The Tax Treatment of Marriage and Its Impact on Family Formation and Labor Supply

Elliott William Isaac
Libertyville, Illinois

B.A. Economics, Wake Forest University, 2011
M.A. Economics, University of Virginia, 2013

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Economics

University of Virginia
April 2018

Abstract

Many developed countries institute progressive tax systems, forcing them to choose between tax equity across marital status or tax equity across family income. Greater equity across family income (i.e., jointness) can exacerbate efficiency costs and labor supply distortions. Another complication in the United States is that tax changes generally involve a change in marriage incentives as well because the United States' tax system creates inequalities between married and unmarried couples. This feature of the United States' tax system, combined with a transfer system targeted largely at single parents, creates an intricate web of family structure incentives: transfer programs for the needy often discourage marriage, while the tax system discourages marriage for some and encourages marriage for others. In this dissertation, I study the tax treatment of the family and its effects on labor supply, marriage, and divorce, in addition to the role of college selectivity in women's marriage and career outcomes.

Chapter 1 examines the impact of the tax and transfer systems on family structure decisions. I use variation from the 1990s in the Earned Income Tax Credit and welfare reform to estimate the effects on marrying and divorcing. I examine flows into and out of marriage, use test scores to predict who is most likely to be affected by the policy changes, and employ a flexible functional form to estimate heterogeneous effects. I find that low-earning single parents are more likely to marry due to the EITC expansion and lower welfare generosity, while mid-earning married parents are less likely to divorce and high-earning married parents are more likely to divorce due to the EITC expansion.

Chapter 2 provides direct evidence of the efficiency costs, tax revenue consequences, and labor supply effects of joint taxation in the United States by leveraging tax variation created by federal same-sex marriage recognition following the 2013 *United States v. Windsor* Supreme Court ruling. I find hours responses to taxation among predicted primary earners and labor force participation responses among predicted secondary earners. I also show that joint taxation decreases efficiency and tax revenue compared to individual taxation, with larger effect sizes for equal-earning couples. My findings suggest that there are efficiency gains to lowering tax rates for secondary earners,

but whether efficiency is worth the lower associated tax equity across households remains an open question.

Chapter 3, co-authored with Suqin Ge and Amalia Miller, estimates the causal effect of college selectivity on marriage and career outcomes for women using the College and Beyond data set and Dale and Krueger's (2002) application of selection on observables and unobservables. We find that attending a more selective college increases women's formal education, the likelihood of remaining single, the likelihood of having positive earnings, and spousal education. We also argue that college selectivity increases women's earnings through the marriage market instead of through human capital directly. Our findings suggest that college selectivity decreases the labor supply gap between married and single women, and that opting out may be operating on the intensive margin of part-time work rather than labor force participation.

JEL codes: D10, D61, H24, H31, I23, I26, I38, J12, J16, J22

Keywords: joint taxation, earned income tax credit, marriage, divorce, labor supply, same-sex marriage, sufficient statistics, college quality

Acknowledgements

A series of influential and caring teachers, role models, and friends led me to this point, and I will never be able to do enough to express my gratitude to them. I want to thank Leora Friedberg for her mentorship and guidance, and for helping me grow as an economist at each opportunity. I want to thank Amalia Miller for her advice and for challenging me to answer difficult questions, setting the standard ever higher. I want to thank Jonathan Colmer for his insight and patience, and for always taking time to offer feedback and answer questions. I am very lucky to have such thoughtful advisors, and I am glad to call myself their student. I am also grateful for the support of several other faculty members, especially Andrew Hayashi, Lee Coppock, Ed Olsen, and Gaurab Aryal.

I also want to thank my fellow students, whose support and community allowed us all to succeed. I am particularly grateful to Bill Johnson, Ben Leyden, Brett Lissenden, Cailin Slattery, and Zach Sullivan. I want to especially thank Zhou Zhang, who has been an incredible resource, role model, and friend. The strength and support of the graduate student community was one of my greatest reasons for enrolling here, and is one of the greatest reasons I will miss Charlottesville.

I want to thank my parents, Amy and John, and my brother, Jonathan, who were an unending source of support and comfort. Thank you to my father, Ron, who provided me the opportunities I needed to be here. Lastly, to Audrey, my wife, best friend, and teammate, thank you for being my sounding board and for your unconditional love, without which I would have been lost long ago.

Contents

Chapter 1

Marriage, Divorce, and Tax and Transfer Policy[1]

1.1 Introduction

The tax and transfer systems in the United States create an intricate web of family structure incentives: transfer programs for the needy often discourage marriage, while the tax system discourages marriage for some and encourages marriage for others. Substantial public debate about the role of family structure incentives in the tax and transfer systems demonstrates a public desire to encourage marriage, and recent changes to the Earned Income Tax Credit and earlier changes to the cash welfare system seek to reduce marriage disincentives. While the same individuals may be eligible for various tax credits and transfer programs depending on whether they marry, much of the literature has dealt separately with either one or the other. I take advantage of policy variations in both tax and transfer programs in the 1990s to estimate how individuals respond to the family structure incentives they face from the tax and transfer systems.

Previous quasi-experimental studies find weak evidence, at best, that individuals respond to family structure incentives contained in the U.S. tax and transfer systems. The most convincing

1. This research was conducted with restricted access to Bureau of Labor Statistics (BLS) data. The views expressed here do not necessarily reflect the views of the BLS.

of these find that there are generally small effects of the marriage tax penalty/subsidy on the probability of marrying, and little to no effect on the probability of divorcing.[2] Many studies in this literature rely on repeated cross sections of data, which do not allow the researchers to distinguish between previously married and unmarried individuals. This aspect of the data combines individuals who were able to respond to the policy by entering into marriage with those who were not, which biases the estimated effect of the tax and transfer systems toward zero due, for instance, to already married individuals appearing to not respond to the policy. Additionally, cross-sectional analysis blurs the marriage and divorce response margin, because if a cross-section shows an increase in the probability of being married it could result either from more people marrying or from fewer divorcing. In essence, the issue is that marriage entry and exit are not necessarily symmetric. Using the 1991–1998 waves of the National Longitudinal Survey of Youth 1979 (NLSY79), I construct marriage and divorce risk samples, which include only individuals who were eligible to marry (because they were unmarried) or divorce (because they were married), respectively. I also distinguish between individuals with and without children because families with children faced substantial changes to their EITC schedules and welfare environments compared to families without children.

Another important feature of the literature is the use of education as a means to separate likely welfare or EITC recipients. For example, Eissa and Liebman (1996), Meyer and Rosenbaum (2001), Eissa and Hoynes (2003, 2004), Schoeni and Blank (2003), and Michelmore (2015) all use education to isolate likely low-earning families in at least some specifications. My approach, on the other hand, uses individuals' Armed Forces Qualification Test (AFQT) z-scores, instead of education, to identify likely welfare or EITC recipients, and uses a flexible spline specification to allow differential effects along the AFQT distribution. This approach offers more variation and identifying power in estimation while still serving the same role as education in other studies.

I employ a difference-in-differences approach, augmented with a flexible spline function in

2. To name a few, Alm and Whittington (1995a) estimate that the elasticity of marrying with respect to the tax cost of marriage is -0.012, Eissa and Hoynes (2003) estimate the elasticity of being married with respect to the tax cost of marriage to be -0.004, and Michelmore (2015) finds that a $1,000 increase in the EITC cost of marriage decreases the probability of marrying by 1.1 percentage points (8.5%).

AFQT z-scores that allows for heterogeneous effects, to estimate the effect of the 1993–1996 EITC expansion on the probability of marrying or divorcing among men and women in the NLSY79, who were 27–41 years old between 1991–1998.[3] I use the presence of at least one EITC-eligible child in the previous period as the dimension of treatment.[4] This treatment dimension captures the differential changes in family structure incentives between individuals with and without children, where individuals with children experienced dramatic increases in EITC generosity and decreases in AFDC/TANF generosity compared to childless individuals. The spline function in AFQT z-score captures the varying changes in family structure incentives between low- and high-earning individuals, and aids in circumventing the endogeneity issues associated with using reported earnings to estimate heterogenous effects; namely, that changes to EITC and AFDC/TANF program parameters cause individuals to manipulate their earnings and, thus, their statuses as treated or non-treated individuals.[5] I also use the marriage and birth histories of women in the 1995 June Current Population Survey to create an alternative panel in order to investigate the parallel trends assumption. I plot new marriage and divorce rates and use an event study approach to examine pre-trends in family structure decisions. I find evidence that the parallel trends assumption is satisfied in the marriage and divorce risk samples.

While the EITC expansion occurred at the national level, AFDC/TANF parameters are set by each state, and therefore changed differentially across states and time. In order to leverage this variation, I measure potential AFDC/TANF benefits or losses using the maximum attainable monthly AFDC/TANF payment for a family of three in the individual's state for the given year. I define AFDC/TANF eligibility using the interaction between an individual's AFQT z-score and

3. I use the term "divorcing" to refer to either a separation or a divorce, which is conventional in the literature. As noted by Whittington and Alm (1997), who study the effect of income taxes on divorce between 1969–1989, it is relatively easy to be able to file under single status or as "head-of-household" if one is separated.

4. An EITC-eligible child must satisfy specific relationship, age, and residence requirements. The child must be related to the tax filer by marriage, blood, or law, must be 18 years old or younger, and must live with the tax filer for at least half of the relevant tax year. The age requirement extends to children 24 years old or younger if the child is a full-time student. There are additional exceptions in the case that a child lives with an extended family member, such as an aunt or uncle, instead of a parent. Therefore, I maintain the assumption that fertility is exogenous, which is common in this literature.

5. The 1993–1996 EITC expansion and AFDC/TANF reform also strengthened labor supply incentives. I discuss the implications of this further in Section 1.6.

the presence of at least one child in the previous period along with an additional indicator variable equal to one if the individual is female.[6] This last interaction reflects the fact that AFDC/TANF is largely targeted at single mothers. Assuming an individual becomes ineligible for AFDC/TANF upon marriage, this variable provides a measure of what the individual would lose in welfare benefits if she were to marry. Variation in eligibility as well as benefit levels across states and time identify the effect of the AFDC/TANF transfer system generosity on the probability of marrying or divorcing.

I find that the 1993–1996 EITC expansion increased the probability of marrying among single parents who were likely EITC eligible, and created heterogenous effects on the probability of divorcing among a similar group who were initially married. I also find evidence of an effect of AFDC/TANF generosity on the probability of marrying, and no effect on the probability of divorcing. I estimate that the expansion of the EITC increased the probability of marrying between 1997–1998 (one to two years after the expansion was complete) by 5.2 percentage points (47.7%) for each standard deviation reduction in AFQT score among likely low-earning single parents. Meanwhile, I estimate that a \$100 decrease in the monthly AFDC/TANF payment for a family of three increases the probability of marrying by 0.4 percentage points (4.5%) for single mothers who are likely eligible for AFDC/TANF benefits, but this evidence is not strong. I also find that the EITC expansion decreased the probability of divorcing in 1994 by 6.0 percentage points (78.9%) for each standard deviation reduction in AFQT score among likely mid-level-earning married parents, but increased the probability of divorcing between 1997–1998 by 10.5 percentage points (96.3%) for each standard deviation reduction in AFQT score among likely high-earning married parents, which is consistent with increased marriage incentives for low-earners and increased divorce incentives for high-earners. Additionally, I find no evidence that AFDC/TANF generosity affects the probability of divorcing for married mothers who would likely be eligible for AFDC/TANF benefits upon divorce. I examine the robustness of my findings by comparing my

6. This eligibility definition is similar to the specification I use to measure the effect of the EITC, but does not take in to account specific eligibility restrictions instituted by the Personal Responsibility and Work Opportunity Reconciliation Act of 1996. Omitting these eligibility requirements likely results in labeling more individuals as being eligible for AFDC/TANF than are truly eligible.

results to those using a sample of repeated cross sections, comparing my results to those using education instead of AFQT score, and using a continuous variable for EITC generosity. I also examine the robustness of my results by estimating the model separately for men and women, where I find some evidence of differential family structure patterns between male and female parents that imply stronger marriage incentives from the EITC among men and stronger divorce incentives from the EITC among women.

These estimates suggest that recent EITC expansions, such as lengthening the plateau range for married families, may have important effects in terms of encouraging marriage among low-earning families. However, these policies likely affect single and married taxpayers differently, creating asymmetric responses for marriage flows and divorce flows as well as asymmetric responses along the income distribution.

The remainder of the paper is organized as follows. Section 1.2 discusses the relevant literature, Section 1.3 discusses policy background, Section 1.4 presents the theoretical motivation, and Section 1.5 discusses the data. Section 1.6 presents the empirical strategy and Section 1.7 presents the main results along with some alternative specifications. Finally, Section 1.8 concludes.

1.2 Relevant Literature

My research differs from most in the literature in three important ways. First, the literature concerning the family structure effects of tax and transfer systems focuses on either taxes or transfers. I connect these strands of literature by incorporating measures of both systems. In addition, many studies examining the effects of the Earned Income Tax Credit (EITC) limit or separate the data by education level as a method to differentiate between likely low- or high-earning families. I employ an alternative method by using an individual's Armed Forces Qualification Test (AFQT) score. Finally, I use longitudinal data to study transitions into and out of marriage, which distinguishes my analysis from prior cross-sectional studies. This type of analysis allows for differential responses upon beginning or ending a marriage (Eissa and Hoynes 2000).

A few studies incorporate one or two of the features that I do. Dickert-Conlin (1999) focuses on both tax and transfer incentives for divorce and uses longitudinal data. Using women from the 1990 Survey of Income and Program Participation (SIPP), she finds that tax penalties for marriage increase the probability of divorce, but that transfer penalty effects are not statistically different from zero: the marginal effect of a $1,000 increase in the tax penalty at the mean is 0.41–0.83 percentage points (15.2–30.7%).[7] The SIPP, however, contains a very short panel, and allows Dickert-Conlin to use only a short time window for analysis spanning 1990–1991. My analysis uses the NLSY79's longer panel to examine a time period of substantial policy change and allows for differential effects of the EITC expansion along the AFQT distribution.[8]

More recently, Light and Omori (2008) use a longitudinal approach with the NLSY79 to estimate a three stage model of cohabitation and marriage. They also jointly incorporate tax and transfer system incentives. The authors estimate that increasing welfare generosity decreases the probability of marrying for a representative white woman, and the size of their estimate is similar to mine.[9] Light and Omori (2008), however, use variation in state income taxation between 1974–2004, whereas I focus specifically on the federal 1993–1996 EITC expansion and welfare reform.[10]

Numerous authors have considered the effect of the EITC in influencing marriage and divorce decisions, including Ellwood (2000), Dickert-Conlin and Houser (2002), Herbst (2011), and Michelmore (2015).[11] Their studies, however, exhibit some important limitations. Herbst uses

7. The mean tax penalty is actually a $498 subsidy. The baseline probability of divorce is 2.7%.

8. I observe individuals for slightly over five years, on average, in my samples.

9. Light and Omori (2008) only report the mean and standard deviation of their AFDC/TANF payment variable. I roughly calculate the 90[th] percentile to be $mean + SD \times 1.282 = 475.44 + 177.39 \times 1.282 = 702.85$, which assumes a normal distribution. Thus, I calculate that an increase in the maximum attainable monthly AFDC/TANF payment for a family of four from the mean to the 90[th] percentile is roughly an increase of $227.41. This equates roughly to claiming that a $100 increase in welfare generosity causes a 0.49 percentage point (10%) decrease in the probability of transitioning from being single to being married.

10. The model of Light and Omori (2008) differs in a number of ways. First, it is an ordered, three stage model in which the first stage considers transitions from being single to either cohabiting or being married, the second stage considers transitions from cohabitation to either being single or being married, and the third stage considers transitions from being married to being single. My model would constitute only their first and third stages. In addition, an individual remains in their sample from 1979 through the end of her first union, whereas an individual exits my sample after the first observation of a family structure change. Finally, my sample only includes 1991–1998, whereas theirs spans 1979–2004.

11. Ellwood (2000) also incorporates welfare reform in to his analysis, although he considers this separately from

Vital Statistics data, which does not allow him to use any individual-level control variables. In the case of the pooled samples of Ellwood, Dickert-Conlin, and Houser, the authors cannot differentiate between individuals who were previously unmarried or previously married. This results in estimates that are biased toward zero, as I demonstrate later. In addition, the samples constructed by Dickert-Conlin and Houser (2002) only extend through 1995. Given that the 1993–1996 EITC expansion was not completely phased in until 1996, and that marriage and divorce decisions may take one to two years to manifest in the data, I use data through 1998. As a result, I find a statistically significant, positive effect of the EITC expansion on the probability of marrying among single individuals, whereas Dickert-Conlin and Houser (2002) find no effect. Michelmore (2015) uses SIPP data from 2001, 2004, and 2008, although the SIPP offers shorter panels than the NLSY79.

Recent research on the EITC has focused on the EITC's additional effects, including its usefulness as a poverty-reduction tool and its effects on the distribution of earnings (Hoynes and Patel 2015) and its effects on children's outcomes (Dahl and Lochner 2012; Hoynes, Miller, and Simon 2015; Bastian and Michelmore 2016). In general, researchers conclude that the EITC is effective at reducing poverty among low-earning families and that the additional family income from the EITC is beneficial to children in those families along numerous dimensions. Aside from additional outcomes, other researchers have examined whether taxpayers' understanding (or lack thereof) of the EITC structure and its incentives contributes to estimates of the EITC's impacts. Chetty and Saez (2013) find experimental evidence that explaining the incentives of the EITC had negligible impacts on an individual's EITC the following tax year. Chetty, Friedman, and Saez (2013) use a quasi-experimental approach and find that maximizing one's EITC amount by bunching at the first EITC kink point can be partially explained by knowledge diffusion from EITC-knowledgable taxpayers to others.

Others have more broadly considered the effect of the tax cost of marriage, including Alm and Whittington (1995a, 1995b), Whittington and Alm (1997), and Eissa and Hoynes (2003). They all use variation from the full tax code (including the EITC), but conduct their studies outside the

the effect of the EITC.

context of the 1993–1996 EITC expansion and AFDC reform. They estimate the effect of income taxation on the probability of marrying, the probability of being married, and the probability of divorcing, respectively. They find generally small elasticities with respect to the tax cost of marriage. As I find below, the authors find stronger responses to taxation when considering the flow into marriages than when considering the stock of marriages. The authors' results also point to the possibility of differential effects of taxation upon beginning, rather than ending, a marriage, further motivating the consideration of flows into and out of marriages instead of stocks.

Another strand of the literature considers the effects of welfare benefits on marriage rates. Results here are mixed. Bitler et al. (2004) find that AFDC waivers and the transition to TANF cause decreases in the probability of both marrying and divorcing, but find only weak evidence of an additional effect of the state's level of welfare generosity. In contrast, Schoeni and Blank (2003) find that AFDC waivers increased the percent married among women without a high school degree, and decreased the percent married among women with exactly a high school degree, again pointing to possible heterogenous effects of the policy. Teitler et al. (2009) find that current TANF participation is positively associated with the probability of marrying, whereas past TANF participation is not influential. I contribute to this strand of the literature by incorporating more specific welfare generosity measures (i.e., the dollar value of a state's maximum attainable welfare payment) and by jointly incorporating tax system effects.

Other researchers, such as Brien, Lillard, and Stern (2006), Sheran (2007), and Gemici and Laufer (2014) estimate structural models of family formation and conduct counterfactual analyses that alter the costs or benefits of marriage. My quasi-experimental approach imposes fewer assumptions on the structure of flows in and out of marriage.

One final, important feature of the literature is the use of education as a means to separate likely welfare or EITC recipients in order to circumvent endogeneity concerns. For example, Eissa and Liebman (1996), Meyer and Rosenbaum (2001), Eissa and Hoynes (2003, 2004), Schoeni and Blank (2003), and Michelmore (2015) all use education to isolate likely low-earning families in at least some specifications. My approach, on the other hand, uses individuals' Armed Forces

Qualification Test (AFQT) scores. AFQT scores offer more variation and identifying power in estimation while still circumventing endogeneity issues associated with earnings.

1.3 Policy Background

Substantial changes to programs that aid low-earning families during the early- to mid-1990s drastically altered the tax and welfare environment these families faced. Both the 1993–1996 EITC expansion and the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA), which replaced the AFDC program with TANF, sought to encourage work, discourage welfare receipt, and encourage marriage among low-earning families.[12]

1.3.1 The Earned Income Tax Credit

Unlike other tax credits, the federal EITC is refundable, meaning that even if the individual's tax liability is zero he will still receive the full amount of his credit from the government. Figure 1 displays the EITC schedules for 1996 as an example of its structure. The amount of a family's credit increases at a constant rate with each additional dollar of earned income in the phase-in range, remains constant in the plateau range, and decreases at a constant rate in the phase-out range until earnings are too high for the family to receive any credit.[13]

Over the first 10 years, the EITC experienced some modest expansions. Since then, three major policies have affected the EITC: the Tax Reform Act of 1986, the 1990 Omnibus Budget

12. I ignore some additional aspects of these policies. For example, many states experimented with AFDC waivers before implementing TANF. I do not consider these waivers nor the individual welfare receipt regulations that may accompany them, which likely leads to labeling more individuals as being eligible for AFDC/TANF than are truly eligible. To the extent that the probability of marrying or divorcing is correlated with being mistakenly labeled as being eligible for AFDC/TANF, the coefficient estimates will be biased. Additionally, the Omnibus Budget Reconciliation Act of 1993 instituted further tax code changes such as introducing higher statutory marginal tax rates of 36% and 39.6% and increased tax rates under the alternative minimum tax. I do not differentiate between these other individual aspects of the act.

13. Tax filers can receive their EITC payment in one of two ways: they may elect to receive their credit as a part of their paycheck throughout the year or as a lump sum after filing their taxes. The option to receive one's credit in installments has been allowed since 1979, but very few tax filers choose this method of recipiency and instead receive their credit as a lump sum (United States General Accounting Office 1992; Romich and Weisner 2000). I ignore this aspect of the EITC because it is unlikely to influence family structure responses.

Reconciliation Act, and the 1993 Omnibus Budget Reconciliation Act.[14] The changes contained in the 1993 act were phased in over the next three years. Table 1 displays the EITC parameters for the sample years I use in this paper. By 1998, the maximum EITC increased by $837 (58%) for families with one child and $2,245 (149%) for families with two or more children, relative to 1993. This expansion constituted the most dramatic EITC expansion to date (Nichols and Rothstein 2015).[15]

1.3.2 Aid to Families with Dependent Children and Temporary Assistance for Needy Families

The AFDC program was originally targeted at single mothers, although in 1961 the program began to provide support to two-parent families in which the principal earner became unemployed.[16] AFDC generosity varies among states and over time because the states set their own benefit levels (Moffitt 2003).[17] In addition, the benefit take-away rate as a function of earnings is generally quite high, and did not vary greatly across states under AFDC.

As a result of Personal Responsibility and Work Opportunity Reconciliation Act of 1996, TANF replaced AFDC, and drastically altered the welfare environment. New TANF policies, some of which were left up to state choice, included lifetime limits for welfare receipt, work requirements, eligibility for two-parent families, and non-cash forms for benefits. The subjective nature of many of these changes makes them difficult, if not impossible, to quantify at the family level. I focus on one measure of AFDC/TANF generosity: the maximum monthly payment available in each state.

14. The 1986 act indexed the EITC to inflation and modestly increased generosity, the 1990 act split the credit to differentiate between families with one and two or more children, and the 1993 act instituted a small EITC for childless families and greatly expanded the maximum credit for families with children. The 1990 Omnibus Budget Reconciliation Act also instituted an additional credit for families with children younger than one year old, which was repealed in the 1993 Omnibus Budget Reconciliation Act (U.S. Congress 2004, 1990).

15. In addition, six states had implemented state EITCs by 1991, for which I account in my analysis.

16. Although the AFDC Unemployed Parent (AFDC-UP) program was created in 1961, states were not required to implement it until 1988.

17. In 1991, for instance, the maximum AFDC payment for a one-parent family with two children in Mississippi was $120 per month, whereas the maximum payment in California was $694 per month.

Overall, the transfer system primarily creates marriage disincentives because the benefit levels for a single-parent family are higher than those for a two-parent family, if such benefits are available at all. TANF's more stringent requirements primarily decreased the benefit of remaining single and on welfare compared to AFDC.

1.3.3 Other Policy Changes

In addition to changes to the EITC and AFDC/TANF, there were two Medicaid expansions in 1989 and 1990 that required states to offer Medicaid coverage to pregnant women and to children in low-income families. These expansions would also have introduced marriage incentives among a similar sample of people as those affected by the 1993–1996 EITC expansion. The 1989 expansion came into effect in April 1990, and so should not introduce much (if any) bias to my estimates because my sample period begins in 1991. The 1990 expansion came into effect in July 1991, and so may introduce a small amount of bias due to the policy change.

Card and Shore-Sheppard (2004) estimate the effects of each of these Medicaid expansions on numerous health insurance outcomes using a regression discontinuity approach. They conclude that the "overall effect of the Medicaid expansions was substantially limited by low takeup rates among the newly eligible children." In addition, I re-estimate my model using only years 1992–1998, so that the 1990 Medicaid expansion would have been in effect for the entire sample period, and find qualitatively and quantitatively similar results. Thus, I find little evidence of bias in my estimated effects of the 1993–1996 EITC expansion due to the preceding Medicaid expansion.

1.4 Marriage and Divorce Predictions

As in Becker (1973, 1974), I model individuals as choosing to marry if the utility from being married is greater than the utility from being single, and analogously for divorce. Since AFDC/TANF is generally collected by single parents, an increase in AFDC/TANF generosity should unambiguously make being single more attractive. However, the incentive effects of the EITC are ambiguous.

The incentives created by the EITC vary dramatically between individuals and may even differ between individuals who are considering marrying each other. I plot six potential situations in Figure 2 an individual could face when considering marriage, which differ by the individual's number of children, the potential spouse's number of children, and the potential spouse's earnings. The figures display the difference between the individual's EITC while married and while single along a range of their own possible earnings, with positive values indicating an increase in EITC, and thus a stronger incentive to marry.[18] I derive these incentives from the EITC schedules in Table 1, and assume that a low-earning potential individual earns $5,000 per year.[19]

Note that the EITC creates marriage disincentives in some situations, such as Figures 2a, 2d, and 2f, and marriage incentives in others. Figures 2a and 2d are situations in which the individual would not move up to a more generous EITC schedule through marriage because he already lives with the family's only EITC-eligible child. Figure 2f displays large sections of marriage disincentives in 1992 and 1994, even though a second EITC-eligible child joins the family via marriage, because the one- and two-child EITC schedules were very similar. The EITC gain through marriage can be as high as $1,500–$2,200, as in Figures 2b, 2c, and 2e, and the EITC loss through marriage can be as low as $700–$900, as in Figures 2d and 2f. However, the EITC difference, whether initially positive or negative, tends to shift in favor of marriage between 1992 and 1998. Conditional on earnings, Figure 2 shows that individuals' EITC differences increased as much as $500–$1,500 during this time period.

Combining the EITC and AFDC/TANF changes increases the magnitude of the average shift in favor of marriage. To illustrate this, Figure 3 displays the change in the marriage penalty/subsidy from 1992 to 1994, 1996, and 1998 due to changes in real AFDC/TANF benefit levels and the EITC, with positive values indicating an increase in the marriage subsidy (or, equivalently, a decrease in the marriage penalty).[20] In these graphs I use the average family AFDC/TANF payment

18. Note that these graphs are from an individual's perspective, meaning they compare the family's EITC while married to the individual's EITC while single. This perspective is comparable to the perspective I use throughout the empirical strategy, and does not consider cohabitation.

19. Earning $5,000 per year during this time frame is roughly the equivalent of working part time at the minimum wage.

20. As an example, the 1994 to 1992 difference plots $(EITC_{married,1994} - EITC_{single,1994} - welfare_{1994}) -$

for the given year as the measure of welfare, and I assume that an individual loses all AFDC/TANF benefits upon marriage.[21] Conditional on earnings, by 1998, individuals' annual gains from EITC and AFDC/TANF had increased \$700–\$2,700 relative to 1992. Figures 2 and 3 lead to my empirical question: did individuals respond to these changes to their marriage and divorce incentives, and, if so, by how much?

I use the presence of at least one EITC-eligible child in the previous period and the individual's Armed Forces Qualification Test (AFQT) score as the two dimensions of treatment. I expect to see an increase in the probability of marrying over this time period among parents relative to childless individuals as a result of the 1993–1996 EITC expansion. I use the individual's AFQT score to separate likely low-earners from likely high-earners. Therefore, I expect to see an increase in the probability of marrying among individuals with low AFQT scores relative to individuals with high AFQT scores due to a higher likelihood of being eligible for the EITC. Overall, I expect to find that the EITC expansion increased the probability of marrying among parents with low AFQT scores relative to parents with high AFQT scores. I also expect to see that lower AFDC/TANF generosity increases the probability of marrying among mothers who are likely eligible for AFDC/TANF, relative to others.

1.5 Data

I use panel data from the National Longitudinal Survey of Youth 1979 (NLSY79) to estimate the marriage and divorce effects of the EITC and AFDC/TANF changes that occurred in the early- to mid-1990s. The NLSY79 allows me to observe individuals over a long time period, to separate individuals based on past marital status, and to use scores from the Armed Forces Qualification Test (AFQT) to separate likely low-earners from likely high-earners.

$(EITC_{married,1992} - EITC_{single,1992} - welfare_{1992})$ for each situation.

21. I also ignore other changes resulting from PRWORA that tightened eligibility for benefits

1.5.1 The NLSY79 Risk Samples

The data come from the 1991–1998 waves of the NLSY79, which offers a longer panel than the Survey of Income and Program Participation (SIPP) and a larger sample of the relevant age cohort than the Panel Study of Income Dynamics (PSID).[22] Through 1994 the NLSY79 surveyed individuals each year, but afterward the survey became biennial. When I begin my analysis in 1991, respondents are between 27–34 years old. By the end of the sample, the respondents are between 33–41 years old.

The marriage and divorce risk samples consist of individuals who are eligible to be married and eligible to be divorced, respectively. Individuals are included in the marriage sample if they are unmarried that year or in their first year of marriage.[23] Subsequent observations of these individuals in their second years of marriage and beyond are excluded from the marriage sample. If the individual divorces he re-enters the marriage sample.[24] I define the divorce sample analogously. Note that some individuals appear intermittently in the data, some appear regularly and then leave, and some remain in the survey for the entire duration, creating an unbalanced panel of 4,500 individuals with 16,474 observations for the full marriage risk sample and 5,640 individuals with 23,335 observations for the full divorce risk sample.

The risk samples ensure that the observations I use to estimate the tax and transfer system effects are for those individuals who can respond to the policies by changing marital status in the specified ways. Previous studies often consider an individual to have responded to the policy if he

22. I begin my analysis in 1991 due to the timing of the policy variation. I use the restricted geocode data in order to link individuals with their state of residence, which is necessary in order to use variation in AFDC/TANF generosity between states and over time. I also limit the sample to observations with reported annual earnings less than $1,000,000.

23. The NLSY79 collects information on the beginning and ending dates of first, second, and third marriages. Therefore, although the survey became biennial in 1994, it is possible to differentiate between marriages that occurred in 1995 (a non-survey year) from those that occurred in 1996 (the next survey year). I do not make this distinction in the empirical strategy because of the absence of other necessary covariates in non-survey years, such as number of children and state of residence. These are particularly important because they determine whether AFDC/TANF is available to the individual and, if so, how much.

24. There are instances in which an individual is married in one wave, divorced in the following wave, and married again in the next wave. In this case, the individual would be included in the divorce sample for all three of those observations. There are 157 (0.7%) such occurrences within the divorce sample. There are 104 (0.6%) analogous occurrences within the marriage sample.

is currently married or currently divorced. However, this process combines individuals who were able to respond the policy by marrying or by divorcing with those who were not. This limitation biases the estimated effects of the tax and transfer systems toward zero, which I illustrate later.

1.5.2 AFQT Z-Scores

Perhaps most importantly, the NLSY79 allows me to use Armed Forces Qualification Test (AFQT) scores to separate likely low-earners from likely high-earners. One empirical issue in determining EITC and AFDC/TANF eligibility is that individuals may manipulate their own earnings in order to become eligible for the programs or to earn a higher amount of assistance, potentially introducing endogeneity due to unobservable characteristics that would influence both earnings manipulation and marital status changes. For this reason, much of the literature uses education to circumvent this endogeneity concern and separate likely EITC and AFDC/TANF eligible individuals from others. I extend this common practice by defining treatment for an individual as their AFQT z-score.[25] AFQT z-scores are unique to these data and offer greater variation and more estimating power than using education in a similar manner. I convert reported AFQT percentile scores to a z-score by assuming a standard normal distribution of AFQT scores and computing AFQT z-score= Φ^{-1}(AFQT percentile), where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function.[26]

An individual with a low AFQT z-score likely has low annual earnings as well, which makes him more likely to be eligible for the EITC and AFDC/TANF. I demonstrate this positive relationship in Figure 5, which plots the average annual earnings by AFQT z-score bins for the marriage and divorce risk samples, along with their 95% confidence intervals. Therefore, my analysis compares individuals who had differing z-scores.

AFQT scores are a portion of the larger Armed Services Vocational Aptitude Battery (ASVAB)

25. In practice I use the negative of the individual's AFQT z-score. Defining the AFQT variable this way makes the interpretation more natural, so that a higher variable value is associated with a higher likelihood of treatment due to a lower AFQT z-score.

26. This approach is common in the literature using AFQT scores from the NLSY79. For example, Neal and Johnson (1996) adjusts AFQT scores in the NLSY79 in a similar manner.

administered to the majority of the respondents of the NLSY79. The AFQT combines arithmetic reasoning, word knowledge, paragraph comprehension, and numeric operations scores into a single measure.[27] The NLSY79 survey administered the ASVAB to its respondents in part to help the military create a standard by which to judge military enlistees because the NLSY79's initial sample was nationally representative.[28]

1.5.3 EITC Eligibility and AFDC/TANF Generosity

I use a spline function in the individual's AFQT z-score to separate individuals who are more likely to be eligible for the EITC and AFDC/TANF from others, and to allow for differential effects of the EITC expansion along the AFQT distribution. I use Figure 5 to help identify trends in the relationship between earnings and z-score and to locate possible notch points in the spline function. Although the relationship between z-score and reported earnings is not overwhelmingly strong, individuals with z-scores less than 0, on average, appear to have earnings lower than approximately $26,000 and therefore fall in the EITC eligibility range. In general, the lines of best fit displayed in Figure 5a (the marriage sample) appear largely linear up to a z-score of 0, with a possible difference in the relationship for individuals with z-scores between 0 and 1 or z-scores above 1. Therefore, I place notches in the spline function at z-scores of 0 and 1, to allow for differential effects of the 1993–1996 EITC expansion along the AFQT distribution.

I extend much of the past literature in this field by incorporating welfare incentives as well. I measure AFDC/TANF generosity in the individual's state using the maximum attainable monthly AFDC/TANF payment for a family of three.[29] It is not sufficient to include this same measure in

27. The other subtests cover general science, coding speed, auto and shop information, mathematics knowledge, mechanical comprehension, and electronics information. Stevenson (2012) provides evidence that the coding speed test measures non-cognitive attributes, such as intrinsic motivation, among the NLSY79 respondents, and is therefore positively correlated with earnings. More generally, because the ASVAB tests were unincentivized, performance on these tests may measure both cognitive and non-cognitive attributes, both of which predict earnings. This aspect of the ASVAB tests in the NLSY79 provides further justification to use AFQT z-scores as a method to separate likely low-earners from likely high-earners.

28. The military primarily uses the ASVAB to sort enlistees into military professions, but also uses the AFQT subset to judge overall trainability and suitability for the military.

29. Bitler et al. (2004), Meyer and Rosenbaum (2001), and others also use the maximum attainable monthly AFDC/TANF payment to estimate the effect of the transfer system.

the regression for all individuals, however, as some individuals are not eligible for AFDC/TANF due to a number of potential reasons. In order to more accurately reflect an individual's potential gain or loss, I interact the maximum attainable monthly AFDC/TANF payment with the individual's z-score, an indicator equal to one if the individual has at least one child this period, and an indicator variable equal to one if the individual is female.[30] This last interaction reflects the fact that AFDC/TANF is largely targeted at single mothers.

1.5.4 Dynamic Selection

Because the sample ages in unison and is not replenished, there is a concern that over time the sample of individuals who are eligible to marry will become biased toward individuals who will never marry due to some unobservable factor. In addition, most transitions into marriage occur in the mid- to late-20s, and so extending the sample period to include older individuals may bias the analysis due to negative duration dependence.[31] In this case, the sample of unmarried individuals becomes increasingly negatively selected over time and we would expect the estimated effects of the 1993–1996 EITC expansion and AFDC/TANF generosity to be biased toward zero in later years.[32] Analogous dynamic selection is also possible in the divorce risk sample. These dynamic selection issues are important considerations, and I therefore limit the sample period and consider only marital transitions between 1991–1998 in order to alleviate, but not completely eliminate, this concern. Although determining the direction and extent of dynamic selection is not feasible in my current framework, I present a simple theoretical model in the online appendix to demonstrate that dynamic selection can lead to a sample that is increasingly unlikely to leave its current state.

30. In specifications that use only women this definition is similar to the difference-in-differences specification with the added interaction of the individual's AFQT z-score instead of the spline function. In specifications that use only men the regression fails. Therefore, I remove the interaction with the female indicator variable in specifications that use only women or only men. In alternative specifications I utilize two other AFDC/TANF eligibility definitions; one based on predicted earnings and the other omitting the female interaction in all specifications. These results are available upon request.

31. Figure 4 displays the hazard rate of new marriages by age among women who were 27–43 years old in 1990, calculated from the 1990 June Current Population Survey. The declining hazard rate at higher ages is another reason I omit survey years beyond 1998 in order to exclude the right-most tail of the age distribution in Figure 4.

32. This is just one possible example of dynamic selection, as any dynamic selection within the samples may work in the opposite direction.

1.5.5 Summary Statistics

Columns 1–3 of Table 2 present the summary statistics for the marriage risk sample as of 1991. I calculate the summary statistics only for individuals whose first observation places them in the marriage sample.[33] Notably, while the marriage and divorce risk samples are similar in terms of gender and age, individuals initially in the marriage risk sample are more likely to be black than individuals initially in the divorce risk sample. Predictably, individuals with lower AFQT z-scores are also more likely to have exactly a high school education or less. The average individual in the marriage risk sample is 30 years old, and 30% of them have at least one child, and so are eligible for a generous EITC.

Columns 4–6 of Table 2 presents the initial summary statistics for the divorce risk sample as of 1991. Again, I calculate the summary statistics only for individuals whose first observation places them in the divorce sample. The divorce risk sample, which is slightly older and has more children than the marriage risk sample, otherwise exhibits trends that are similar to those of the marriage risk sample. Notably, individuals with lower AFQT z-scores are more likely to have exactly a high school education or less and be black or hispanic than are individuals with higher z-scores. 50% of individuals in the divorce sample have at least one child. In addition, individuals in the divorce risk sample are more likely to have children, and have, on average, approximately one more child, than individuals in the marriage risk sample.[34]

1.6 Empirical Strategy

My goal is to estimate the effects of the EITC and AFDC/TANF on the probability of marrying or divorcing using policy changes that occurred in the early- to mid-1990s. I use a difference-

33. These statistics do not include individuals who enter the marriage sample via divorce. Including individuals who enter the marriage sample via divorce may result in individuals being included twice or individuals being included in both the marriage and divorce sample summary statistics. The summary statistics in Table 2 are not fully representative of the panel because they only appear once.

34. This aspect of the divorce sample is likely due to the fact that being married and having children are correlated, and is unlikely due to age. The average individual in the divorce sample is 0.4 years older than the average individual in the marriage sample, yet 50% of the divorce sample has at least one child whereas only 30% of the marriage sample has at least one child.

in-differences approach comparing parents to childless individuals before and after the policies. Using the presence of children as a treatment dimension means that the analysis focuses on those individuals in Figures 2–3, panels d–f, in which a single parent with one child considers marriage, and more precisely isolates individuals who would most likely be affected by the EITC expansion.

Expanding upon the general method of Eissa and Liebman (1996), I use a spline function in AFQT z-score to estimate heterogeneous effects of the EITC expansion along the AFQT (and resulting earnings) distribution.[35] I use the individual's AFQT z-score, rather than reported earnings, to avoid endogeneity concerns associated with using reported earnings directly.[36] Finally, I use the state's maximum attainable monthly AFDC/TANF payment to measure the effects of changes in the transfer system, using variation between states and over time. All models are linear probability models due to the large number of fixed effects.

Recall that the NLSY79 becomes biennial in 1994. Thus, I can observe behavioral responses in 1994, 1996, and 1998 only, compared to the pre-policy years of 1991, 1992, and 1993. In light of this, I estimate the following equation:

$$\begin{aligned} Married_{it} = \beta_0 &+ \beta_1 f(\text{AFQT Z-Score}_i) \times HasChild_{it} \times PostYear_t \\ &+ \beta_2 \text{AFQT Z-Score}_i \times HasChild_{it} \times Female_i \times maxAFDC_{it} \\ &+ \beta_3 Z_{it} + \beta_4 X_{it} + \varepsilon_{it} \end{aligned} \tag{1.1}$$

The specification also includes (denoted by the vector Z_{it}) the $f(\text{AFQT Z-Score}_i)$, $HasChild_{it}$, and year fixed effect variables entered individually along with pairwise interactions between each of them. $Married_{it}$ is an indicator for being married in year t, $f(\text{AFQT Z-Score}_i)$ is a linear spline function in AFQT z-score with notches at z-scores of 0 and 1, $HasChild_{it}$ is an indicator of having at least one EITC-eligible child in the previous period, and $PostYear_t$ is a vector of

35. I convert reported AFQT percentile scores to a z-score by assuming a standard normal distribution of AFQT scores and computing AFQT z-score $= \Phi^{-1}$(AFQT percentile), where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function. Additionally, in practice I use the negative of the individual's AFQT z-score, so that a higher value of this variable reflects a higher likelihood of being eligible for the EITC and AFDC/TANF due to a lower AFQT score.

36. Specifically, the concern is that there are unobservable factors that would influence the individual to manipulate his eligibility for the EITC or AFDC/TANF and to change his marital status.

indicator variables for the post-EITC expansions years 1994–1998. $maxAFDC_{it}$ is the maximum attainable monthly AFDC/TANF payment for a family of three (measured in hundreds of dollars) in the individual's state in year t. X_{it} is a vector of other covariates in year t that likely influence marriage decisions, including age group, race, gender, whether the state has a state EITC in year t, state fixed effects, and current, one-period, and two-period lagged number of children.[37]

I use age group and educational level dummies instead of the standard measure of age, age squared, or years of education due to possible non-linear effects on marital outcomes. Age groups begin with 28–31 years old and advance in three year groups up to 40 years old or older (27 years old or younger is the omitted category). Education groups are less than high school, high school degree, and some college (college degree or more is the omitted category). The spline function in AFQT, $f(\text{AFQT Z-Score}_i)$, flexibly allows parents to have differential responses to the EITC expansion along the AFQT distribution. The coefficients of interest are the vector β_1, which measures the effect of the 1993–1996 EITC expansion in 1994, 1996, and 1998, as well as β_2, which measures the effect of AFDC/TANF generosity in the individual's state.

