckman, C. Meghir, and E. Vytlacil, "Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, 2009, *76*, 1191–1206.

Gahvari, Firouz, "In-kind Versus Cash Transfers In The Presence Of Distortionary Taxes," *Economic Inquiry*, 1995, *33* (1), 45–53.

Grinstead, Charles M. and J. Laurie Snell, *Introduction to Probability*, second revised ed., American Mathematical Society, July 1997.

Gruber, Jonathan, "Medicaid," in Robert A. Moffitt, ed., *Means-Tested Transfer Programs in the United States*, The University of Chicago Press, 2003, chapter 1.

_ **and Kosali Simon**, "Crowd-out 10 years later: Have recent public insurance expansions crowded out private health insurance?," *Journal of Health Economics*, 2008, *27* (2), 201 – 217.

Guryan, Jonathan, "Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts," Working Paper WP8269, National Bureau of Economic Research 2001.

_ , "Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts," Working Paper WP8269, National Bureau of Economic Research 2003.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, "Identification and estimation of treatment effects with a regression-discontinuity design," *Econometrica*, 2001, *69(1)*, 201–209.

Ham, John C., "Estimation of a Labour Supply Model with Censoring Due to Unemployment and Underemployment," *The Review of Economic Studies*, 1982, *49* (3), pp. 335–354.

_ **, Xianghong Li, and Lara Shore-Sheppard**, "Seam Bias, Multiple-State, Multiple-Spell Duration Models and the Employment Dynamics of Disadvantaged Women," Working Paper 15151, National Bureau of Economic Research July 2009.

Hamermesh, Daniel S., *Jobless Pay and the Economy*, Baltimore: The Johns Hopkins University Press, 1977.

Hausman, Jerry A., Whitney K. Newey, Hidehiko Ichimura, and James L. Powell, "Identification and estimation of polynomial errors-in-variables models," *Journal of Econometrics*, 1991, *50* (3), 273 – 295.

Heckman, J. J. and R. Robb, "Alternative Methods for Evaluating the Impact of Interventions," *Journal of Econometrics*, 1985, *30*, 239–267.

Hullegie, Patrick and Tobias J. Klein, "The effect of private health insurance on medical care utilization and self-assessed health in Germany," *Health Economics*, 2010, *19* (9), 1048–1062.

Imbens, Guido W. and Whitney K. Newey, "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 2009, *77(5)*, 1481–1512.

Irvin, Carol, Deborah Peikes, Chris Trenholm, and Nazmul Khan, "Discontinuous Coverage in Medicaid and the Implications of 12-Month Continuous Coverage for Children: Final Report," Technical Report, Mathematica Policy Research 2001.

Jacob, Brian A. and Lars Lefgren, "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis," *The Review of Economics and Statistics*, 2004, *86* (1), 226–244.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan, "Earnings Losses of Displaced Workers," *The American Economic Review*, 1993, *83* (4), pp. 685–709.

Kabbani, Nader S. and Parke E. Wilde, "Short Recertification Periods in the U.S. Food Stamp Program," *The Journal of Human Resources*, 2003, *38*, pp. 1112–1138.

Kahn, Shulamit and Kevin Lang, "The Effect of Hours Constraints on Labor Supply Estimates," *The Review of Economics and Statistics*, 1991, *73* (4), pp. 605–611.

Kaiser, "A 50 State Update on Eligibility Rules, Enrollment and Renewal Procedures, and Cost-Sharing Practices in Medicaid and SCHIP," Technical Report, The Henry J. Kaiser Family Foundation 2002-2011.

Kleven, Henrik Jacobsen and Mazhar Waseem, "Tax Notches in Pakistan: Tax Evasion, Real Responses, and Income Shifting," Working Paper, London School of Economics 2011.

Koch, Thomas G., "Using RD Design to Understand Heterogeneity in Health Insurance Crowd-Out," Technical Report September 2010.

Kodde, D. A., F. C. Plam, and G. A. Pfann, "Asymptotic least-squares estimation efficiency considerations and applications," *Journal of Applied Econometrics*, 1990, *5* (3), 229–243.

Kornfeld, Robert, "Explaining Recent Trends in Food Stamp Program Caseload," E-FAN-02-008, U.S. Department of Agriculture, Economic Research Service 2002.

Lee, David S., "Randomized Experiments from Non-random Selection in U.S. House Elections," *Journal of Econometrics*, February 2008, *142* (2), 675–697.

_ **and David Card**, "Regression discontinuity inference with specification error," *Journal of Econometrics*, 2008, *142* (2), 655 – 674. The regression discontinuity design: Theory and applications.

_ **and Thomas Lemieux**, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, June 2010, *48*, 281–355.

_ **and** _ , "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, forthcoming.

Li, Tong and Quang Vuong, "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 1998, *65* (2), 139 – 165.

MaCurdy, Thomas E., "An Empirical Model of Labor Supply in a Life-Cycle Setting," *Journal of Political Economy*, 1981, *89* (6), pp. 1059–1085.