The equations I estimate for the divorce sample are analogous to equation 1.1. The difference is that the dependent variable in these equations is equal to one if the individual reports being either divorced or separated within the last year, which is conventional in this literature. In addition, I include the number of years the individual has been married as an additional regressor in these models.[38]

Note that my empirical strategy does not distinguish between differing sources of incentives originating from the 1993–1996 EITC expansion nor AFDC/TANF reform. Specifically, both of these policies strengthened labor supply incentives among low-earning families. Ellwood (2000) provides an illustrative example concerning the EITC. If a married couple with two earners faces a marriage disincentive due to the EITC then they may be better off divorcing. However, expanding

37. I include one- and two-period lags of the number of children because family structure decisions may take time to manifest in the data. Herbst (2011) lags his main EITC variable by two years for this reason.

38. Including the number of years married as an explanatory variable implies dynamic selection in the sample. This means that individuals who are married for longer are more likely to stay married. To the extent that dynamic selection effects are different between individuals with different AFQT percentile scores, the results will be biased.

the EITC increases the family's income and creates labor supply disincentives for the secondary earner if the family falls in the phase-out range of the schedule. The secondary earner may specialize more in household production, which could benefit the family overall and counteract the divorce incentive. My empirical strategy does not differentiate between a mechanism such as in Ellwood's (2000) example from any other possible mechanism.

Assuming a mother may only receive AFDC/TANF if she is single, then I expect the effect of AFDC/TANF generosity on the probability of marrying to be negative. On the other hand, I expect the effect of the 1993–1996 EITC expansion on the probability of marrying to be positive among low earners. This prediction is based on Figure 3, which shows that the combined incentives from the EITC and AFDC shifted in favor of marriage over time. I expect the opposite-signed effects on the probability of divorcing.

1.7 Results

I estimate the effects of the 1993–1996 EITC expansion and of AFDC/TANF generosity on the probability of marrying and divorcing using risk samples including only unmarried or married individuals, respectively. I compare a treatment group of individuals who had at least one child in the previous period to a control group of childless individuals while taking into account differences in Armed Forces Qualification Test (AFQT) z-scores, which I use to indicate differing potential levels of treatment. This approach circumvents endogeneity issues associated with using earnings to directly classify eligible individuals such as potential earnings manipulation in order to alter their statuses as treated or non-treated individuals. This approach also enables me to use a more flexible functional form compared to studies that use education to classify likely-eligible individuals, since I can use a flexible spline function that allows for identification of different responses to the EITC expansion within groups of people that have similar z-scores. I also allow the effects of the EITC expansion to vary across time, which can capture both the fact that the EITC grows in magnitude each year and that family structure decisions may manifest in years after the initial policy change.

I find that single parents with the lowest z-scores are more likely to marry after the EITC expansion. I also find that married parents with mid-range z-scores are less likely to divorce, while married parents with the highest z-scores are more likely to divorce as a result of the 1993–1996 EITC expansion. These effects are consistent with the hypothesized effects in Section 1.4 based on likely earnings of individuals in these z-score ranges. In addition, I find subtle differences in family structure responses between men and women, which suggest that male and female parents may marry different types of people. Single fathers, for example, are more likely to increase their number of children upon marriage, which can strengthen marriage incentives from the EITC by increasing the family's tax credit. I also find some evidence that declines in AFDC/TANF generosity increase the likelihood of marriage.

1.7.1 Main Results

Table 3 presents my main results. Among single parents who are the most likely to be eligible for the EITC, I estimate that the EITC expansion significantly increased the probability of marrying between 1997–98 by 5.2 percentage points for each standard deviation reduction in AFQT score. In other words, I estimate that due to the EITC expansion, an unmarried parent with a z-score of -1 is 5.2 percentage points (47.7%) more likely to marry one to two years after the expansion was complete than an unmarried childless individual with a z-score of 0. I find no other statistically significant effect of the EITC expansion on the probability of marrying among any other AFQT z-score groups or during any other years. This finding is consistent with the theoretical predictions in Section 1.4, which show potentially substantial gains in EITC amount through marriage among low-earning individuals. Although imprecise in earlier years, the coefficient estimate among this subgroup grows over time, providing some evidence that the effects of taxes on the decision to marry may manifest years after marriage incentives have strengthened.

Among married parents who are more likely to be in the phase-out range of the EITC or ineligible, I estimate that the EITC expansion significantly decreased the probability of divorcing in 1994 by 6.0 percentage points for each standard deviation reduction in AFQT score. A standard deviation reduction in AFQT score among individuals in this range is more likely to be the difference between being eligible for the EITC rather than ineligible, which may lead to the large point estimate (139.5%). This finding is consistent with the theoretical predictions in Section 1.4: likely mid-earners divorce less frequently as a result of the EITC expansion because the substantial expansion of the one- and two-child EITC schedules provided additional income to families with children as the base of the credit schedule grew. This is sometimes referred to as the "stabilization effect," where additional family income stabilizes a marriage that may otherwise have been close to separating. In addition, the estimated effect of the EITC expansion on the decision to marry among this AFQT subgroup is positive in 1994, and drops to near 0 in subsequent years. Although imprecise, this pattern suggests that, in addition to a decreased propensity to divorce as a result of the EITC expansion among married parents, single parents who are more likely to be in the phase-out range of the EITC or ineligible faced stronger marriage incentives.

Finally, among married parents who are the most likely to be ineligible for the EITC, I find that the EITC expansion significantly increased the probability of divorcing between 1997–98 by 10.5 percentage points (172.1%) for each standard deviation reduction in AFQT score. Individuals with the highest z-scores are most likely ineligible for the EITC, but by divorcing and losing one source of income they stand to gain a larger EITC since they are more likely to become eligible for the credit due to the expansion. It is also unsurprising that the timing of this effect is delayed, since the highest earners would not have faced these incentives from the EITC until late in the expansion period when the credit's base was largest.

Note, also, that the effects of the EITC expansion are often the same sign when comparing the probability of marrying to the probability of divorcing. Although many of these estimates are imprecise, their signs suggest that family structure incentives in the tax system create asymmetric responses along the marriage and divorce margins.

To relate the estimates to the size of the EITC expansion, I perform a back-of-the-envelope calculation to determine the effect of a $500 increase in the maximum EITC. I scale the estimates in Table 3 to represent a $500 increase in the maximum EITC based on the EITC schedules in

Table 1. I calculate that, for each standard deviation reduction in AFQT score, an additional $500 in EITC increases the probability of marrying by 1.2–3.4 percentage points (11.0–31.2%) among single parents with the lowest AFQT z-scores, decreases the probability of divorcing by 2.5–5.7 percentage points (62.8–132.6%) among married parents with mid-range AFQT z-scores, and increases the probability of divorcing by 2.3–6.9 percentage points (37.7–113.1%) among married parents with the highest AFQT z-scores. These calculations suggest similar magnitudes across parents with low- to mid-range AFQT z-scores. Recall, however, that the timing differs, perhaps because the impact on the divorce decision is immediate, while it may take some time to find a potential spouse and/or decide to marry.

Lastly, there is some evidence of a negative effect of AFDC/TANF generosity on the probability of marrying. The estimate reveals that, all else equal, a $100 decrease in the maximum monthly attainable AFDC/TANF payment for a family of three increases the probability of marrying by 0.4 percentage points (4.5%) among single mothers who are the most likely to be eligible for AFDC/TANF benefits, relative to others.[39] This estimate is statistically significant at the 90% confidence level.

1.7.2 Comparing Risk Samples to Cross-Sections

Risk samples ensure that I estimate policy effects on individuals who can respond, say, by marrying (because they were previously unmarried). Cross-sections, on the other hand, blur the marriage and divorce response margins, pooling together individuals who can respond by marrying and by not divorcing, and analogously for divorce. This limitation biases the estimated effect of the tax and transfer systems toward zero due, for instance, to already married individuals appearing to not respond to the policy. I use the NLSY79 to create repeated cross-section samples in order to compare my empirical estimates with a similar repeated cross-section sample.

Table 4 displays estimates from the NLSY79 risk samples (reproduced from Table 3) and estimates from cross-section samples. The estimates from the cross-section samples are the effects

39. This is a $1,200 decrease in the maximum annual attainable AFDC/TANF payment for a family of three.

of the policies on the probability of being married or being divorced.[40] The dependent variable in the marriage regressions is an indicator equal to 1 if the individual is married, and the dependent variable in the divorce regressions is an indicator equal to 1 if the individual is divorced or separated.

The cross-section samples in Table 4 generally display muted effects of the 1993–1996 EITC expansion on the probability of being married or divorced compared to the risk sample estimates, although some coefficient estimates using cross-sections are larger in absolute value than the risk sample estimates. Muted effects may occur because the sample includes individuals who are unable to respond to the policy by marrying because they were already married, and analogously for divorce. Larger coefficient estimates may occur because an individual needs to pass through marriage in order to divorce, meaning a cross-section sample that shows an increase in the probability of being married should also show a decrease in the probability of being divorced. This balancing out is evident among the cross-section estimates in Table 4, and blurs the line between the response marriage and divorce response margins. This feature highlights another weakness of using cross-section data to analyze family structure transitions.

1.7.3 Robustness and Alternative Specifications

One of the contributions of this paper is to use AFQT z-scores to distinguish who is most likely to have low earnings and be eligible for the EITC and AFDC/TANF, whereas earlier papers often use education instead. Table 6 compares the earlier estimates using z-scores to estimates using education group indicators. In these specifications, I replace the AFQT variable in Equation 1.1 with a vector of education group indicator variables. In the full specification the omitted education group is “college degree or more,” and in the two subsequent sample restrictions the control group becomes “some college” and “exactly high school degree,” respectively. In addition, in these specifications I replace the AFQT z-score interaction in the AFDC/TANF eligibility definition

40. Each cross-section sample uses all observations from both the marriage and divorce risk samples. Note that the total number of observations in the cross-section samples will not be equal to the simple sum of observations in the risk samples because a newly married individual is included in both the marriage and divorce risk sample.

with an indicator variable equal to one if the individual has less than a high school degree.

I present the results of this alternative specification along with the original estimates in Table 5. The estimates using education corroborate the original estimates, revealing that single parents with less than a high school degree are more likely to marry as a result of the EITC expansion, and that married parents with exactly a high school degree are less likely to divorce as a result of the EITC expansion. These estimates also show that single mothers with less than a high school degree are more likely to marry as AFDC/TANF generosity decreases, which is consistent with my earlier findings. However, the precision of the estimates in this alternative specification is strongest using the full sample, in which the control group includes individuals with a college degree or more. Using very highly educated individuals as a control group for high school drop outs, however, may violate the parallel trends assumption between treatment and control groups, and reduces the reliability of the estimates in Table 5 that use education indicators. These results suggest that using education may lead to less reliable estimates about the impact of the EITC expansion.

In addition, much of the previous literature focuses on women alone, whereas my analysis includes both men and women. However, I would expect to see responses in a sample of men who are most likely to marry such women. Indeed, although the point estimates differ somewhat in Table 6, where I estimate the model separately for men and women, they tell a similar story as the main results in Table 3; namely, single parents with the lowest AFQT scores are more likely to marry, married parents with mid-range AFQT scores are less likely to divorce, and married parents with the highest AFQT scores are more likely to divorce. When compared to the main results, the point estimates in Table 6 are larger in absolute value among women in the divorce sample, and smaller in absolute value among women in the marriage sample, while the opposite is true for men.

Differences in the point estimates for men and women may result from the types of spouses they choose. Table 7 reports some summary statistics of individuals in each sample who change family structure in order to gain a better understanding of who individuals marry and how their spouses may influence the point estimates. On average, single fathers are more likely than single mothers to increase their number of children upon marriage. This can occur either due to marrying a single mother, or by having a new child with their new spouse. Thus, single fathers are more likely to move from the one-child EITC schedule to the two-child EITC schedule upon marriage, creating a greater incentive to marry. In addition, women are more likely to retain custody of their children upon divorce, creating a stronger incentive for women to divorce in response to the EITC expansion than it would for men. These patterns in the marriage and divorce samples help explain why point estimates among women in the marriage sample are smaller in absolute value than among men, and vice versa in the divorce sample.

In order to more explicitly quantify the effect of a higher EITC amount on the probability of marrying and divorcing, Table 8 presents estimates of an alternative specification of Equation 1.1 that parameterizes the effect of the EITC as linear in $MaxEITC_t$, the maximum credit available each year on the one-child schedule. Note, however, that this specification does not allow for gradual effects of the EITC expansion. I find that a \$1,000 increase in the maximum EITC payment causes a 4.5 percentage point increase (51.1%) in the probability of marrying for each standard deviation reduction in AFQT score among single parents with the lowest AFQT z-scores relative to single childless individuals. This estimate is comparable in size to my back-of-the-envelope calculation using the main estimates in Table 3. The remaining coefficient estimates are also comparable to the back-of-the-envelope calculations, although they are not as precise in this specification. Finally, I also estimate that a \$100 decrease in AFDC/TANF generosity increases the probability of marrying by 0.5 percentage points (5.7%) among single mothers who are likely eligible for AFDC/TANF benefits, relative to others.

However, note that the maximum EITC available each year may not be an appropriate measure of an individual's actual or potential EITC because the range of family income within the plateau range is relatively narrow and because the majority of EITC recipients do not fall within the plateau range. Although the point estimates provide additional insight into the magnitude of the effect of the EITC expansion, this method may ignore much of the key variation in how the EITC creates differing incentives across the budget constraint.

1.7.4 Investigating the Parallel Trends Assumption

Finally, the identification assumption throughout is that marriage and divorce decisions of individuals with and without children would have evolved similarly in the absence of the EITC expansion and welfare reform. To investigate this assumption, I use the marriage and birth histories of women surveyed in the 1995 June Current Population Survey to create a panel back to 1987 in order to examine the parallel trends assumption before the EITC expansions in 1990 and 1993–1996.[41] I restrict the sample of women to the same birth cohorts in the NLSY79 and estimate Equation 1.1 using education indicator variables instead of the AFQT variables (as in Table 5). I plot in Figure 6 the coefficient estimates and 95% confidence intervals of the triple interactions, $EduGroup_i \times Year_t \times HasChild_{it}$. The coefficient estimates in the pre-policy years (1987–1993) are not statistically different from zero, with the exception of the 1992 coefficient among individuals with less than a high school education in the divorce sample, but this coefficient is barely statistically significant at the 95% level. Overall, this event study approach provides some evidence that the parallel trends assumption holds in both the marriage and divorce samples. Note, however, that precision of these estimates is an issue, especially in the marriage sample.

In addition, I use the panel of women from the 1995 June CPS to plot the overall rate of new marriages and divorces in each risk sample between individuals with and without children. Figure 7 shows that the trends in new marriages and divorces appear mostly parallel, providing further supporting evidence in favor of the parallel trend assumption in both samples. Note, however, that the downward trend in new marriage rates, which likely reflects the aging of these cohorts, may be slightly steeper among women with children than among women without children.

1.8 Conclusion

The Earned Income Tax Credit and the Aid to Families with Dependent Children program generate a wide array of family structure incentives, some of them conflicting. Much of the literature on

41. The June CPS is helpful because it offers a substantially larger sample than the NLSY79, although the birth and marital histories are collected only for women.

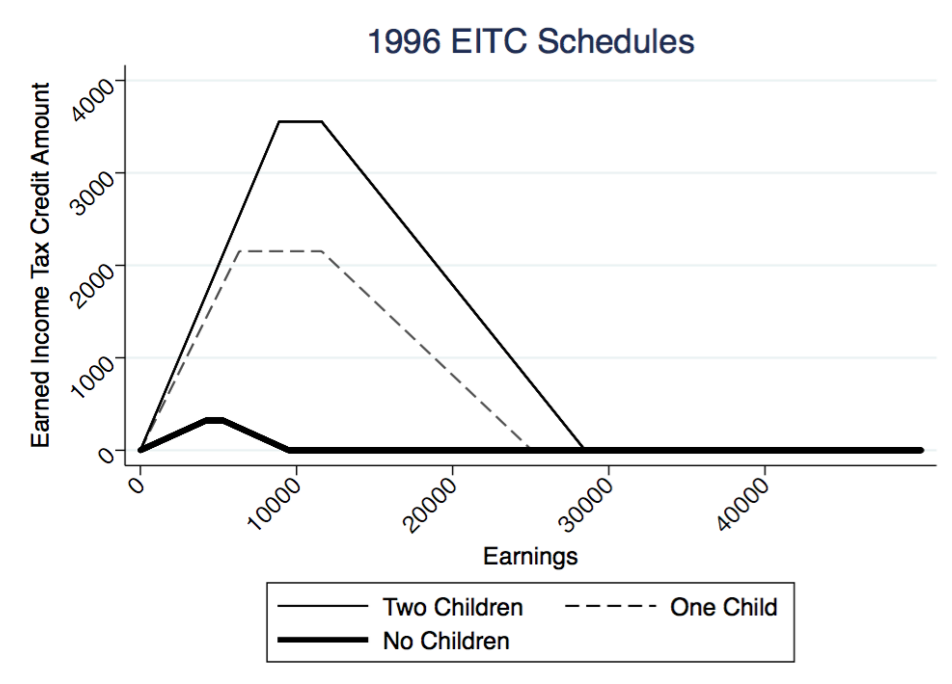
the EITC and AFDC/TANF focuses on labor supply responses, but marriage and divorce offer other avenues through which an individual can respond to the programs. For very low earning couples, the EITC provides an incentive to be married, but for couples outside this range the EITC provides an incentive to be single. These incentives apply regardless of whether we consider marriage responses or divorce responses. I use longitudinal data and a triple-difference approach with a flexible spline function to estimate the effect of the 1993–1996 EITC expansion on the probability of marrying and divorcing, which allows for non-constant effects as the expansion is implemented. In the main specification, the two dimensions of treatment are 1) the individual's AFQT z-score and 2) whether the individual had an EITC-eligible child in the previous period. Z-scores aid in circumventing the endogeneity issues associated with earnings in this context; namely, that changes to EITC and AFDC/TANF program parameters cause individuals to manipulate their earnings and, thus, their statuses as treated or non-treated individuals. Prior studies often use education to differentiate between low- and high-earning individuals, but z-scores offer greater estimating power than education and are unique to the NLSY.

In addition, I show that using a cross-section sample instead of a risk sample mutes the estimated effect of the 1993–1996 EITC expansion on the probability of marrying and divorcing using the difference-in-differences approach. This is because a cross-section sample limits the researcher's ability to distinguish between individuals who were previously unmarried from those who were already married, thereby grouping together individuals who are able and unable to respond the policy, respectively. In addition, cross-section estimates that show an increase in the probability of being married should also show a decrease in the probability of being divorced, creating ambiguity as to which margin of behavior changes as a result of the policy. Finally, I compare estimates that use z-scores to differentiate between likely EITC-eligible and -ineligible individuals to those that use education instead. The results show that using education decreases the precision of the estimates, and provides weak evidence, at best, of any effects on the probability of marrying and divorcing.

I conclude that individuals do respond to the family structure incentives contained in the EITC and AFDC/TANF, and that the response to the 1993–1996 EITC expansion, in particular, was substantial. The main results show that the 1993–1996 EITC expansion increased the probability of marrying between 1997–1998 by 5.2 percentage points (47.7%) for each standard deviation reduction in AFQT score among unmarried parents who are most likely to be eligible for the EITC. In addition, I find that, all else equal, a $100 decrease in the maximum monthly attainable AFDC/TANF payment for a family of three increases the probability of marrying by 0.4–0.5 percentage points (4.5–5.7%) among single mothers who are likely eligible for AFDC/TANF benefits, relative to others. I also find that the EITC expansion decreased the probability of divorcing in 1994 by 6.0 percentage points (139.5%) for each standard deviation reduction in AFQT score among married parents who are more likely to be in the phase-out range of the EITC or ineligible, but increased the probability of divorcing between 1997–1998 by 10.5 percentage points (172.1%) among married parents who are the most likely to be ineligible for the EITC. Finally, I find no evidence of an effect of AFDC/TANF generosity on the probability of divorcing. These estimates imply that, for each standard deviation reduction in AFQT score, a $500 decrease in tax liability causes a 1.2–3.4 percentage point (11.0–31.2%) increase in the probability of marrying among single parents who are most likely eligible for the EITC, a 2.7–5.7 percentage point (62.8–132.6%) decrease in the probability of divorcing among married parents who are more likely to be in the phase-out range of the EITC or ineligible, and a 2.3–6.9 percentage point (37.7–113.1%) increase in the probability of divorcing among married parents who are most likely ineligible for the EITC.

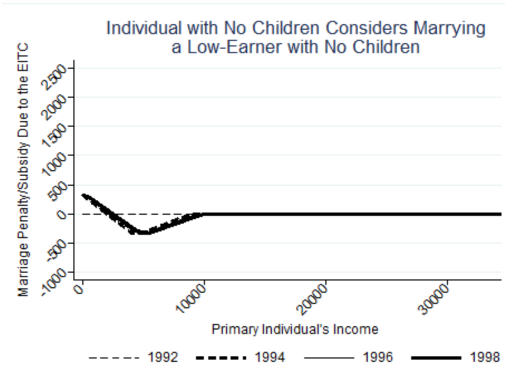
Future work in this field should continue to incorporate both tax and transfer system incentives into the analysis and utilize longitudinal data to study the flows into and out of marriages, which decouples the response margin between marriage and divorce. Although the transfer system largely discourages marriage, the tax system can either encourage or discourage marriage. The interaction is important because individuals who are eligible for one program are often eligible for others, and thus face oftentimes conflicting incentives. The effects of these programs are likely different due to different take-up rates and, possibly, different methods of receipt. Thus, one dollar in tax relief is not necessarily the same as one dollar of welfare assistance.

Figure 1: 1996 EITC Schedules

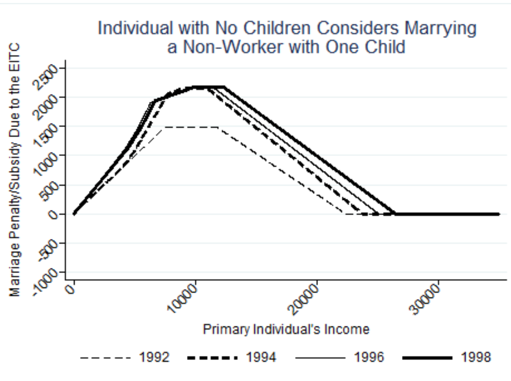


Notes: The data come from the EITC parameters presented in Table 1 and are taken from the U.S. Congress Joint Committee on Taxation's 2004 Green Book. All dollar values have been converted to real values using 1996 as the base year.

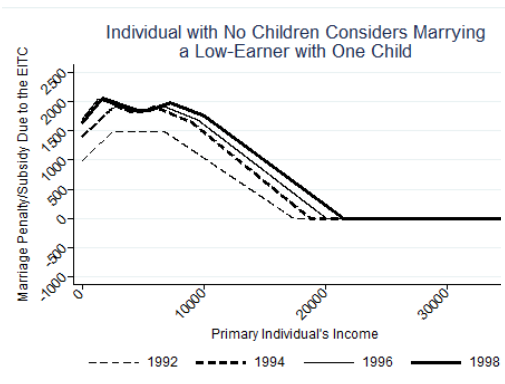
Figure 2: Difference Between EITC Under Marriage and EITC Under Single Status



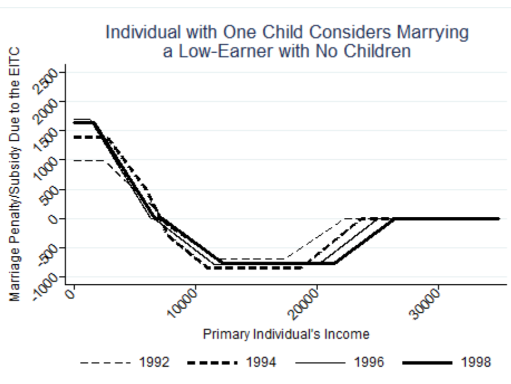
(a)



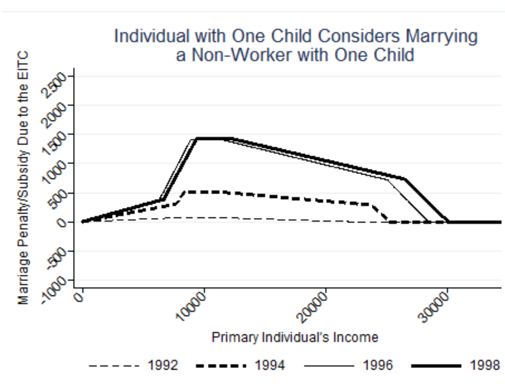
(b)



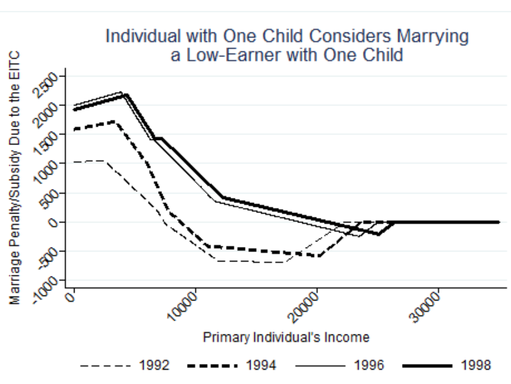
(c)



(d)



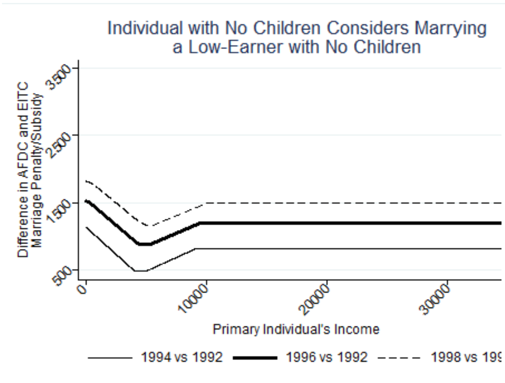
(e)



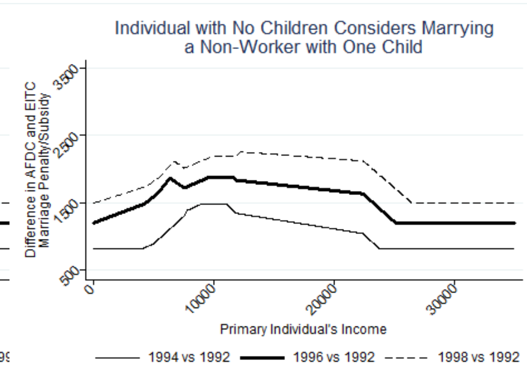
(f)

Notes: The data come from the EITC parameters presented in Table 1 and are taken from the U.S. Congress Joint Committee on Taxation's 2004 Green Book. All dollar values have been converted to real values using 1996 as the base year. As an example, the 1994 difference plots $(EITC_{married,1994} - EITC_{single,1994})$ for each situation. A "low-earner" is assumed to earn $5,000 per year. These situations are meant to be representative of the incentives faced by low-income individuals. Some situations result in zero effects, such as when an individual with no children considers marrying a non-worker with no children or when an individual with one child considers marrying a non-worker with no children.

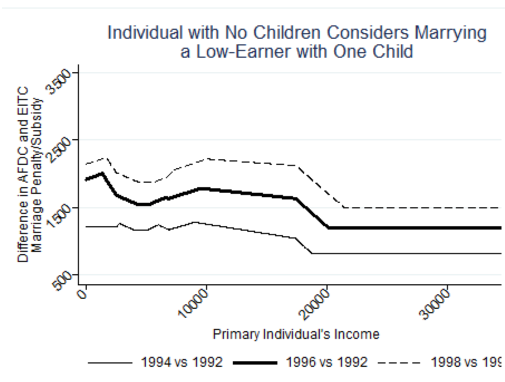
Figure 3: Difference Between AFDC and EITC While Married and AFDC and EITC Under While Single



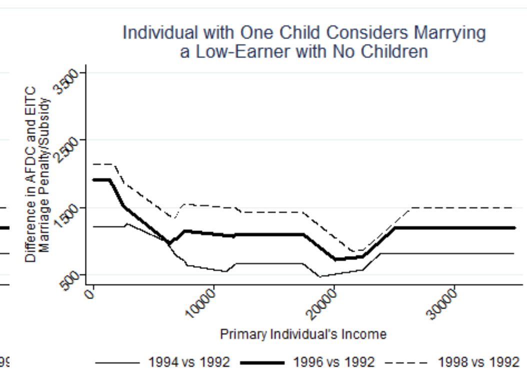
(a)



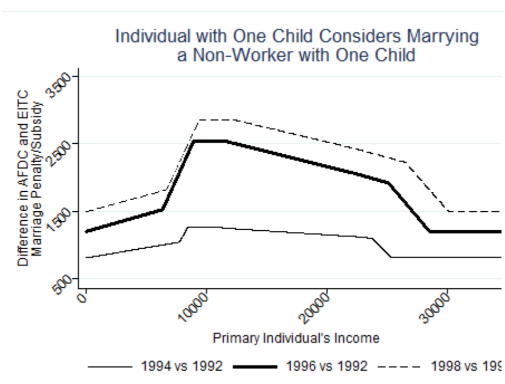
(b)



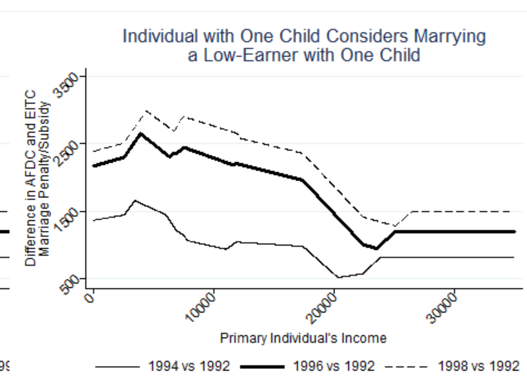
(c)



(d)



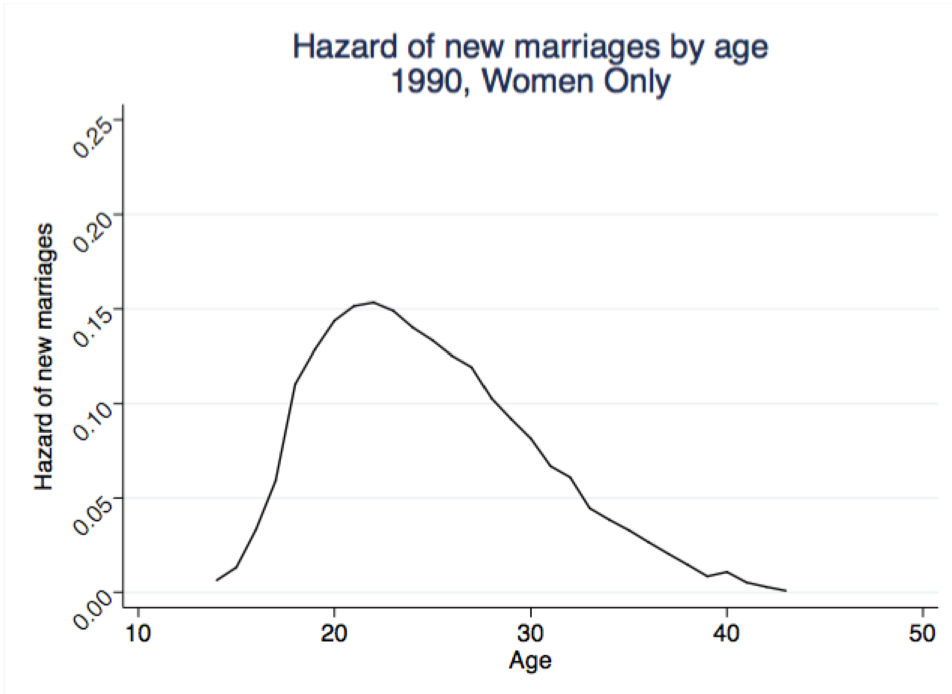
(e)



(f)

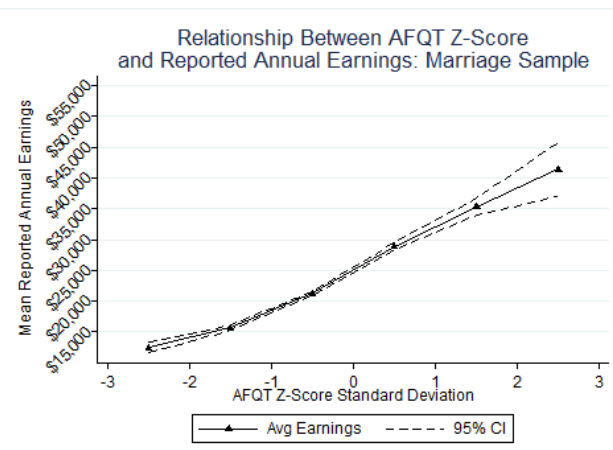
Notes: The data come from the EITC parameters presented in Table 1 and taken from the U.S. Congress Joint Committee on Taxation's 2004 Green Book. All dollar values have been converted to real values using 1996 as the base year. As an example, the 1994 to 1992 difference plots $(EITC_{married,1994} - EITC_{single,1994} - AFDC_{1994}) - (EITC_{married,1992} - EITC_{single,1992} - AFDC_{1992})$ for each situation. A "low-earner" is assumed to earn \$5,000 per year. These situations are meant to be representative of the incentives faced by low-income individuals. Some situations result in zero effects, such as when an individual with no children considers marrying a non-worker with no children or when an individual with one child considers marrying a non-worker with no children.

Figure 4: Hazard of New Marriages By Age Among Women Ages 27–43 in 1990

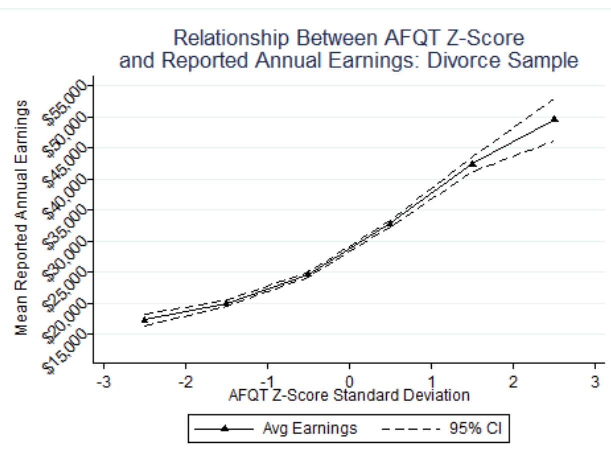


Notes: The data come from the June 1990 Current Population Survey and represent the author's own calculations.

Figure 5: The Relationship Between AFQT Z-Score and Average Annual Earnings



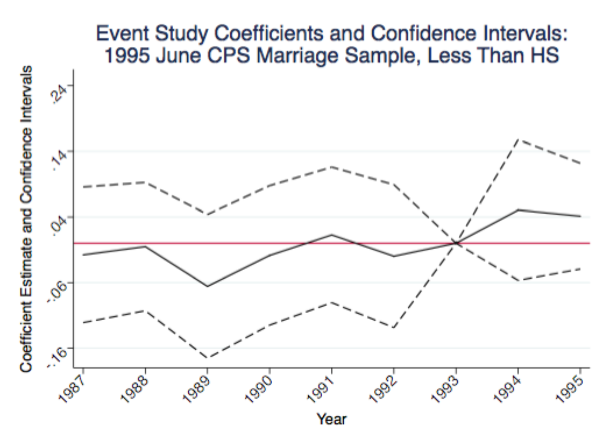
(a)



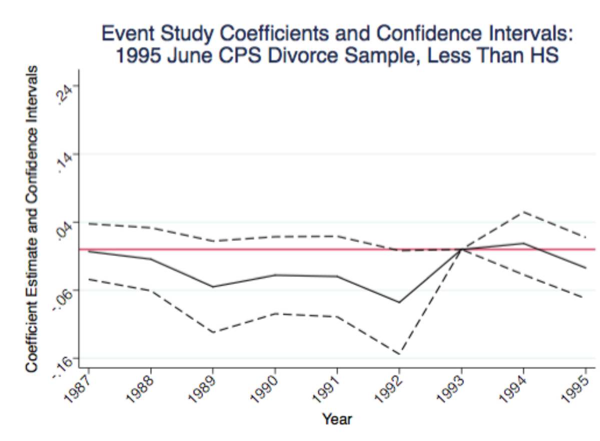
(b)

Notes: The data come from the 1991–1998 waves of the NLSY79. The y-axis displays the average annual earnings of individuals in the marriage and divorce risk samples within a one standard deviation bin. The x-axis displays their AFQT z-score bin.

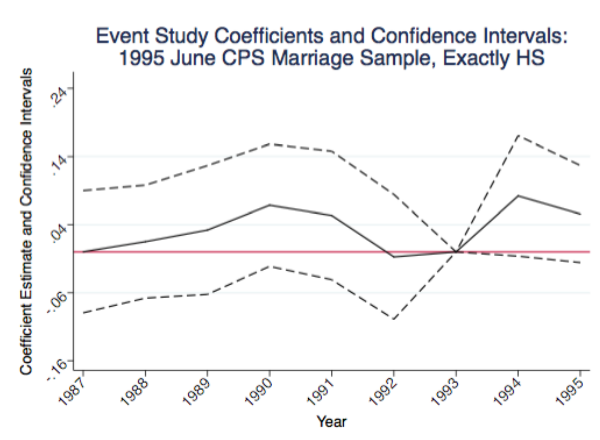
Figure 6: Event Study Coefficients and Confidence Intervals: 1995 June CPS



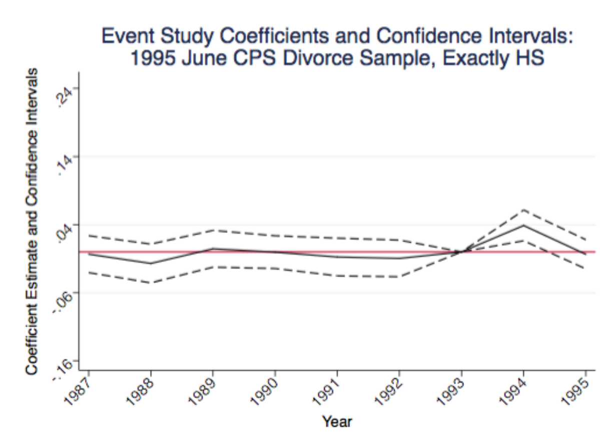
(a) Marriage Sample: 1995 June CPS



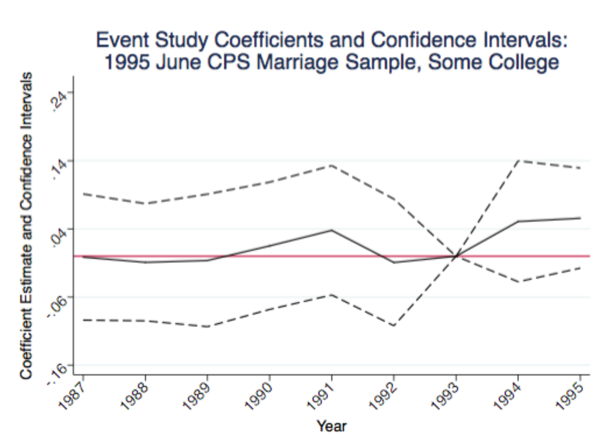
(b) Divorce Sample: 1995 June CPS



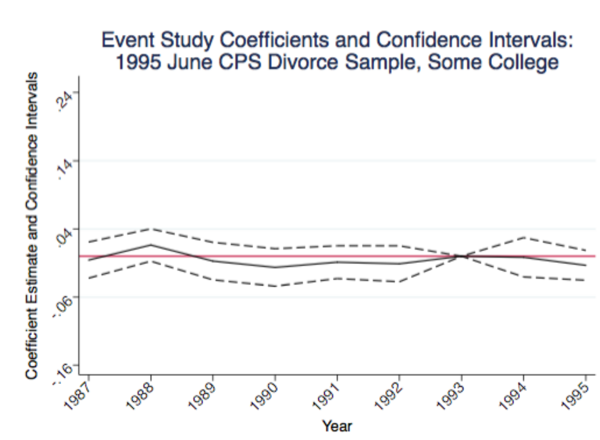
(c) Marriage Sample: 1995 June CPS



(d) Divorce Sample: 1995 June CPS



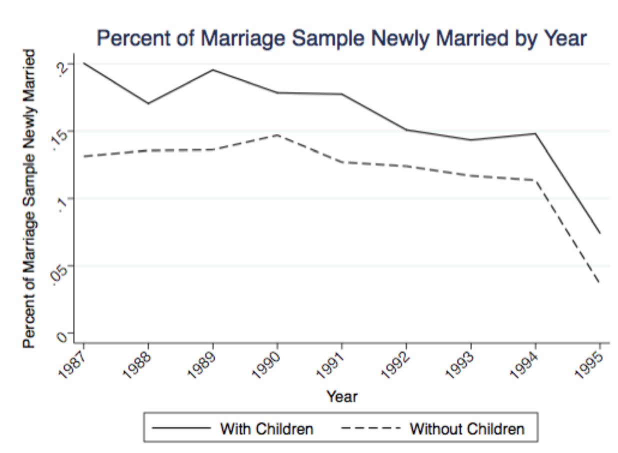
(e) Marriage Sample: 1995 June CPS



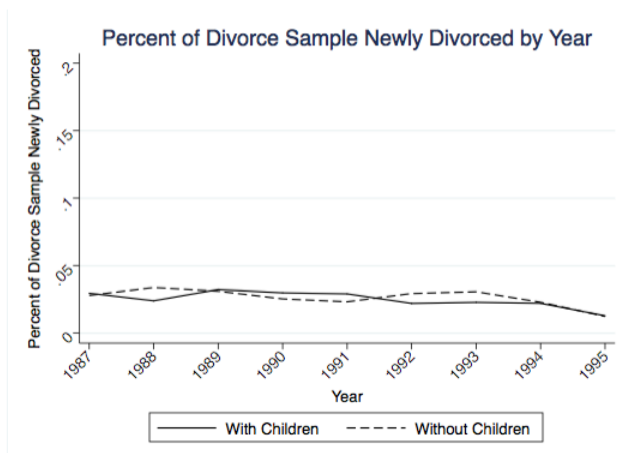
(f) Divorce Sample: 1995 June CPS

Notes: The data come from the 1995 June Current Population Survey, which collects marital and birth histories of women. Each panel plots the coefficient estimates and 95% confidence intervals of the $EduGroup_i \times Year_t \times HasChild_{it}$ variable from the specification of Equation 1.1 that uses education group indicator variables instead of the AFQT z-score variables.

Figure 7: Family Structure Changes in the Marriage and Divorce Samples: 1995 June CPS



(a) 1995 June CPS Marriage Sample



(b) 1995 June CPS Divorce Sample

Notes: The data come from the June 1995 Current Population Survey, which collects marital and birth histories of women. Panels a and b plot the percent of new marriages and divorces occurring in the marriage and divorce risk samples, respectively.

Table 1: EITC Parameters

Calendar Year	Phase-In Rate (%)	Min Income For Max Credit	Max Credit	Phase-Out Rate (%)	Phase-Out Range: Beginning Income	Phase-Out Range: Ending Income
1991						
One child	16.7	7,140	1,192	11.93	11,250	21,250
Two children	17.3	7,140	1,235	12.36	11,250	21,250
1992						
One child	17.6	7,520	1,324	12.57	11,840	22,370
Two children	18.4	7,520	1,384	13.14	11,840	22,370
1993						
One child	18.5	7,750	1,434	13.21	12,200	23,050
Two children	19.5	7,750	1,511	13.93	12,200	23,050
1994						
No children	7.65	4,000	306	7.65	5,000	9,000
One child	26.3	7,750	2,038	15.98	11,000	23,755
Two children	30	8,425	2,528	17.68	11,000	25,296
1995						
No children	7.65	4,100	314	7.65	5,130	9,230
One child	34	6,160	2,094	15.98	11,290	24,396
Two children	36	8,640	3,110	20.22	11,290	26,673
1996						
No children	7.65	4,220	323	7.65	5,280	9,500
One child	34	6,330	2,152	15.98	11,610	25,078
Two children	40	8,890	3,556	21.06	11,610	28,495
1997						
No children	7.65	4,340	332	7.65	5,430	9,770
One child	34	6,500	2,210	15.98	11,930	25,750
Two children	40	9,140	3,656	21.06	11,930	29,290
1998						
No children	7.65	4,460	341	7.65	5,570	10,030
One child	34	6,680	2,271	15.98	12,260	26,473
Two children	40	9,390	3,756	21.06	12,260	30,095

Notes: The data come from the U.S. Congress Joint Committee on Taxation's 2004 Green Book.

Table 2: Initial Summary Statistics

	Marriage sample			Divorce Sample		
Variable	Mean	Min	Max	Mean	Min	Max
Age	30.6	27	41	31.0	27	41
Any Children	0.3	0	1	0.5	0	1
Number of Children	0.5	0	8	1.2	0	8
AFQT Z-Score	-0.4	-3.2	3.1	-0.2	-3.1	3.2
	Frequency	AFQT Z-Score		Frequency	AFQT Z-Score	
Education						
Less than HS	22.9%	-1.1		17.3%	-1.0	
HS Degree	34.9%	-0.6		39.3%	-0.4	
Some College	22.4%	-0.1		22.1%	0.1	
College Degree	12.0%	0.5		12.9%	0.7	
More than College	7.8%	0.6		8.3%	0.8	
Gender						
Male	54.9%	-0.4		48.9%	-0.1	
Female	45.1%	-0.4		51.1%	-0.2	
Race						
Hispanic	17.7%	-0.6		20.1%	-0.6	
Black	39.1%	-0.9		19.7%	-0.7	
Other	43.2%	0.1		60.2%	0.2	
Number of Individuals	3,939			4,762		

Notes: The data come from the National Longitudinal Survey of Youth 1979. The statistics are for individuals who are included in the indicated sample in the first year they are observed. The statistics do not include individuals who enter the marriage sample via divorce or who enter the divorce sample via marriage. Hence, these summary statistics may not be representative of the full panel of individuals in the marriage and divorce samples. Naturally, individuals who enter the marriage sample via divorce will alter the composition of the sample, and analogously for individuals who enter the divorce sample via marriage.

Table 3: All Individuals: Effects of the EITC and AFDC/TANF on Marriage and Divorce

	Outcome: Married		Outcome: Divorced	
	Mean of dependent variable		Mean of dependent variable	
HasChild × 1994×				
1(AFQT Z-Score ≤ 0)	0.076	-0.004	0.043	0.022
		(0.023)		(0.019)
1(0 < AFQT Z-Score ≤ 1)		0.058		-0.060**
		(0.062)		(0.029)
1(1 < AFQT Z-Score)		0.169		0.037
		(0.116)		(0.041)
HasChild × 1994		0.027		-0.031*
		(0.025)		(0.018)
HasChild × 1996×				
1(AFQT Z-Score ≤ 0)	0.111	0.028	0.061	0.017
		(0.024)		(0.023)
1(0 < AFQT Z-Score ≤ 1)		-0.008		0.006
		(0.066)		(0.035)
1(1 < AFQT Z-Score)		-0.018		-0.023
		(0.109)		(0.036)
HasChild × 1996		-0.004		-0.017
		(0.027)		(0.021)
HasChild × 1998×				
1(AFQT Z-Score ≤ 0)	0.109	0.052**	0.061	0.029
		(0.026)		(0.024)
1(0 < AFQT Z-Score ≤ 1)		-0.007		-0.060
		(0.066)		(0.037)
1(1 < AFQT Z-Score)		0.034		0.105**
		(0.089)		(0.050)
HasChild × 1998		-0.007		-0.014
		(0.027)		(0.022)
Maximum AFDC/TANF for Family of Three	0.088	-0.004*	0.048	0.000
		(0.002)		(0.001)
Number of Observations		16474		23335
Number of Individuals		4500		5640
R^2		0.058		0.102

Notes: Standard errors are in parentheses and are clustered at the individual level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The data come from the National Longitudinal Survey of Youth 1979 and cover survey years 1991 through 1998. The dependent variable is equal to one if the individual is married (divorced) and zero if he is not. The definitions of AFDC eligibility and *TreatChild* are explained in the text. Other control variables included in all regressions, but whose coefficients are not reported, are dummies for educational level group, dummies for age group, race, gender, current, one-period, and two-period lagged number of children, state dummies, year dummies, and pairwise interactions of the treatment variables and year dummies. An individual is included in the marriage sample if he is not currently married, but is included in his first year of marriage. An individual is included in the divorce sample if he is currently married, but is included in the first year of divorce. If an individual marries and divorces during the sample, he re-enters the marriage sample.

Table 4: Comparing Risk Samples to Cross-Section Samples: Effects of the EITC and AFDC/TANF on Marriage and Divorce

	Outcome: Married				Outcome: Divorced			
	Mean of dependent variable				Mean of dependent variable			
	Risk sample	Cross-section	Risk sample	Cross-section	Risk sample	Cross-section	Risk sample	Cross-section
HasChild × 1994×								
1(AFQT Z-Score ≤ 0)	0.076	0.592	-0.004	-0.018	0.043	0.148	0.022	-0.001
			(0.023)	(0.018)			(0.019)	(0.016)
1(0 < AFQT Z-Score ≤ 1)			0.058	0.026			-0.060**	-0.023
			(0.062)	(0.033)			(0.029)	(0.026)
1(1 < AFQT Z-Score)			0.169	0.035			0.037	0.004
			(0.116)	(0.051)			(0.041)	(0.033)
HasChild × 1994			0.027	0.009			-0.031*	-0.035**
			(0.025)	(0.019)			(0.018)	(0.015)
HasChild × 1996×								
1(AFQT Z-Score ≤ 0)	0.111	0.604	0.028	-0.002	0.061	0.162	0.017	0.010
			(0.024)	(0.021)			(0.023)	(0.018)
1(0 < AFQT Z-Score ≤ 1)			-0.008	-0.036			0.006	0.025
			(0.066)	(0.040)			(0.035)	(0.031)
1(1 < AFQT Z-Score)			-0.018	0.099*			-0.023	-0.029
			(0.109)	(0.056)			(0.036)	(0.040)
HasChild × 1996			-0.004	-0.031			-0.017	-0.029
			(0.027)	(0.022)			(0.021)	(0.018)
HasChild × 1998×								
1(AFQT Z-Score ≤ 0)	0.109	0.618	0.052**	-0.002	0.061	0.175	0.029	0.014
			(0.026)	(0.025)			(0.024)	(0.021)
1(0 < AFQT Z-Score ≤ 1)			-0.007	-0.016			-0.060	-0.014
			(0.066)	(0.046)			(0.037)	(0.036)
1(1 < AFQT Z-Score)			0.034	-0.017			0.105**	0.010
			(0.089)	(0.062)			(0.050)	(0.046)
HasChild × 1998			-0.007	-0.036			-0.014	-0.057***
			(0.027)	(0.025)			(0.022)	(0.021)
Maximum AFDC/TANF for	0.088	0.586	-0.004*	-0.012***	0.048	0.145	0.000	0.002
Family of Three			(0.002)	(0.002)			(0.001)	(0.002)
Number of Observations			16474	37903			23335	37903
Number of Individuals			4500				5640	
R^2			0.058	0.285			0.102	0.065

Notes: Standard errors are in parentheses and are clustered at the individual level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The data come from the National Longitudinal Survey of Youth 1979 and cover survey years 1991 through 1998. The dependent variable is equal to one if the individual is married (divorced) and zero if he is not. The definitions of AFDC eligibility and *TreatChild* are explained in the text. Other control variables included in all regressions, but whose coefficients are not reported, are dummies for educational level group, dummies for age group, race, gender, current, one-period, and two-period lagged number of children, state dummies, year dummies, and pairwise interactions of the treatment variables and year dummies. An individual is included in the marriage sample if he is not currently married, but is included in his first year of marriage. An individual is included in the divorce sample if he is currently married, but is included in the first year of divorce. If an individual marries and divorces during the sample, he re-enters the marriage sample.

Table 5: Comparing the Use of AFQT and Education Groups as Methods for Differentiating Likely Low- and High-Earning Families

	Outcome: Married					Outcome: Divorced				
	Mean of dependent variable	Full risk sample		Edu ≤ some college risk sample	Edu ≤ exactly HS risk sample	Mean of dependent variable	Full risk sample		Edu ≤ some college risk sample	Edu ≤ exactly HS risk sample
		AFQT	Education				AFQT	Education		
$HasChild \times 1994 \times$										
$LessThanHS$	0.076	-0.004	0.009	-0.014	-0.053*	0.043	0.022	-0.014	-0.034	-0.028
		(0.023)	(0.051)	(0.032)	(0.030)		(0.019)	(0.036)	(0.039)	(0.037)
$ExactlyHS$		0.058	0.065	0.041			-0.060**	0.013	-0.005	
		(0.062)	(0.050)	(0.031)			(0.029)	(0.019)	(0.025)	
$SomeCollege$		0.169	0.024				0.037	0.017		
		(0.116)	(0.052)				(0.041)	(0.023)		
$HasChild \times 1994$		0.027	-0.015	0.012	0.043*		-0.031*	-0.016	-0.000	0.001
		(0.025)	(0.047)	(0.025)	(0.023)		(0.018)	(0.013)	(0.021)	(0.019)
$HasChild \times 1996 \times$										
$LessThanHS$	0.111	0.028	0.106**	0.040	0.045	0.061	0.017	0.068**	0.050	0.115***
		(0.024)	(0.053)	(0.039)	(0.034)		(0.023)	(0.029)	(0.034)	(0.034)
$ExactlyHS$		-0.008	0.054	-0.006			0.006	-0.047*	-0.066**	
		(0.066)	(0.050)	(0.034)			(0.035)	(0.025)	(0.030)	
$SomeCollege$		-0.018	0.061				-0.023	0.020		
		(0.109)	(0.053)				(0.036)	(0.025)		
$HasChild \times 1996$		-0.004	-0.042	0.023	0.006		-0.017	-0.008	0.007	-0.053**
		(0.027)	(0.047)	(0.028)	(0.023)		(0.021)	(0.015)	(0.022)	(0.023)
$HasChild \times 1998 \times$										
$LessThanHS$	0.109	0.052**	0.101**	0.040	0.065*	0.061	0.029	-0.009	-0.033	-0.037
		(0.026)	(0.050)	(0.040)	(0.037)		(0.024)	(0.039)	(0.043)	(0.041)
$ExactlyHS$		-0.007	0.037	-0.023			-0.060	0.025	0.002	
		(0.066)	(0.046)	(0.036)			(0.037)	(0.024)	(0.029)	
$SomeCollege$		0.034	0.059				0.105**	0.025		
		(0.089)	(0.049)				(0.050)	(0.027)		
$HasChild \times 1998$		-0.007	-0.026	0.036	0.001		-0.014	-0.011	0.010	0.019
		(0.027)	(0.041)	(0.029)	(0.025)		(0.022)	(0.016)	(0.024)	(0.021)
Maximum AFDC/TANF for Family of Three	0.088	-0.004*	-0.014***	-0.014***	-0.015***	0.048	0.000	0.001	0.002	0.001
		(0.002)	(0.004)	(0.004)	(0.004)		(0.001)	(0.002)	(0.002)	(0.003)
Number of Observations		16474	17165	13790	9669		23335	24404	18602	13022
Number of Individuals		4500	4710	3880	2816		5640	5931	4674	3361
R^2		0.058	0.059	0.058	0.065		0.102	0.104	0.114	0.125

Notes:Standard errors are in parentheses and are clustered at the individual level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The data come from the National Longitudinal Survey of Youth 1979 and cover survey years 1991 through 1998. The dependent variable is equal to one if the individual is married (divorced) and zero if he is not. Columns labeled "AFQT" estimate models following Equation 1.1, and columns labeled "Education" replace the AFQT variable with a vector of education level indicator variables. The definitions of AFDC eligibility and $TreatChild$ are explained in the text. Other control variables included in all regressions, but whose coefficients are not reported, are dummies for educational level group, dummies for age group, race, gender, current, one-period, and two-period lagged number of children, state dummies, year dummies, and pairwise interactions of the treatment variables and year dummies. An individual is included in the marriage sample if he is not currently married, but is included in his first year of marriage. An individual is included in the divorce sample if he is currently married, but is included in the first year of divorce. If an individual marries and divorces during the sample, he re-enters the marriage sample.

Table 6: Men and Women Separately: Effects of the EITC and AFDC/TANF on Marriage and Divorce

	Outcome: Married				Outcome: Divorced			
	Women only		Men only		Women only		Men only	
	Mean of dependent variable		Mean of dependent variable		Mean of dependent variable		Mean of dependent variable	
HasChild × 1994×								
1(AFQT Z-Score ≤ 0)	0.075	0.076	0.001	0.023	0.046	0.041	0.045	0.007
			(0.033)	(0.046)			(0.028)	(0.024)
1(0 < AFQT Z-Score ≤ 1)			0.011	0.056			-0.087**	-0.037
			(0.077)	(0.128)			(0.041)	(0.040)
1(1 < AFQT Z-Score)			0.144	0.323			0.004	0.046
			(0.120)	(0.345)			(0.054)	(0.056)
HasChild × 1994			0.006	0.007			-0.040	-0.022
		(0.034)	(0.051)			(0.026)	(0.025)	
HasChild × 1996×								
1(AFQT Z-Score ≤ 0)	0.119	0.104	0.015	0.037	0.068	0.054	-0.041	0.048*
			(0.039)	(0.039)			(0.043)	(0.026)
1(0 < AFQT Z-Score ≤ 1)			-0.007	0.021			0.012	-0.001
			(0.087)	(0.136)			(0.051)	(0.047)
1(1 < AFQT Z-Score)			0.007	-0.085			-0.014	-0.028
			(0.118)	(0.282)			(0.059)	(0.044)
HasChild × 1996			0.017	-0.075*			0.004	-0.030
		(0.038)	(0.045)			(0.032)	(0.027)	
HasChild × 1998×								
1(AFQT Z-Score ≤ 0)	0.107	0.111	0.025	0.071	0.062	0.060	0.040	0.023
			(0.041)	(0.051)			(0.034)	(0.032)
1(0 < AFQT Z-Score ≤ 1)			-0.090	0.109			-0.096*	-0.040
			(0.089)	(0.127)			(0.050)	(0.052)
1(1 < AFQT Z-Score)			0.028	0.175			0.165**	0.053
			(0.105)	(0.238)			(0.082)	(0.057)
HasChild × 1998			-0.030	-0.006			-0.008	-0.019
		(0.038)	(0.056)			(0.031)	(0.031)	
Maximum AFDC/TANF for Family of Three	0.090	0.086	0.004	-0.002	0.051	0.046	-0.004*	0.000
			(0.004)	(0.005)			(0.002)	(0.002)
Number of Observations			7577	8897			11293	12042
Number of Individuals			2108	2392			2868	2772
R^2			0.045	0.090			0.040	0.195

Notes: Standard errors are in parentheses and are clustered at the individual level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The data come from the National Longitudinal Survey of Youth 1979 and cover survey years 1991 through 1998. The dependent variable is equal to one if the individual is married (divorced) and zero if he is not. The definitions of AFDC eligibility and *TreatChild* are explained in the text. Other control variables included in all regressions, but whose coefficients are not reported, are dummies for educational level group, dummies for age group, race, current, one-period, and two-period lagged number of children, state dummies, year dummies, and pairwise interactions of the treatment variables and year dummies. An individual is included in the marriage sample if he is not currently married, but is included in his first year of marriage. An individual is included in the divorce sample if he is currently married, but is included in the first year of divorce. If an individual marries and divorces during the sample, he re-enters the marriage sample.

Table 7: Summary Statistics of Men and Women who Change Family Structure

Variable	Mean
Men who marry	
# of children (current period)	1.918
# of children (last period)	1.630
Age of youngest child (current period)	4.972
Age of youngest child (last period)	4.822
Women who marry	
# of children (current period)	1.971
# of children (last period)	1.824
Age of youngest child (current period)	7.824
Age of youngest child (last period)	7.433
Men who divorce	
# of children (current period)	0.639
# of children (last period)	2.103
Women who divorce	
# of children (current period)	1.968
# of children (last period)	2.169

Notes: The data come from the National Longitudinal Survey of Youth 1979 and cover survey years 1991 through 1998.

Table 8: The Effects of Taxes and Transfers: Using EITC Generosity as a Continuous Variable

	Outcome: Married		Outcome: Divorced	
	Mean of dependent variable		Mean of dependent variable	
$HasChild \times MaxEITC_t \times$ (AFQT Z-Score$_i \leq 0$)	0.088	0.045**	0.048	0.022
		(0.020)		(0.018)
$HasChild \times MaxEITC_t \times$ ($0 <$ AFQT Z-Score$_i \leq 1$)		-0.010		-0.021
		(0.054)		(0.028)
$HasChild \times MaxEITC_t \times$ ($1 <$ AFQT Z-Score$_i$)		0.016		0.027
		(0.084)		(0.035)
$HasChild \times MaxEITC_t$		-0.000		-0.000
		(0.000)		(0.000)
Maximum AFDC/TANF for Family of Three		-0.005**		0.001
		(0.002)		(0.001)
Number of Observations		16474		23335
Number of Individuals		4500		5640
R^2		0.055		0.097

Notes: Standard errors are in parentheses and are clustered at the individual level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The data come from the National Longitudinal Survey of Youth 1979 and cover survey years 1991 through 1998. The dependent variable is equal to one if the individual is married (divorced) and zero if he is not. The definitions of AFDC eligibility and *TreatChild* are explained in the text. Other control variables included in all regressions, but whose coefficients are not reported, are dummies for educational level group, dummies for age group, race, current and one-period lagged number of children, state dummies, year dummies, and pairwise interactions of the treatment variables and year dummies. An individual is included in the marriage sample if he is not currently married, but is included in his first year of marriage. An individual is included in the divorce sample if he is currently married, but is included in the first year of divorce. If an individual marries and divorces during the sample, he re-enters the marriage sample.

1.A Appendix: Theoretical Model of Dynamic Selection

The model considers a single individual who has a "propensity to marry" (θ) that is randomly drawn from a uniform distribution between $[0,1]$.[42] Each period an unmarried individual is randomly matched with a potential partner and receives a match quality shock (ε), where $\varepsilon \sim$ i.i.d. $N(0,1)$. When deciding whether to marry, the individual compares his utility while single (U^s) to his utility while married ($U^m + \varepsilon$). If $U^m + \varepsilon > U^s$ then he marries, and if $U^m + \varepsilon < U^s$ then he remains single. I assume that U^s is strictly decreasing in θ and that U^m is strictly increasing in θ. In other words, utility of being single decreases as one's propensity to marry increases, and utility of being married increases as one's propensity to marry increases. I illustrate this relationship as well as a hypothetical increase in marriage incentives in Figure A1.

These assumptions imply that there is some threshold level of $\hat{\theta}$, where individuals with $\theta > \hat{\theta}$ will marry and individuals with $\theta < \hat{\theta}$ will remain single, but the i.i.d. match quality shock will provide some heterogeneity in the decision. Specifically:

$$\begin{aligned} Pr[marry] = & \quad Pr[U^m + \varepsilon > U^s] \\ = & \quad Pr[\varepsilon > -(U^m - U^s)] \\ = & \quad 1 - \Phi(-(U^m - U^s)) \end{aligned}$$

Without any change in marriage incentives, the differential probability of marrying along the distribution of θ will skew the sample of unmarried individuals over time to contain relatively more individuals with low levels of θ. I illustrate this effect in Figure A2, which displays the density of θ within the unmarried population over time. In Period 0 everyone is unmarried, so the distribution of θ is the uniform distribution. Each period the random match quality shock causes more people with higher values of θ to marry than it does people with lower values of θ, creating a sample that, over time, will contain a larger proportion of individuals with low levels of θ even without any changes in marriage incentives.

An increase in marriage incentives will shift the U^m line up to $U^{m'}$ and create a new threshold

42. θ could also be described as a "desire to marry" or a "desire to start a family."

level of theta ($\hat{\theta}'$). Combined with the already existing trend of the sample, this change in marriage incentives will exacerbate the issue of dynamic selection due to an even larger probability of marrying among individuals with higher levels of θ. Therefore, the distribution skews even more heavily to the left as marriage incentives increase.

This dynamic selection into single status creates important implications for the sample of unmarried individuals in my empirical analysis and may introduce bias to the estimates in later years. If this dynamic selection story is true then we would expect to see decreases in the probability of marrying in later sample years due to the policy. This is because the sample of unmarried individuals in later years contains relatively more individuals with low levels of θ relative to the sample of unmarried individuals in earlier years. The policy encouraged marginal individuals to marry, leaving behind a population of unmarried individuals who are less likely to marry compared to the earlier sample.

Figure A1: Illustrated Theoretical Model

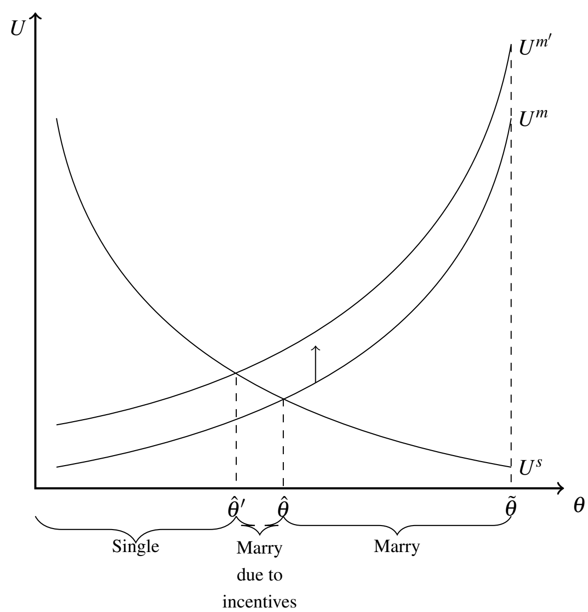
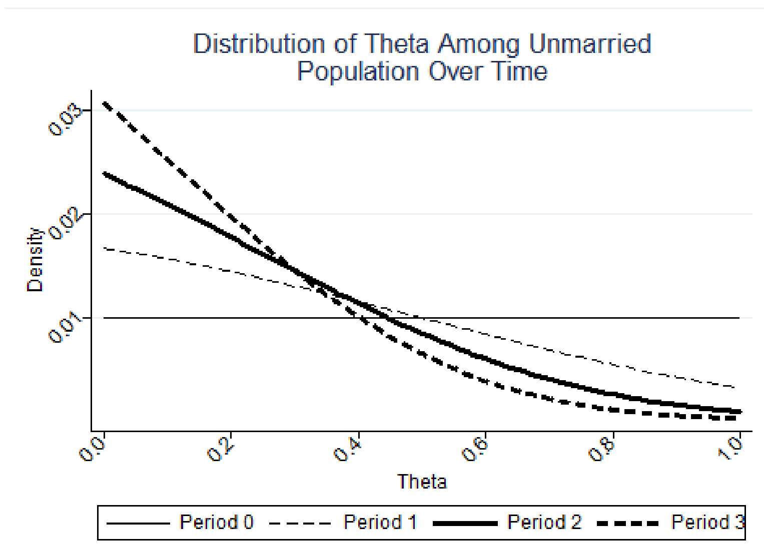


Figure A2: Distribution of Theta Over Time



Chapter 2

Suddenly Married: Joint Taxation and the Labor Supply of Same-Sex Married Couples After *U.S. v. Windsor*

2.1 Introduction

Many developed countries institute progressive tax systems, forcing them to choose between tax equity across marital status or tax equity across family income (Rosen 1977). Greater equity across family income (i.e., jointness) can exacerbate efficiency costs and labor supply distortions, yet direct evidence of the efficiency costs or labor supply effects of joint taxation is relatively rare due to a lack of natural experiments involving a large-scale switch in systems. Another complication is that such switches generally involve a change in marriage incentives as well, introducing another margin of distortion. In this paper, I provide direct evidence of the efficiency costs and labor supply effects of joint taxation in the United States by leveraging tax variation created by federal recognition of existing same-sex marriages following the 2013 *United States v. Windsor* Supreme Court ruling.

The United States' 1996 Defense of Marriage Act prevented same-sex marriages from being recognized at the federal level, but states retained the authority to permit same-sex marriages at the state level. This legislative environment meant that same-sex married couples were still required to file federal taxes as two single individuals through 2012. In June 2013 over 71,000 marriages recognized by states since 2004 were suddenly recognized by the United States federal government as a result of the *United States v. Windsor* ruling (DeSilver 2013). The ruling required same-sex married couples to file federal taxes as married.[1]

This shift of already-married couples from the individual to the family as the unit of taxation is unusual because governments usually employ only one system at any particular time. I leverage separate variation in household after-tax income and marginal tax rates generated by federal same-sex marriage recognition to separately identify the income and substitution effects of taxation and, therefore, compensated (Hicksian) labor supply elasticities. I then use my estimates in a sufficient statistic framework to calculate the additional deadweight loss and tax revenue created by joint taxation relative to individual taxation.

I use the 2012–2015 waves of the American Community Survey, which are the first of the U.S. Census Bureau surveys to explicitly identify same-sex married couples. I use a generalized difference-in-differences framework, which compares predicted changes in marginal tax rates and household after-tax income for a treatment group of individuals in same-sex couples who married before the *United States v. Windsor* ruling to a control group of individuals in same-sex cohabiting couples.[2,3] I focus on same-sex couples who had already married before the Supreme Court ruling in order to exploit the plausibly exogenous shock to federal marital status. I also extend work by Baldwin, Allgrunn, and Ring (2011) and McClelland, Mok, and Pierce (2014) by predicting higher or lower earning status and separately analyzing the labor supplies of each partner, which

1. This paper does not examine the 2015 *Obergefell v. Hodges* Supreme Court ruling, which required all states to permit same-sex marriages. The *United States v. Windsor* ruling only required the federal government to recognize at the federal level same-sex marriages that were permitted by states.

2. My generalized difference-in-differences approach can also be thought of as a treatment intensity specification, in which treated individuals received differing intensities of treatment reflected in the changes in their tax rates and tax liabilities. The implicit assumption throughout is that there are parallel trends between the treatment and control groups, and I employ an event study approach to provide evidence that the parallel trends assumption is satisfied.

3. I use opposite-sex married couples as an alternative control group in one of my robustness checks and find that the estimates are similar to my main results.

is otherwise commonly accomplished by separating the couple by gender. Higher or lower earners may respond to taxation differently due, for example, to differing work preferences or attachment to the labor force. Separately analyzing each partner in same-sex couples yields separate estimates of the effects of taxation by predicted earning status while abstracting away from gender.

Distinct variation in tax rates and household income is crucial in estimating compensated labor supply elasticities. To estimate the effect of changes in tax rates on labor supply, I use the predicted percentage change in the federal marginal net-of-tax rate.[4] Variation in an individual's tax rates due to the Supreme Court ruling originates through two primary channels: differences in tax bracket definitions between the single and joint schedules, and the addition of both partners' earnings to taxable income rather than only the individual's earnings.[5] To estimate the income effect, I measure changes in the couple's predicted marriage subsidy (or penalty), which is a common measure of marriage incentives under joint taxation (Alm and Whittington 1999; Eissa and Hoynes 2003; Isaac 2017), but which is new to the literature estimating income effects of taxation because shifting from the single to joint tax schedule is endogenous in most circumstances. The marriage subsidy is defined as the difference between the sum of the individuals' tax liabilities if they are single and the couple's joint tax liability if they are married. To avoid reverse causality issues associated with using reported earnings to measure tax changes, I use predicted earnings and the NBER TAXSIM model to quantify the variation in predicted tax rates and household after-tax income.

On average, predicted higher earners in my sample faced a 2% increase and predicted lower earners faced a 5% decrease in their federal marginal net-of-tax rates as a result of the *United States v. Windsor* ruling. In addition, the average same-sex married couples experienced a decrease of $642.09 in household after-tax income due to the marriage subsidy, with substantial variation across households depending on total household earnings and how those earnings are split between partners.

In my estimation results, I find that predicted lower earners respond to taxation along the exten-

4. The marginal net-of-tax rate is one minus the marginal tax rate.

5. I do not treat the spouse's earnings as one's own unearned income in my empirical specification because the partners are taxed separately.

sive margin, but predicted higher earners do not. I estimate significant extensive margin Hicksian (wage) and income participation elasticities among predicted lower earners, which also vary by gender, with higher responsiveness to changes in marginal net-of-tax rates among male predicted lower earners compared to female predicted lower earners. I estimate relatively large extensive margin Hicksian participation elasticities of of 1.236 and 0.400 for male and female predicted lower earners, respectively, and income participation elasticities of -0.022 and -0.044 for male and female predicted lower earners, respectively.

In contrast, I find that predicted higher earners respond to taxation along the intensive margin, but predicted lower earners do not. I calculate that the income and substitution effects largely off-set each other, resulting in a small uncompensated (Marshallian) hours elasticity among predicted higher earners. The intensive margin elasticity estimates are very similar when using a sample of both men and women, men only, or women only. I estimate significant Hicksian hours elasticities between 0.360 and 0.382 and significant income hours elasticities between -0.010 and -0.013 among predicted higher earners. These Hicksian elasticity estimates are in-line with others in the literature using different populations of interest and sources of variation (see Keane's (2011) review of estimated labor supply responses to taxation).[6]

Finally, I derive a sufficient statistic formula to calculate the changes in deadweight loss and tax revenue imposed by joint taxation relative to individual taxation (Feldstein 1999; Immervoll et al. 2007; Chetty 2009). My estimates suggest that the shift to joint taxation among same-sex married couples induced \$398 million in additional deadweight loss and cost \$294 million in tax revenue, which, compared to the total amount of personal income tax collected by the federal government from same-sex married couples, amounts to an additional efficiency cost of approximately 26.2%

6. I examine several potential influences that might confound my estimates including classification error between same-sex and opposite-sex married couples in the American Community Survey, compliance barriers to filing jointly following *United States v. Windsor*, sample selection bias among the treatment group, and state-year specific shocks that may affect labor supply. I find that my original results are robust to all of these alternative specifications.

and tax revenue losses of approximately 19.4%.[7,8] If my elasticity estimates apply to all married couples in the United States, I find that joint taxation creates $79.0 billion in additional deadweight loss and generates $67.9 billion less in tax revenue relative to individual taxation, which, compared to the total amount of personal income tax collected by the federal government from all married couples filing jointly, amounts to an additional efficiency cost of approximately 8.1% and tax revenue losses of approximately 6.9%.[9]

My research is grounded in traditional labor supply and taxation questions, but also adds to the small, but growing, literature concerning same-sex couples. Prior research of same-sex couples and LGBT individuals has focused on workplace discrimination (Badgett 1995; Carpenter 2007; Plug, Webbink, and Martin 2014), health outcomes (Buchmueller and Carpenter 2010; Gonzales and Blewett 2014), or differences in labor market behavior between same- and opposite-sex couples (Tebaldi and Elmslie 2006; Oreffice 2011; Antecol and Steinberger 2013). This paper is the first, to the best of my knowledge, to leverage tax variation among same-sex married couples to identify the effects of taxation on the labor supply of married couples. My analysis provides direct evidence of the additional efficiency costs and reduced tax revenue of joint taxation relative to individual taxation, and suggests that the efficiency of the United States tax system can be improved by lowering tax rates among secondary earners so as to mitigate the efficiency costs along the extensive margin from those workers. However, whether increased efficiency is worth the lower associated tax equity across families remains an open question.

The remainder of the paper is organized as follows. Section 2.2 discusses policy background and prior research. Section 2.3 discusses the data and Section 2.4 presents the empirical strategy along with the results. Section 2.5 presents robustness checks and alternative specifications.

7. 26.2% and 19.4% are relative to approximately $1.5 billion in individual income tax revenue from same-sex married couples. I calculate total personal income tax from same-sex married couples using tax liabilities calculated by NBER's TAXSIM model and the population estimates of same-sex married couples in the American Community Survey.

8. My conclusion that *United States v. Windsor* decreased federal tax revenue by $294 million is consistent with predictions by Alm, Leguizamon, and Leguizamon (2014).

9. 8.1% and 6.9% are relative to approximately $984 billion in personal income tax collected by the federal government from all married couples filing jointly in fiscal year 2014 (U.S. Department of the Treasury, Internal Revenue Service 2016).

Finally, Section 2.6 presents the deadweight loss and tax revenue analysis, and Section 2.7 concludes.

2.2 Background

In this section I present a brief overview of same-sex marriage legislation in the United States and prior research concerning labor supply responses to taxation and economic outcomes for same-sex couples and LGBT individuals.

2.2.1 Same-Sex Marriage Legislation in the United States

The Defense of Marriage Act (DOMA) was established in 1996 and defined "marriage," for federal government purposes, as the union between one man and one woman; it defined "spouse" as a member of the opposite sex who is a husband or wife. Despite this definition at the federal level, states were allowed to decide for themselves whether they would recognize same-sex marriages. Some states, beginning with Vermont in 2000, began recognizing civil unions or domestic partnerships for same-sex couples who were interested in a marriage-like commitment. These institutions, however, are not entirely identical to marriages even at the state level, are not recognized as marriages under federal law, and may not be recognized by other states.

The first state to recognize same-sex marriages was Massachusetts in 2004, and after that 13 other states passed same-sex marriage legalization provisions before the Supreme Court ruling in *United States v. Windsor* in June 2013.[10] Figure 1 presents a timeline of same-sex marriage legalization in the U.S.[11] In addition, DOMA did not require states to recognize same-sex marriages granted by other states. Therefore, despite the fact that some states recognized same-sex mar-

10. California allowed same-sex marriages beginning in May 2008, but Proposition 8 (passed and enacted in November 2008) prevented new same-sex marriage licenses until the federal ruling in *United States v. Windsor*. However, the state continued to recognize marriages among same-sex couples in California who married between May 2008 and November 2008.

11. Figure 1 does not include civil union or domestic partnership legislation, as these institutions are not legally identical to marriage. Those states listed as recognizing same-sex marriages in 2013 are only those that had recognized or voted to recognize same-sex marriages before the Supreme Court ruling in June.

riages, state of residency determined which state benefits same-sex married couples could utilize. DOMA's definitions of marriage and spouse also prevented same-sex married couples from obtaining any federal benefits of marriage available to opposite-sex married couples. This changed with the Supreme Court decision in *United States v. Windsor* in June 2013.

The Supreme Court ruled that DOMA's definitions of marriage and spouse were unconstitutional, thereby requiring the federal government to recognize same-sex marriages allowed by the states. While many federal statutes recognize marriages according to where they occurred (the place-of-celebration rule), many others only recognize marriages that are, in turn, recognized by the couple's state of residence (the domicile rule). The federal tax code recognizes marriages by the place-of-celebration rule, meaning that the Supreme Court ruling affected same-sex married couples regardless of where they reside or whether their state of residence recognized same-sex marriages. This had immediate effects on the tax environment faced by same-sex couples who had already married by the time of the ruling. Whereas previously these couples filed federal tax returns as two single individuals, beginning in tax year 2013 they were required to file as either married, filing jointly or married, filing separately (U.S. Department of the Treasury, Internal Revenue Service 2013).[12]

It is also worth mentioning the case of *Obergefell v. Hodges*, a second Supreme Court ruling concerning same-sex marriage legislation that was decided in 2015. Following the Supreme Court ruling in *United States v. Windsor*, many states began recognizing same-sex marriages, but many others continued enforcing or enacted new bans on same-sex marriages. The case of *Obergefell v. Hodges* concerned the constitutionality of bans on same-sex marriages, and that ruling determined that marriage is a fundamental right, and therefore required states to allow and recognize same-sex marriages. My analysis ends in 2015 and focuses on same-sex couples who married before 2013,

12. Fisher, Gee, and Looney (2016) note that take-up of joint filing among same-sex married couples may have been less than 100% following *United States v. Windsor* due to compliance barriers in some states. In particular, ten states in 2014 required married couples to file jointly on their state tax returns if they filed jointly on their federal tax returns, while simultaneously not allowing same-sex married couples to file jointly on state tax returns. These states are Alabama, Georgia, Kansas, Kentucky, Louisiana, Michigan, Missouri, Nebraska, North Dakota, and Ohio. I estimate a model excluding individuals living in these states from the sample and find that compliance barriers do not affect my results.

and is therefore not affected by the *Obergefell v. Hodges* ruling.

2.2.2 Literature on Labor Supply and Taxation

Labor supply and responses to taxation have garnered a substantial amount of attention in the economics literature. As described by Keane's (2011) survey, research on the topic can be divided between studies using static and dynamic models as well as between studies of male and female labor supply. Past researchers have generally found small, if any, effects of taxation on male labor supply, with compensated (Hicksian) elasticities ranging from 0.05–0.84 with an average of 0.31. Past researchers have generally estimated larger labor supply responses to taxation among women. My analysis assumes a static model due to a lack of panel data, but I expand upon prior work by distinguishing between predicted higher and predicted lower earners within the couple rather than differentiating by gender, and by using new and unique variation to estimate labor supply elasticities.

Many past studies exploit tax variation caused by small to moderate tax changes among a subset of tax filers. For example, Crossley and Jeon (2007), Saez, Slemrod, and Giertz (2012), and Saez (2016) use variation in tax rates at the top of the earnings distribution to estimate labor supply and earnings responses among the highest earners. Others, such as Eissa and Liebman (1996), Ellwood (2000), Eissa and Hoynes (2004), and Moulton, Graddy-Reed, and Lanahan (2016) use variation in the Earned Income Tax Credit to estimate hours and participation responses to taxes among likely low earners. Fetter and Lockwood (2017) exploit the introduction of the Old Age Assistance Program to estimate labor force participation decisions among older workers. Relative to the literature, the income taxation changes I study in this paper, resulting from the Supreme Court ruling in *United States v. Windsor*, were unusually wide-ranging as they affected all same-sex married couples across the earnings and age distributions.

The *United States v. Windsor* ruling created large-scale variation that is more similar to that exploited by LaLumia (2008), who uses the introduction of joint taxation in the United States in the 1940s to estimate labor force participation responses among married couples, Selin (2014), who estimates the effects of a switch from joint to individual taxation in Sweden in 1971, or Kalíšková (2014), who estimates the impact of moving to joint taxation using a voluntary switch to joint taxation in the Czech Republic in 2005. These are the only studies, to the best of my knowledge, offering direct evidence of the effects of joint taxation through natural experiments from changes in tax systems. Each study exploits changes across the earnings distribution due to introducing or removing joint taxation, but, with the exception of the Czech Republic's policy change, these natural experiments occurred at least 46 years ago.[13] The Supreme Court ruling effectively shifted same-sex married couples from an individual taxation system to one of joint taxation, and, as a result, in this paper, I am able to exploit a recent natural experiment to study changes in both tax rates and tax liabilities.

It is also worth noting that most of the previously mentioned studies distinguish between male and female labor supply, and many assume that the wife is the secondary earner in the household while the husband supplies labor relatively inelastically. The secondary earner's labor supply decision is often assumed to be made after the primary earner's decision, where the secondary earner takes as given her partner's decision. Assuming that the wife is the secondary earner was largely driven by the observation that this was often the case in the data. Blau and Kahn (2007) and Heim (2007), however, document shrinking labor supply elasticities among women over the last four decades, and suggest that work preferences have strengthened among women over time, increasing their earnings and shifting their status as secondary earners. Baldwin, Allgrunn, and Ring (2011) and McClelland, Mok, and Pierce (2014) consider primary or secondary earner status regardless of gender in order to compare the results to those using the male-female split. McClelland, Mok, and Pierce (2014) conclude that the marginal worker is more likely to be the lower earner in the couple rather than assuming the marginal worker is the wife. In this paper, I predict individual earnings using a similar procedure as McClelland, Mok, and Pierce (2014) in order to separate same-sex couples into a sample of predicted primary earners and a sample of predicted secondary earners.

13. The Czech Republic's voluntary switch to joint taxation was relatively short-lived. The Czech Republic instituted a flat tax system beginning in 2008 (Kalíšková 2014).

This paper is also connected to the optimal taxation literature. Kleven, Kreiner, and Saez (2009) analyze optimal income tax rates among married couples and conclude that optimal rates can, in some cases, exhibit negative jointness, meaning that tax rates on secondary earners decrease as their spouse's earnings increase. Feldstein (1999) develops a sufficient statistic approach to estimate the deadweight loss created by income taxation. He calculates deadweight loss equal to 32% of total personal income tax revenue. Immervoll et al. (2007) document comparable levels of deadweight loss in their study of European welfare programs. The authors find deadweight loss ranging from 19% in low-tax countries to 82% in high-benefit countries. I use a sufficient statistic approach similar to Feldstein (1999) and discussed more generally by Chetty (2009) to calculate the additional deadweight loss and tax revenue created by joint taxation relative to individual taxation, concluding that joint taxation creates approximately 8.0% more deadweight loss and 6.9% less tax revenue relative to individual taxation.

2.2.3 Economic Research on Same-Sex Couples and LGBT Individuals

Economic research concerning same-sex couples and LGBT individuals is scarce due primarily to few available sources of data. Until recently, data editing procedures in prior waves of the census and American Community Surveys, such as changing the gender or marital status of same-sex partners, made it difficult to identify same-sex couples in the data.[14]

Oreffice (2011) uses the 2000 decennial census to test whether same-sex cohabiting couples respond to bargaining power shifts in a similar manner as opposite-sex married or cohabiting couples.[15] She finds that relatively richer partners hold more bargaining power in both opposite- and same-sex couples. However, she concludes that older partners in opposite-sex married couples hold more bargaining power, whereas younger partners in same- and opposite-sex cohabiting couples

14. In the 1990 census, if a couple appeared to be a same-sex married couple then one partner's gender was changed so that the couple appeared to be an opposite-sex married couple. In the 2000 and 2010 censuses and in pre-2012 waves of the American Community Survey, if a couple appeared to be a same-sex married couple then their marital status was changed to "unmarried partner," sometimes without an accompanying data quality flag (U.S. Census Bureau 2009).