Marquis, Kent H. and Jefferey C. Moore, "Measurement Errors in SIPP Program Reports," Technical Report, U.S. Census Bureau 1990.

Matsudaira, J., "Mandatory Summer School and Student Achievement," *Journal of Econometrics*, February 2008, *142* (2), 829–850.

McCrary, Justin, "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, February 2008, *142* (2), 698–714.

Meyer, Bruce D., "Unemployment Insurance and Unemployment Spells," *Econometrica*, 1990, *58:4*, 757–782.

Mirrlees, J. A., "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 1971, *38* (2), pp. 175–208.

Moffitt, Robert, "An Economic Model of Welfare Stigma," *The American Economic Review*, 1983, *73* (5), pp. 1023–1035.

_ **and Walter Nicholson**, "The Effect of Unemployment Insurance on Unemployment: The Case of Federal Supplemental Benefits," *The Review of Economics and Statistics*, 1982, *64:1*, 1–11.

NGA, "Maternal and Child Health Update," Technical Report, National Governors Association 2000-2008.

Nichols, Albert L. and Richard J. Zeckhauser, "Targeting Transfers through Restrictions on Recipients," *The American Economic Review*, 1982, *72* (2), pp. 372–377.

Nielsen, Helena S., Torben Sorensen, and Christopher R. Taber, "Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform," NBER Working Paper WP14535, National Bureau of Economic Research 2009.

Nielsen, Helena Skyt, Torben Sørensen, and Christopher R. Taber, "Estimating the Effect of Student Aid of College Enrollment: Evidence from a Government Grant Policy Reform," Technical Report, NBER WP 14535 2008.

_ **,** _ **, and Christopher Taber**, "Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform," *American Economic Journal: Economic Policy*, forthcoming.

Olson, Lynn M., Suk fong S. Tang, and Paul W. Newacheck, "Children in the United States with Discontinuous Health Insurance Coverage," *New England Journal of Medicine*, 2005, *353* (4), 382–391.

Pischke, Jörn-Steffen, "Individual Income, Incomplete Information, and Aggregate Consumption," *Econometrica*, 1995, *63* (4), pp. 805–840.

Saez, Emmanuel, "Do Taxpayers Bunch at Kink Points?," Working Paper 7366, National Bureau of Economic Research, September 1999.

_ , "Do Taxpayers Bunch at Kink Points?," Working Paper, NBER-TAPES Conference, 2002.

_ , "Do Taxpayers Bunch at Kink Points?," *American Economic Journal: Economic Policy*, 2010, *2(3)*, 180 – 212.

_ , "Do Taxpayers Bunch at Kink Points?," *American Economic Journal: Economic Policy*, forthcoming.

Salanie, Bernard, *The Economics of Taxation*, The MIT Press, 2003.

Satchachai, Panutat and Peter Schmidt, "GMM with more moment conditions than observations," *Economics Letters*, 2008, *99* (2), 252 – 255.

Schanzenbach, Diane Whitmore, "Do School Lunches Contribute to Childhood Obesity?," *J. Human Resources*, 2009, *44* (3), 684–709.

Schennach, Susanne M., "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 2004, *72* (1), pp. 33–75.

Schwarz, Maik and Sebastien Van Bellegem, "Consistent density deconvolution under partially known error distribution," *Statistic and Probability Letters*, 2010, *80*, 236–241.

Simonsen, Marianne, Lars Skipper, and Niels Skipper, "Price Sensitivity of Demand of Prescription Drugs: Exploiting a Kinked Reimbursement Scheme," Technical Report November 2009.

Singh, Nirvikar and Ravi Thomas, "Welfare Policy: Cash versus Kind, Self-Selection and Notches," *Southern Economic Journal*, 2000, *66* (4), pp. 976–990.

Slemrod, Joel, "Buenas Notches: Lines and Notches in Tax System Design," Working Paper September 2010. First Draft: March 2010.

Snyder, Stephen E. and William N. Evans, "The Effect of Income on Mortality: Evidence from the Social Security Notch," *Review of Economics and Statistics*, 2006, *88(3)*, 482–495.

Spiegelman, Robert G., Christopher J. O'Leary, and Kenneth J. Kline, "The Washington Reemployment Bonus Experiment Final Report," Technical Report, W.E. Upjohn Institute for Employment Research 1992.

Stata Press, *Base Reference Manual: Stata 11 Documentation* 2010.

Thistlethwaite, Donald L. and Donald T. Campbell, "Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment," *Journal of Educational Psychology*, December 1960, *51*, 309–317.

Urquiola, Miguel, "Identifying Class Size Effects in Developing Countries: Evidence from Rural Bolivia," *Review of Economics and Statistics*, 2006, *88*, 171–177.

_ **and Eric Verhoogen**, "Class-Size Caps, Sorting, and the Regression-Discontinuity Design," *American Economic Review*, 2009, *99(1)*, 179–215.

Yelowitz, Aaron S., "The Medicaid Notch, Labor Supply, and Welfare Participation: Evidence from Eligibility Expansions," *The Quarterly Journal of Economics*, 1995, *110(4)*, 909–939.