15. The 2000 census includes an "unmarried partner" option as an individual's relationship to the household head. It was not possible at the time for same-sex couples to marry in the United States.

hold more bargaining power.

Antecol and Steinberger (2013) use the 2000 decennial census to examine how much of the labor supply gap between married women and partnered lesbians can be explained by the presence of children. They find that the presence of children explains 56% of the unconditional annual hours gap between married women in opposite-sex couples and secondary earners in female same-sex couples, and 9% of the gap between married women in opposite-sex couples and primary earners in female same-sex couples.

Stevenson (2012) uses the 2003–2004 American Community Survey to predict the expected labor supply and federal tax revenue consequences of federal same-sex marriage legalization by exploiting variation caused by the Jobs and Growth Tax Relief Reconciliation Act of 2003. The author estimates statistically insignificant uncompensated own-wage elasticities among male primary earners in both opposite- and same-sex couples, and estimates opposite-signed elasticities among female secondary earners in same-sex couples. The author predicts a $20–40 million increase in federal tax revenue resulting from same-sex marriage legalization.

Alm, Leguizamon, and Leguizamon (2014) use the 2010 American Community Survey, and also predict the state and federal income tax consequences of federal same-sex marriage legalization. The authors' preferred estimates conclude that the federal government could gain $5.7 million or lose up to -$315.8 million. My calculations of the change in tax revenue as a result of *United States v. Windsor* are in line with estimates from Alm, Leguizamon, and Leguizamon (2014).

Other research on LGBT individuals has focused on workplace discrimination (Badgett 1995; Carpenter 2007; Plug, Webbink, and Martin 2014), health outcomes (Buchmueller and Carpenter 2010; Gonzales and Blewett 2014), or differences in labor market behavior between same- and opposite-sex couples (Tebaldi and Elmslie 2006). Overall, while the literature concerning same-sex couples and LGBT individuals has grown, labor supply studies have traditionally focused on differences between same-sex and opposite-sex couples' behavior or estimating tax revenue consequences of same-sex marriage legalization. This has largely been due to few available sources of data concerning same-sex couples, especially considering that the same-sex couple population is very small relative to the opposite-sex couple population. The Supreme Court ruling in *United States v. Windsor* and improved data quality allow me to exploit new variation among this understudied population in order to address classic economic questions concerning labor supply responses to income taxation.

2.3 Data

I use the 2012–2015 waves of the American Community Survey to estimate the effects of taxation on annual hours of work and labor force participation. The 2012 wave of the ACS was the first of the U.S. Census Bureau surveys to explicitly identify same-sex married couples, whereas data editing procedures in prior waves presented substantial obstacles to identifying same-sex couples.[16]

My main sample includes same-sex married couples who married in 2012 or earlier and same-sex cohabiting couples in which both partners are between 25 and 54 years old (inclusive).[17] I identify cohabiting couples as those who label themselves as an "unmarried partner" of the household head, which is distinct from renters, roommates, or friends. I include same-sex couples with at least one earner when examining labor force participation. I further restrict the sample to couples in which both members work when examining annual hours of work.[18] This leaves me with a sample of 13,220 same-sex couples in my extensive margin sample (4,116 married couples and 9,104 cohabiting couples), and a sample of 10,343 same-sex couples in my intensive margin sample (3,022 married couples and 7,321 cohabiting couples).

16. The 2013–2015 waves of the American Community Survey contain a new variable, *SSMC*, explicitly identifying same-sex couples. The 2012 wave contains a data quality flag that indicates whether a couple's status was changed from "same-sex married couple" to "same-sex cohabiting couple," allowing me to identify same-sex married couples in this year as well. Previous waves of the American Community Survey do not contain the *SSMC* variable or the marital status data quality flag.

17. I can observe when a couple married, and so can condition on marrying before the Supreme Court ruling. I cannot, however, observe when a cohabiting couple began their relationship or their cohabitation. This limitation means that the control group of same-sex cohabiting couples may contain some very recent couples, although I provide graphical evidence that mean demographic variables remained stable over time. To the extent that relationship duration affects labor supply, my results may be biased. I also consider an alternative control group of opposite-sex married couples in which I can also condition on marrying in 2012 or earlier and obtain qualitatively similar coefficient estimates compared to by main results, suggesting that issues with the cohabiting couples control group are not a large concern.

18. This restriction eliminates 2,877 same-sex couples, 1,094 of whom are married.

2.3.1 Predicted Earnings

I seek to estimate the effects of taxation on labor supply, but an individual's labor supply decisions affect her marginal tax rate and tax liability through her earnings. Using reported earnings, therefore, to calculate marginal tax rates or tax liabilities would introduce endogeneity through reverse causality, since higher earnings are associated with higher marginal tax rates. I therefore generate a plausibly exogenous measure of an individual's taxes by predicting individual earnings based on predetermined characteristics and use predicted earnings, rather than reported earnings, to estimate each individual's tax rates and liabilities.

I predict individual earnings within the sample of all same-sex couples using age and its square, years of education and its square, race, gender, occupation, college major, and state and year fixed effects.[19,20] These predicted earnings slightly understate reported earnings.[21,22] In particular, the prediction tends to miss the right tail of the earnings distribution. 67.6% of the extensive margin sample are labeled with a predicted primary or secondary earning status that matches their observed earning status.[23,24]

The earnings prediction also allows me to assign predicted primary or secondary earning status in the couple and analyze each group separately, which is common in this field. However, past studies often separate the couple by gender and assume that the wife is the secondary earner,

19. The occupation variable offers information for individuals who had worked within the previous five years. I cannot observe occupation information for individuals who are currently unemployed and who have never worked before or for individuals who have not been in the labor force for the past five years. Approximately 3.6% of the full, extensive margin sample are individuals with no occupation information.

20. Some of these variables, such as state of residence, year, and gender, will not vary between partners. These variables offer no additional information in separating the couple into the two predicted earner samples, but they offer greater explanatory power in obtaining a plausibly exogenous measure of earnings, which will enter into their tax rate and tax liability calculations.

21. The adjusted R^2 of this regression is 0.3064.

22. The predicted earnings procedure allows me to assign policy variables to non-workers, but it raises the Heckman concern that non-workers may have systematically lower wages than workers. This concern is not a substantial problem in my context because I use tax rates, rather than earnings, to explain labor supply behavior. Another issue is that earnings depend on both the wage and hours of work, so that my prediction process is akin to a Mincer wage equation where I use the predicted wage and full-time hours to predict tax changes. However, my process may capture systematic differences in hours of work, as well.

23. In other words, 67.6% of the extensive margin sample are observed to be secondary (primary) earners using reported earnings and are also predicted to be secondary (primary) earners using predicted earnings. This statistic decreases to 65.1% of the intensive margin sample.

24. The presence of mis-classified individuals in terms of earning status would bias my estimates toward zero.

which is not possible in this paper because I focus on same-sex couples, and which may, in fact, be a stronger assumption today for opposite-sex couples than in the past (Baldwin, Allgrunn, and Ring 2011; McClelland, Mok, and Pierce 2014). Separating each couple by predicted earning status allows individual responses to taxation to differ between predicted higher and lower earners, such as different coefficient signs, magnitudes, or intensive or extensive response margins.

Table 1 presents summary statistics of the predicted higher and predicted lower earners in same-sex married and cohabiting couples in the 2012–2015 American Community Survey. Notably, male same-sex couples are slightly less likely to be married.[25] Same-sex married couples are slightly older and earn slightly more than same-sex cohabiting couples. It is also worth noting that same-sex partners have, on average, at least some college education. Conditional on working, predicted higher earners generally work full-time, full-year, whereas predicted lower earners work less. Predicted lower earners in cohabiting couples work about 167 more hours per year than predicted lower earners in married couples.

Figure 2 displays means of the demographic variables in Table 1 to show their stability over time. Similar trends over time between same-sex married and cohabiting couples would provide visual evidence that the parallel trends assumption may be satisfied between the two groups. The presence of children is the only demographic variable with a noticeably divergent trend between 2012–2013, after which the means appear parallel between same-sex married and cohabiting couples. The remaining demographic variables do not exhibit visual differences in mean trends over time, lending some support in favor of the parallel trends assumption.

2.3.2 Tax Change Measures

Two couples with the same total income may experience different changes in their federal tax liabilities as a result of the ruling due to the fact that a progressive, household-based tax system is not marriage neutral. The Supreme Court ruling shifted same-sex married couples to the joint tax

25. The gender ratio between likely higher and lower earners need not be exactly 50%. A larger number of female, rather than male, same-sex married couples would shift this ratio above 50% in favor of women.

schedule and suddenly altered their federal marginal tax rates and their total taxes owed. Generally speaking, a couple with two relatively equal earners will pay more in federal taxes upon marriage (a marriage penalty), whereas a couple with two relatively unequal earners will pay less upon marriage (a marriage subsidy). By focusing on same-sex couples who had already married by the time of the Supreme Court ruling, I can abstract away from the marriage decision and use tax rate and household income changes associated with the change in federal marriage recognition.[26]

The variation in marginal tax rates and in total taxes paid are due to not only the *United States v. Windsor* ruling but also over-time (i.e., schedule movement) and cross-sectional sources (i.e., partner's earnings). Panel data would be ideal in this situation, but to my knowledge no panel survey of labor supply yet exists with sufficient observations of same-sex married couples. The 2012–2015 waves of the American Community Survey are the first of the U.S. Census Bureau surveys to explicitly identify same-sex married couples, which, along with their size, make them the best sources of data to examine the labor supply effects of taxation among same-sex married couples. To overcome the limitation of using repeated cross-sections rather than panel data, I simulate the expected federal marginal tax rate and tax liability changes a same-sex married couple would have faced as a result of the Supreme Court ruling.

When simulating the changes in marginal tax rates and tax liabilities, I define tax units in the data based on reported marital status, school enrollment, age, and whether the federal government recognized same-sex marriage. I divide same-sex married couples in the 2012–2013 waves (tax years 2011–2012) into two tax units because these couples were required to file as two single individuals for their federal tax returns.[27] I assume that the predicted higher earner claims any

26. Same-sex married couples would not have faced state tax changes due to the ruling in *United States v. Windsor* because the ruling affected only the federal definitions of "marriage" and "spouse" contained in the Defense of Marriage Act. The 2015 Supreme Court ruling in *Obergefell v. Hodges*, however, established marriage as a fundamental right, which would have altered the state legislative landscapes faced by same-sex married couples. It is possible that same-sex couples faced marriage incentives from their state tax code if they lived in a state that legalized same-sex marriage before the *United States v. Windsor* ruling. I do not consider this in my empirical strategies, but these incentives would, in general, be substantially smaller than those created by the federal tax code. Indeed, Light and Omori (2008) find that marriage penalties created by state taxes do not have significant effects on family structure decisions.

27. I implicitly assume that the couple's reported marital status is correct. In other words, if a same-sex couple reports themselves to be married even though they reside in a state that does not recognize same-sex marriages then I assume the couple married in a state that did recognize same-sex marriages.

dependent children for tax purposes in this case and files as "head of household." I calculate predicted individual and couple federal tax rates and liabilities using NBER's TAXSIM model, and I discuss the details of these calculations below.

The Marginal Net-of-Tax Rate

Variation in marginal tax rates due to the Supreme Court ruling comes from two sources: the shift from the single to joint tax schedule and the addition of both partners' earnings into taxable income. The variation due to the shift in tax schedules is over-time, whereas the variation due to the addition of both partners' earnings is cross-sectional.

In order to capture both sources of variation using cross-sectional data, I first use predicted earnings and the NBER TAXSIM model to calculate the federal marginal net-of-tax rate each couple faced after the Supreme Court ruling.[28] I then adjust the tax year to 2012 (the tax year before the policy came into effect), adjust the individual's filing status, dependents, and predicted earnings to reflect the pre-ruling tax environment, and calculate each individual's simulated pre-ruling federal marginal net-of-tax rate. Finally, I express the percentage change in the marginal net-of-tax rate from 2013 to the observation year in order to estimate elasticities:

$$\%\Delta(1-\hat{\tau}_{it}) = \begin{cases} \frac{[1-\hat{\tau}_{it}]-[1-\hat{\tau}_{i2013}]}{1-\hat{\tau}_{i2013}}, & \text{if year } \geq 2014 \\ 0, & \text{if year } \leq 2013 \end{cases}$$

where $1-\hat{\tau}_{it}$ is individual i's predicted marginal net-of-tax rate in year $t \geq 2014$ and $1-\hat{\tau}_{i2013}$ is the predicted marginal net-of-tax rate in 2013 (tax year 2012). Table 2 provides illustrative examples of how the variable $\%\Delta(1-\hat{\tau}_{it})$ appears in the data.

The variable $\%\Delta(1-\hat{\tau}_{it})$ is able to capture post-ruling variation in federal marginal net-of-tax rates due to both the shift in tax schedules and the addition of both spouses' earnings.[29] Table 1

28. The marginal net-of-tax rate is one minus the marginal tax rate.

29. I calculate the $\%\Delta(1-\hat{\tau}_{it})$ variable for all individuals in the sample rather than only for same-sex married couples. Doing so allows me to capture additional variation in marginal net-of-tax rates due to other changes to the single filing schedule that is unrelated to the Supreme Court ruling. This additional variation in marginal net-of-tax rates should aid in identification.

shows that, on average, predicted primary earners in same-sex married couples faced a 2% increase in their federal marginal (last-dollar) net-of-tax rates with a standard deviation of 9 percentage points. Marginal net-of-tax rate changes varied from a decrease of 35% to an increase of 83% among predicted primary earners in same-sex married couples. On average, predicted secondary earners in same-sex married couples faced a 5% decrease in their federal marginal (last-dollar) net-of-tax rates with a standard deviation of 10 percentage points. Marginal net-of-tax rates varied from a decrease of 57% to an increase of 55% among predicted secondary earners in same-sex married couples as a result of the *United States v. Windsor* ruling.

To demonstrate how marginal tax rates vary upon marriage relative to filing individually, I plot in Figure 3 the federal marginal tax rates before and after *United States v. Windsor*, which depends on total household earnings and how the earnings are split within the household. Each data point represents a simulated couple and is weighted by the number of individuals within that particular pre- and post-Windsor tax movement bin.[30] The figures show, in general, that secondary earners face higher marginal tax rates under joint taxation relative to individual taxation. Variation among primary earners is more muted, exhibiting smaller changes in marginal tax rates and with some primary earners experiencing lower tax rates and others experiencing higher rates. The presence of children somewhat alters the landscapes for primary and secondary earners, but the main relationships remain unchanged.

The Marriage Subsidy

The nature of the Supreme Court ruling allows me to use the marriage subsidy as a new measure of household income changes in order to identify the income effect, and, by extension, the Hicksian wage elasticity. The marriage subsidy is defined as the difference between the sum of the individuals' tax liabilities if they are single and the couple's joint tax liability if they are married. The marriage subsidy is a common measure of tax incentives to marry or divorce in the family structure literature (Alm and Whittington 1995b, 1999; Ellwood 2000; Eissa and Hoynes 2003; Isaac 2017),

30. These figures come from simulated data to provide a visual sense of how marginal tax rates can change following a switch from individual to joint taxation. They do not come from the sample I use for estimation.

but is new to the literature seeking to estimate income effects of taxation because shifting from the single to joint tax schedule is endogenous in most circumstances. The marriage subsidy exists in the United States tax system because the tax system is not neutral with respect to marriage, and is an inevitable feature of a progressive, household-based tax system. Two unmarried couples who have equal total earnings may end up paying different amounts in income taxes after they marry depending on how those earnings are split between the partners. Generally speaking, two relatively equal earners will end up paying more in taxes upon marriage, while two relatively unequal earners will end up paying less in taxes upon marriage.

I follow a procedure parallel to my calculation of marginal net-of-tax rates, above. I first use predicted earnings and the NBER TAXSIM model to calculate the federal tax liability each same-sex married couple faced after the Supreme Court ruling. I then adjust the tax year to 2012 (the tax year before the policy came into effect), adjust the individual's filing status, dependents, and predicted earnings to reflect the pre-ruling tax environment, and calculate each individual's simulated pre-ruling federal tax liability. I then calculate the marriage subsidy variable as:

$$\Delta\hat{T}_{ij} = \begin{cases} [\hat{T}_i + \hat{T}_j] - [\hat{T}_{ij}], & \text{if year } \geq 2014 \text{ and same-sex married couple} \\ 0, & \text{otherwise} \end{cases}$$

where $\hat{T}_i$ is individual i's predicted tax liability in 2013 (tax year 2012) under the single federal tax schedule and $\hat{T}_{ij}$ is the couple's predicted tax liability in year $t \geq 2014$. A positive value indicates an increase in household income due to a lower joint tax liability after the Supreme Court ruling. Note that the value of the variable $\Delta\hat{T}_{ij}$ will be the same for both partners in a same-sex married couple.

To demonstrate how the marriage subsidy varies with household earnings and the split in earnings within the household, I plot in Figure 4 the simulated marriage subsidy as a percentage of household earnings, which depends on total household earnings and how the earnings are split within the household.[31,32] In general, couples with a single earner will receive a marriage subsidy, but, conditional on household earnings, a more even split in earnings between partners decreases the marriage subsidy and can also turn it into a penalty. For example, a childless couple earning \$74,000 split evenly between the partners would receive a small marriage subsidy of \$98 (0.1% of total household earnings), whereas if all household earnings were earned by a single partner then the couple would receive a marriage subsidy of \$4,885 (6.6% of total household earnings). Figure 4b demonstrates variation in the marriage subsidy among couples with one child. Marriage penalties are more pervasive in families with one child compared to marriage penalties among childless couples. For example, a couple with one child earning \$74,000 split evenly between the partners would face a marriage penalty of \$2,653 (3.4% of total household earnings), whereas if all household earnings were earned by a single partner then the couple would receive a marriage subsidy of \$2,448 (3.3% of total household earnings). 48% of same-sex married couples and 8% of same-sex cohabiting couples have children, which will enter into my calculations of the tax rate and tax liability changes.

Figure 5 displays the distribution of predicted household earnings and predicted split in earnings within the household using the sample of same-sex couples I analyze below in order to provide a sense of where couples in the sample may fall in Figure 4. Household earnings and earnings splits are concentrated most heavily around \$120,000 split relatively evenly between the two partners. These predicted earnings will enter into my calculations of the predicted change in marginal net-of-tax rates and the predicted change in household income due to the marriage subsidy.

Table 1 shows that, on average, same-sex married couples experienced a decrease of \$642.09 in household income due to the marriage subsidy, with substantial variation across households. Table 2 provides illustrative examples of how the variable $\Delta\hat{T}_{ij}$ appears in the data.

31. Figure 4 was inspired by a similar figure published by Amanda Cox for *The New York Times* (Cox 2015). I calculate the marriage subsidy in Figure 4 in 2013 (i.e., married in 2013, single in 2012) assuming that all household income is earned through wages. Other inputs into the TAXSIM model, such as other deductions or credits, will create additional variation in the marriage subsidy in practice.

32. These figures come from simulated data to provide a visual sense of how marginal tax rates can change following a switch from individual to joint taxation. They do not come from the sample I use for estimation.

2.4 Empirical Strategy

I use tax variation among same-sex married couples caused by the Supreme Court ruling in *United States v. Windsor* to estimate the effects of taxation on labor supply between 2012–2015, considering both the extensive and intensive margins. I use a generalized difference-in-differences, or treatment intensity, framework where individuals in same-sex married couples comprise the treatment group and individuals in same-sex cohabiting couples comprise the control group.[33] I quantify the predicted tax changes couples would have faced due to the Supreme Court ruling in order to separately identify income and substitution effects of taxation on labor supply.

2.4.1 Using Tax Change Measures to Separate Income and Substitution Effects

Higher or lower earners may respond to taxation differently due, for example, to differing work preferences or attachment to the labor force. To allow heterogenous responses to taxation by predicted earning status, I estimate the following generalized difference-in-differences equation separately for predicted higher earners and predicted lower earners:

$$Y_{it} = \gamma_0 + \gamma_1 \%\Delta(1-\hat{\tau}_{it}) + \gamma_2 \Delta\hat{T}_{ij} + \gamma_3 SSMC_i + \gamma_4 X_{it} + \delta_t + \mu_s + \varepsilon_{it} \tag{2.1}$$

Where Y_{it} is either annual hours of work or labor force participation, $\%\Delta(1-\hat{\tau}_{it})$ is the individual's predicted percentage change in federal marginal net-of-tax rate from before the *United States v. Windsor* ruling to the observation year, $\Delta\hat{T}_{ij}$ is the couple's predicted change in household income due to the marriage subsidy, $SSMC_i$ is equal to one if the couple is a same-sex married couple, X_{it} is a vector of additional covariates including age, race, education, gender, and the age difference between the predicted higher and lower earners, and δ_t and μ_s are year and state fixed effects, respectively. State fixed effects will capture, among other characteristics, overall state

33. In my main sample, I further restrict the data to couples who married in 2012 or earlier to avoid selection issues related to same-sex couples who chose to delay marriage until after the Supreme Court ruling.

attitudes toward same-sex relationships and local labor market discrimination against LGBT individuals, which may be correlated with the state's decision to recognize same-sex marriages (Gao and Zhang 2016). I also include the partner's predicted earnings to control for the level of predicted earnings that enter into the tax change variables, and bootstrap standard errors to account for the predicted regressors.

The coefficients of interest are γ_1, which determines the total uncompensated (Marshallian) elasticity, and γ_2, which determines the income effect. Identification of γ_1 and γ_2 comes from individual- and couple-level tax rate and liability variation across individuals in same-sex married couples compared to individuals in same-sex cohabiting couples before and after the Supreme Court ruling conditional on other observable covariates. The identifying assumption I must make in this generalized difference-in-differences framework is that annual hours of work and labor force participation of same-sex married and cohabiting couples would have evolved parallel to each other in the absence of the Supreme Court ruling.[34]

Although the parallel trends assumption is not explicitly testable, I use an event study approach to examine whether there is evidence of differential pre-trends between same-sex married and cohabiting couples. I omit the tax change variables and instead include interactions between $SSMC_i$ and each observation year in the data.[35] A statistically significant coefficient on the interaction term in 2012 would indicate pre-trends between same-sex married and cohabiting couples that would violate the parallel trends assumption. Figure 6 presents the event study coefficient estimates and confidence intervals, showing that there is no significant evidence at conventional levels of pre-trends between same-sex married and cohabiting couples.[36] Also, as mentioned earlier, Figure 2 shows that the mean demographic variables exhibit similar trends among the treatment and control

34. Figure 2 shows that most demographic variables for the treatment and control groups do not exhibit visual differences in mean trends over time, lending some support in favor of the parallel trends assumption. The exception is the presence of children, which exhibits a divergent trend from 2012–2013, but appears parallel between the two groups thereafter.

35. The interaction $SSMC_i \times 2013$ is omitted as the base year.

36. The coefficient estimates in Figure 6 are also traditional difference-in-differences coefficient estimates. The absence of significant effects of the *United States v. Windsor* ruling using this specification indicates that the overall changes in labor force participation and annual hours of work were small among same-sex married couples relative to same-sex cohabiting couples. Full results from this difference-in-differences specification are available upon request.

groups, providing some additional evidence in favor of the parallel trends assumption.

The empirical specification in Equation 2.1 is consistent with the collective model of labor supply studied by Chiappori (1992) and Chiappori, Fortin, and Lacroix (2002). In Appendix A I describe the collective model, its connection to my empirical strategy, and present parameter and elasticity estimates implied by the collective model. The elasticity estimates I present with my main findings below are very similar to the elasticity estimates implied by the collective model in Appendix A.

2.4.2 Estimating Structural Elasticities

I use the estimated coefficients from Equation 2.1 to estimate uncompensated (Marshallian) elasticities, income elasticities, and compensated (Hicksian) elasticities along the intensive and extensive margins. The Marshallian elasticity is a combination of a substitution effect (holding utility constant) and an income effect (holding net-of-tax rates constant). Following Gruber and Saez (2002) and Keane (2011), the total effect of a wage rate change is given by the Slutsky equation below, which I have multiplied by $\frac{w}{H}$ to convert into elasticities:

$$\frac{\partial H}{\partial w}\frac{w}{H} = \left.\frac{\partial H}{\partial w}\frac{w}{H}\right|_{u} + \left[\frac{\partial H}{\partial Y}\frac{Y}{H}\right]\frac{wH}{Y} \tag{2.2}$$

Where H is hours worked, w is the hourly wage rate, and Y is total income. The left-hand side term in Equation 2.2 is the Marshallian elasticity. The first term on the right-hand side is the Hicksian elasticity and the second term on the right-hand side is the income effect.

To estimate these elasticities, I interpret γ_1 as the effect of a percentage change in the net-of-tax hourly wage and γ_2 as the effect of a change in household non-labor income. Given these interpretations, I estimate the Marshallian, income, and Hicksian elasticities along the intensive margin as:

$$\text{Marshallian: } \varepsilon_M = \frac{\partial H}{\partial w}\frac{w}{H} = \frac{\partial H}{\partial ln(w)}\frac{1}{H} = \hat{\gamma}_1\frac{1}{\bar{H}} \tag{2.3a}$$

$$\text{Income: } \varepsilon_I = \frac{\partial H}{\partial Y}\frac{Y}{H} = \hat{\gamma}_2\frac{\bar{Y}}{\bar{H}} \tag{2.3b}$$

$$\text{Hicksian: } \varepsilon_H = \frac{\partial H}{\partial w}\frac{w}{H} - \frac{\partial H}{\partial Y}w = \hat{\gamma}_1\frac{1}{\bar{H}} - \hat{\gamma}_2\bar{w} \tag{2.3c}$$

where $\bar{H}$ is the average hours worked in the sample, $\bar{Y}$ is the average household non-labor income in the sample, and $\bar{w}$ is the average hourly wage rate in the sample.

Economic theory predicts that the Hicksian elasticity is positive and, if leisure is a normal good, that the income effect is negative. However, the sum of the two effects along the intensive may be either positive or negative depending on which effect is larger. Therefore, I expect $\hat{\gamma}_2 < 0$ and $\hat{\gamma}_1\frac{1}{\bar{H}} - \hat{\gamma}_2\bar{w} > 0$, but cannot predict the sign of $\hat{\gamma}_1$ along the intensive margin a priori.

It is also possible to calculate the Hicksian and income elasticities along the extensive margin. In this case, the Hicksian and Marshallian elasticity concepts are the same because the income effect term in Equation 2.2 is zero. The Hicksian and income elasticities along the extensive margin are:

$$\text{Hicksian: } \eta_H = \frac{\partial LFP}{\partial w}\frac{w}{LFP} = \frac{\partial LFP}{\partial ln(w)}\frac{1}{LFP} = \hat{\gamma}_1\frac{1}{\overline{LFP}} \tag{2.4a}$$

$$\text{Income: } \eta_I = \frac{\partial LFP}{\partial Y}\frac{Y}{LFP} = \hat{\gamma}_2\frac{\bar{Y}}{\overline{LFP}} \tag{2.4b}$$

where $\overline{LFP}$ is the average labor force participation rate in the sample. Along the extensive margin, economic theory predicts that $\hat{\gamma}_1 > 0$ and $\hat{\gamma}_2 < 0$.

2.4.3 Results

Tables 3–4 present extensive and intensive margin generalized difference-in-differences coefficient estimates, respectively, from Equation 2.1, in which I use the predicted percentage change in the marginal net-of-tax rate and the predicted change in household income due to the marriage subsidy in order to separate the income and substitution effects. I estimate small and statistically insignificant effects of taxation on the labor force participation of predicted primary earners. In contrast, I estimate that a 10% increase in the marginal net-of-tax rate increases labor force participation by 5.9 percentage points (7.1%) and a $1,000 increase in household income decreases labor force participation by 1.0 percentage points (1.2%) among predicted secondary earners in same-sex married couples relative to same-sex cohabiting couples. These coefficients imply a significant Hicksian participation elasticity of 0.679 and a significant income participation elasticity of -0.029 among predicted secondary earners. I also estimate the model separately for male and female same-sex couples, and find that a 10% increase in the marginal net-of-tax rate increases labor force participation by 10.9 and 3.5 percentage points among men and women, respectively, and that a $1,000 increase in household income decreases labor force participation by 0.8 and 1.4 percentage points among men and women, respectively. These coefficients imply significant Hicksian participation elasticities of 1.236 and 0.400 for men and women, respectively, and significant income participation elasticities of -0.022 and -0.044 for men and women, respectively.

Along the intensive margin, I estimate significant Hicksian and income hours elasticities only among predicted primary earners. I calculate that the income and substitution effects largely offset each other, resulting in a small (and sometimes significant) Marshallian hours elasticity. The elasticity estimates are very similar when using the sample of both men and women, the sample of men, or the sample of women. I estimate Hicksian hours elasticities between 0.360 and 0.382 and income hours elasticities between -0.010 and -0.013 among predicted primary earners. I estimate opposite-signed, but statistically insignificant elasticities among predicted secondary earners, which is largely driven by female predicted secondary earners.

These estimates are in-line with many other estimates in the literature, as reported by Keane

(2011) in his review of the literature on labor supply responses to taxation. Keane (2011) reports male Hicksian hours elasticities between 0.02 and 1.32 across 22 studies, with an average of 0.31, and female Hicksian participation elasticities between 0.01 to 1.60 across 4 studies. Chetty (2012), in his examination of micro and macro labor supply elasticities, places bounds of [0.28,0.54] on intensive margin Hicksian elasticities while accounting for optimization frictions. Chetty (2012) also notes that larger percentage changes in the marginal net-of-tax rates can lead to larger elasticity estimates, as may be the case here because many same-sex married couples faced dramatic changes in their tax rates and liabilities as a result of the Supreme Court ruling.[37] Finally, at the mean values of the tax change variables, I estimate that the introduction of joint taxation decreased labor force participation of predicted secondary earners by 2.335 percentage points, increased labor force participation of predicted primary earners by 0.141 percentage points, and increased hours worked by predicted primary earners by 13.574 hours per year. All of these overall effects are consistent with and similar in magnitude to LaLumia's (2008) estimates of the effects of joint taxation on labor supply.

2.5 Robustness Checks and Alternative Specifications

I examine the robustness of my estimates to four potential confounding factors: misclassification of opposite-sex married couples in the data, compliance barriers to filing jointly following *United States v. Windsor*, sample selection bias within the treatment group, and state-year specific shocks that may affect labor supply. I also estimate a specification using an alternative control group of opposite-sex married couples to address concerns about the suitability of same-sex cohabiting couples as a control group for same-sex married couples. Finally, I explore several alternative specifications that allow for additional effects or controls. In all cases I find quantitatively and qualitatively similar results, indicating that my main findings are robust to all of these potential concerns.

37. The author's argument is that labor supply frictions are less likely to attenuate short-run responses when the marginal net-of-tax rate changes are very large because of the larger utility cost of not adjusting labor supply.

There are two main concerns in the literature about same-sex couples that need to be explored. First, Black et al. (2007) and Gates and Steinberger (2010) document substantial measurement error of same-sex couples in the 2000 decennial census and the 2005–2007 American Community Survey, respectively, mostly due to opposite-sex couples mis-marking one of the partner's genders so that the couple appears to be a same-sex couple in the data. The studies' main concern is in estimating the size of the same-sex couple population, where even a small amount of measurement error greatly influences the population estimate since the same-sex couple population is very small relative to the opposite-sex couple population. Measurement error in identifying same-sex couples can introduce bias into the coefficient estimates.[38] Note, however, that the 2012–2015 waves of the American Community Survey are the first Census Bureau surveys to explicitly identify same-sex couples in the data, whereas prior surveys employed editing practices that made it substantially more difficult to identify same-sex couples on top of the mis-measurement issues.[39]

To further address the measurement error concern, I estimate the model by restricting the sample to couples who responded using a computer assisted telephone or personal interview process (CATI/CAPI) rather than the traditional mail-in form. Gates and Steinberger (2010), Kreider and Lofquist (2015), and Lofquist (2015) all document much lower gender mis-marking among CATI/CAPI respondents compared to mail-in respondents due to an automatic response check. Tables 5–6, columns 1–2, present the generalized difference-in-differences coefficient estimates using the sample of CATI/CAPI respondents. None of the coefficient estimates are statistically different from the main results, and the qualitative conclusions of the main results in Tables 3–4 remain the same, indicating little effect of mis-measurement in the data.

The second concern from the same-sex couple literature is that there may not have been 100% take-up of joint filing among same-sex married couples after the *United States v. Windsor* ruling due to compliance barriers in some states (Fisher, Gee, and Looney 2016). Ten states in 2014

38. This type of measurement error likely biases my estimates toward zero.

39. The 2013–2015 waves of the American Community Survey contain a new variable, *SSMC*, explicitly identifying same-sex couples. The 2012 wave contains a data quality flag that indicates whether a couple's status was changed from "same-sex married couple" to "same-sex cohabiting couple," allowing me to identify same-sex married couples in this year as well. Previous waves of the American Community Survey do not contain the *SSMC* variable or the marital status data quality flag.

required married couples to file jointly on their state tax returns if they filed jointly on their federal tax returns, while simultaneously not allowing same-sex married couples to file jointly on state tax returns.[40] Such inconsistent tax policies surrounding same-sex married couples create compliance barriers to filing jointly in some states. These barriers may bias my coefficient estimates toward zero if couples do not respond to their tax changes due to compliance barriers.

To address the concern about compliance barriers, I estimate the model by excluding from the sample same-sex couples residing in states with inconsistent tax policies. Tables 5–6, columns 3–4, present the generalized difference-in-differences coefficient estimates using the restricted sample of same-sex couples. The coefficient estimates using this sample restriction remain stable, and the qualitative conclusions of the main results in Tables 3–4 remain the same, indicating little bias in the estimates due to inconsistent tax policies.

Sample selection with regards to married couples is also a notable concern. My main sample includes same-sex cohabiting couples and same-sex married couples who married in 2012 or earlier, but restricting married couples in this way means that married couples observed in 2015 must have been married for at least three years, whereas married couples observed in 2012 could be newly married. To address this concern, I estimate the model on a sample of same-sex couples that restricts married couples to those who have been married for at least three years in each survey year.[41] Furthermore, missing variables in the 2012 wave of the American Community Survey allow me to use only observations from 2013 or later in this sample.[42] Tables 5–6, columns 5–6, present the generalized difference-in-differences coefficient estimates using this restricted sample of same-sex couples. The coefficient estimates using this sample restriction remain stable, and the qualitative conclusions of the main results in Tables 3–4 remain the same, indicating little bias in

40. These states are Alabama, Georgia, Kansas, Kentucky, Louisiana, Michigan, Missouri, Nebraska, North Dakota, and Ohio.

41. There is no additional restriction on same-sex cohabiting couples. I cannot observe when these couples began their relationship or cohabitation, and so cannot further restrict this group in the sample.

42. I rely on a marriage data quality flag in 2012 to identify same-sex married couples, which indicates whether the couple's marital status was changed from "same-sex married couple" to "same-sex cohabiting couple." The variable for "year of marriage" exists only for married couples, and since the marital status of same-sex married couples was edited in 2012, the "year of marriage" variable is not defined for same-sex married couples in this wave of the American Community Survey. The absence of the year of marriage does not allow me to determine how long the couple has been married, and so I exclude 2012 observations of all couples in this specification.

the estimates due to sample selection related to the length of marriages among married couples.

I also estimate a specification of Equation 2.1 using an alternative control group of opposite-sex married couples. It is not immediately clear whether a control group of same-sex cohabiting couples is the best choice rather than a control group of opposite-sex married couples. Same-sex cohabiting couples have similar pre-ruling tax environments, which may make them the more suitable control group. Opposite-sex married couples, however, share a similar legal relationship status and a similar post-ruling tax environment. I am also unable to observe the duration of a cohabiting couple's relationship, limiting my ability to restrict the sample in the same way as married couples.[43] I re-estimate my model using opposite-sex married couples as an alternative control group. Tables 5–6, columns 7–8, present the generalized difference-in-differences coefficient estimates using the opposite-sex married couples control group. I estimate coefficients that are smaller in magnitude, with a significant Hicksian participation elasticity of 0.460 for predicted lower earners and a significant Hicksian hours elasticity of 0.241 for predicted higher earners. These results remain qualitatively similar to my main results, providing evidence that my main findings are not driven by the choice of control group.

In addition, it is possible to include state-by-year fixed effects because the *United States v. Windsor* ruling affected federal tax policy for same-sex married couples and treatment was not tied to state of residence. This specification provides greater control for idiosyncratic shocks that affect labor supply and vary at the state-year level, such as variation in state tax policy or parameters. Tables 5–6, columns 9–10, present the generalized difference-in-differences coefficient estimates using state-by-year fixed effects. The coefficient estimates are essentially unchanged, suggesting that my main findings are robust even when controlling for state-year level shocks.

Finally, I examine several alternative specifications of Equation 2.1 that allow for additional effects or controls. Table 7 presents the elasticity estimates from specifications that include the

43. Table A1 presents summary statistics of the same-sex and opposite-sex married couples in this specification. Summary statistics may vary between this sample and my main sample of same-sex couples (even among the same-sex married couples, who are in both samples) because I re-do my earnings prediction process using the full sample of all married couples. This leads to some re-sorting of predicted primary and secondary earners, although the means change very little.

spouse's percentage change in marginal net-of-tax rates, an indicator for the presence of children, or a group-specific linear time trend. In all cases the labor supply elasticities remain quantitatively similar to the main estimates in Tables 3–4. The cross-wage Hicksian elasticities are negative, indicating that spousal labor supply is substitutable to some degree, which is consistent with findings in the literature (Chiappori, Fortin, and Lacroix 2002; Blau and Kahn 2007; Heim 2009; Gayle and Shephard 2016). Controlling for the presence of children addresses concerns about the comparability of the treatment and control groups along this dimension, but does not significantly change the estimates. Including a group-specific linear time trend relaxes the parallel trends assumption needed for identification in my main analysis. The results in this specification are unchanged, suggesting that the parallel trends assumption is not restrictive in my main findings and that the results hold even under a weaker assumption.

2.6 Deadweight Loss and Tax Revenue Implications

In this section, I apply my Hicksian elasticities and the relative tax changes created by *United States v. Windsor* to an existing sufficient statistic framework explored in more detail by Feldstein (1999), Immervoll et al. (2007), and Chetty (2009). My calculations below provide further insight into the welfare and tax revenue consequences implied by my Hicksian elasticity estimates.[44]

2.6.1 Deriving the Sufficient Statistic Formula

The total change in deadweight loss due to a small tax change is $dDWL = dW - dTR$, where dW is the change in consumer welfare due to the tax change and dTR is the change in tax revenue. Figure 7 displays how deadweight loss and tax revenue change following a small tax increase. A higher tax rate increases tax revenue through a mechanical revenue effect, dM, but decreases tax revenue through behavioral responses, dB, so that the total change in tax revenue is $dTR = dM + dB$.

44. Note that Feldstein (1999) and Immervoll et al. (2007) use a similar framework as below and calculate the additional deadweight loss at various Hicksian elasticity values taken from the literature. My calculations below amount to a similar calibration exercise using my estimated Hicksian elasticities.

Appealing to the envelope theorem, the change in consumer welfare due to taxation, dW, is equal to the mechanical revenue effect, dM, as displayed in Figure 7.[45] Substituting these expressions into $dDWL$, we obtain $dDWL = dM - (dM + dB) = -dB$, which describes how deadweight loss changes following a change in tax rates. In order to calculate dB, I consider Chetty's (2009) simple model to derive the sufficient statistic formula.

Each partner ($i = 1, 2$) in a married couple chooses hours of work, h^i, to maximize her utility, $u^i = C^i - \psi(h^i)$, which is a function of consumption, C^i, and a disutility of work, $\psi(h^i)$, subject to a constant marginal tax rate, τ, an hourly wage, w_i, and household non-labor income, y.[46] The price of consumption is normalized to 1. The household budget constraint is, therefore, $(1-\tau)[w_1h^1 + w_2h^2] + y + T \geq C^1 + C^2$. Finally, assume that the government redistributes the tax revenue through a lump sum transfer, $T = \tau(w_1h^1 + w_2h^2)$, to each household.

This utility function ignores income effects in order to make the derivation tractable, and is also money metric so that social welfare, below, is measured in dollars. This simplification is common in the sufficient statistic literature studying income taxation, and means that the Marshallian and Hicksian elasticity concepts are the same along the intensive margin. With income effects, these formulas would apply "with the Hicksian elasticity in place of the Marshallian elasticity" so that the excess burden is defined using compensating variation (Chetty 2009). This utility function also means that my calculations below amount to a calibration exercise of existing sufficient statistic formulas using my estimated Hicksian elasticities.

I consider a utilitarian welfare function for a representative household, in which consumer welfare is equal to the sum of the partners' utilities subject to a household budget constraint. Total welfare, $\mathbf{W}$, is the sum of consumer welfare and the tax revenue raised by the government. Substituting in the budget constraint to each individual's utility function, the social welfare function

45. Chetty (2009) explains the intuition behind this. Appealing to the envelope theorem, an infinitesimal change in marginal tax rates will have no first-order effects on welfare because the individual is already optimizing, meaning that if behavioral responses did have first-order effects on consumer welfare when taxes change then the individuals would not have been optimizing. The change in deadweight loss, therefore, will be due to the behavioral responses operating through the budget constraint.

46. This formulation of individual utility is akin to a collective model of labor supply without caring (Chiappori, Fortin, and Lacroix 2002). I plan to explore the collective model's welfare and tax revenue implications in the context of the *United States v. Windsor* ruling in future research.

becomes:

$$\mathbf{W}(\tau) = \left\{(1-\tau)(w_1h^1+w_2h^2)+y+T-\psi(h^1)-\psi(h^2)\right\}+\underbrace{\tau(w_1h^1+w_2h^2)}_{T=\text{Tax Revenue}} \tag{2.5}$$

The term in braces above is consumer welfare. Taking the derivative of $\mathbf{W}(\tau)$ with respect to τ, we can appeal to the envelope theorem, meaning that behavioral responses cannot have first-order effects on consumer welfare, which corresponds to $dW = dM$. Thus, the change in the welfare cost of taxation is equal to:

$$\begin{aligned} dW = dM &= \frac{d[(1-\tau)(w_1h^1+w_2h^2)+y+T-\psi(h^1)-\psi(h^2)]}{d\tau} \\ &= -(w_1h^1+w_2h^2)d\tau \end{aligned} \tag{2.6}$$

The change in tax revenue is due to the mechanical change in tax revenue, dM, and the behavioral change, dB:

$$dTR = \frac{d[\tau(w_1h^1+w_2h^2)]}{d\tau} = \underbrace{(w_1h^1+w_2h^2)d\tau}_{dM}+\underbrace{\tau\left(w_1\frac{dh^1}{d\tau}+w_2\frac{dh^2}{d\tau}\right)d\tau}_{dB} \tag{2.7}$$

From equations 2.6 and 2.7, we obtain:

$$\begin{aligned} dDWL &= -d\mathbf{W}(\tau) = -(dW+dTR) \\ &= \left[w_1h^1+w_2h^2-w_1h^1-w_2h^2-\tau\left(w_1\frac{dh^1}{d\tau}+w_2\frac{dh^2}{d\tau}\right)\right]d\tau \\ &= \frac{\tau}{1-\tau}\left(w_1h^1\varepsilon_H^1+w_2h^2\varepsilon_H^2\right)d\tau \end{aligned} \tag{2.8}$$

Where $\varepsilon_H^i = \frac{dh^i}{d(1-\tau)}\frac{1-\tau}{h^i}$ is the compensated Hicksian hours elasticity for individual i. Equation 2.8 shows that the change in deadweight loss along the intensive margin due to a tax change can be expressed as a function of the marginal tax rate, earnings, and the compensated Hicksian hours elasticity. A symmetric equation holds for extensive margin responses with the extensive compensated elasticity, η_H^i, substituted for ε_H^i, which is similar to the formulation derived by Immervoll et al. (2007). In this case, the change in deadweight loss is due not to changing hours but instead to changing labor force participation.

2.6.2 Empirical Implementation and Findings

I use the sample means of τ and $w_i h^i$ along with my compensated Hicksian elasticity estimates to compute $dDWL$ in Equation 2.8. These calculations represent measures of the marginal deadweight loss along the intensive and extensive margins due to a small change in tax rates. However, primary and secondary earners faced different directions and magnitudes of tax changes, which can create conflicting changes in deadweight loss along the intensive and extensive margins. I weight both margins' $dDWL$ measures by the relative direction and size of the marginal tax rate changes experienced by same-sex married couples following *United States v. Windsor* in order to obtain a measure of the total change in deadweight loss due to the introduction of joint taxation. This weighting process produces a measure of the additional deadweight loss due to joint taxation relative to individual taxation.[47] Finally, I calculate the additional deadweight loss due to joint taxation for households with different earnings splits.

I plot my calculations in Figure 8, where a positive value indicates more deadweight loss under joint taxation relative to individual taxation and a negative value indicates less deadweight loss.[48] The figure shows that, in general, joint taxation creates more deadweight loss than individual taxation, and that this additional deadweight loss generally decreases as the couple moves closer to a single-earner household. Joint taxation creates, on average, $9,343 in additional deadweight loss per family relative to individual taxation among households with relatively even earnings splits. The additional deadweight loss created by joint taxation decreases as households' earnings splits

47. The weights represent the marginal tax rate changes faced by same-sex married couples due to *United States v. Windsor*, and are therefore relative to the system of individual taxation these couples previously faced. The interpretation of the net additional deadweight loss measure using the weights is, therefore, the additional deadweight loss created by joint taxation relative to individual taxation due to both intensive and extensive margin effects.

48. The earnings splits displayed in Figure 8 are binned in intervals of 10. In other words, the 50-50 split bin contains households where the earnings split is 50-50 up to 59-41, the 60-40 split bin contains households where the earnings split is 60-40 up to 69-31, etc.

become more uneven. Joint taxation reduces deadweight loss, on average, by $2,852 per family relative to individual taxation among households with a single earner because primary earners in these households generally experienced decreases in marginal tax rates.[49]

I use the household weights in the American Community Survey to obtain population estimates of all married couples by earnings split in order to compute the total additional deadweight loss due to joint taxation relative to individual taxation. Figure 9 displays the distribution of earnings splits among same-sex and opposite-sex married couples in my samples. The overall shape of earnings splits is similar between same-sex and opposite-sex married couples, with the exception that there are more single-earner households among opposite-sex married couples, which may reflect the labor supply incentives for married couples due to joint taxation. These distributions understate the absolute number of households in each earnings split category because I use only observations of single- or dual-earner families who responded to the American Community Survey using computer assisted telephone or personal interviews in order to avoid the possibility of misclassification between same-sex and opposite-sex married couples. Table 8 displays the data points in Figures 8 and 9 as well as the calculations of total additional deadweight loss created by joint taxation relative to individual taxation. I find that *United States v. Windsor* created $398 million in additional deadweight loss for same-sex married couples, while joint taxation in general creates $79.0 billion in additional deadweight loss for all married couples relative to individual taxation.[50] Relative to the government's receipts of $1.5 billion in individual income taxes from same-sex married couples and $984 billion from all married couples in fiscal year 2014, this represents approximately 26.2% of additional tax revenue lost due to inefficiencies for same-sex married couples and 8.0% lost for all married couples relative to individual taxation (U.S. Department of the Treasury, Internal

49. Stevenson (2012) provides evidence that intensive margin elasticities are similar between married men in opposite-sex couples and primary earners in same-sex cohabiting couples. He also finds that married women in opposite-sex couples appear to have larger intensive and extensive margin elasticities than secondary earners in same-sex cohabiting couples, although the comparison is to both male and female secondary earners in same-sex cohabiting couples.

50. $79.0 billion of total additional deadweight loss created by joint taxation relative to individual taxation is the sum of the additional deadweight loss created by same-sex and opposite-sex married couples.

Revenue Service 2016).[51,52] The large decreases in efficiency concentrated among relatively equal earning couples may be due to the sizable participation response among predicted secondary earners. A secondary earner leaving the labor market when she earns about half of her family's income is a meaningful loss of labor market productivity, leading to large inefficiencies. Note, however, that this model does not incorporate welfare improvements through increased home production, and only considers efficiencies in the labor market.

It is also possible to compute the change in tax revenue, dTR, as a result of joint taxation. The sufficient statistic approach shows that $dTR = (w_1 h^1 + w_2 h^2)d\tau + \tau\left(w_1 \frac{dh^1}{d\tau} d\tau + w_2 \frac{dh^2}{d\tau} d\tau\right)$. I plot the average dTR by household earnings split in Figure 10, which takes into account mechanical and behavioral effects of joint taxation using sample averages of same-sex married couples. A positive value in Figure 10 represents an increase in tax revenue, whereas a negative value indicates a decrease in tax revenue. The figure shows that, in general, joint taxation reduces tax revenue relative to individual taxation, and that the average loss in tax revenue generally decreases as the couple moves closer to a single-earner household. Joint taxation reduces tax revenue by \$6,271 per family, on average, relative to individual taxation among households with relatively even earnings splits. The amount of reduced tax revenue created by joint taxation decreases as households' earnings splits become more uneven. Joint taxation increases tax revenue, on average, by \$1,662 per family relative to individual taxation among dual-earner households with a very uneven earnings split, but reduces tax revenue among single-earner households due to lower marginal tax rates among primary earners. Table 9 displays the data points in Figure 10 and the population estimates from Figure 9 as well as the calculations of the reduction in tax revenue under joint taxation relative to individual taxation. I find that federal tax revenue was \$294 million less for same-sex married couples as a result of *United States v. Windsor* relative to individual taxation, which is consistent with

51. The \$1.5 billion in individual income tax revenue from same-sex married couples represents my calculations using tax liabilities calculated by NBER's TAXSIM model and the population estimates of same-sex married couples in the American Community Survey.

52. This calculation does not tell us the total efficiency cost of taxation, only the additional cost of joint taxation relative to individual taxation. Feldstein (1999) finds that the personal income tax creates deadweight loss equal to 32.2% of tax revenue. Immervoll et al. (2007) find efficiency losses in European countries ranging from 19% in countries with low taxes and benefits to 82% in Denmark, which has a high level of benefits.

simulations by Alm, Leguizamon, and Leguizamon (2014). Considering all married couples, I find that federal tax revenue is $67.9 billion lower under joint taxation relative to individual taxation. Relative to the government's receipts of $1.5 billion in individual income taxes from same-sex married couples and $984 billion from all married couples in fiscal year 2014, this amounts to approximately 19.4% and 6.9%, respectively, of total government receipts from the personal income tax in fiscal year 2015. Again, the large decreases in tax revenues concentrated among relatively equal earning couples may be due to the sizable participation response among predicted secondary earners. A secondary earner leaving the labor market when she earns about half of her family's income is a meaningful loss of taxable income, leading to reductions in tax revenue.

Finally, the change in tax revenue above is due only to behavioral responses to changes in tax rates, and does not include any measure of the marriage penalty associated with the *United States v. Windsor* ruling. To address this, I compute the overall change in federal tax revenues due only to the marriage penalty each couple experienced along with the household sampling weights in the American Community Survey. I calculate that federal tax revenue declined by $30.8 million due to the marriage penalty following the Supreme Court ruling, which amounts to approximately 10.5% of the change in tax revenue due to behavioral responses calculated above.[53] There is substantial variation in the marriage penalty across households, as well. In Figure 11 I plot the household marriage penalty by a continuous measure of the household earnings split using observations of same-sex married couples in the 2014–2015 waves of the American Community Survey. As expected, households with more even earnings split are more likely to face a penalty, while households with a more unequal earnings split are more likely to face subsidies. Note that this "instantaneous" measure of the marriage penalty assumes no labor supply responses to taxation.

My estimates of the change in federal tax revenue due to the marriage penalty are consistent with other calculations of the federal tax revenue consequences of same-sex marriage legalization, such as those by Stevenson (2012) and Alm, Leguizamon, and Leguizamon (2014). When as-

53. As in my previous calculations of the change in deadweight loss and the change in tax revenue due to behavioral responses, I only use observations of same-sex married couples who responded to the American Community Survey using a computer-assisted telephone or personal interview. This approach minimizes the probability that I include opposite-sex married couples who are mis-classified as same-sex married couples in the data.

suming a 50% marriage rate among same-sex couples, Stevenson (2012) simulates that same-sex marriage legalization would lead to a $38 million decrease in federal tax revenue using a measure of the marriage penalty that is similar to mine. Alm, Leguizamon, and Leguizamon (2014), also when assuming a 50% marriage rate, simulate a $95–237 million decrease in federal tax revenue. Changes in tax revenue due to the marriage penalty shrink as the assumed marriage rate decreases, and only 33% of same-sex couples in my sample are married.[54] Stevenson (2012) also introduces a measure of the "endogenous" marriage penalty that incorporates labor supply responses to taxation. I believe the endogenous marriage penalty offers useful insight into the interaction between the marriage penalty and labor supply responses to taxation when calculating changes in tax revenues, and I plan to explore this measure further.

2.7 Conclusion

The June 2013 Supreme Court decision in *United States v. Windsor* has been heralded as a landmark civil rights case. While many commentators have focused on the legal and social effects of defining who can marry, the Supreme Court ruling also had immediate consequences for the tax environment faced by same-sex couples who were already married by the time of the ruling. I leverage improved data quality of same-sex married couples and new tax variation among this understudied population to examine classic economic questions concerning the labor supply effects, efficiency costs, and tax revenue consequences of joint taxation.

I use the 2012–2015 waves of the American Community Survey, which are the first of the Census Bureau surveys to explicitly identify same-sex married couples. I also predict individual earnings to construct a sample of predicted higher earners and a sample of predicted lower earners. My main specifications use a generalized difference-in-differences framework in order to compare

54. It is also worth noting that I calculate the change in federal tax revenue due to the marriage penalty among same-sex couples who had married before *United States v. Windsor*, and, therefore, my calculations represent the immediate impact on tax revenue from the treated population. Stevenson (2012) and Alm, Leguizamon, and Leguizamon (2014), however, consider the question, "how would federal tax revenue change if all states were to simultaneously legalize same-sex marriage, conditional on an assumed marriage rate?"

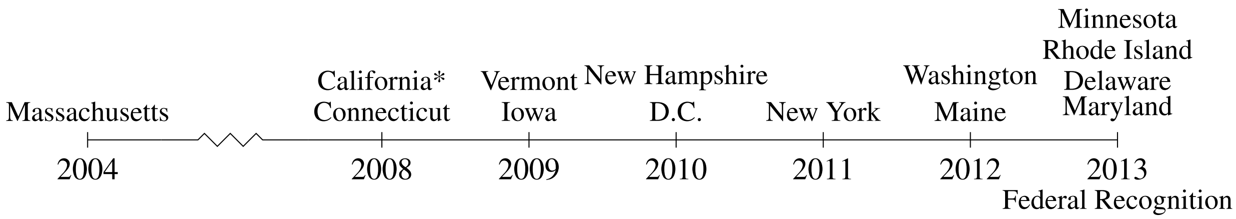
a treatment group of same-sex married couples to a control group of same-sex cohabiting couples. I quantify two sources of variation created by the Supreme Court ruling. The first is the shift in federal marginal tax rates due to the sudden movement from the single tax schedule to the joint filing tax schedule, and the second is the change in household after-tax income due to the marriage subsidy. Variation in tax rates originates through two primary channels: differences in tax bracket definitions between the single and joint schedules, and the addition of both partners' earnings in taxable income rather than only the individual's earnings. Separate variation in household after-tax income exists because the United States tax system is not equal across marital status. Two couples' total federal tax liability may change differently as a result of the ruling depending on total household earnings and the split in earnings between partners. These sources of variation allow me to separate the income and substitution effects of taxation.

I estimate that a 10% increase in the marginal net-of-tax rate increases labor force participation by 5.9 percentage points and a $1,000 increase in household income decreases labor force participation by 1.0 percentage points among predicted secondary earners in same-sex married couples relative to same-sex cohabiting couples. These coefficients imply a significant compensated (Hicksian) participation elasticity of 0.679 and a significant income participation elasticity of –0.029 among predicted secondary earners. I find small and statistically insignificant effects of taxation on the labor force participation of predicted primary earners. Along the intensive margin, I estimate significant Hicksian hours elasticities between 0.360 and 0.382 and significant income hours elasticities between -0.010 and -0.013 among predicted primary earners. The substitution and income effects largely off-set each other, resulting in small and sometimes significant uncompensated (Marshallian) hours elasticities. Overall, I conclude that predicted primary earners in married couples respond to joint taxation along the intensive margin, whereas predicted secondary earners respond along the extensive margin. These results are robust to a number of potential confounding factors and alternative specifications.

Finally, I use my Hicksian elasticity estimates in a sufficient statistic framework to calculate the additional deadweight loss and tax revenue created by joint taxation relative to individual taxation, which is equal across marital status rather than family income. I find that *United States v. Windsor* induced \$398 million in additional deadweight loss and cost \$294 million in tax revenue from same-sex married couples relative to individual taxation, which amounts to approximately 26.2% and 19.4% of total federal government income tax receipts from same-sex married couples, respectively. If my elasticity estimates apply to all married couples in the United States, I find that joint taxation creates \$79.0 billion in additional deadweight loss and reduces tax revenue by \$67.9 billion relative to individual taxation, which amounts to approximately 8.0% and 6.9% of total federal government income tax receipts from all married couples, respectively. My analysis provides direct evidence of the additional efficiency costs and reduced tax revenue of joint taxation relative to individual taxation, and suggests that we may improve efficiency of the United States tax system by lowering tax rates among secondary earners so as to mitigate the efficiency costs along the extensive margin.

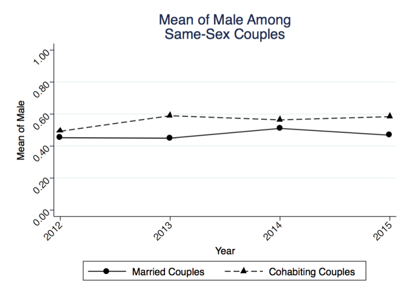
Future work in this field should continue to explore direct effects of varied tax systems. Labor supply responses to different tax structures can then provide insight into the efficiency and tax revenue advantages or disadvantages in a straightforward way. In addition, improved data quality combined with important legal victories for same-sex couples in the United States offer the possibility of further understanding not only the particular economic changes and challenges faced by same-sex couples, but also an opportunity to learn more about married couples' responses to legislation more generally. Despite my conclusion that joint taxation is less efficient and generates less revenue than individual taxation, whether increased efficiency is worth the lower associated tax equity remains an open question.

Figure 1: Timeline of Same-Sex Marriage Legalization

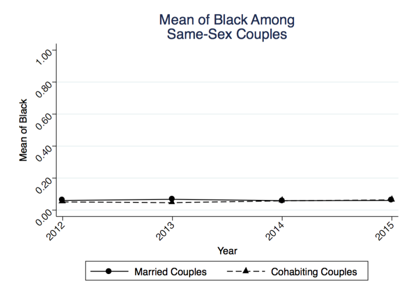


Notes: California legalized same-sex marriage in 2008, but the statute was suspended by Proposition 8 until the Supreme Court decision in 2013. Same-sex marriage licenses issued in California before Proposition 8 continued to be honored as legal. States listed in 2013 are only those states that had either enacted or voted to enact same-sex marriage recognition before the Supreme Court ruling in June.

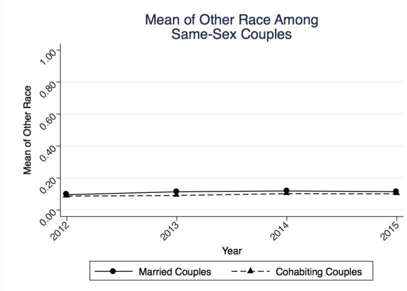
Figure 2: Mean Demographic Variables Among Same-Sex Couples Over Time



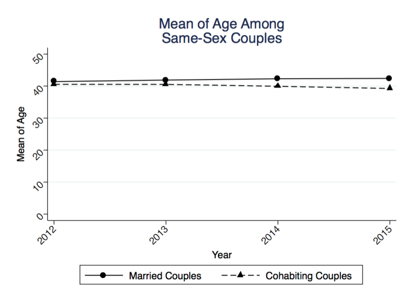
(a) Male



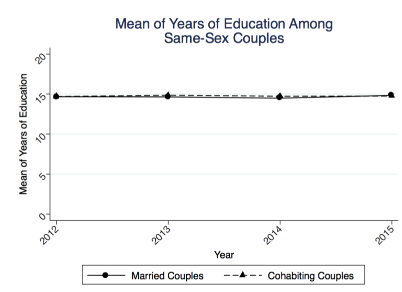
(b) Black



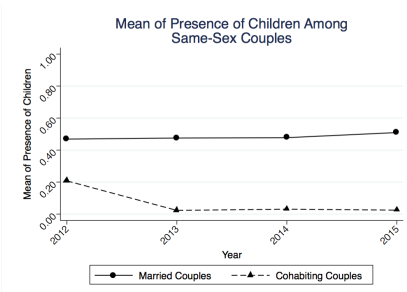
(c) Other race



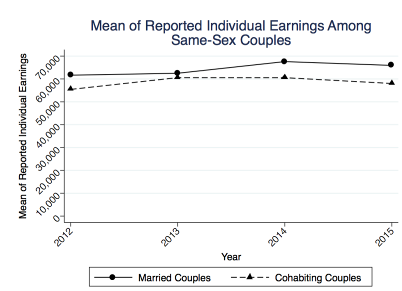
(d) Age



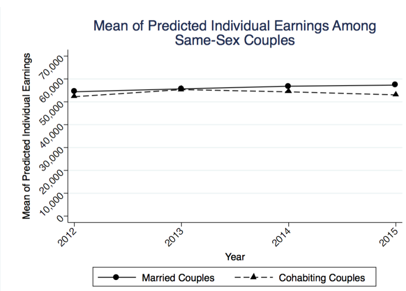
(e) Years of education



(f) Presence of children



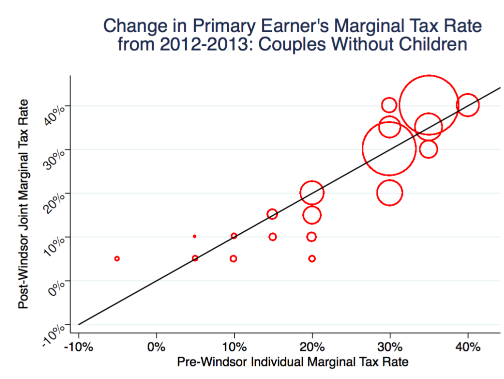
(g) Reported earnings



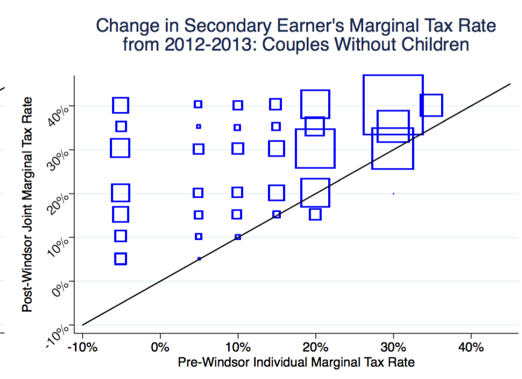
(h) Predicted earnings

Notes: The data come from the 2012–2015 waves of the American Community Survey. Each data point presents the mean value of the demographic variable in that year. The samples includes same-sex cohabiting couples and same-sex married couples who married in 2012 or earlier, who do not live with any other couple, and who have at least one earner in the household.

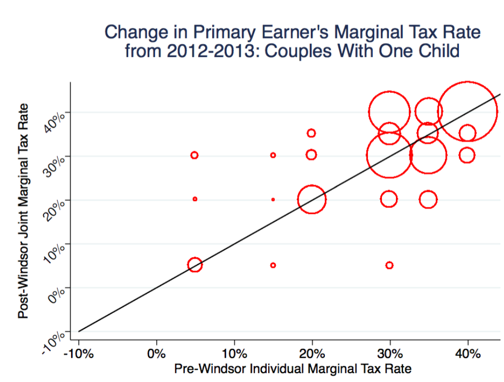
Figure 3: Variation in Marginal Tax Rates Among Primary and Secondary Earners



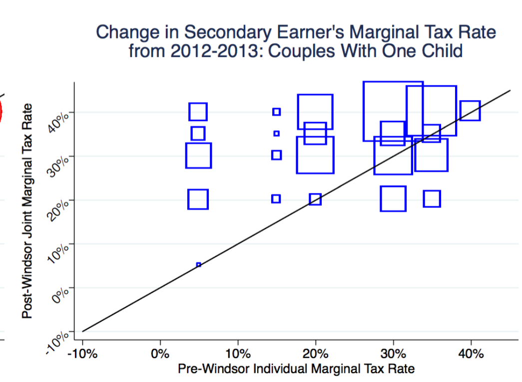
(a) Primary earners, no children



(b) Secondary earners, no children



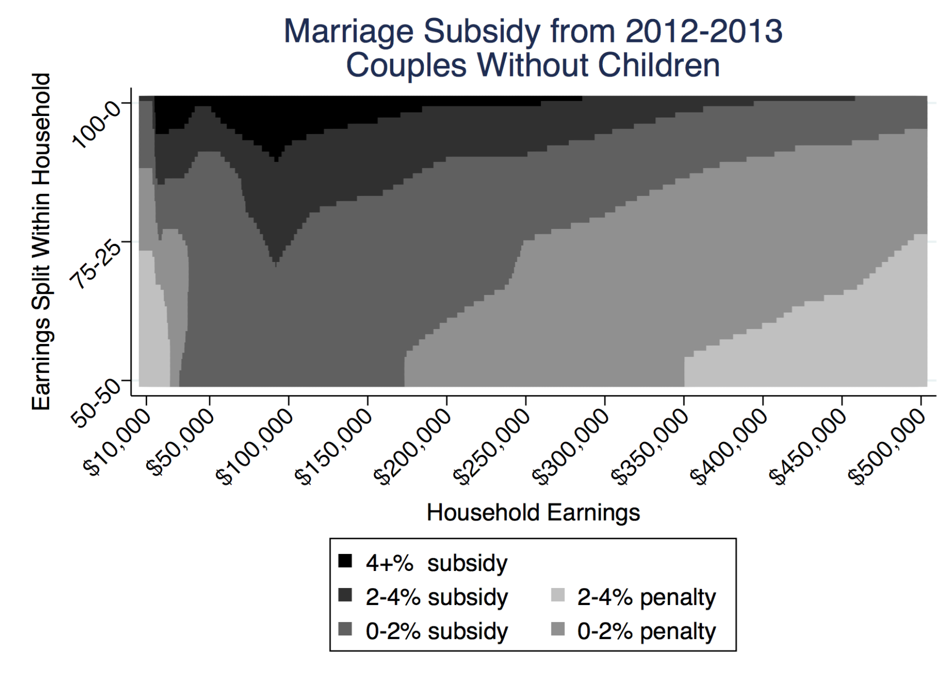
(c) Primary earners, one child



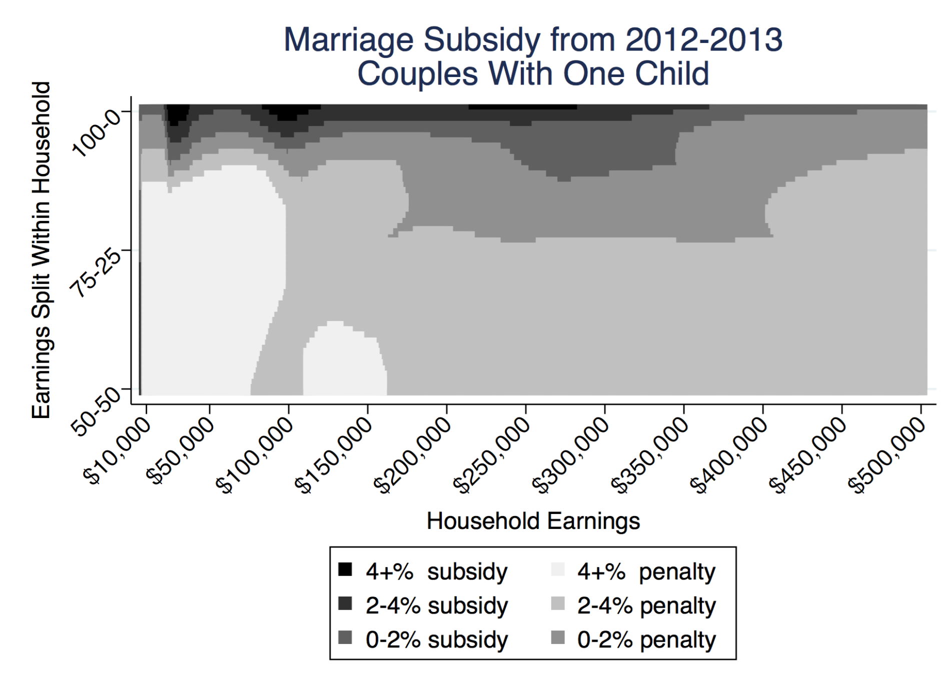
(d) Secondary earners, one child

Notes: The data in Figure 3 are generated, and do not originate from the sample I analyze. I group marginal tax rates into 5 percentage point bins. Figures 3a and 3b demonstrate how marginal tax rates vary following *United States v. Windsor* among primary and secondary earners in childless couples, respectively. Figures 3c and 3d demonstrate how marginal tax rates vary following *United States v. Windsor* among primary and secondary earners in childless couples, respectively. Each data point is weighted by the number of individuals in the same pre- and post-Windsor tax movement bin.

Figure 4: Variation in the Marriage Penalty Among Childless Couples and Couples with One Child



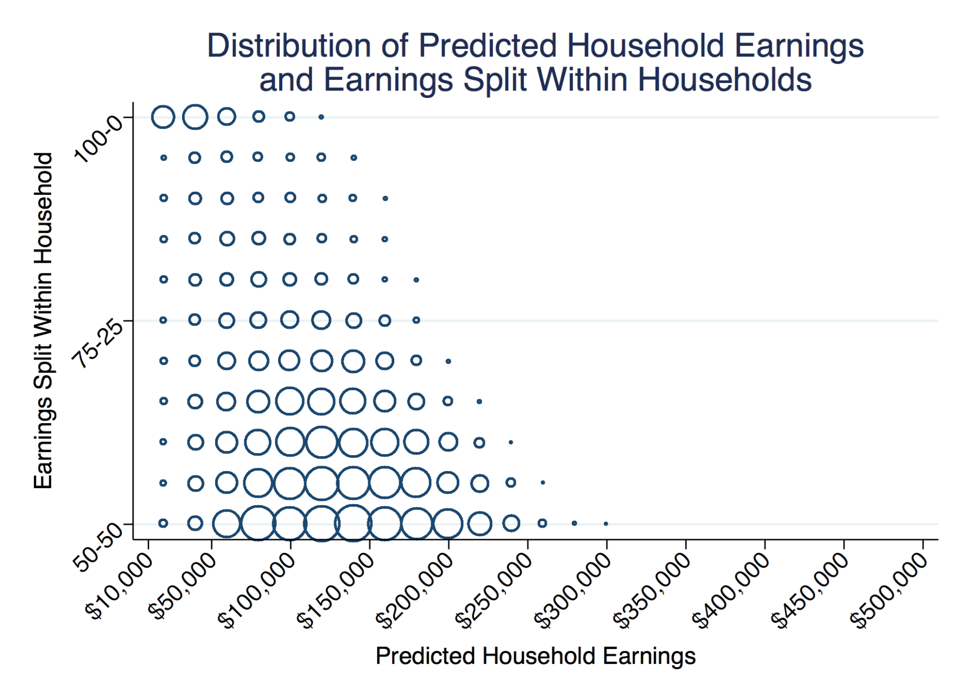
(a)



(b)

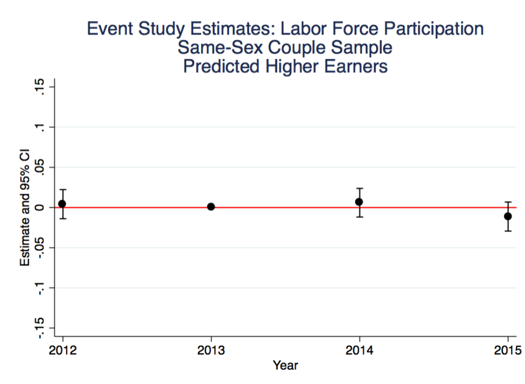
Notes: The data in Figure 4 are generated, and do not originate from the sample I analyze. Figure 4a demonstrates how the marriage subsidy varies for tax year 2013 among childless couples as a function of total household earnings and the split in earnings within the household. Figure 4b demonstrates how the marriage subsidy varies for tax year 2013 among couples with one child as a function of total household earnings and the split in earnings within the household.

Figure 5: Distribution of Predicted Household Earnings and Earnings Splits Among Same-Sex Couples

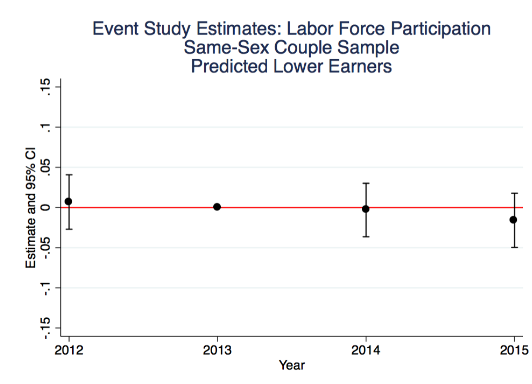


Notes: The data in Figure 5 come from the sample of same-sex couples I analyze in this paper, and are weighted to reflect the number of observations in each predicted household earnings-earnings split bin. Observations of same-sex married and cohabiting couples are included in the figure.

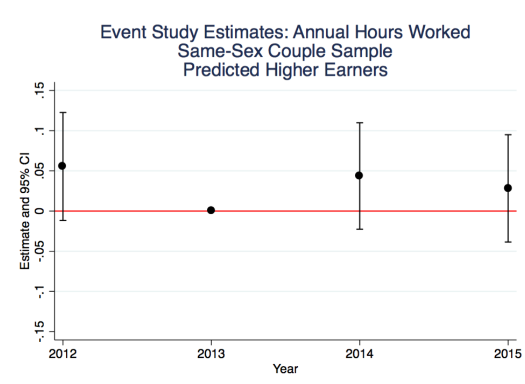
Figure 6: Event Study Results



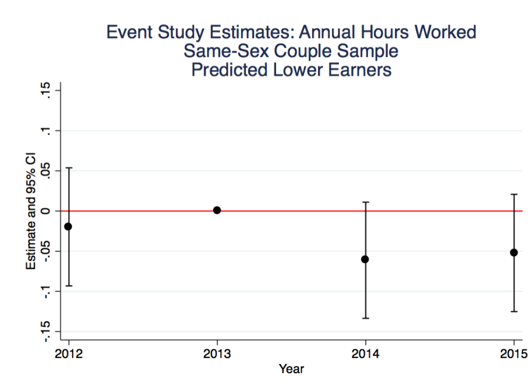
(a)



(b)



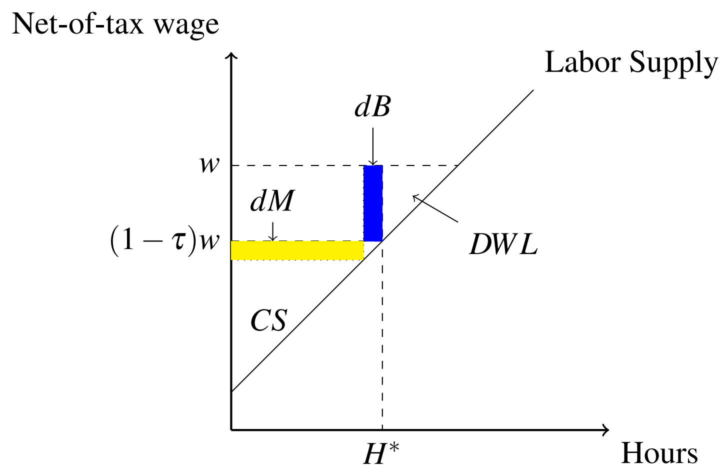
(c)



(d)

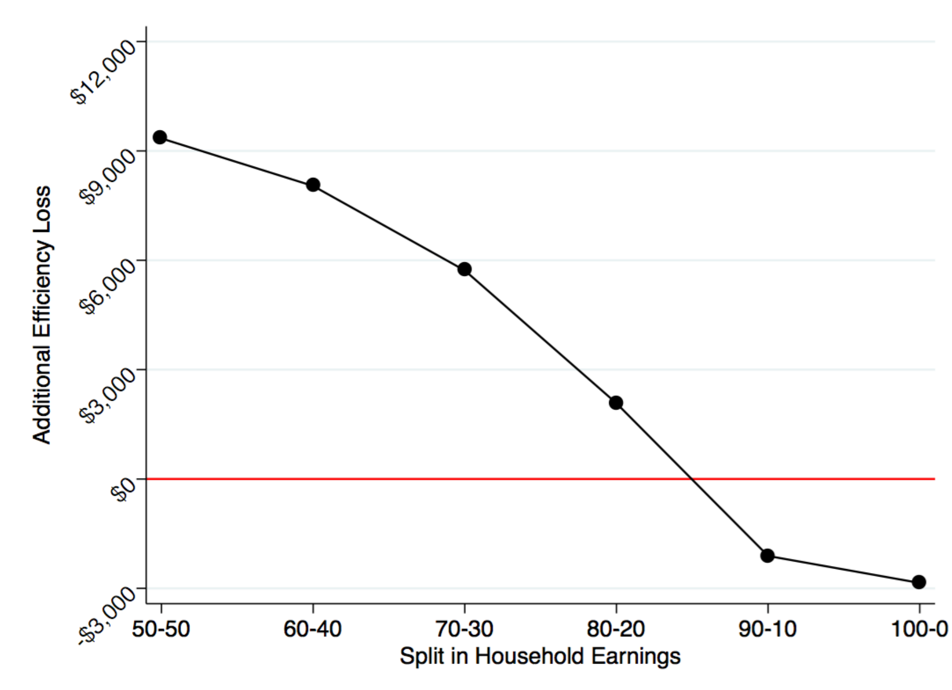
Notes: The data come from the 2012–2015 waves of the American Community Survey. The figures plot coefficient estimates and 95% confidence intervals of θ_1^t from the regression $Y_{it} = \theta_0 + \theta_1^{2012} SSMC_i \times 2012 + \theta_1^{2014} SSMC_i \times 2014 + \theta_1^{2015} SSMC_i \times 2015 + \theta_2 SSMC_i + \theta_3 X_{it} + \delta_t + \mu_s + \varepsilon_{it}$.

Figure 7: Welfare Consequences of a Tax Increase



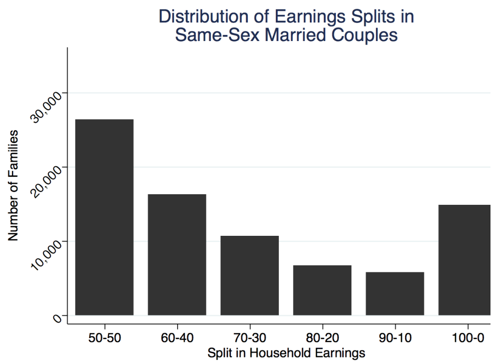
Notes: The figure graphs the effect of a small increase in the marginal tax rate from τ to $\tau+\varepsilon$. The total change in tax revenue is $dTR = dM + dB$, where dM is the mechanical change in tax revenue and dB is the behavioral change in tax revenue following a small tax increase. The change in deadweight loss is equal to $-dB$, as described in Section 2.6.

Figure 8: Additional Deadweight Loss Under Joint Taxation Relative to Individual Taxation

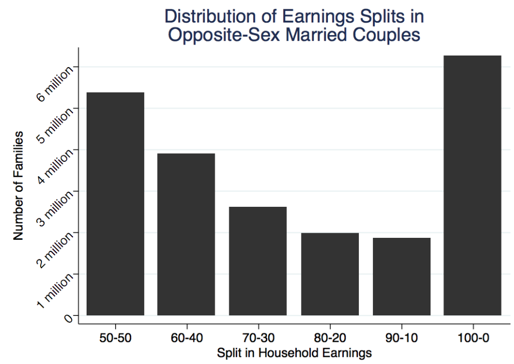


Notes: Each data point represents the average net additional deadweight loss created by joint taxation relative to individual taxation along the intensive and extensive margins per family for households in the same earnings split bins. Households with earnings splits between 50-50 and 59-41 are included in the 50-50 split bin, households with earnings splits between 60-40 and 69-31 are included in the 60-40 split bin, and so on.

Figure 9: Distribution of Married Couples by Earnings Split



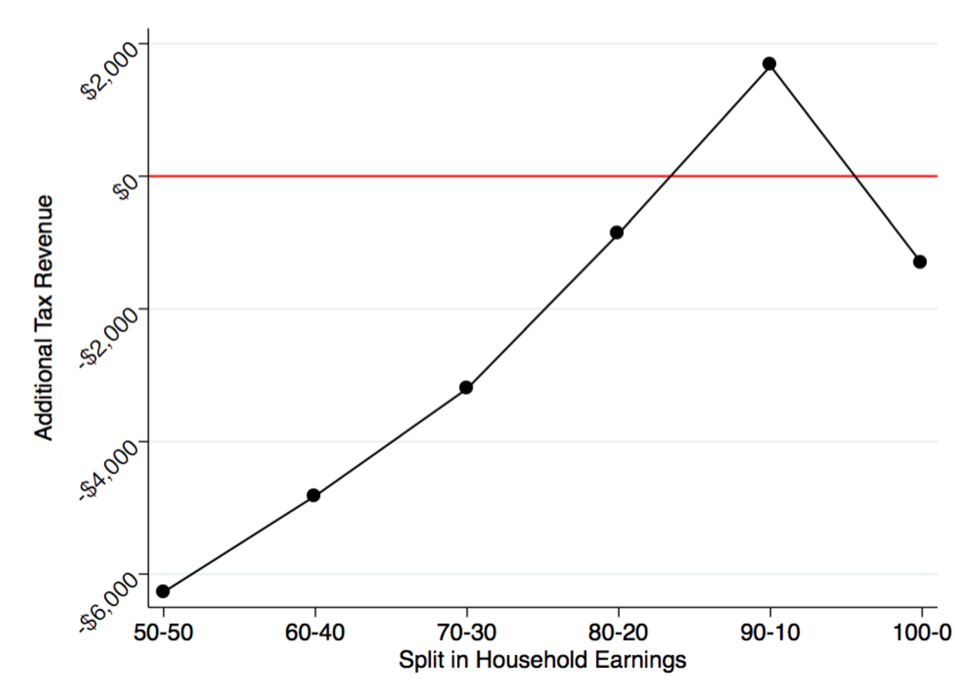
(a)



(b)

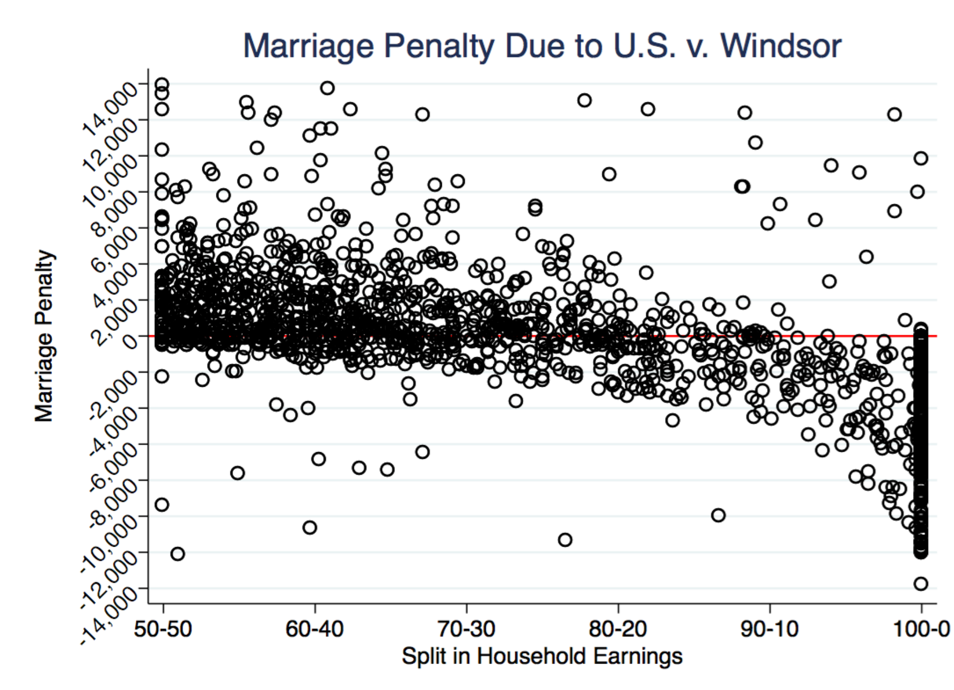
Notes: The data come from the 2012–2015 waves of the American Community Survey. Population estimates are computed using household weights among households in the main analysis sample, which includes couples who married in 2012 or earlier in which at least one partner works and both partners are between 25–54 years old. I further limit the population estimates to households who responded to the American Community Survey using a computer assisted telephone or personal interview to reduce the risk of misclassification between same-sex and opposite-sex married couples.

Figure 10: Additional Tax Revenue Under Joint Taxation Relative to Individual Taxation



Notes: Each data point represents the average net additional tax revenue created by joint taxation relative to individual taxation along the intensive and extensive margins per family for households in the same earnings split bins. Households with earnings splits between 50-50 and 59-41 are included in the 50-50 split bin, households with earnings splits between 60-40 and 69-31 are included in the 60-40 split bin, and so on.

Figure 11: Marriage Penalty Due to *United States v. Windsor* by Household Earnings Split



Notes: The data come from the 2014–2015 waves of the American Community Survey and include same-sex married couples who married in 2012 or earlier and who responded to the American Community Survey using a computer-assisted telephone or personal interview. Each data point represents the marriage penalty faced by a same-sex married household with a particular household earnings split. I have also removed households from this figure that experienced a marriage penalty or subsidy greater than $14,000 simply for presentation purposes (100 observations amounting to 3.0% of the sample).

Table 1: Summary Statistics of Couples in the 2012–2015 American Community Survey

	Same-sex couples		Opposite-sex couples	
	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners
Male	0.47	0.47	0.55	0.55
	(0.50)	(0.50)	(0.50)	(0.50)
Black	0.06	0.07	0.05	0.06
	(0.24)	(0.25)	(0.22)	(0.24)
Other race	0.10	0.12	0.08	0.11
	(0.31)	(0.32)	(0.27)	(0.31)
Age	42.50	41.51	40.79	39.43
	(7.34)	(8.04)	(8.22)	(8.81)
Education	15.20	14.12	15.39	14.11
	(3.04)	(3.05)	(2.35)	(2.47)
Any children	0.48	0.48	0.08	0.08
	(0.50)	(0.50)	(0.27)	(0.27)
Conditional number	1.93	1.93	1.53	1.53
of children	(1.00)	(1.00)	(0.76)	(0.76)
Labor force	0.97	0.83	0.97	0.89
participation	(0.18)	(0.38)	(0.17)	(0.31)
Conditional	2,044.66	1,582.79	2,039.77	1,749.91
annual hours	(757.91)	(978.73)	(723.89)	(873.57)
Reported earnings	81,847.25	47,102.58	70,485.46	46,414.47
	(91,956.16)	(62,610.46)	(74,954.99)	(54,424.22)
Predicted earnings	68,637.44	43,519.05	66,086.05	41,213.28
	(28,365.59)	(27,292.93)	(27,576.06)	(25,464.46)
% change in marginal	0.02	-0.05	0.00	0.00
net-of-tax rate[a]	(0.09)	(0.10)	(0.03)	(0.04)
Change in HH income	-642.09	-642.09	0	0
due to marriage subsidy[a]	(5,214.21)	(5,214.21)		
Observations	4,116	4,116	9,104	9,104
Worker Observations	3,022	3,022	7,321	7,321

Notes: The data come from the 2012–2015 waves of the American Community Survey. The samples includes same-sex cohabiting couples and same-sex married couples who married in 2012 or earlier, who do not live with any other couple, and who have at least one earner in the household. Annual hours worked is the product of "usual hours worked per week" and "weeks worked last year." There is some variation within the first-dollar marginal net-of-tax rate variable among same-sex cohabiting couples due to other changes in the tax code between tax years 2012–2013. The number of children statistics are conditional on having any children.
a: These summary statistics are for post-period observations in 2014–2015.

Table 2: Illustration of How the Taxation Variables Are Constructed

Couple	Individual	Type	Year	Predicted Earnings	$1-\hat{\tau}_{it}$	$1-\hat{\tau}_{i2013}$	$\%\Delta(1-\hat{\tau}_{it})$	$\hat{T}^{ij}$	$\hat{T}^{i}+\hat{T}^{j}$	$\Delta\hat{T}_{ij}$
A	1	Cohabiting	2012	$14,182			0			0
A	2	Cohabiting	2012	$43,773			0			0
B	1	Married	2012	$ 32,693			0			0
B	2	Married	2012	$ 39,350			0			0
C	1	Cohabiting	2013	$34,163			0			0
C	2	Cohabiting	2013	$114,550			0			0
D	1	Married	2013	$54,592			0			0
D	2	Married	2013	$58,529			0			0
E	1	Cohabiting	2014	$37,873	0.85	0.85	0			0
E	2	Cohabiting	2014	$59,853	0.75	0.75	0			0
F	1	Married	2014	$9,208	0.85	1	-0.15	$5,207	$5,877	$670
F	2	Married	2014	$62,018	0.85	0.70	0.21	$5,207	$5,877	$670
G	1	Cohabiting	2015	$70,266	0.75	0.75	0			0
G	2	Cohabiting	2015	$80,397	0.75	0.75	0			0
H	1	Married	2015	$67,084	0.77	0.77	0	$21,617	$17,048	-$4,569
H	2	Married	2015	$88,271	0.77	0.72	0.07	$21,617	$17,048	-$4,569

Notes: Recall that $\%\Delta(1-\hat{\tau}_{it}) = \frac{1-\hat{\tau}_{it}-1-\hat{\tau}_{i2013}}{1-\hat{\tau}_{i2013}}$ if the year is 2014 or later (and 0 otherwise), and $\Delta\hat{T}_{ij} = [\hat{T}^{i}+\hat{T}^{j}] - [\hat{T}^{ij}]$ if the year is 2014 or later and the couple is a same-sex married couple (and 0 otherwise). The couple type indicates whether the couple is treated (married) or untreated (cohabiting). I describe the predicted earnings process in Section 2.3. $1-\hat{\tau}_{it}$ is individual i's predicted marginal net-of-tax rate in year $t \geq 2014$ and $1-\hat{\tau}_{i2013}$ is the predicted marginal net-of-tax rate in 2013 (tax year 2012). A positive value of the change in marginal net-of-tax rates, $\%\Delta(1-\hat{\tau}_{it})$, indicates an increase in the individual's marginal net-of-tax rate due to the Supreme Court ruling. $\hat{T}^{i}$ is individual i's predicted tax liability in 2013 (tax year 2012) under the single federal tax schedule and $\hat{T}^{ij}$ is the couple's predicted tax liability in year $t \geq 2014$. A positive value of the marriage subsidy, $\Delta\hat{T}_{ij}$, indicates an increase in household income due to a lower joint tax liability after the Supreme Court ruling. I calculate $\hat{\tau}_{it}$, T^{ij}, and $T^{i}+T^{j}$ variables using an individual's predicted earnings and the NBER TAXSIM model.

Table 3: Generalized Difference-in-Differences Effects of the *United States v. Windsor* Ruling on Labor Force Participation Among Same-Sex Couples

	Men and Women		Men		Women	
	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners
Outcome: LFP						
% change in marginal net-of-tax rate (10s)	0.005	0.059***	0.002	0.109***	0.008	0.035***
	(0.004)	(0.011)	(0.006)	(0.017)	(0.006)	(0.013)
Change in HH income due to marriage subsidy ($1,000s)	-0.001	-0.010***	-0.000	-0.008**	-0.002	-0.014**
	(0.001)	(0.003)	(0.001)	(0.003)	(0.002)	(0.007)
Spouse's predicted earnings ($10,000s)	0.000	0.012***	0.000	0.012***	0.000	0.013***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Same-sex married couple	-0.002	-0.046***	-0.003	-0.030***	-0.001	-0.057***
	(0.004)	(0.007)	(0.005)	(0.010)	(0.005)	(0.011)
Year = 2013	0.003	0.008	0.010*	0.017	-0.003	0.001
	(0.004)	(0.008)	(0.006)	(0.011)	(0.007)	(0.011)
Year = 2014	0.007	0.005	0.011*	0.027**	0.004	-0.014
	(0.004)	(0.008)	(0.006)	(0.011)	(0.007)	(0.012)
Year = 2015	0.014***	0.006	0.023***	0.014	0.006	-0.000
	(0.004)	(0.008)	(0.005)	(0.011)	(0.007)	(0.012)
Age	-0.000	-0.005***	-0.000	-0.005***	-0.000	-0.005***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Age difference	0.000	-0.002***	0.000	-0.003***	0.000	-0.002**
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
Years of education	0.004***	0.015***	0.004***	0.013***	0.005***	0.016***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Black	-0.035***	0.000	-0.049***	-0.003	-0.023*	0.004
	(0.010)	(0.013)	(0.016)	(0.017)	(0.013)	(0.018)
Other race	-0.012*	-0.018*	-0.020**	-0.024*	-0.001	-0.011
	(0.006)	(0.010)	(0.009)	(0.013)	(0.010)	(0.015)
Female	-0.006**	0.001				
	(0.003)	(0.006)				
Hicksian elasticity (substitution effect)	0.056	0.679***	0.017	1.236***	0.087	0.400***
	(0.043)	(0.130)	(0.059)	(0.188)	(0.061)	(0.154)
Income elasticity (income effect)	-0.002	-0.029***	-0.001	-0.022**	-0.006	-0.044**
	(0.002)	(0.008)	(0.002)	(0.009)	(0.006)	(0.021)
Number of Observations	13,220	13,220	6,973	6,973	6,247	6,247

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Bootstrapped standard errors are in parentheses. The data come from the 2012-2015 waves of the American Community Survey. The sample includes same-sex cohabiting couples and same-sex married couples who married in 2012 or earlier, who do not live with any other couple, and who have at least one earner in the household. Labor force participation is equal to 1 if the individual has positive annual hours of work. All specifications include state fixed effects.

Table 4: Generalized Difference-in-Differences Effects of the *United States v. Windsor* Ruling on Annual Hours of Work Among Same-Sex Couples

	Men and Women		Men		Women	
	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners
Outcome: Annual hours of work/1000						
% change in marginal	0.037**	0.006	0.045	0.034	0.032*	-0.006
net-of-tax rate (10s)	(0.016)	(0.017)	(0.028)	(0.031)	(0.019)	(0.021)
Change in HH income due	-0.012***	0.000	-0.010*	-0.010	-0.013**	0.009
to marriage subsidy ($1,000s)	(0.003)	(0.006)	(0.005)	(0.008)	(0.006)	(0.005)
Spouse's predicted	0.026***	0.016***	0.026***	0.007*	0.027***	0.026***
earnings ($10,000s)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)
Same-sex married couple	-0.016	-0.068***	-0.015	-0.057***	-0.015	-0.081***
	(0.013)	(0.015)	(0.019)	(0.021)	(0.018)	(0.022)
Year = 2013	0.002	0.008	0.017	-0.015	-0.008	0.031
	(0.016)	(0.017)	(0.022)	(0.023)	(0.022)	(0.024)
Year = 2014	-0.012	0.031*	0.020	0.027	-0.042*	0.033
	(0.016)	(0.017)	(0.020)	(0.023)	(0.023)	(0.023)
Year = 2015	0.003	0.040**	0.007	0.016	0.000	0.065**
	(0.016)	(0.017)	(0.021)	(0.024)	(0.021)	(0.026)
Age	-0.000	0.002***	-0.002*	0.001	0.002	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age difference	0.002*	-0.005***	0.004***	-0.006***	-0.001	-0.003**
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Years of education	0.020***	0.021***	0.022***	0.027***	0.018***	0.013***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Black	-0.073***	-0.081***	-0.095***	-0.105**	-0.053	-0.059
	(0.025)	(0.029)	(0.036)	(0.044)	(0.035)	(0.036)
Other race	-0.020	-0.041**	-0.041	-0.016	0.018	-0.073**
	(0.021)	(0.020)	(0.028)	(0.028)	(0.031)	(0.032)
Female	-0.053***	-0.039***				
	(0.012)	(0.013)				
Marshallian elasticity	0.002**	0.000	0.002	0.002	0.002*	-0.000
(income + substitution effects)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Hicksian elasticity	0.382***	-0.009	0.360*	0.277	0.366**	-0.198
(substitution effect)	(0.113)	(0.143)	(0.185)	(0.214)	(0.168)	(0.120)
Income elasticity	-0.012***	0.000	-0.010*	-0.008	-0.013**	0.009
(income effect)	(0.003)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)
Number of Observations	10,343	10,343	5,431	5,431	4,912	4,912

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Bootstrapped standard errors are in parentheses. The data come from the 2012-2015 waves of the American Community Survey. The sample includes same-sex cohabiting couples and same-sex married couples who married in 2012 or earlier, who do not live with any other couple, and who have two earners in the household. Annual hours of work is equal to usual hours worked per week multiplied by weeks worked last year. All specifications include state fixed effects.

Table 5: Robustness Specifications: Generalized Difference-in-Differences Effects of the *United States v. Windsor* Ruling on Labor Force Participation Among Same-Sex Couples

	CATI/CAPI interview		Inconsistent laws		Married 3 years		Opposite-Sex		State-by-year FE	
	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners
Outcome: LFP										
% change in marginal net-of-tax rate (10s)	0.003	0.059***	0.002	0.070***	0.007*	0.057***	0.001	0.035***	0.005	0.060***
	(0.005)	(0.014)	(0.005)	(0.011)	(0.004)	(0.012)	(0.001)	(0.002)	(0.004)	(0.011)
Change in HH income due to marriage subsidy ($1,000s)	-0.000	-0.008**	-0.001	-0.010***	-0.001	-0.010***	-0.001	-0.008***	-0.001	-0.010***
	(0.001)	(0.003)	(0.001)	(0.004)	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)	(0.003)
Spouse's predicted earnings ($10,000s)	-0.000	0.012***	0.001	0.012***	0.001	0.013***	-0.002***	-0.007***	0.000	0.012***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.000)	(0.000)	(0.001)	(0.002)
Same-sex married couple	-0.002	-0.046***	-0.001	-0.042***	-0.004	-0.054***	0.003	0.022***	-0.002	-0.046***
	(0.004)	(0.010)	(0.004)	(0.008)	(0.004)	(0.009)	(0.003)	(0.006)	(0.004)	(0.007)
Year = 2013	0.013	0.053***	0.004	0.005	-0.011**	-0.003	0.001	-0.003***	0.021	0.003
	(0.009)	(0.016)	(0.005)	(0.008)	(0.004)	(0.008)	(0.000)	(0.001)	(0.027)	(0.231)
Year = 2014	0.018*	0.051***	0.008*	0.006	-0.007*	-0.002	0.003***	-0.006***	0.016	-0.055
	(0.009)	(0.016)	(0.005)	(0.009)	(0.004)	(0.008)	(0.000)	(0.001)	(0.027)	(0.254)
Year = 2015	0.031***	0.054***	0.016***	0.011	0.000	0.000	0.005***	-0.009***	0.020	0.165**
	(0.009)	(0.016)	(0.004)	(0.009)	(0.000)	(0.000)	(0.000)	(0.001)	(0.028)	(0.070)
Age	-0.000	-0.004***	0.000	-0.005***	-0.000*	-0.004***	-0.000***	0.002***	-0.000	-0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age difference	0.000	-0.002***	0.000	-0.002***	0.000	-0.002***	-0.001***	0.001***	0.000	-0.002***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
Years of education	0.004***	0.013***	0.003***	0.014***	0.003***	0.012***	0.004***	0.026***	0.004***	0.015***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.000)	(0.000)	(0.001)	(0.001)
Black	-0.025**	-0.002	-0.032***	-0.002	-0.026**	-0.005	-0.022***	0.038***	-0.035***	0.000
	(0.011)	(0.016)	(0.011)	(0.014)	(0.010)	(0.015)	(0.001)	(0.002)	(0.010)	(0.013)
Other race	-0.009	-0.015	-0.011*	-0.020**	-0.018**	-0.016	-0.006***	-0.048***	-0.012**	-0.017*
	(0.007)	(0.012)	(0.006)	(0.010)	(0.008)	(0.012)	(0.001)	(0.001)	(0.006)	(0.010)
Female	-0.007*	-0.002	-0.005	0.001	-0.008**	-0.004	-0.028***	-0.144***	-0.007**	-0.000
	(0.004)	(0.008)	(0.003)	(0.006)	(0.004)	(0.008)	(0.001)	(0.001)	(0.003)	(0.006)
Hicksian elasticity (substitution effect)	0.027	0.586***	0.024	0.797***	0.072*	0.649***	0.006	0.460***	0.054	0.683***
	(0.050)	(0.136)	(0.053)	(0.129)	(0.043)	(0.140)	(0.007)	(0.025)	(0.044)	(0.126)
Income elasticity (income effect)	-0.000	-0.017**	-0.002	-0.029***	-0.003	-0.036***	-0.006	-0.080***	-0.002	-0.029***
	(0.001)	(0.007)	(0.002)	(0.011)	(0.003)	(0.011)	(0.006)	(0.023)	(0.002)	(0.009)
Number of Observations	8,380	8,380	11,335	11,335	9,181	9,181	1,082,995	1,082,995	13,220	13,220

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Bootstrapped standard errors are in parentheses. The data come from the 2012-2015 waves of the American Community Survey. The sample includes same-sex cohabiting couples and same-sex married couples who married in 2012 or earlier, who do not live with any other couple, and who have at least one earner in the household. Labor force participation is equal to 1 if the individual has positive annual hours of work. The CATI/CAPI sample includes only couples who responded to the American Community Survey using computer assisted telephone or personal interviews. The "Inconsistent Laws" sample excludes couples residing in states with inconsistent tax policies surrounding same-sex married couples. The "Married 3 Years" sample includes same-sex married couples who have been married for at least three years and same-sex cohabiting couples. The "Opposite-Sex" sample uses an alternative control group of opposite-sex married couples who married in 2012 or earlier in which both partners are between 25–54 years old.

Table 6: Robustness Specifications: Generalized Difference-in-Differences Effects of the *United States v. Windsor* Ruling on Annual Hours of Work Among Same-Sex Couples

	CATI/CAPI interview		Inconsistent laws		Married 3 years		Opposite-Sex		State-by-year FE	
	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners
Outcome: Annual hours of work/1000										
% change in marginal net-of-tax rate (10s)	0.031*	0.011	0.035*	0.010	0.038**	-0.006	0.001	0.012***	0.038**	0.005
	(0.018)	(0.020)	(0.019)	(0.020)	(0.016)	(0.017)	(0.003)	(0.004)	(0.016)	(0.017)
Change in HH income due to marriage subsidy ($1,000s)	-0.011***	-0.003	-0.012***	0.000	-0.012***	0.000	-0.008**	0.002	-0.012***	0.001
	(0.004)	(0.006)	(0.004)	(0.006)	(0.003)	(0.005)	(0.003)	(0.006)	(0.003)	(0.006)
Spouse's predicted earnings ($10,000s)	0.023***	0.012***	0.025***	0.015***	0.026***	0.012***	0.016***	-0.011***	0.026***	0.015***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.000)	(0.000)	(0.003)	(0.003)
Same-sex married couple	-0.024	-0.084***	-0.012	-0.070***	-0.024	-0.075***	0.011	0.008	-0.014	-0.067***
	(0.016)	(0.020)	(0.014)	(0.016)	(0.016)	(0.019)	(0.011)	(0.013)	(0.013)	(0.015)
Year = 2013	-0.037	-0.032	0.001	0.006	0.002	-0.034*	0.004**	0.006***	0.042	0.215**
	(0.028)	(0.031)	(0.017)	(0.018)	(0.016)	(0.018)	(0.002)	(0.002)	(0.125)	(0.102)
Year = 2014	-0.047*	-0.002	0.000	0.030	-0.013	-0.007	0.014***	0.011***	-0.353	-0.440
	(0.028)	(0.031)	(0.016)	(0.018)	(0.016)	(0.017)	(0.002)	(0.002)	(0.353)	(0.588)
Year = 2015	-0.029	0.001	-0.001	0.042**	0.000	0.000	0.021***	0.017***	0.114	-0.205
	(0.029)	(0.031)	(0.016)	(0.019)	(0.000)	(0.000)	(0.002)	(0.002)	(0.116)	(0.244)
Age	0.001	0.003***	-0.001	0.002**	0.000	0.002**	0.002***	0.005***	-0.000	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
Age difference	0.001	-0.006***	0.002*	-0.005***	0.002*	-0.006***	-0.003***	0.000**	0.002*	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
Years of education	0.022***	0.022***	0.020***	0.020***	0.020***	0.022***	0.018***	0.029***	0.020***	0.021***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.000)	(0.000)	(0.003)	(0.003)
Black	-0.093***	-0.081**	-0.075***	-0.094***	-0.081***	-0.093***	-0.104***	0.035***	-0.073***	-0.084***
	(0.029)	(0.033)	(0.028)	(0.030)	(0.030)	(0.034)	(0.003)	(0.003)	(0.025)	(0.027)
Other race	-0.024	-0.046*	-0.021	-0.041*	-0.009	-0.037	-0.091***	0.014***	-0.020	-0.042**
	(0.025)	(0.025)	(0.021)	(0.022)	(0.024)	(0.024)	(0.002)	(0.003)	(0.021)	(0.021)
Female	-0.076***	-0.053***	-0.053***	-0.044***	-0.061***	-0.030**	-0.272***	-0.382***	-0.054***	-0.041***
	(0.014)	(0.016)	(0.012)	(0.013)	(0.014)	(0.015)	(0.002)	(0.002)	(0.011)	(0.013)
Marshallian elasticity (income + substitution effects)	0.001*	0.001	0.002*	0.001	0.002**	-0.000	0.000	0.001***	0.002**	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
Hicksian elasticity (substitution effect)	0.372***	0.067	0.383***	-0.008	0.402***	-0.004	0.241**	-0.041	0.387***	-0.013
	(0.115)	(0.149)	(0.121)	(0.151)	(0.113)	(0.140)	(0.101)	(0.142)	(0.105)	(0.144)
Income elasticity (income effect)	-0.013***	-0.003	-0.012***	0.000	-0.014***	0.000	-0.023**	0.006	-0.012***	0.000
	(0.004)	(0.007)	(0.004)	(0.006)	(0.004)	(0.006)	(0.009)	(0.022)	(0.003)	(0.005)
Number of Observations	6,564	6,564	8,905	8,905	7,176	7,176	737,219	737,219	10,343	10,343

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Bootstrapped standard errors are in parentheses. The data come from the 2012-2015 waves of the American Community Survey. The sample includes same-sex cohabiting couples and same-sex married couples who married in 2012 or earlier, who do not live with any other couple, and who have two earners in the household. Annual hours of work is equal to usual hours worked per week multiplied by weeks worked last year. The CATI/CAPI sample includes only couples who responded to the American Community Survey using computer assisted telephone or personal interviews. The "Inconsistent Laws" sample excludes couples residing in states with inconsistent tax policies surrounding same-sex married couples. The "Married 3 Years" sample includes same-sex married couples who have been married for at least three years and same-sex cohabiting couples. The "Opposite-Sex" sample uses an alternative control group of opposite-sex married couples who married in 2012 or earlier in which both partners are between 25–54 years old.

Table 7: Labor Force Participation and Hours Elasticities Using Alternative Specifications

	Original		Spouse's $\%\Delta(1-\tau)$		Include Children		Linear Time Trend	
	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners
Outcome: LFP								
Hicksian elasticity	0.056	0.679***	0.065	0.722***	0.057	0.653***	0.061	0.692***
(substitution effect)	(0.043)	(0.130)	(0.043)	(0.127)	(0.044)	(0.127)	(0.043)	(0.135)
Income elasticity	-0.002	-0.029***	-0.002	-0.029***	-0.003	-0.032***	-0.003	-0.029***
(income effect)	(0.002)	(0.008)	(0.002)	(0.008)	(0.002)	(0.010)	(0.002)	(0.008)
Cross-wage Hicksian			-0.065*	-0.483***				
elasticity			(0.036)	(0.107)				
Number of Observations	13,220	13,220	13,220	13,220	13,220	13,220	13,220	13,220
	Original		Spouse's $\%\Delta(1-\tau)$		Include Children		Linear Time Trend	
	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners
Outcome: Annual hours of work/1000								
Marshallian elasticity	0.002**	0.000	0.002**	0.001	0.002**	0.000	0.002**	0.000
(income + substitution effects)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hicksian elasticity	0.382***	-0.009	0.378***	-0.018	0.427***	0.066	0.390***	0.012
(substitution effect)	(0.113)	(0.143)	(0.109)	(0.139)	(0.123)	(0.140)	(0.118)	(0.140)
Income elasticity	-0.012***	0.000	-0.011***	0.001	-0.013***	-0.002	-0.012***	-0.000
(income effect)	(0.003)	(0.005)	(0.003)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)
Cross-wage Hicksian			0.295***	-0.025				
elasticity			(0.086)	(0.177)				
Number of Observations	10,343	10,343	10,343	10,343	10,343	10,343	10,343	10,343

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Bootstrapped standard errors are in parentheses. The data come from the 2012-2015 waves of the American Community Survey. The sample includes same-sex cohabiting couples and same-sex married couples who married in 2012 or earlier, who do not live with any other couple, and who have two earners in the household. Annual hours of work is equal to usual hours worked per week multiplied by weeks worked last year. The column labeled "Spouse's $\%\Delta(1-\tau)$" includes the spouse's percentage change in marginal net-of-tax rate to allow for substitutability in labor supply. The column labeled "Include Children" includes an indicator variable equal to one if the couple has any children to control for their effect on labor supply. The column labeled "Linear Time Trend" includes a group-specific linear time trend, which relaxes the parallel trends assumption needed for identification.

Table 8: Additional Deadweight Loss Due to Joint Taxation Relative to Individual Taxation

		Same-sex married couples		Opposite-sex married couples	
Earnings split	Additional deadweight loss per family	Population	Total deadweight loss (millions)	Population	Total deadweight loss (millions)
50-50 split	9,343	26,355	$246	5,360,795	$50,089
60-40 split	8,038	16,298	$131	3,894,849	$31,307
70-30 split	5,721	10,698	$61	2,616,340	$14,967
80-20 split	2,077	6,684	$14	1,969,614	$4,091
90-10 split	-2,112	5,801	-$12	1,870,553	-$3,950
100-0 split	-2,852	14,819	-$42	6,265,819	-$17,871
			$398		$78,632

Notes: The dollar figures above represent the additional deadweight loss under joint taxation relative to individual taxation. Population estimates of married couples are computed using household weights in the 2014 American Community Survey among married couples who married in 2012 or earlier, who responded using a computer assisted telephone or personal interview, and in which both parters are between 25–54 years old. The additional deadweight loss, therefore, is a conservative measure based on these population estimates.

Table 9: Additional Tax Revenue Due to Joint Taxation Relative to Individual Taxation

		Same-sex married couples		Opposite-sex married couples	
Earnings split	Additional tax revenue per family	Population	Total tax revenue (millions)	Population	Total tax revenue (millions)
50-50 split	-6,271	26,355	-$165	5,360,795	-$33,618
60-40 split	-4,819	16,298	-$79	3,894,849	-$18,768
70-30 split	-3,202	10,698	-$34	2,616,340	-$8,378
80-20 split	-862	6,684	-$6	1,969,614	-$1,698
90-10 split	1,662	5,801	$10	1,870,553	$3,109
100-0 split	-1,324	14,819	-$20	6,265,819	-$8,298
			-$294		-$67,651

Notes: The dollar figures above represent the additional tax revenue under joint taxation relative to individual taxation. Population estimates of married couples are computed using household weights in the 2014 American Community Survey among married couples who married in 2012 or earlier, who responded using a computer assisted telephone or personal interview, and in which both parters are between 25–54 years old. The additional tax revenue, therefore, is a conservative measure based on these population estimates.

A Appendix: The Collective Labor Supply Model

My generalized difference-in-differences specification is consistent with a collective model of family decision making. In this section I outline the collective model and reproduce the relevant propositions and proofs, as presented by Chiappori (1992) and Chiappori, Fortin, and Lacroix (2002). I then offer a parametric specification that corresponds to my generalized difference-in-differences specification in Section 2.4 and derive and present estimates of the Marshallian labor supply parameters, the derivatives of the sharing rule, and the Marshallian, Hicksian, and income elasticities implied by the collective model.

A.1 The Collective Model

There are two individuals in the household, and each has a distinct utility function. Assume that decisions are Pareto efficient, regardless of the decision-making process in the household. Partners can be expected to know each other's preferences well and that they interact and make decisions often. If this is the case then it seems unlikely that any decision made by the household would leave Pareto improvements available, providing informal justification for this assumption (Chiappori, Fortin, and Lacroix 2002).

Let h^i and C^i be individual i's ($i = 1,2$) labor supply and private Hicksian consumption, respectively. Assume that the price of the consumption good and the individual's available time are normalized to 1 (i.e., $0 \leq h^i \leq 1$). Let w_1, w_2, and y be each individual's respective hourly wage rate and the household non-labor income. In addition, let $\mathbf{s}$ be a vector of distribution factors, which are defined as "variables that affect the household members bargaining position but not preferences or the joint budget set" (Chiappori, Fortin, and Lacroix 2002). Let $\mathbf{z}$ be a vector of preference factors.

The collective model can allow for altruistic preferences of individuals, meaning that one individual's utility can be a function of her partner's leisure and consumption. For simplicity, however, I make the following assumption:

Assumption 1. *Individuals have egoistic preferences, meaning each utility function takes the form*

$U^i(1-h^i,C^i,\mathbf{z})$*, where* U^i *is strictly quasi-concave, increasing, and continuously differentiable for* $i=1,2$*.*

Under Assumption 1, household behavior must be a solution to the following optimization problem:

$$\begin{aligned} \max_{h^1,C^1} \quad & U^1(1-h^1,C^1,\mathbf{z}) \\ \text{subject to} \quad & \delta : U^2(1-h^2,C^2,\mathbf{z}) \geq \bar{u}^2 \\ & \lambda : w_1h^1+w_2h^2+y \geq C^1+C^2 \end{aligned} \tag{$\bar{P}$}$$

for some utility level, $\bar{u}^2$, which is, in general, a function of w_1, w_2, y, $\mathbf{s}$, and $\mathbf{z}$.

Now consider an alternative interpretation of the optimization problem $\bar{P}$, in which the household members first split non-labor income, y, between them according to a sharing rule, so that individual 1 receives $\phi(w_1,w_2,y,\mathbf{s},\mathbf{z})$ and individual 2 receives $y-\phi(w_1,w_2,y,\mathbf{s},\mathbf{z})$. If this is the case then individual i's optimization problem can be written as:

$$\begin{aligned} \max_{h^1,C^1} \quad & U^i(1-h^i,C^i,\mathbf{z}) \\ \text{subject to} \quad & \lambda : w_ih^i+\phi^i(w_1,w_2,y,\mathbf{s},\mathbf{z}) \geq C^i \end{aligned} \tag{P$_i$}$$

where $\phi^1 \equiv \phi$ and $\phi^2 \equiv y-\phi$. Chiappori (1992) shows that problem P_i is equivalent to problem $\bar{P}$ under Assumption 1. The proposition is reproduced below.

Proposition 1. *Let* $h^1(w_1,w_2,y,\mathbf{s},\mathbf{z})$ *and* $h^2(w_1,w_2,y,\mathbf{s},\mathbf{z})$ *be arbitrary functions. There exists a function* $\bar{u}^2(w_1,w_2,y,\mathbf{s},\mathbf{z})$ *such that* h^1 *and* h^2 *are solutions to problem* $\bar{P}$ *if and only if there exists a function* $\phi(w_1,w_2,y,\mathbf{s},\mathbf{z})$ *such that* h^i *is a solution of* P_i *for* $i=1,2$*.*

Proof. This proof is reproduced from Chiappori (1992). First assume that h^1 and h^2 together with consumptions C^1 and C^2 are a solution to problem $\bar{P}$ for some function $\bar{u}^2$. Define $\phi(w_1,w_2,y,\mathbf{s},\mathbf{z}) = C^1(w_1,w_2,y,\mathbf{s},\mathbf{z})-w_1h^1(w_1,w_2,y,\mathbf{s},\mathbf{z})$. Then h^1 is a solution to problem P_1; otherwise it would be possible to increase individual 1's utility without changing individual 2's expenditures, which would be a contradiction. The same argument is true for individual 2.

Now assume that h^1, h^2, C^1, and C^2 are solutions to problems P_1 and P_2 for some function ϕ. Define $\bar{u}^2(w_1, w_2, y, \mathbf{s}, \mathbf{z}) = U^2(1 - h^2(w_1, w_2, y, \mathbf{s}, \mathbf{z}), C^2(w_1, w_2, y, \mathbf{s}, \mathbf{z}))$. Then h^1 and h^2 are solutions to problem $\bar{P}$. We know that $w_2(1 - h^2) + C^2 = e^2(\bar{u}^2)$, where e^2 is the expenditure function associated with U^2. Therefore, any pair (L'^2, C'^2) that also provides utility $\bar{u}^2$ will be at least $w_2 + y - \phi(w_1, w_2, y, \mathbf{s}, \mathbf{z})$. In particular, if (L'^1, L'^2, C'^1, C'^2) is a solution to problem $\bar{P}$ such that $U^1(L'^1, C'^1) > U^1(L^1, C^1)$ then the pair (L'^1, C'^1) costs at most the same amount as (L^1, C^1), which is a contradiction. □

Intuitively, Proposition 1 states that the decision process can be thought of as a two stage process, with the first stage being that the household members divide non-labor income between them according to a sharing rule and the second stage being that each household member maximizes individual utility by choosing labor supply and private consumption according to her corresponding budget constraint.

Assuming an interior solution, optimization problem P_i implies that the individual unrestricted labor supply functions are:

$$h^1 = H^1(w_1, \phi(w_1, w_2, y, \mathbf{s}, \mathbf{z}), \mathbf{z}) \tag{2.9a}$$

$$h^2 = H^2(w_2, y - \phi(w_1, w_2, y, \mathbf{s}, \mathbf{z}), \mathbf{z}) \tag{2.9b}$$

where $H^i(\cdot)$ is individual i's Marshallian labor supply function. Using equations 2.9a and 2.9b, define the following expressions:

$$A = \frac{\frac{\partial h^1}{\partial w_2}}{\frac{\partial h^1}{\partial y}} \qquad B = \frac{\frac{\partial h^2}{\partial w_1}}{\frac{\partial h^2}{\partial y}}$$

$$C = \frac{\frac{\partial h^1}{\partial s}}{\frac{\partial h^1}{\partial y}} \qquad D = \frac{\frac{\partial h^2}{\partial s}}{\frac{\partial h^2}{\partial y}}$$

whenever $\frac{\partial h^1}{\partial y} \cdot \frac{\partial h^2}{\partial y} \neq 0$. Chiappori, Fortin, and Lacroix (2002) provide a set of necessary conditions that allow identification of the sharing rule up to an additive constant. The proposition and proof are reproduced below.

Proposition 2. *Take any point such that* $\frac{\partial h^1}{\partial y} \cdot \frac{\partial h^2}{\partial y} \neq 0$. *If there exists exactly one distribution factor such that* $C \neq D$ *then the following conditions are necessary for any pair* (h^1, h^2) *to be a solution to problem* P_i *for some sharing rule* ϕ*:*

$$\frac{\partial}{\partial s}\left(\frac{D}{D-C}\right) = \frac{\partial}{\partial y}\left(\frac{CD}{D-C}\right) \tag{2.10a}$$

$$\frac{\partial}{\partial w_1}\left(\frac{D}{D-C}\right) = \frac{\partial}{\partial y}\left(\frac{BC}{D-C}\right) \tag{2.10b}$$

$$\frac{\partial}{\partial w_2}\left(\frac{D}{D-C}\right) = \frac{\partial}{\partial y}\left(\frac{AD}{D-C}\right) \tag{2.10c}$$

$$\frac{\partial}{\partial w_1}\left(\frac{CD}{D-C}\right) = \frac{\partial}{\partial s}\left(\frac{BC}{D-C}\right) \tag{2.10d}$$

$$\frac{\partial}{\partial w_2}\left(\frac{CD}{D-C}\right) = \frac{\partial}{\partial s}\left(\frac{AD}{D-C}\right) \tag{2.10e}$$

$$\frac{\partial}{\partial w_2}\left(\frac{BC}{D-C}\right) = \frac{\partial}{\partial w_1}\left(\frac{AD}{D-C}\right) \tag{2.10f}$$

$$\frac{\partial h^1}{\partial w_1} - \frac{\partial h^1}{\partial y}\left(h^1 + \frac{BC}{D-C}\right)\left(\frac{D-C}{D}\right) \geq 0 \tag{2.10g}$$

$$\frac{\partial h^2}{\partial w_2} - \frac{\partial h^2}{\partial y}\left(h^2 - \frac{AD}{D-C}\right)\left(-\frac{D-C}{C}\right) \geq 0 \tag{2.10h}$$

Under the assumption that conditions 2.10a–2.10h hold and for a given $\mathbf{z}$*, the sharing rule is defined up to an additive function* $\kappa(\mathbf{z})$ *depending only on the preference factors* $\mathbf{z}$*. The partial derivatives of the sharing rule with respect to wages, non-labor income, and the distribution factor* *are given by:*

$$\begin{aligned}
\frac{\partial\phi}{\partial w_1} &= \frac{BC}{D-C} \\
\frac{\partial\phi}{\partial w_2} &= \frac{AD}{D-C} \\
\frac{\partial\phi}{\partial y} &= \frac{D}{D-C} \\
\frac{\partial\phi}{\partial s} &= \frac{CD}{D-C}
\end{aligned} \tag{2.11}$$

Proof. This proof is reproduced from Chiappori, Fortin, and Lacroix (2002). Start with equations 2.9a and 2.9b:

$$h^1 = H^1(w_1, \phi(w_1, w_2, y, \mathbf{s}, \mathbf{z}), \mathbf{z})$$

$$h^2 = H^2(w_2, y - \phi(w_1, w_2, y, \mathbf{s}, \mathbf{z}), \mathbf{z})$$

Then:

$$\begin{aligned}
A = \frac{\frac{\partial h^1}{\partial w_2}}{\frac{\partial h^1}{\partial y}} &= \frac{\frac{\partial\phi}{\partial w_2}}{\frac{\partial\phi}{\partial y}} \\
B = \frac{\frac{\partial h^2}{\partial w_1}}{\frac{\partial h^2}{\partial y}} &= -\frac{\frac{\partial\phi}{\partial w_1}}{1-\frac{\partial\phi}{\partial y}} \\
C = \frac{\frac{\partial h^1}{\partial s}}{\frac{\partial h^1}{\partial y}} &= \frac{\frac{\partial\phi}{\partial s}}{\frac{\partial\phi}{\partial y}} \\
D = \frac{\frac{\partial h^2}{\partial s}}{\frac{\partial h^2}{\partial y}} &= -\frac{\frac{\partial\phi}{\partial s}}{1-\frac{\partial\phi}{\partial y}}
\end{aligned}$$

Assuming that $C \neq D$ then the last two equations lead to:

$$\frac{\partial\phi}{\partial y} = \frac{D}{D-C}$$
$$\frac{\partial\phi}{\partial s} = \frac{CD}{D-C}$$

And the first two equations lead to:

$$\frac{\partial\phi}{\partial w_1} = \frac{BC}{D-C}$$
$$\frac{\partial\phi}{\partial w_2} = \frac{AD}{D-C}$$

These partial derivatives are compatible if and only if they satisfy the usual cross-derivative restrictions. Therefore, the following conditions are necessary and sufficient:

$$
\begin{aligned}
\frac{\partial}{\partial s}\left(\frac{D}{D-C}\right) &= \frac{\partial}{\partial y}\left(\frac{CD}{D-C}\right) \\
\frac{\partial}{\partial w_1}\left(\frac{D}{D-C}\right) &= \frac{\partial}{\partial y}\left(\frac{BC}{D-C}\right) \\
\frac{\partial}{\partial w_2}\left(\frac{D}{D-C}\right) &= \frac{\partial}{\partial y}\left(\frac{AD}{D-C}\right) \\
\frac{\partial}{\partial w_1}\left(\frac{CD}{D-C}\right) &= \frac{\partial}{\partial s}\left(\frac{BC}{D-C}\right) \\
\frac{\partial}{\partial w_2}\left(\frac{CD}{D-C}\right) &= \frac{\partial}{\partial s}\left(\frac{AD}{D-C}\right) \\
\frac{\partial}{\partial w_2}\left(\frac{BC}{D-C}\right) &= \frac{\partial}{\partial w_1}\left(\frac{AD}{D-C}\right)
\end{aligned}
$$

If these conditions are fulfilled, then ϕ is defined up to an additive function $\kappa(\mathbf{z})$ depending only on the preference factors $\mathbf{z}$. The inequalities 2.10g and 2.10h follow from standard integrability arguments. Finally, the knowledge of Marshallian labor supplies allows one to recover preferences for any given value of $\kappa(\mathbf{z})$. □

The above propositions and proofs establish the equivalence of optimization problems $\bar{P}$ and P_i as well as identification of the sharing rule, ϕ.

A.2 Parametric Specification of the Collective Labor Supply Model

In order to estimate and test a collective labor supply model, I must first specify a functional form for the unrestricted individual labor supply functions. To reflect my empirical specification in Section 2.4, let h^1 and h^2 be given as follows:

$$h^1 = a_0 + a_1 ln(w_1) + a_2 y + a_3 s + \mathbf{a}_4'\mathbf{z} \quad (2.12a)$$

$$h^2 = b_0 + b_1 ln(w_2) + b_2 y + b_3 s + \mathbf{b}_4'\mathbf{z} \quad (2.12b)$$

I consider the generalized difference-in-differences regression equation in Section 2.4 to correspond to the unrestricted labor supply functions above. The connection between the theoretical model and the empirical specification is not one-to-one, however, because I seek to use tax variation among same-sex married couples, rather than cross-sectional variation in wage rates or non-labor income, in order to separately identify the income and substitution effects. I utilize the difference in age between the predicted primary and secondary earners as the distribution factor, s. I argue that my taxation variables represent exogenous shocks to the individual's net-of-tax wage rate and household income, and therefore marginal changes in these variables will accurately estimate a_1 through a_3 and b_1 through b_3 from the theoretical model, which are necessary to recover individual preferences and the sharing rule.

Note that the partial derivative conditions 2.10a–2.10f are automatically satisfied with this functional form because A, B, C, and D are constants. The partial derivatives of ϕ are then given by:

$$\begin{aligned}\frac{\partial\phi}{\partial y} &= \frac{D}{D-C} = \frac{a_2 b_3}{\Delta}\\ \frac{\partial\phi}{\partial s} &= \frac{CD}{D-C} = \frac{a_3 b_3}{\Delta}\\ \frac{\partial\phi}{\partial w_1} &= \frac{BC}{D-C} = 0\\ \frac{\partial\phi}{\partial w_2} &= \frac{AD}{D-C} = 0\end{aligned} \tag{2.13}$$

where $\Delta \equiv a_2 b_3 - a_3 b_2$. Integrating these partial derivatives produces the sharing rule:

$$\phi = \frac{b_3}{\Delta}(a_2 y + a_3 s) + \kappa(\mathbf{z}) \tag{2.14}$$

It is also possible to recover individual Marshallian labor supply functions that have functional forms consistent with equations 2.9a and 2.9b. These functions take the following form:

$$H^1 = \alpha_1 ln(w_1) + \alpha_2 \phi + \alpha_3 \mathbf{z} \tag{2.15a}$$

$$H^2 = \beta_1 ln(w_2) + \beta_2 (y - \phi) + \beta_3 \mathbf{z} \tag{2.15b}$$

Using equation 2.14 to transform the unrestricted labor supply functions into Marshallian labor supply functions, I recover the following parameters: $\alpha_1 = a_1$, $\alpha_2 = \frac{\Delta}{b_3}$, $\beta_1 = b_1$, and $\beta_2 = -\frac{\Delta}{a_3}$.[55]

The final conditions on the compensated individual labor supplies (conditions 2.10g and 2.10h in proposition 2) are given by:

$$\frac{\alpha_1}{w_1} - \alpha_2 h^1 \geq 0$$
$$\frac{\beta_1}{w_2} - \beta_2 h^2 \geq 0$$

These conditions can be globally satisfied if $\alpha_1 \geq 0$, $\alpha_2 \leq 0$, $\beta_1 \geq 0$, and $\beta_2 \leq 0$.

A.3 Labor Supply Elasticities

I can quantify the Marshallian, income, and Hicksian hours elasticities implied by the collective model using the Marshallian labor supply functions 2.15a and 2.15b. Reproducing the argument from Section 2.4.2 the total effect of a wage rate change is given by the Slutsky equation below, which I have multiplied by $\frac{w}{H}$ to convert into elasticities:

$$\frac{\partial H}{\partial w}\frac{w}{H} = \left.\frac{\partial H}{\partial w}\frac{w}{H}\right|_u + \left[\frac{\partial H}{\partial Y_i}\frac{Y_i}{H}\right]\frac{wH}{Y_i} \tag{2.16}$$

Where H is hours worked, w is the hourly wage rate, and Y is unearned income. The left-hand side term in Equation 2.16 is the Marshallian elasticity. The first term on the right-hand side is the Hicksian elasticity and the second term on the right-hand side is the income effect.

Using the collective model's Marshallian labor supply functions, where $Y_1 = \phi$ and $Y_2 = y - \phi$,

55. Stern (1986) shows that the labor supply functions 2.15a and 2.15b imply the following indirect utility functions $v^1(w_1, \phi, \mathbf{z}) = \left[\frac{e^{\alpha_2 w_1}}{\alpha_2}\right](\alpha_1 ln(w_1) + \alpha_2\phi + \alpha_3'\mathbf{z}) - \frac{\alpha_1}{\alpha_2}\int_{-\infty}^{\alpha_2 w_1}\frac{e^t}{t}dt$ and $v^2(w_2, y-\phi, \mathbf{z}) = \left[\frac{e^{\beta_2 w_2}}{\beta_2}\right](\beta_1 ln(w_2) + \beta_2(y-\phi) + \beta_3'\mathbf{z}) - \frac{\beta_1}{\beta_2}\int_{-\infty}^{\beta_2 w_2}\frac{e^t}{t}dt$

the Marshallian, income, and Hicksian elasticities along the intensive margin are given by:

$$\text{Marshallian: } \varepsilon_M = \frac{\partial H}{\partial w}\frac{w}{H} = \frac{\partial H}{\partial ln(w)}\frac{1}{H} = \alpha_1\frac{1}{H} \tag{2.17a}$$

$$\text{Income: } \varepsilon_I = \frac{\partial H}{\partial Y_i}\frac{Y_i}{H} = \alpha_2\frac{Y_i}{H} \tag{2.17b}$$

$$\text{Hicksian: } \varepsilon_H = \frac{\partial H}{\partial w}\frac{w}{H} - \frac{\partial H}{\partial Y_i}w = \alpha_1\frac{1}{H} - \alpha_2 w \tag{2.17c}$$

where β is substituted for α when considering individual 2.

A.4 Results

I find empirically that the inequality conditions 2.10g and 2.10h above appear to be globally satisfied for the combined sample of men and women, although this may largely be driven by men, for whom the conditions also appear to be globally satisfied. I go further than the global conditions, and calculate the expressions in equation 2.10g and 2.10h at the sample means of hourly wages and annual hours worked.[56] I do not reject the hypothesis that the Slutsky conditions are greater than or equal to zero in any sample. In the following discussion, I consider the pooled sample of men and women.

Identification of the sharing rule rests on the existence of a distribution factor, which is a factor that shifts the bargaining power within the household but does not affect the budget constraint. I follow Oreffice (2011) and use the age difference between the predicted primary earner and the predicted secondary earner. The coefficient of the distribution factor should have opposite signs between the samples of predicted higher and lower earners, which would indicate that the age difference increases the bargaining power of one partner and decreases the bargaining power of the other. This relationship holds for the combined sample of men and women, although it appears to

56. I obtain inequality estimates of 24.10 (chi-squared test statistic of 5.24) for primary earners and 60.02 (chi-squared test statistic of 1.87) for secondary earners within the sample of men and women.

be driven by men. The estimates indicate that the younger partner holds more bargaining power, and therefore works less, which Oreffice (2011) also finds.

Table A2 presents estimates of α_1, α_2, β_1, and β_2, estimates of the Marshallian, Hicksian, and income elasticities implied by the collective model, and estimates of the parameters of the sharing rule, ϕ, as derived above. I assign predicted primary earners as "individual 1" in the theoretical model and predicted secondary earners as "individual 2." I obtain positive estimates of α_1 and β_1, indicating that an increase in the marginal net-of-tax rate increases hours of work. I also obtain negative estimates of α_2 and β_2, indicating that an increase in household non-labor income decreases hours of work, as expected by economic theory. The derivatives of the sharing rule, ϕ, indicate that the primary earner enjoys a 1.2% premium in the family. If household income increases by \$1 then the predicted primary earner receives an additional \$1.01 of the family's non-labor income. I also estimate that if the predicted primary earner is one year older than the predicted secondary earner then the predicted secondary earner receives an additional \$143.77 out of household non-labor income.

The elasticity estimates implied by the collective model are strikingly similar to the main elasticity estimates in Tables 3–4. Among predicted primary earners, I estimate a small Marshallian hours elasticity of 0.002 due to largely off-setting income and substitution effects, resulting in a Hicksian hours elasticity of 0.378 and an income hours elasticity of -0.011. The elasticity estimates for predicted secondary earners are insignificant, although the Marshallian and income elasticities are tightly estimated around zero and are similar to the elasticity estimates in Tables 3–4. The Hicksian hours elasticity for predicted secondary earners is 0.798, which differs from my earlier estimate and is larger than the elasticity for predicted primary earners and imprecisely estimated. The larger Hicksian elasticity estimate is due to the small value of β_1 and larger value of β_2 for predicted secondary earners compared to α_1 and α_2. Smaller sensitivity to changes in $ln(w_2)$ and larger sensitivity to changes in $y-\phi$ imply a larger Hicksian elasticity to counteract the income effect and justify a small Marshallian elasticity. It should be noted that this collective model is suited for dual earner families, and I conclude in Section 2.4 that predicted secondary earners generally respond to tax changes along the participation margin rather than the hours margin. Therefore, this version of the collective model may describe behavior for predicted primary earners better than behavior for predicted secondary earners, leading to similar elasticities among predicted primary earners but not among predicted secondary earners.

Table A1: Summary Statistics of Couples in the 2012–2015 American Community Survey

	Same-sex couples		Opposite-sex couples	
	Predicted higher earners	Predicted lower earners	Predicted higher earners	Predicted lower earners
Male	0.47	0.47	0.85	0.15
	(0.50)	(0.50)	(0.35)	(0.35)
Black	0.06	0.07	0.06	0.06
	(0.24)	(0.25)	(0.23)	(0.23)
Other race	0.10	0.12	0.12	0.14
	(0.30)	(0.33)	(0.33)	(0.34)
Age	42.43	41.58	41.98	40.60
	(7.36)	(8.03)	(7.59)	(7.79)
Education	15.21	14.11	14.16	13.75
	(3.02)	(3.06)	(2.99)	(2.88)
Any children	0.48	0.48	0.78	0.78
	(0.50)	(0.50)	(0.41)	(0.41)
Conditional number	1.93	1.93	2.06	2.06
of children	(1.00)	(1.00)	(1.01)	(1.01)
Labor force	0.97	0.83	0.97	0.76
participation	(0.18)	(0.37)	(0.17)	(0.43)
Conditional	2037.95	1589.50	2116.10	1342.85
annual hours	(760.83)	(979.58)	(723.91)	(989.56)
Reported earnings	81,676.72	47,273.11	72,862.59	32,247.04
	(92,554.78)	(61,817.58)	(77,529.69)	(43,387.57)
Predicted earnings	67,803.75	43,841.69	66,048.75	31,340.04
	(30,170.36)	(27,568.97)	(26,727.38)	(24,645.46)
% change in marginal	0.01	-0.04	0.00	0.00
net-of-tax rate[a]	(0.09)	(0.09)	(0.03)	(0.03)
Change in HH income	-762.37	-762.37	0	0
due to marriage subsidy[a]	(5,082.70)	(5,082.70)		
Observations	4,116	4,116	1,078,879	1,078,879
Worker Observations	3,022	3,022	734,197	734,197

Notes: The data come from the 2012–2015 waves of the American Community Survey. The samples includes same-sex married couples and opposite-sex married couples who married in 2012 or earlier, who do not live with any other couple, and who have at least one earner in the household. Annual hours worked is the product of "usual hours worked per week" and "weeks worked last year." There is some variation within the first-dollar marginal net-of-tax rate variable among opposite-sex married couples due to other changes in the tax code between tax years 2012–2013. The number of children statistics are conditional on having any children.
a: These summary statistics are for post-period observations in 2014–2015.

Table A2: Collective Model Parameter Estimates and Elasticities

	Marshallian labor supplies: $H^1 = \alpha_1 ln(w_1) + \alpha_2\phi + \alpha_3\mathbf{z}$ $H^2 = \beta_1 ln(w_2) + \beta_2(y-\phi) + \beta_3\mathbf{z}$				
	α_1	α_2	Marshallian elasticity	Hicksian elasticity	Income elasticity
Primary earners	3.651**	-0.012***	0.002**	0.378***	-0.011***
	(1.594)	(0.004)	(0.001)	(0.114)	(0.003)
	β_1	β_2	Marshallian elasticity	Hicksian elasticity	Income elasticity
Secondary earners	0.612	-0.031	0.000	0.798	0.003
	(1.701)	(0.023)	(0.001)	(0.584)	(0.002)
	$\frac{\partial\phi}{\partial y}$	$\frac{\partial\phi}{\partial s}$			
Sharing rule	1.012***	-143.770			
	(0.173)	(98.809)			

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Robust standard errors are in parentheses. The data come from the 2012-2015 waves of the American Community Survey. The sample includes same-sex cohabiting couples and same-sex married couples who married in 2012 or earlier, who do not live with any other couple, and couples in which both partners have positive annual hours of work. Annual hours worked is the product of "usual hours worked per week" and "weeks worked last year." The corresponding equations are $H^1 = \alpha_1 ln(w_1) + \alpha_2\phi + \alpha_3\mathbf{z}$ and $H^2 = \beta_1 ln(w_2) + \beta_2(y-\phi) + \beta_3\mathbf{z}$.

Chapter 3

Elite Schools and Opting-Out: Estimating the Effects of College Selectivity on Women's Marriage and Career Outcomes

with Suqin Ge and Amalia Miller

3.1 Introduction

Women experienced dramatic changes in the labor and marriage markets over the past half century, giving rise to several well-documented trends. Women's college attendance and graduation increased, eventually surpassing men's figures by 1978, and women's shares at elite or highly selective colleges has grown (Goldin 2006). In addition, greater shares of women in high status and high-pay occupations, along with greater formal human capital, is widely accepted as a major contributor to the increase in women's earnings relative to men over the last fifty years (Blau and Kahn 1997; Bailey, Hershbein, and Miller 2012). Yet researchers have noted the leveling off and slight decline in female labor force participation in the United States since the mid-1990s as well as limited relative wage gains for women over that period, leaving unclear the impact of the increase in college quality on outcomes for female students. (Blau and Kahn 2006, 2007; Goldin 2006). Marriage rates have also consistently declined since the 1970s and the median age at first marriage has risen over the same time period (Parker and Stepler 2017). These phenomena are likely interrelated. In this paper, we estimate the causal effects of college selectivity on women's family and career outcomes in order to determine its role in the labor supply and marriage trends of the past half century.

Some have speculated that the recent plateau and decline in female labor force participation might be due, in particular, to elite women "opting out" of the labor force (Belkin 2003; Story 2005; Hersch 2013).[1] This hypothesis has received limited support from the academic literature (Bertrand, Goldin, and Katz 2010; Antecol 2011; Herr and Wolfram 2012; Hersch 2013), with others finding no evidence of opting out (Goldin 2006; Goldin and Katz 2008; Antecol 2011). In this paper, we focus on highly educated women and use Dale and Krueger's (2002) innovative application of selection on observables and unobservables to control for selection into colleges and provide further insight into women's labor supply decisions and the opt-out hypothesis. We emphasize the interaction between marriage and earnings and highlight various channels through which college selectivity can affect women's earnings other than through human capital directly, and we consider a wide range of family and career outcomes that adds to our understanding of the role of college selectivity along many dimensions.

Prior studies provide important evidence on the correlates of opting out, but do not address the question of how attending a more selective college affects women's family and career outcomes, or whether opting-out may be more pronounced among women who attend more selective colleges. It is possible that selection into colleges makes college selectivity appear to predict greater labor force attachment when, instead, women who attend more selective colleges differ from others along unobservable dimensions that may affect both family and career outcomes. Bertrand, Goldin, and Katz (2010) find significant increases in likelihood that female MBAs work part-time in the first

1. Belkin's 2003 *New York Times* article, cited in several academic studies, combines statistical evidence with several anecdotes of women with undergraduate and advanced degrees from elite schools who were taking time out of the labor force.

15 years after graduating, but consider opting-out to be not working at all. Antecol (2011) finds no evidence of opting-out when examining labor force participation without considering marital status or race, but finds some evidence of opting-out on the intensive margin among white, college educated, married women in male-dominated occupations. Goldin (2006) uses data from the College and Beyond survey, which we use in this paper, to show that women with advanced degrees took less time off of work than women with only a bachelors degree. She also considers opting-out to be not working at all, but does not consider the intensive margin of part-time work. Goldin and Katz (2008) consider a broader "non-full-year full-time" category, but do not find evidence that women from elite universities leave the labor force at greater rates than others.[2]

The theoretical effects of attending a more selective college on labor market outcomes are ambiguous, but we argue that the possibility of a negative effect may be more pronounced for women who are weakly attached to the labor market. As we discuss in Section 3.2, school attendance may affect marriage market outcomes by changing the pool of available partners (lower search costs for partners at the school you attend), the set of peers "competing" for the same partners, or by changing the value of marriage and therefore the quality threshold for an acceptable mate through labor market or preference effects.

The potential interaction between marital outcomes and labor market outcomes is likely to be far stronger for women than for men, and is central to the opt-out story. For example, if women who attend better schools marry men who attend better schools and who earn more, then they may reduce their labor supply because of the income effect.[3] At the same time, it is possible for the reverse to be true. If attending a more selective college makes a woman more attached to the labor force, possibly due to better career options, then she may be less invested in marriage and family or less willing to leave the labor market conditional on being married. If women with elite degrees are more likely to marry high-earning men, then the income effect would be a reason why the opt-out story could be more common for these women rather than others, but greater labor force

2. Goldin and Katz (2008) present mean labor force participation rates for 1970, 1980, and 1990 Harvard entering cohorts and and mean labor force participation rates by highest degree earned, but do not report means by cohort-degree, leaving unclear any potential interaction between them.

3. Here just the correlation between school quality and earnings matters, rather than a causal effect.

attachment due to attending a more selective college would work against this story.

Due to the theoretical ambiguity, we focus on empirical evidence to address the question of how attending a more selective college affects family and career outcomes for women. Our data set and identification approach are based on the innovative study by Dale and Krueger (2002), who account for the selection of individuals into different colleges to estimate the causal effects of college selectivity on career outcomes using data from the Andrew W. Mellon Foundation's College and Beyond survey on full-time full-year workers. The key feature of their approach is that they exploit information about applications and admissions to match and then compare individuals to others in their cohort who applied to, were accepted to, and were rejected from similar colleges. We conduct our analysis using individuals in the College and Beyond data set, without the restriction that respondents be working full-time and full-year in the market. We focus on women, unlike Dale and Krueger (2002) who use pooled samples of both sexes, but also estimate our models for men.[4] We first examine educational and marriage market effects using the full sample of women, regardless of labor force participation.

We find that women who attend more selective colleges get more years of schooling, are more likely to obtain advanced degrees, and are less likely to be married 15 years after graduating. We also find that attending a more selective colleges makes women more likely to have positive earnings, but less likely to work full-time (and more likely to work part-time) conditional on working. These results are consistent with trends in women's college attendance, graduation, and marriage rates over the past fifty years. They also suggest that women who attend more selective colleges are, in fact, more likely to reduce labor supply, although we find that this decision operates more on the intensive margin of part-time work rather than the extensive margin of labor force participation.

We also find some evidence that attending a more selective college increases women's earnings, which stands in contrast with Dale and Krueger's (2002) original results but is consistent with other studies (Brewer, Eide, and Ehrenberg 1999; Black and Smith 2004, 2006; Heim 2009; Li et al. 2012; Long 2008, 2010; Sekhri 2014). We then attempt to determine the mechanism by

4. We do not report our results for men here, but these results are available upon request.

which college selectivity affects women's earnings by considering three possibilities. We argue that women's earnings do not increase directly due to greater education. Instead, we argue that college selectivity increases women's earnings through marriage decisions. We find significant evidence that women who attend more selective colleges are more likely to be single, and that single women have higher earnings than married women, which suggests that attending a more selective college increases women's earnings by altering marriage decisions. We also find significant evidence that women who attend more selective colleges marry men who have more years of schooling and are more likely to obtain advanced degrees, suggesting that attending a more selective college improves spousal characteristics that would be rewarded in the labor market. These findings are consistent with income effects (created by stronger spousal characteristics and earning potential) reducing women's labor supply conditional on marital status.

This paper focuses on the marriage decisions, labor supply, and earnings of women who attend highly selective colleges. College tuition has grown substantially over the past fifty years, implying that the potential returns to college are pervasive given that students are still willing to expend money and effort to attend higher quality colleges.[5] In examining these potential returns to attending more selective colleges, we hope to add further insight to the rich literature studying the effects of attending more selective colleges (or attending college at all) on earnings (Brewer, Eide, and Ehrenberg 1999; Dale and Krueger 2002, 2014; Black and Smith 2004, 2006; Heim 2009; Li et al. 2012; Long 2008, 2010; Sekhri 2014) as well as studies that examine the marriage market returns to college (Lefgren and McIntyre 2006; Ge 2011; Kaufmann, Messner, and Solis 2013; Lafortune 2013; Wang 2013; Bruze 2015). We provide evidence that there are real labor market returns for women attending more selective colleges, and that these returns materialize through altered marriage decisions. The interaction of labor supply, gender, and marriage are likely more important for women than men, and may play a leading role in determining the slow-down of convergence in the pay gap and the plateau and slight decline in female labor force participation.

The remainder of the paper is organized as follows. Section 3.2 outlines our theoretical frame-

5. Financial aid may also be more generous at more selective schools despite the larger sticker price, which might change the calculation for students with greater need-based eligibility.

work, Section 3.3 details our empirical strategy, and Section 3.4 discusses the data. Section 3.5 presents our results. Finally, Section 3.6 concludes.

3.2 Theoretical Framework

The theoretical effects of attending a more selective college on labor market outcomes are ambiguous. In this section, we provide a broad overview of the possible effects of attending a more selective college on labor force participation, hours, and marriage. We also consider three hypotheses about how attending a more selective college may influence women's earnings.

It is possible that attending a more selective college increases labor force attachment among women. Further years of education can increase women's labor market returns and, therefore, decrease the value of their outside option of not working. Peer effects may also play a role. One's peers at a more selective college may be more career oriented, which could influence women to adopt similar values. Additionally, peer effects may be an important determinant of success at college, which could lead to larger labor market returns and a lower value of the outside option of not working. Finally, stronger networks at more selective schools may reduce the search costs for jobs, which would increase labor force participation. All of these factors may help explain how, in isolation, attending a more selective college may increase women's labor supply.

Similarly, school attendance may affect marriage market outcomes by changing the pool of available partners (lower search costs for partners at the school you attend), the set of peers "competing" for the same partners, or by changing the value of marriage and therefore the quality threshold for an acceptable mate through labor market or preference effects. Although attending a more selective college may increase labor force attachment in isolation, the interaction of college selectivity and marriage outcomes can create the opposite effect, which is central to the opt-out story. For example, if women who attend better schools marry men who attend better schools and who earn more, then they may reduce their labor supply because of the income effect.[6] At the

6. Here just the correlation between school quality and earnings matters, rather than a causal effect.

same time, it is possible for the reverse to be true. If attending a more selective college makes a woman more attached to the labor force, then she may be less invested in marriage and family, and if married women are less willing to leave the labor market entirely then they may choose to work part-time instead of full-time.

We also consider three hypotheses about the effect that attending a more selective college may have on women's earnings. The most common hypothesis in the literature is that attending a more selective college increases women's education, which directly increases women's earnings. We call this the "education mechanism." This is the hypothesis underlying Dale and Krueger's (2002) analysis, as well as, for instance, Brewer, Eide, and Ehrenberg (1999), Black and Smith (2004, 2006), Heim (2009), Li et al. (2012), Long (2008, 2010), and Sekhri (2014). Dale and Krueger's (2002) original results and our replication of their results already suggest that this hypothesis is unlikely to hold among the College and Beyond students.[7] Recall, however, that the population of students in the College and Beyond survey all attended relatively selective colleges, and our analysis is unlikely to be applicable to the full population of college-goers in 1976. This may be one reason why other researchers estimate positive returns to attending a more selective college on earnings whereas Dale and Krueger (2002) do not. In order to test this hypothesis, we estimate the returns to attending a more selective college on own years of schooling, whether the individual earned an advanced degree, and log earnings. If, for instance, evidence indicates that attending a more selective college increases own years of schooling and/or whether the individual earned an advanced degree, but does not affect log earnings, then we would be able to rule out this mechanism as a possibility for the effect of college selectivity on earnings.

Another hypothesis we consider is that college selectivity affects women's earnings through altered marriage decisions. We call this the "marriage mechanism." In this case, attending a more selective college may affect whether a woman marries. Spouses then bargain over how much labor to supply, which would affect women's labor supply and earnings. Therefore, attending a more selective college alters marriage decisions, which, through household bargaining, affects women's

7. Table A1 in the appendix presents our replication of Dale and Krueger's (2002) results to illustrate that we are able to replicate, to a large degree, their sample selection, matching process, and empirical specifications.

labor supply and earnings. This is one possible explanation for the opt-out story, wherein women who attended more selective colleges may be more likely to marry. The most likely outcome through household bargaining would be a reduction in female labor supply, as has been documented extensively in the literature, leading to a reduction in hours of work and/or labor force participation among women who attended more selective colleges. In order to test this hypothesis, we estimate the returns to attending a more selective college on whether the individual is married and estimate heterogeneous effects of college selectivity on earnings depending on marital status. If, for instance, evidence indicates that attending a more selective college increases women's probability of remaining single and that single women have higher earnings than married women, then this would provide support for the hypothesis that attending a more selective college increases women's earnings by decreasing the probability of marrying. This finding would be consistent with recent evidence from Miller (2011) and Kleven, Landais, and Søgaard (2018), who document significant, substantial, and persistent child earnings penalties for women. If attending a more selective college decreases women's probabilities of experiencing a child earnings penalty by increasing the likelihood of remaining single, then attending a more selective college increases women's earnings overall.

A final hypothesis we consider is that college selectivity affects women's earnings through their spouse's characteristics and earning potential. We call this the "spousal mechanism." Attending a more selective college may affect the threshold women consider for potential partners or affect the pool of potential partners through network effects. The characteristics and earning potential of spouses may then affect women's labor supply through income effects. This is another possible explanation for the opt-out story, wherein income effects due to spousal earnings (materialized through higher spousal thresholds or a stronger marriage market network) create substantial decreases in women's labor supply. In order to test this hypothesis, we estimate the returns to attending a more selective college on spousal years of schooling, whether the spouse has an advanced degree, spousal log earnings, and whether the individual has positive earnings. If, for instance, evidence indicates that attending a more selective college increases spouses' years of schooling, the probability that the spouse has an advanced degree, spouse's earnings, and the likelihood that married women have positive earnings with no effect among single women, then this would provide support for the hypothesis that income effects through spousal earnings may decrease women's earnings. This analysis may also provide insight into whether the opt-out story is largely about labor force participation or whether there may also be intensive margin effects manifesting as part-time work rather than full-time.

3.3 Empirical Strategy

We use Dale and Krueger's (2002) application of selection on observables to construct matched samples of students who applied to, were accepted to, and were rejected from similar colleges. Below we provide a brief outline of their model and empirical approach as it pertains to our analysis.

Assume that each college weighs characteristics of each applicant and uses a threshold to determine admissions decisions. Some of these characteristics are observable to the econometrician (e.g., GPA and SAT score) and some are unobservable (e.g., career ambition and ability). We can represent individual i's admissions decision at college j as:

$$\begin{cases} Z_{ij} = \gamma_1 X_{1i} + \gamma_2 X_{2i} + e_{ij} > C_{ij}, & \text{admit to college } j \\ \text{otherwise,} & \text{reject from college } j \end{cases} \tag{3.1}$$

Where Z_{ij} is the student's quality as judged by the admissions committee, X_{1i} are characteristics observable to the econometrician, X_{2i} are unobservable characteristics, γ_1 and γ_2 are weights given to each set of characteristics, e_{ij} represents idiosyncratic views of the admissions committee, and C_{ij} is the admissions threshold.

Then suppose that an individual's outcomes after college can be explained as:

$$Y_{ij} = \beta_0 + \beta_1 SAT_{j^*} + \beta_2 X_{1i} + \underbrace{\beta_3 X_{2i} + \varepsilon_{ij}}_{u_{ij}}, \tag{3.2}$$

where Y_{ij} is the individual's outcome, SAT_{j^*} is the average SAT score of students at the college the individual attended, and X_{1i} and X_{2i} are the same as in Equation 3.1. β_1 captures the effect of attending a more selective institution, which could, for example, reflect peer effects or prestige.

Since X_{2i} is unobservable, coefficient estimates of β_1 in Equation 3.2 will be biased upward with X_{2i} included in the error term, u_{ij}, if X_{2i} is positively correlated with the outcome. Appealing to Equation 3.1, X_{2i} is positively related to whether a student is accepted to college j, meaning that individuals with higher values of X_{2i} will also have higher values of u_{ij}, which upwardly biases coefficient estimates of β_1.

Dale and Krueger's (2002) innovation is to include a set of indicator variables representing groups of students who applied to, were accepted to, and were rejected from similar colleges. Students who submitted similar applications and received the same admissions decisions are likely very similar in their unobservable characteristics, X_{2i}. Including these group indicator variables will absorb at least some of the effect of unobservable characteristics of X_{2i} on outcomes.[8]

The critical assumption is that, conditional on acceptances, students enrollment decisions are uncorrelated with X_{2i} and ε_{ij}. It is possible that this assumption is more likely when considering some outcomes but not others. For example, if, conditional on acceptances, students are more likely to enroll at colleges that coincide with their marital expectations or prospects then estimates of the effects of attending a more selective college on marriage outcomes may be biased depending on the relationship between marital expectations and college choice.[9] Dale and Krueger (2002) note that even if students make enrollment decisions that are in part determined by X_{2i}, then the estimate of β_1 may still be unbiased after including group indicator variables if students applied to a fine enough range of colleges.

We make one change to Dale and Krueger's (2002) matching process by additionally requiring

8. It is also possible for this strategy to amplify bias in estimates of β_1 if X_{2i} is positively related to which college a student enrolls at conditional on acceptances. Dale and Krueger (2002), however, note that if the range of colleges students applied to are fine enough then this procedure will still produce unbiased estimates of β_1.

9. For example, if students who want to marry enroll at more selective colleges, and if wanting to marry is positively related to being married later, then our estimates of the effect of college selectivity on marriage would be positively biased. On the other hand, if students who want to marry enroll at less selective colleges then our estimates of β_1 would be negatively biased.

that students be of the same gender.[10] This additional matching consideration actually relaxes Dale and Krueger's (2002) original assumption about student unobservables because it allows for men and women with the same application and admissions decisions to differ in other unobservable ways according to gender. In practice, and following Dale and Krueger (2002), our main matching process considers colleges to be equivalent if their average SAT scores are within the same 25 point interval.

We also utilize Dale and Krueger's (2002) self-revelation model. The innovation here is that students may have a strong understanding of their own ability, and may reveal their ability through their college applications. If students with greater ability apply to colleges that require higher SAT scores on average, then including the average SAT score of the colleges to which the student applied could also control for selection into more selective colleges. However, it is also possible that the self-revelation model overstates the effect of attending a more selective college on outcomes if students with greater ability are more likely to attend more selective colleges. We estimate the self-revelation on the sample of matched-applicants, rather than on the full sample of students. The idea here is that the matched-applicant model includes only students who are similar to each other in terms of applications and admissions decisions, and by using the full sample to estimate the self-revelation model we are including students who are less similar to each other because they were not matched. Therefore, estimating the self-revelation model on the matched-applicant sample can provide additional information because the sample contains students who are more comparable to each other based on observables and unobservables.[11]

10. Dale and Krueger's (2002) matching process is not gender-specific, so that men and women could be in the same matched group.

11. Table A2 displays summary statistics of the matched applicants compared to the unmatched applicants. The applicants do not appear to differ substantially. The most notable differences are that unmatched applicants are more likely to have submitted only one application and are more likely to attend a public university.

3.4 Data

In this paper we use data from the College and Beyond survey administered by Mathematica Policy Research for the Andrew W. Mellon Foundation, which Bowen and Bok (1998) and Dale and Krueger (2002) describe in more detail. We focus here on the 1976 college entering cohort, although the survey also covers the 1951 and 1989 college entering cohorts.[12] This survey includes rich self-reported data collected between 1996–1997, when the respondents were 38 or 39 years old, combined with administrative data from 34 colleges and universities describing each individual's institutional record. We also combine these data with information provided by the Higher Education Research Institute and by the College Entrance Examination Board.

We exclude individuals who attended four historically black colleges and universities, individuals with missing income information, and individuals with missing college application information. These restrictions leave us with 9,917 women and 9,738 men. Institutional administrative data on these students includes, for example, SAT score and GPA, and was collected for every matriculant at the 30 private colleges and a subsample of the entering cohorts at the four public universities.[13] Data from the College Entrance Examination Board includes information from the Student Descriptive Questionnaire, such as high school class rank and parental income. We also use supplementary data on parental occupation and parental education from a questionnaire of college freshmen administered by the Higher Education Research Institute and the Cooperative Institutional Research Program.

The College and Beyond survey collected data on, among other things, 1995 annual earnings, occupation, demographics, education, some spousal characteristics, and satisfaction measures. It also collected information on other schools that individuals applied to, along with self-reported acceptances or rejections from those schools. Following Dale and Krueger (2002), this information, combined with data on school average SAT scores, allows us to match students together who

12. We focus on the 1976 college entering cohort because their age in 1996, when the survey was conducted, is most appropriate to study family formation and career decisions. The 1951 college entering cohort would have been 63 at this time, and the 1989 college entering cohort would have been 25.

13. At public universities, data were collected for all minority students, all varsity letter winners, all students with a combined SAT score of 1,350 or higher, and a random sample of other students.

applied to, were accepted by, and were rejected from similar schools. Our main matching process considers schools to be similar if their average SAT score is in the same 25 point interval.

We reproduce Table 1 from Dale and Krueger (2002) to illustrate how their matching process works using a hypothetical group of students. We do not include in our matched sample students who applied to only one school or students whose school set did not coincide with others' school sets. In addition to the matching conditions in Table 1 we impose the additional restriction that individuals must be of the same gender. This additional restriction does not alter our results, as we demonstrate below, but adds a further dimension of similarity between matched applicants, allowing for the possibility that women who share a school set differ in unobservable ways from men who share the same school set.

Table 2 displays weighted summary statistics from our full sample and our sample of matched applicants. For completeness, we also reproduce the available weighted summary statistics of Dale and Krueger's (2002) samples in order to highlight differences and similarities arising from their restriction to full-year full-time workers. Relaxing this restriction leads to a larger full sample that, on average, earns less, includes more women, earned slightly lower SAT scores, and submitted slightly fewer college applications. Our primary motivation is to estimate effects of attending a more selective college on both career and family outcomes, and restricting to full-year full-time workers unnecessarily restricts the sample, especially when considering outcomes such as own education, marriage, spousal characteristics, and labor force participation.[14]

Finally, Table 3 compares our replication estimates using Dale and Krueger's (2002) matching process with estimates using our additional same-gender requirement and the replication sample in order to demonstrate that our matching process does not meaningfully affect the estimates.[15,16] In-

14. Table A1 in the appendix presents our replication of Dale and Krueger's (2002) results to illustrate that we are able to replicate, to a large degree, their sample selection, matching process, and empirical specifications.

15. Estimates of the basic model, basic model with restricted sample, and self-revelation model in Table 3 should not differ because these models do not use the matched student indicator variables and do not depend on the matching process.

16. Note that there is some loss in sample size when we estimate the models using our matching process on Dale and Krueger's (2002) samples because some observations are no longer matched. For example, a woman matched to a man would be included in Dale and Krueger's (2002) sample, but would be excluded when we use our matching process on the same sample.

stead, we argue that including part-time and non-workers allows us to examine opt-out hypotheses and mechanisms.

Including more women who are part-time or non-workers is, in particular, useful in studying the effects of college selectivity on women's labor supply and investigating the opt-out hypothesis. Compared to men, women in our full sample, on average, earn less, earned slightly lower SAT scores, applied to slightly less selective colleges, were more highly ranked in their college class, and are less likely to have positive earnings. We were able to match 40.8% of individuals in our full sample to other individuals of the same gender who applied to, were accepted to, and were rejected from similar schools (40.8% of women and 40.7% of men).

It should also be noted that the population of students in the College and Beyond survey all attended relatively selective colleges.[17] Therefore, our analysis is unlikely to be applicable to the full population of college-goers in 1976. As Goldin (2006) notes, however, the College and Beyond survey is well-suited to examine the opt-out theory because this trend may be specific to or more pronounced among graduates from elite colleges, such as those in these data. The key empirical question is whether women who attended *elite* colleges are more or less likely to opt out of the labor force.

3.5 Results

We estimate the effects of college selectivity on women's family and career outcomes. We first examine the effect of attending a more selective college on education, marriage, labor force participation, and intensive margin labor supply while controlling for selection into colleges. We then estimate the effect of college selectivity on women's earnings, and we examine three hypotheses of how attending a more selective college may influence women's earnings. First, we consider the traditional hypothesis that attending a more selective college increases women's years of education, which directly affects earnings (the "education mechanism"). Second, we examine whether

17. The two lowest school average SAT scores of schools in the sample are 1020 (Dennison University) and 1038 (Pennsylvania State University).

attending a more selective colleges affects women's earnings through altered marriage decisions (the "marriage mechanism"). Finally, we examine whether attending a more selective college affects women's earnings through their spouse's characteristics and earning potential (the "spousal mechanism").

3.5.1 Education, Marriage, and Labor Supply

We first examine the effect of attending a more selective college on education, marriage, labor force participation, and intensive margin labor supply while controlling for selection into colleges. For brevity, we report estimates for the matched-applicant model and the self-revelation model, which are our main models.[18]

Table 4 presents coefficient estimates of the matched-applicant and self-revelation models.[19] After controlling for selection, we estimate that attending a college with a 100 point higher average SAT score increases women's years of schooling by 0.105–0.252 years (0.62–1.49%), increases the likelihood that a woman earns an advanced degree by 3.6–4.8 percentage points (7.09–9.45%), and decreases the likelihood that a woman marries by 3.9–4.2 percentage points (5.26–5.66%). The matched-applicant model also reveals a positive effect of college selectivity on the likelihood that the respondent has positive earnings, on the order of a 2.3 percentage point (2.77%) increase per 100 point increase in school average SAT score, but the self-revelation model shows no significant effect.[20] In addition, the matched-applicant and self-revelation model reveal no significant effect of attending a more selective college on whether women work full-time or part-time, conditional on working.[21] The point estimate from the matched-applicant model is negative, however, which is

18. Coefficient estimates for the basic models are available upon request.

19. We use weights to make the population representative of the 1976 college entering cohort at the schools in the College and Beyond survey. All matriculants at private colleges were sampled, but only a subpopulation of matriculants at the four public universities were sampled.

20. The College and Beyond survey asked respondents to report their 1995 pretax annual earnings in one of the following ten intervals: less than $1,000; $1,000-$9,999; $10,000-$19 999; $20,000-$29,999; $30,000-$49,999; $50,000-$74,999; $75,000-$100,000; $100,000-$149,999; $150,000-$199,999; and more than $200,000. Following Dale and Krueger (2002), we converted the lowest nine earnings categories to a cardinal scale by assigning values equal to the midpoint of each range, and then calculated the natural log of earnings. All individuals, therefore, have a positive value for own earnings. Therefore, we define having "positive" earnings as being in the second category of earnings or above.

21. We consider a woman to be a full-time worker if she answered "yes" to the question "Were you working full-time

consistent with women from elite colleges opting out of full-time work in favor of part-time work even though the effects are insignificant.

Our findings show that women who attend a more selective college earn more education, providing evidence of the first portion of the education mechanism, and are less likely to marry, providing evidence of the first portion of the marriage and spousal mechanisms. We examine these mechanisms in more detail below.

Our labor supply findings suggest opposing stories in relation to the opt-out hypothesis. On one hand, we find some evidence that attending a more selective college increases the likelihood that women have positive earnings, which suggests that college selectivity increases labor force attachment, and that women are opting into the labor force rather than opting out. On the other hand, although our standard errors are large, our point estimates suggest that college selectivity decreases the likelihood that women work full-time (conditional on working), suggesting that working women who attend more selective colleges are opting out of full-time work in favor of part-time work. Our measures of part-time and full-time work, however, are coarse, and any opt-out effect of college selectivity may also manifest in earnings. We now turn to the effect of college selectivity on women's earnings.

3.5.2 Earnings

Table 5 presents coefficient estimates of the matched-applicant and self-revelation models. After controlling for selection, and without considering heterogeneity, we estimate that attending a college with a 100 point higher average SAT score increases women's earnings by 8.4–13.9%.

We orient our earnings analysis by examining three hypotheses for how college selectivity may affect earnings. First, we consider the traditional hypothesis that attending a more selective college increases women's years of education, which directly affects earnings (the "education mecha-

for pay or profit during all of 1995?" Dale and Krueger (2002) restrict their sample only to individuals who answered "yes" to this question, but we include women who answered "no" as well. When examining full-time or part-time work conditional on working, we restrict the sample to women with positive earnings. Those who answered "no" to the previous question, therefore, we consider to be part-time workers.

nism"). Second, we examine whether attending a more selective college affects women's earnings through altered marriage decisions (the "marriage mechanism"). Finally, we examine whether attending a more selective college affects women's earnings through their spouse's characteristics and earning potential (the "spousal mechanism").

Hypothesis 1: The Education Mechanism

The education mechanism is the most common hypothesis considered in the literature examining the effects of college selectivity on earnings.[22] If this mechanism is true then attending a more selective college increases women's years of school and the likelihood of obtaining and advanced degree, and women's higher levels of education directly affect their earnings.

To test this hypothesis, we estimate heterogeneous effects of attending a more selective college on log earnings. The education mechanism predicts positive returns to college quality on earnings for all women through further education. Specifically, we examine whether this is the case for both married and single women.

Columns 3–4 of Table 5 present coefficient estimates of the matched-applicant and self-revelation models allowing for heterogeneity depending on marital status. We find that the positive effect of college selectivity on women's log earnings is driven by married women in the sample. We do not find a similar positive effect of attending a more selective college on log earnings for single women, casting doubt on further education being the primary mechanism for increased women's earnings. Instead, we argue that the earnings effects are driven through marriage and spousal mechanisms, described below.

Hypothesis 2: The Marriage Mechanism

The marriage mechanism is that college selectivity affects women's earnings through altered marriage decisions. In this case, attending a more selective college may affect whether a woman marries. Spouses then bargain over how much labor to supply, which would affect women's labor

22. This is the hypothesis underlying Dale and Krueger's (2002) original analysis.

supply and earnings. We have already presented estimates of the effects of attending a more selective college on the likelihood of being married and heterogenous effects on earnings depending on marital status. These two sets of results allow us to test the marriage mechanism hypothesis.

As mentioned above, we find that attending a more selective college decreases women's probabilities of marrying by 3.9–4.2 percentage points (5.26–5.66%), on average, after controlling for selection effects (columns 5–6 of Table 4). We also find that married women have lower earnings than single women (columns 3–4 of Table 5) and are less likely to have positive earnings (columns 5–6 of Table 5). At the sample average of school SAT score, single women's earnings are 1.2–1.8 times higher than married women's earnings and single women are 27.5–33.9 percentage points (33.1–40.8%) more likely to have positive earnings. Attending a more selective college also decreases the gaps between married and single women by about 5.6–5.9% for earnings and 2.8–3.5 percentage points (5.6–5.9%) for the likelihood of positive earnings per 100 point increase in school average SAT score, depending on the model. We do not find positive effects of attending a more selective college on single women's earnings after controlling for selection, and actually estimate a significant negative effects using the matched-applicant and self-revelation models. We interpret the lack of a positive effect of college selectivity on earnings for single women as consistent with Dale and Krueger's (2002) null findings of any direct effect of college selectivity on earnings.

Instead, these results suggest that attending a more selective college increases women's earnings by decreasing the probability that they marry and therefore have lower earnings. This finding is consistent with recent evidence from Miller (2011) and Kleven, Landais, and Søgaard (2018), who document significant, substantial, and persistent child earnings penalties for women. Since attending a more selective college decreases women's probabilities of experiencing a child earnings penalty by increasing the likelihood of remaining single, then attending a more selective college increases women's earnings overall. In addition these results highlight the advantage of including part-time and non-workers in our sample, unlike Dale and Krueger (2002) who include only full-year full-time workers, in order to examine the marriage mechanism and its connection to women's earnings.

Hypothesis 3: The Spousal Mechanism

The spousal mechanism is that college selectivity affects women's earnings through their spouse's characteristics and earning potential. Attending a more selective college may affect the threshold women consider for potential partners or affect the pool of potential partners through network effects. The characteristics and earning potential of spouses may then affect women's labor supply through income effects. To test this hypothesis, we combine prior findings with additional estimates of the returns to attending a more selective college on spousal years of schooling, whether the spouse has an advanced degree, and spousal log earnings. Table 6 presents our weighted least squares coefficient estimates.

We find that attending a more selective college increases spouse's years of education by 0.054–0.311 years (0.31–1.80%) and increases the likelihood that the spouse earned an advanced degree by 2.4–8.0 percentage points (4.03–13.45%) per 100 point increase in school average SAT score, on average and after controlling for selection. We find no significant effect of college selectivity on spousal log earnings after controlling for selection, with point estimates between -2.5% and 4.3%, consistent with Dale and Krueger's (2002) findings.

In examining the spousal mechanism, we also refer back to our estimates of college selectivity on whether the respondent has positive earnings. This outcome variable can provide further insight into how college selectivity affects labor force participation decisions, which examines directly the opt-out story. We find some evidence using the matched-applicant model that attending a more selective college increase the likelihood that the respondent has positive earnings (columns 7–8 of Table 4). The self-revelation model does not reveal significant coefficients even though the effect is positive.

It is also informative to refer back to our estimates of heterogeneous effects of attending a more selective college on earnings and on the likelihood the respondent has positive earnings depending on marital status. We find that married women have significantly lower earnings are significantly less likely to have positive earnings relative to single women despite the fact that, from an education standpoint, college selectivity improves spousal qualities that would be rewarded in the labor market. Our results are consistent with the spousal mechanism, which states that married women may reduce labor supply due to income effects since college selectivity improves the earning potential of their spouse.

3.5.3 Discussion

We find evidence of non-monetary effects of attending a more selective college. Women who attend more selective colleges obtain more years of schooling, are more likely to earn an advanced degree, and less likely to marry. We also find that married women who attend more selective colleges are more likely to marry a spouse with more years of education and a higher likelihood of earning an advanced degree. In relation to women's earnings, however, our findings largely confirm Dale and Krueger's (2002) results that there are no direct effects of attending a more selective college on earnings. Recall, however, that the colleges included in the College and Beyond survey are all very selective to begin with, and so our sample is not representative of the full population of college-goers in 1976. There may still be positive earnings returns to attending college relative to not attending or positive returns to attending a more selective college relative to colleges in the middle or bottom of the selectivity distribution.

Although we do not find direct effects of college selectivity on earnings, we propose two mechanisms by which marriage can affect women's earnings and we find evidence in favor of both. First, the marriage mechanism claims that college selectivity affects women's marriage decisions and may influence earnings through bargaining in marriage. We find that attending a more selective college increases the likelihood that women remain single and find that single women have higher earnings relative to married women. Our findings, therefore, suggest that college selectivity does affect women's earnings through altered marriage decisions.

Second, the spousal mechanism claims that college selectivity affects married women's spouse's characteristics, which may affect earnings through income effects to due changes in spousal earnings potential. We find that attending a more selective college increases spousal education and that married women are more likely to work part-time or not work relative to single women. Our findings, therefore, suggest that income effects due to more positive spousal characteristics may reduce female labor supply and earnings for married women. College selectivity may also affect women's earnings through the spousal mechanism by decreasing the likelihood of marrying.

Our findings are consistent with evidence from Miller (2011) and Kleven, Landais, and Søgaard (2018), who document significant, substantial, and persistent child earnings penalties for women. Since attending a more selective college decreases women's probabilities of experiencing a child earnings penalty by increasing the likelihood of remaining single, then attending a more selective college increases women's earnings overall.

The opt-out story posits that women who graduate from elite colleges are more likely to leave the labor market entirely. Our findings show that married women have significantly lower earnings and are significantly less likely to have positive earnings relative to single women, but we find that college selectivity decreases the gaps between married and single women. This suggests that attending a more selective college increases labor force attachment for married women despite lower earnings than single women, implying that opting out may be operating on the intensive margin of part-time work rather than on the participation margin.

One corollary of our analysis is that our findings also suggest that attending a more selective college increases assortative matching on education, which Geruso and Royer (2018) also find when examining compulsory schooling laws in the United Kingdom. The coefficient estimates for own and spousal years of schooling and whether the respondent or spouse obtained an advanced degree demonstrate that attending a more selective college increases both partners' levels of education. We confirm this finding in Table 7, which uses as an outcome variable the absolute difference in years of education between the respondent and their spouse. We report these coefficient estimates for the pooled sample of both women and men because any effect of college selectivity on assortative matching would be reflected among both female and male students.[23] We find that a

23. The implicit assumption here is that women generally marry men, so assortative matching on education would be reflected among both women and men since they pair together. This abstracts away from same-sex marriage, which

100 point increase in school average SAT score decreases the absolute difference between spouses' years of education by 0.014–0.123 years, on average, after controlling for selection. Greenwood et al. (2014) conclude that increases in assortative matching over time has led to greater income inequality, whereas Hryshko, Juhn, and McCue (2017) conclude that assortative matching plays only a minor role. We leave further analysis of assortative matching and how it may have changed to future work.

3.6 Conclusion and Future Work

The past half century has been characterized by women's dramatic increase in college attendance and graduation, which is widely accepted as a major contributor to the narrowing of the gender pay gap since the 1970s. Despite this trend, researchers have noted the leveling off and slight decline in female labor force participation in the United States since the mid-1990s as well as the limited relative wage gains for women over that period, speculating that elite women in particular might be "opting out" of the labor force. Researchers have also noted a steady decline in marriage rates since the 1970s. We use the College and Beyond survey and Dale and Krueger's (2002) methodology to control for selection on observables and unobservables to estimate the causal effects of attending a more selective college on these interrelated phenomena by examining women's education, marriage, labor supply, and earnings outcomes.

Our main specifications match women together who applied to, were accepted to, and were rejected from similar schools. Including indicator variables for each group of matched applicants controls for unobservable characteristics such as motivation (i.e., number of additional applications) and ability (i.e., admissions decisions) in order to at least partially control for selection into colleges based on unobservables. Dale and Krueger (2002) additionally restrict to full-year full-time workers, whereas we do not. Relaxing this restriction produces a full sample that includes more women who work part-time or do not work, making our approach more suited toward testing

was not recognized by any state at the time the College and Beyond survey was conducted in 1996, and was explicitly not recognized by the federal government under the Defense of Marriage Act.

the opt-out story. We also require that students be the same gender in order to be matched together, which allows for the possibility that men and women who apply to similar school sets may have different unobservable characteristics.

We find that attending a more selective college increases women's years of schooling, increases the likelihood that a woman obtains an advanced degree, and decreases the likelihood that women marry. We also find that positive returns of college selectivity on earnings are not the same for married and single women. This suggests that the commonly examined hypothesis that attending a more selective college increases human capital and, therefore, earnings does not appear to hold among students who attended the colleges in the College and Beyond survey.

We do, however, propose two alternative mechanisms by which college selectivity may affect earnings through marriage. First, we find that attending a more selective college increases the likelihood that women remain single. Combined with the fact that single women have significantly higher earnings than married women, this suggests that attending a more selective college does increase women's earnings through a decreased likelihood of marrying (and therefore earning less). Second, among married women we find that attending a more selective college increases spousal years of schooling and the likelihood that the spouse obtained an advanced degree. We also find no significant effects of college selectivity on spousal log earnings, but that being married is associated with a decrease in the likelihood of women having positive earnings. Taken together, these findings suggest another potential mechanism: women who attend more selective colleges marry men of greater quality from an education standpoint. Therefore, our results suggest that, conditional on marriage, college selectivity increases women's earnings because married women are more likely to reduce labor supply, perhaps due to income effects created by more positive spousal attributes that would be rewarded in the labor market. Finally, we also find evidence that attending a more selective college increases assortative matching on education.

Our results show that married women have significantly lower earnings and are significantly less likely to have positive earnings relative to single women, but we find that college selectivity decreases the gaps between married and single women. This suggests that attending a more selective college increases labor force attachment for married women despite lower earnings, implying that opting out may be operating on the intensive margin of part-time work rather than on the participation margin.

We plan to compliment our existing findings through future work. In particular, it is important to establish the determinants of enrollment decisions conditional on multiple acceptances. Dale and Krueger (2002) note that selection into more selective colleges conditional on multiple acceptances should not be an issue in the matching models so long as students applied to a fine enough range of schools. However, it is still possible that, conditional on multiple acceptances, students select into different colleges based on unobservable marriage expectations or prospects, which may bias our results.[24]

It may also be useful in future work to explore differences between the 1976 college entering cohort, which we examine in this paper, and the 1951 and 1989 college entering cohorts from the College and Beyond survey. One limitation, however, is the substantial differences in ages between the cohorts at the time of the survey, which may limit comparability between groups.

24. For example, conditional on multiple acceptances, women may sort into single-gender colleges rather than co-ed colleges based on unobservable marriage preferences or expectations. We do find larger decreases in the probability of marrying among due to college selectivity students who attending single-gender colleges, but we continue to find significant decreases in the probability of marrying even among students who attended co-ed colleges.

Table 1: Dale and Krueger (2002) - Table I: Illustration of How Matched-Applicant Groups Were Constructed

		Student applications to college							
		Application 1		Application 2		Application 3		Application 4	
Student	Matched-applicant group	School average SAT	School admissions decision	School average SAT	School admissions decision	School average SAT	School admissions decision	School average SAT	School admissions decision
Student A	1	1280	Reject	1226	Accept*	1215	Accept	na	na
Student B	1	1280	Reject	1226	Accept	1215	Accept*	1155	Accept
Student C	2	1360	Accept	1310	Reject	1270	Accept*	1155	Accept
Student D	2	1355	Accept	1316	Reject	1270	Accept*	1160	Accept
Student E	2	1370	Accept*	1316	Reject	1260	Accept	1150	Accept
Student F	Excluded	1180	Accept*	na	na	na	na	na	na
Student G	Excluded	1180	Accept*	na	na	na	na	na	na
Student H	3	1360	Accept	1308	Accept*	1260	Accept	1160	Accept
Student I	3	1370	Accept*	1311	Accept	1255	Accept	1155	Accept
Student J	3	1350	Accept	1316	Accept*	1265	Accept	1155	Accept
Student K	4	1245	Reject	1217	Reject	1180	Accept*	na	na
Student L	4	1235	Reject	1209	Reject	1180	Accept*	na	na
Student M	5	1140	Accept	1055	Accept*	na	na	na	na
Student N	5	1145	Accept*	1060	Accept	na	na	na	na
Student O	No match	1370	Reject	1038	Accept*	na	na	na	na

Notes: This table is reproduced from Dale and Krueger (2002), Table I.
* Denotes school attended.
na = did not report submitting application.
The data shown on this table represent hypothetical students. Students F and G would be excluded from the matched-applicant subsample because they applied to only one school (the school they attended). Student O would be excluded becaus no other student applied to an equivalent set of institutions.

Table 2: Summary Statistics of the College and Beyond Survey: 1976 College Entering Cohort

	Full sample				Matched-applicant sample			
	D&K statistics	All	Women	Men	D&K statistics	All	Women	Men
Log(earnings)	11.096	10.475	9.818	11.152	11.148	10.504	9.836	11.192
	(0.747)	(1.580)	(1.819)	(0.872)	(0.737)	(1.593)	(1.834)	(0.866)
Annual earnings (1995 dollars)	84,219	68,429	45,062	92,510	88,276	70,796	46,339	96,018
	(60,841)	(61,897)	(46,940)	(66,080)	(62,598)	(63,627)	(47,976)	(67,805)
Female	0.392	0.507	1.000	0.000	0.385	0.508	1.000	0.000
	(0.488)	(0.500)	(0.000)	(0.000)	(0.487)	(0.500)	(0.000)	(0.000)
Black	0.050	0.042	0.051	0.034	0.050	0.042	0.050	0.035
	(0.218)	(0.202)	(0.219)	(0.182)	(0.219)	(0.201)	(0.217)	(0.183)
Hispanic	0.013	0.011	0.010	0.011	0.014	0.012	0.012	0.012
	(0.115)	(0.102)	(0.101)	(0.103)	(0.117)	(0.107)	(0.107)	(0.107)
Asian	0.023	0.020	0.019	0.022	0.027	0.022	0.022	0.022
	(0.150)	(0.140)	(0.135)	(0.145)	(0.163)	(0.147)	(0.145)	(0.148)
Other race	0.003	0.003	0.004	0.003	0.003	0.003	0.003	0.002
	(0.059)	(0.059)	(0.059)	(0.056)	(0.057)	(0.052)	(0.054)	(0.050)
Predicted log(parental) income)	9.984	9.919	9.928	9.909	9.997	9.928	9.941	9.915
	(0.353)	(0.302)	(0.303)	(0.300)	(0.349)	(0.298)	(0.299)	(0.296)
Own SAT/100	11.672	11.170	10.933	11.414	11.875	11.439	11.170	11.717
	(1.634)	(2.795)	(2.824)	(2.744)	(1.632)	(2.649)	(2.717)	(2.547)
School average SAT/100	11.655	11.642	11.632	11.653	11.812	11.784	11.762	11.807
	(0.943)	(0.941)	(0.948)	(0.934)	(0.943)	(0.938)	(0.942)	(0.932)
Net tuition (1976 dollars)	2,454	2,455	2,422	2,489	2,651	2,640	2,609	2,672
	(1,145)	(1,147)	(1,163)	(1,130)	(1,094)	(1,105)	(1,125)	(1,083)
Log(net tuition)	7.647	7.648	7.627	7.671	7.749	7.742	7.722	7.763
	(0.622)	(0.619)	(0.634)	(0.602)	(0.582)	(0.587)	(0.606)	(0.566)
High school top 10 percent	0.418	0.395	0.393	0.397	0.427	0.405	0.392	0.419
	(0.493)	(0.489)	(0.488)	(0.489)	(0.495)	(0.491)	(0.488)	(0.493)
High school rank missing	0.356	0.388	0.398	0.377	0.355	0.380	0.403	0.355
	(0.479)	(0.487)	(0.489)	(0.485)	(0.478)	(0.485)	(0.491)	(0.479)
College athlete	0.078	0.079	0.060	0.099	0.085	0.083	0.060	0.107
	(0.268)	(0.270)	(0.237)	(0.298)	(0.279)	(0.276)	(0.238)	(0.309)
Average SAT/100 of schools applied to	11.513	11.495	11.448	11.543	11.601	11.545	11.466	11.626
	(0.940)	(0.943)	(0.943)	(0.941)	(0.991)	(1.001)	(0.998)	(0.998)
One additional application	0.225	0.234	0.239	0.228	0.490	0.566	0.576	0.556
	(0.417)	(0.423)	(0.426)	(0.420)	(0.500)	(0.496)	(0.494)	(0.497)

Notes: Means are weighted to reflect sampling procedures at public universities. Means and standard deviations from Dale and Krueger's (2002) sample are reported when available. Dale and Krueger's (2002) sample includes both men and women and includes only full-year full-time workers. Our sample includes men and women, although we largely focus on women in our analysis, and also includes part-time or non-workers.

Continued: Summary Statistics of the College and Beyond Survey: 1976 College Entering Cohort

	Full sample				Matched-applicant sample			
	D&K statistics	All	Women	Men	D&K statistics	All	Women	Men
Two additional applications	0.225	0.207	0.202	0.213	0.490	0.337	0.342	0.331
	(0.417)	(0.405)	(0.402)	(0.409)	(0.500)	(0.473)	(0.475)	(0.471)
Three additional applications	0.156	0.139	0.127	0.152	0.134	0.092	0.080	0.104
	(0.363)	(0.346)	(0.333)	(0.359)	(0.340)	(0.289)	(0.272)	(0.306)
Four additional applications	0.040	0.032	0.026	0.037	0.011	0.005	0.001	0.008
	(0.196)	(0.175)	(0.159)	(0.190)	(0.104)	(0.069)	(0.038)	(0.090)
Undergraduate percentile rank in class	50.791	51.214	53.515	48.844	51.666	52.247	54.682	49.743
	(28.267)	(28.034)	(27.118)	(28.760)	(28.268)	(28.140)	(27.079)	(28.981)
Attained advanced degree	0.542	0.515	0.482	0.549	0.573	0.545	0.508	0.582
	(0.498)	(0.500)	(0.500)	(0.498)	(0.495)	(0.498)	(0.500)	(0.493)
Graduated from college	0.839	0.836	0.834	0.839	0.862	0.855	0.850	0.859
	(0.367)	(0.370)	(0.372)	(0.367)	(0.345)	(0.352)	(0.357)	(0.348)
Public college	0.413	0.416	0.427	0.404	0.329	0.337	0.346	0.328
	(0.492)	(0.493)	(0.495)	(0.491)	(0.470)	(0.473)	(0.476)	(0.470)
Private college	0.442	0.428	0.360	0.498	0.523	0.501	0.426	0.578
	(0.497)	(0.495)	(0.480)	(0.500)	(0.500)	(0.500)	(0.494)	(0.494)
Liberal arts college	0.145	0.157	0.213	0.098	0.148	0.162	0.228	0.093
	(0.353)	(0.364)	(0.410)	(0.298)	(0.355)	(0.368)	(0.420)	(0.291)
Married		0.756	0.741	0.772		0.759	0.742	0.777
		(0.429)	(0.438)	(0.419)		(0.428)	(0.438)	(0.416)
Never married		0.129	0.133	0.126		0.129	0.132	0.125
		(0.336)	(0.339)	(0.332)		(0.335)	(0.339)	(0.331)
Divorced		0.057	0.065	0.049		0.053	0.061	0.044
		(0.233)	(0.247)	(0.217)		(0.224)	(0.240)	(0.205)
Years of education		16.972	16.825	17.125		17.132	16.954	17.315
		(2.345)	(2.275)	(2.405)		(2.302)	(2.243)	(2.347)
1(positive earnings)		0.911	0.833	0.992		0.911	0.831	0.992
		(0.285)	(0.373)	(0.089)		(0.285)	(0.374)	(0.087)
Spousal years of education		16.781	17.146	16.420		16.902	17.282	16.527
		(2.289)	(2.399)	(2.114)		(2.276)	(2.387)	(2.094)
Spousal advanced degree		0.485	0.570	0.402		0.506	0.595	0.418
		(0.500)	(0.495)	(0.490)		(0.500)	(0.491)	(0.493)
1(positive spousal earnings)		0.699	0.924	0.476		0.695	0.923	0.471
		(0.459)	(0.265)	(0.499)		(0.460)	(0.267)	(0.499)
Log(spousal earnings)		7.617	10.246	5.051		7.600	10.268	5.005
		(5.066)	(3.036)	(5.327)		(5.100)	(3.069)	(5.338)
N	14,238	19,658	9,917	9,738	6,223	8,012	4,049	3,963

Notes: Means are weighted to reflect sampling procedures at public universities. Means and standard deviations from Dale and Krueger's (2002) sample are reported when available. Dale and Krueger's (2002) sample includes both men and women and includes only full-year full-time workers. Our sample includes men and women, although we largely focus on women in our analysis, and also includes part-time or non-workers.

Table 3: Comparison of Dale and Krueger (2002) Replication and Estimates Using an Alternative Matching Process

	Basic model: full sample		Basic model: restricted sample		Matched-applicant model		Exact matched-applicant model		*Barron's* matched-applicant model		Self-revelation model	
	D&K replication	Our matches on D&K sample	D&K replication	Our matches on D&K sample	D&K replication	Our matches on D&K sample	D&K replication	Our matches on D&K sample	D&K replication	Our matches on D&K sample	D&K replication	Our matches on D&K sample
School-average SAT score/100	0.078***	0.078***	0.083***	0.083***	0.018	0.018	0.019	-0.101***	0.048**	0.050**	0.001	0.001
	(0.016)	(0.016)	(0.013)	(0.013)	(0.020)	(0.028)	(0.059)	(0.036)	(0.018)	(0.019)	(0.017)	(0.017)
Predicted log(parental income)	0.179***	0.179***	0.177***	0.177***	0.135***	0.163***	0.147**	0.217***	0.152***	0.172***	0.154***	0.154***
	(0.027)	(0.027)	(0.046)	(0.046)	(0.048)	(0.042)	(0.064)	(0.068)	(0.032)	(0.034)	(0.026)	(0.026)
Own SAT score/100	0.019***	0.019***	0.002	0.002	-0.012	-0.011	-0.014	-0.012	0.005	-0.001	0.010	0.010
	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.012)	(0.018)	(0.007)	(0.010)	(0.006)	(0.006)
Female	-0.401***	-0.401***	-0.403***	-0.403***	-0.393***	0.000	-0.388***	0.000	-0.383***	0.000	-0.394***	-0.394***
	(0.015)	(0.015)	(0.017)	(0.017)	(0.026)	(.)	(0.035)	(.)	(0.017)	(.)	(0.015)	(0.015)
Black	-0.042	-0.042	-0.065	-0.065	-0.095*	-0.070	-0.102	-0.079	-0.065*	-0.057	-0.051	-0.051
	(0.033)	(0.033)	(0.046)	(0.046)	(0.050)	(0.057)	(0.082)	(0.109)	(0.034)	(0.037)	(0.033)	(0.033)
Hispanic	0.028	0.028	0.076	0.076	0.018	0.007	0.074	0.086	0.023	0.044	0.018	0.018
	(0.056)	(0.056)	(0.083)	(0.083)	(0.085)	(0.111)	(0.174)	(0.249)	(0.070)	(0.095)	(0.055)	(0.055)
Asian	0.173***	0.173***	0.237***	0.237***	0.221***	0.190***	0.277***	0.281***	0.168***	0.119**	0.158***	0.158***
	(0.031)	(0.031)	(0.054)	(0.054)	(0.059)	(0.056)	(0.064)	(0.100)	(0.041)	(0.052)	(0.032)	(0.032)
Other/missing race	-0.133	-0.133	-0.076	-0.076	-0.100	-0.168	-0.377**	-0.378	-0.052	-0.111	-0.138	-0.138
	(0.086)	(0.086)	(0.099)	(0.099)	(0.135)	(0.162)	(0.178)	(0.363)	(0.098)	(0.079)	(0.084)	(0.084)
High school top 10 percent	0.057***	0.057***	0.091***	0.091***	0.070**	0.069**	0.063	0.087	0.072***	0.077**	0.060***	0.060***
	(0.020)	(0.020)	(0.028)	(0.028)	(0.026)	(0.031)	(0.042)	(0.071)	(0.024)	(0.030)	(0.020)	(0.020)
High school rank missing	-0.002	-0.002	0.036	0.036	0.002	-0.013	0.024	-0.001	0.020	0.033	-0.012	-0.012
	(0.025)	(0.025)	(0.030)	(0.030)	(0.038)	(0.035)	(0.045)	(0.061)	(0.025)	(0.030)	(0.023)	(0.023)
Athlete	0.114***	0.114***	0.097***	0.097***	0.080**	0.065**	0.133***	0.109	0.108***	0.096***	0.105***	0.105***
	(0.023)	(0.023)	(0.029)	(0.029)	(0.029)	(0.030)	(0.047)	(0.077)	(0.026)	(0.033)	(0.023)	(0.023)
Average SAT score/100 of schools applied to											0.087***	0.087***
											(0.013)	(0.013)
One additional application											0.062***	0.062***
											(0.012)	(0.012)
Two additional applications											0.064***	0.064***
											(0.023)	(0.023)
Three additional applications											0.118***	0.118***
											(0.025)	(0.025)
Four additional applications											0.135***	0.135***
											(0.033)	(0.033)
Adjusted R^2	0.104	0.104	0.098	0.098	0.118	0.129	0.139	0.177	0.110	0.115	0.110	0.110
N	14,238	14,238	6,200	6,200	6,200	5,647	2,994	2,127	9,717	7,192	14,238	14,238

Notes: The data come from the College and Beyond survey for the 1976 college entering cohort of both women and men. Standard errors are in parentheses and are clustered at the school-of-matriculation level. Estimates are weighted using sampling weights from the College and Beyond survey to reflect sampling procedures at the four public universities. Our matching process additionally requires that matched students by of the same gender, as opposed to Dale and Krueger (2002) who do not impose this requirement. Estimates in this table using our matching process are estimated with this additional gender requirement using the Dale and Krueger (2002) replication sample. Estimates of the basic model, basic model with restricted sample, and self-revelation model should not differ above because these models do not use the matched student indicator variables.

Table 4: The Effects of College Selectivity on Women's Education, Marriage, and Labor Supply

	Own years of education		Respondent earned advanced degree		Married		Respondent has positive earnings		Working full-time conditional on working	
	Matched-applicant model	Self-revelation model	Matched-applicant model	Self-revelation model	Matched-applicant model	Self-revelation model	Matched-applicant model	Self-revelation model	Matched-applicant model	Self-revelation model
School-average SAT score/100	0.252**	0.105	0.048***	0.036***	-0.039**	-0.042***	0.023**	0.014	-0.014	0.006
	(0.108)	(0.080)	(0.015)	(0.012)	(0.019)	(0.013)	(0.011)	(0.009)	(0.027)	(0.017)
Predicted log(parental income)	0.429**	0.466***	0.105**	0.112***	0.032	0.036	-0.081***	-0.080***	-0.052	-0.076**
	(0.200)	(0.162)	(0.040)	(0.028)	(0.032)	(0.022)	(0.027)	(0.023)	(0.036)	(0.032)
Own SAT score/100	0.214***	0.203***	0.040***	0.037***	-0.008	-0.010*	0.003	0.002	-0.005	-0.007
	(0.029)	(0.026)	(0.006)	(0.006)	(0.007)	(0.005)	(0.006)	(0.005)	(0.006)	(0.004)
Own SAT score missing	2.247***	2.053***	0.490***	0.474***	-0.179**	-0.172**	0.117	0.076	0.044	0.025
	(0.431)	(0.434)	(0.094)	(0.102)	(0.074)	(0.066)	(0.076)	(0.059)	(0.078)	(0.046)
Black	0.465**	0.464***	0.143***	0.130***	-0.248***	-0.229***	0.141***	0.116***	0.174***	0.136***
	(0.191)	(0.152)	(0.048)	(0.037)	(0.033)	(0.024)	(0.024)	(0.021)	(0.034)	(0.029)
Hispanic	0.229	0.294	-0.033	0.015	-0.040	-0.051	-0.027	-0.045	0.090	0.093**
	(0.234)	(0.272)	(0.060)	(0.053)	(0.087)	(0.062)	(0.065)	(0.060)	(0.060)	(0.042)
Asian	0.963***	0.845***	0.177***	0.139***	-0.068	-0.057	0.108***	0.096***	0.125***	0.102***
	(0.260)	(0.239)	(0.056)	(0.048)	(0.052)	(0.037)	(0.024)	(0.018)	(0.041)	(0.034)
Other/missing race	-1.612	-1.048	-0.317	-0.156	-0.299	-0.320**	0.064	0.084	0.356***	0.292***
	(1.506)	(1.262)	(0.243)	(0.165)	(0.182)	(0.153)	(0.116)	(0.078)	(0.109)	(0.020)
High school top 10 percent	0.473***	0.570***	0.072***	0.093***	0.019	0.028	0.035**	0.029*	0.017	0.019
	(0.110)	(0.133)	(0.026)	(0.028)	(0.031)	(0.022)	(0.016)	(0.015)	(0.023)	(0.020)
High school rank missing	0.202*	0.297**	0.021	0.034	0.013	0.011	0.028*	0.017	-0.038**	-0.030
	(0.112)	(0.138)	(0.025)	(0.027)	(0.030)	(0.023)	(0.016)	(0.015)	(0.017)	(0.020)
Athlete	0.538***	0.434***	0.074**	0.051*	-0.029	-0.018	0.045**	0.013	0.013	0.002
	(0.165)	(0.131)	(0.032)	(0.028)	(0.021)	(0.020)	(0.018)	(0.020)	(0.044)	(0.036)
Average SAT score/100 of schools applied to		0.227***		0.055***		0.010		0.015		0.011
		(0.052)		(0.014)		(0.008)		(0.009)		(0.011)
One additional application		-0.196		-0.160		0.008		-0.157***		0.241
		(0.829)		(0.108)		(0.159)		(0.015)		(0.181)
Two additional applications		-0.215		-0.147		0.009		-0.157***		0.227
		(0.830)		(0.110)		(0.162)		(0.015)		(0.180)
Three additional applications		0.076		-0.117		0.044		-0.145***		0.236
		(0.833)		(0.114)		(0.163)		(0.021)		(0.189)
Adjusted R^2	0.121	0.101	0.087	0.091	0.021	0.021	0.032	0.015	-0.000	0.013
N	4,049	4,049	4,049	4,049	4,049	4,049	4,049	4,049	3,381	3,381

Notes: The data come from the College and Beyond survey for the 1976 college entering cohort of women. Standard errors are in parentheses and are clustered at the school-of-matriculation level. Estimates are weighted using sampling weights from the College and Beyond survey to reflect sampling procedures at the four public universities.

Table 5: The Effects of College Selectivity on Women's Earnings

	Log earnings				Respondent has positive earnings	
	Matched-applicant model	Self-revelation model	Matched-applicant model	Self-revelation model	Matched-applicant model	Self-revelation model
School-average SAT score/100	0.139*	0.084**				
	(0.069)	(0.039)				
Married × school average SAT			0.155**	0.088*	0.026*	0.014
			(0.075)	(0.049)	(0.014)	(0.010)
Single × school average SAT			-0.023	-0.069**	-0.008	-0.014*
			(0.072)	(0.031)	(0.013)	(0.007)
Married			-3.039***	-2.791***	-0.581***	-0.504***
			(0.579)	(0.443)	(0.156)	(0.093)
Predicted log(parental income)	-0.358**	-0.368***	-0.328**	-0.334***	-0.076**	-0.073***
	(0.139)	(0.108)	(0.146)	(0.113)	(0.030)	(0.024)
Own SAT score/100	0.006	0.008	-0.002	-0.001	0.002	-0.000
	(0.035)	(0.026)	(0.033)	(0.026)	(0.006)	(0.005)
Own SAT score missing	0.451	0.395	0.279	0.224	0.084	0.043
	(0.397)	(0.263)	(0.394)	(0.256)	(0.074)	(0.056)
Black	0.817***	0.663***	0.593***	0.456***	0.098***	0.076***
	(0.144)	(0.123)	(0.142)	(0.121)	(0.023)	(0.019)
Hispanic	-0.111	-0.203	-0.163	-0.251	-0.037	-0.054
	(0.279)	(0.254)	(0.307)	(0.251)	(0.070)	(0.061)
Asian	0.797***	0.792***	0.728***	0.732***	0.095***	0.085***
	(0.141)	(0.117)	(0.141)	(0.103)	(0.024)	(0.017)
Other/missing race	0.646	0.592	0.365	0.294	0.010	0.026
	(0.631)	(0.363)	(0.656)	(0.383)	(0.117)	(0.071)
High school top 10 percent	0.262***	0.233***	0.274***	0.254***	0.037***	0.033**
	(0.094)	(0.079)	(0.086)	(0.072)	(0.014)	(0.014)
High school rank missing	0.118	0.074	0.124	0.081	0.029**	0.019
	(0.089)	(0.074)	(0.084)	(0.064)	(0.014)	(0.013)
Athlete	0.273***	0.146	0.243***	0.127	0.040*	0.009
	(0.072)	(0.103)	(0.074)	(0.103)	(0.019)	(0.020)
Average SAT score/100 of schools applied to		0.136***		0.144***		0.017*
		(0.039)		(0.038)		(0.009)
One additional application		-0.824*		-0.835*		-0.159***
		(0.421)		(0.413)		(0.025)
Two additional applications		-0.835*		-0.843*		-0.159***
		(0.427)		(0.421)		(0.025)
Three additional applications		-0.695		-0.678		-0.141***
		(0.417)		(0.403)		(0.028)
Adjusted R^2	0.046	0.031	0.095	0.079	0.074	0.058
N	4,049	4,049	4,049	4,049	4,049	4,049

Notes: The data come from the College and Beyond survey for the 1976 college entering cohort of women. Standard errors are in parentheses and are clustered at the school-of-matriculation level. Estimates are weighted using sampling weights from the College and Beyond survey to reflect sampling procedures at the four public universities.

Table 6: The Effects of College Selectivity on Spousal Characteristics

	Spouse's years of education		Spouse earned advanced degree		Spouse's log earnings	
	Matched-applicant model	Self-revelation model	Matched-applicant model	Self-revelation model	Matched-applicant model	Self-revelation model
School-average SAT score/100	0.311***	0.054	0.080***	0.024*	0.043	-0.025
	(0.076)	(0.066)	(0.017)	(0.012)	(0.031)	(0.036)
Predicted log(parental income)	0.812***	0.730***	0.160***	0.133***	0.279***	0.273***
	(0.216)	(0.143)	(0.041)	(0.030)	(0.087)	(0.051)
Own SAT score/100	0.139***	0.154***	0.030***	0.026***	-0.026	-0.019
	(0.040)	(0.035)	(0.009)	(0.007)	(0.018)	(0.013)
Own SAT score missing	1.064	1.449**	0.119	0.175	-0.342**	-0.249**
	(0.632)	(0.561)	(0.119)	(0.105)	(0.165)	(0.119)
Black	-0.423	-0.309	-0.032	-0.038	-0.177**	-0.194**
	(0.291)	(0.211)	(0.073)	(0.045)	(0.065)	(0.071)
Hispanic	-0.231	-0.449	-0.012	-0.098	0.048	0.044
	(0.428)	(0.316)	(0.128)	(0.086)	(0.129)	(0.136)
Asian	0.322	0.219	0.037	0.007	-0.111	-0.052
	(0.349)	(0.210)	(0.076)	(0.054)	(0.135)	(0.100)
Other/missing race	1.683**	1.174***	0.276	0.232	-0.322	-0.409*
	(0.817)	(0.325)	(0.222)	(0.139)	(0.342)	(0.231)
High school top 10 percent	0.113	0.058	0.033	0.025	-0.063	-0.055
	(0.115)	(0.096)	(0.029)	(0.024)	(0.048)	(0.046)
High school rank missing	0.182	0.034	0.045*	0.012	-0.039	-0.015
	(0.132)	(0.110)	(0.026)	(0.023)	(0.067)	(0.063)
Athlete	-0.035	0.097	0.016	0.033	0.004	0.072
	(0.141)	(0.110)	(0.039)	(0.038)	(0.079)	(0.053)
Average SAT score/100 of schools applied to		0.329***		0.067***		0.096***
		(0.067)		(0.010)		(0.031)
One additional application		-0.207		-0.041		-0.090
		(0.603)		(0.134)		(0.255)
Two additional applications		-0.169		-0.039		-0.051
		(0.617)		(0.137)		(0.254)
Three additional applications		-0.281		-0.055		0.009
		(0.625)		(0.140)		(0.274)
Adjusted R^2	0.094	0.084	0.082	0.070	0.047	0.031
N	2,954	2,954	2,954	2,954	2,674	2,674

Notes: The data come from the College and Beyond survey for the 1976 college entering cohort of women. Standard errors are in parentheses and are clustered at the school-of-matriculation level. Estimates are weighted using sampling weights from the College and Beyond survey to reflect sampling procedures at the four public universities.

Table 7: Assortative Matching: Effects of College Selectivity on Educational Differences Between Married Partners

	Absolute difference in years of education	
	Matched-applicant model	Self-revelation model
School-average SAT score/100	-0.123*	-0.014
	(0.061)	(0.057)
Predicted log(parental income)	-0.131	-0.171
	(0.168)	(0.105)
Own SAT score/100	-0.027	-0.018
	(0.026)	(0.019)
Own SAT score missing	-0.283	-0.258
	(0.332)	(0.246)
Black	0.211	0.275**
	(0.152)	(0.116)
Hispanic	0.325	0.255
	(0.374)	(0.233)
Asian	-0.053	-0.129
	(0.160)	(0.115)
Other/missing race	0.267	0.479
	(0.763)	(0.566)
High school top 10 percent	0.161	0.040
	(0.119)	(0.101)
High school rank missing	0.093	-0.034
	(0.104)	(0.088)
Athlete	-0.018	-0.060
	(0.103)	(0.053)
Average SAT score/100 of schools applied to		-0.053*
		(0.027)
One additional application		0.251
		(0.248)
Two additional applications		0.219
		(0.246)
Three additional applications		0.112
		(0.240)
Adjusted R^2	0.014	0.007
N	6,009	6,009

Notes: The data come from the College and Beyond survey for the 1976 college entering cohort of both women and men. Standard errors are in parentheses and are clustered at the school-of-matriculation level. Estimates are weighted using sampling weights from the College and Beyond survey to reflect sampling procedures at the four public universities. Specifications are restricted to married women and men.

3.A Appendix: Additional Tables

Table A1 displays Dale and Krueger's (2002) results along with our replication analysis. We are able to replicate, to a large degree, their sample selection, matching process, and empirical specifications.

Table A2 displays summary statistics of the matched applicants compared to the unmatched applicants. The applicants do not appear to differ substantially, alleviating some concerns that the self-revelation model estimated using the matched-applicant sample will introduce sample selection concerns.

Table A1: Replication of Dale and Krueger's (2002) Results

	Basic model: no selection controls				Matched-applicant model		Alternative matched-applicant models				Self-revelation model	
	Full sample		Restricted sample		Similar school-SAT matches		Exact school-SAT matches		*Barron's* matches			
Outcome: ln(Own Earnings)	D&K	Replication	D&K	Replication	D&K	Replication	D&K	Replication	D&K	Replication	D&K	Replication
School-average SAT score/100	0.075	0.078	0.082	0.083	-0.016	0.018	-0.106	0.019	0.004	0.048	-0.001	0.001
	(0.016)	(0.016)	(0.014)	(0.013)	(0.022)	(0.020)	(0.036)	(0.059)	(0.016)	(0.018)	(0.018)	(0.017)
Predicted log(parental income)	0.187	0.179	0.190	0.177	0.163	0.135	0.232	0.147	0.154	0.152	0.161	0.154
	(0.027)	(0.027)	(0.033)	(0.046)	(0.033)	(0.048)	(0.079)	(0.064)	(0.028)	(0.032)	(0.025)	(0.026)
Own SAT score/100	0.015	0.019	0.006	0.002	-0.011	-0.012	0.003	-0.014	-0.005	0.005	0.009	0.010
	(0.007)	(0.006)	(0.007)	(0.008)	(0.007)	(0.008)	(0.014)	(0.012)	(0.005)	(0.007)	(0.006)	(0.006)
Female	-0.405	-0.401	-0.410	-0.403	-0.395	-0.393	-0.476	-0.388	-0.400	-0.383	-0.396	-0.394
	(0.015)	(0.015)	(0.018)	(0.017)	(0.024)	(0.026)	(0.049)	(0.035)	(0.017)	(0.017)	(0.014)	(0.015)
Black	-0.048	-0.042	-0.026	-0.065	-0.057	-0.095	-0.028	-0.102	-0.057	-0.065	-0.034	-0.051
	(0.035)	(0.033)	(0.053)	(0.046)	(0.053)	(0.050)	(0.049)	(0.082)	(0.039)	(0.034)	(0.035)	(0.033)
Hispanic	0.035	0.028	0.070	0.076	0.020	0.018	-0.248	0.074	0.036	0.023	0.007	0.018
	(0.056)	(0.056)	(0.076)	(0.083)	(0.099)	(0.085)	(0.206)	(0.174)	(0.066)	(0.070)	(0.053)	(0.055)
Asian	0.168	0.173	0.245	0.237	0.241	0.221	0.368	0.277	0.163	0.168	0.155	0.158
	(0.032)	(0.031)	(0.054)	(0.054)	(0.064)	(0.059)	(0.141)	(0.064)	(0.049)	(0.041)	(0.037)	(0.032)
Other/missing race	-0.087	-0.133	-0.048	-0.076	0.060	-0.100	-0.072	-0.377	-0.050	-0.052	-0.192	-0.138
	(0.069)	(0.086)	(-0.143)	(0.099)	(0.180)	(0.135)	(0.083)	(0.178)	(0.134)	(0.098)	(0.116)	(0.084)
High school top 10 percent	0.064	0.057	0.091	0.091	0.079	0.070	0.091	0.063	0.079	0.072	0.063	0.060
	(0.020)	(0.020)	(0.022)	(0.028)	(0.026)	(0.026)	(0.032)	(0.042)	(0.024)	(0.024)	(0.019)	(0.020)
High school rank missing	0.002	-0.002	0.040	0.036	0.016	0.002	0.029	0.024	0.025	0.020	-0.009	-0.012
	(0.022)	(0.025)	(0.026)	(0.030)	(0.038)	(0.038)	(0.066)	(0.045)	(0.027)	(0.025)	(0.022)	(0.023)
Athlete	0.111	0.114	0.088	0.097	0.104	0.080	0.169	0.133	0.093	0.108	0.094	0.105
	(0.024)	(0.023)	(0.030)	(0.029)	(0.039)	(0.029)	(0.096)	(0.047)	(0.033)	(0.026)	(0.024)	(0.023)
Average SAT score/100 of schools applied to											0.090	0.087
											(0.013)	(0.013)
One additional application											0.064	0.062
											(0.011)	(0.012)
Two additional applications											0.074	0.064
											(0.022)	(0.023)
Three additional applications											0.112	0.118
											(0.028)	(0.025)
Four additional applications											0.085	0.135
											(0.027)	(0.033)
Adjusted R^2	0.100	0.104	0.110	0.098	0.112	0.118	0.142	0.139	0.106	0.110	0.113	0.110
N	14,238	14,238	6,335	6,200	6,335	6,200	2,330	2,994	9,202	9,717	14,238	14,238

Notes: The data come from the College and Beyond survey for the 1976 college entering cohort of both women and men. Standard errors are in parentheses and are clustered at the school-of-matriculation level. Estimates are weighted using sampling weights from the College and Beyond survey to reflect sampling procedures at the four public universities.

Table A2: Summary Statistics of the College and Beyond Survey: Matched vs. Unmatched Applicants

	All		Women		Men	
	Unmatched applicants	Matched applicants	Unmatched applicants	Matched applicants	Unmatched applicants	Matched applicants
Log(earnings)	10.456	10.504	9.806	9.836	11.126	11.192
	(1.572)	(1.593)	(1.809)	(1.834)	(0.875)	(0.866)
Annual earnings	66,911	70,796	44,242	46,339	90,262	96,018
(1995 dollars)	(60,714)	(63,627)	(46,248)	(47,976)	(64,856)	(67,805)
Female	0.507	0.508	1.000	1.000	0.000	0.000
	(0.500)	(0.500)	(0.000)	(0.000)	(0.000)	(0.000)
Black	0.043	0.042	0.051	0.050	0.034	0.035
	(0.202)	(0.201)	(0.220)	(0.217)	(0.181)	(0.183)
Hispanic	0.010	0.012	0.010	0.012	0.010	0.012
	(0.099)	(0.107)	(0.097)	(0.107)	(0.101)	(0.107)
Asian	0.019	0.022	0.017	0.022	0.021	0.022
	(0.136)	(0.147)	(0.128)	(0.145)	(0.143)	(0.148)
Other race	0.004	0.003	0.004	0.003	0.004	0.002
	(0.062)	(0.052)	(0.062)	(0.054)	(0.059)	(0.050)
Predicted log(parental)	9.913	9.928	9.919	9.941	9.906	9.915
income)	(0.304)	(0.298)	(0.305)	(0.299)	(0.303)	(0.296)
Own SAT/100	10.997	11.439	10.780	11.170	11.219	11.717
	(2.872)	(2.649)	(2.881)	(2.717)	(2.847)	(2.547)
School average SAT/100	11.551	11.784	11.549	11.762	11.554	11.807
	(0.932)	(0.938)	(0.942)	(0.942)	(0.923)	(0.932)
Net tuition (1976 dollars)	2,336	2,640	2,303	2,609	2,371	2,672
	(1,159)	(1,105)	(1,171)	(1,125)	(1,144)	(1,083)
Log(net tuition)	7.588	7.742	7.566	7.722	7.611	7.763
	(0.631)	(0.587)	(0.645)	(0.606)	(0.617)	(0.566)
High school top 10 percent	0.389	0.405	0.393	0.392	0.384	0.419
	(0.487)	(0.491)	(0.489)	(0.488)	(0.486)	(0.493)
High school rank missing	0.393	0.380	0.394	0.403	0.391	0.355
	(0.488)	(0.485)	(0.489)	(0.491)	(0.488)	(0.479)
College athlete	0.076	0.083	0.059	0.060	0.094	0.107
	(0.265)	(0.276)	(0.236)	(0.238)	(0.291)	(0.309)
Average SAT/100 of	11.462	11.545	11.436	11.466	11.490	11.626
schools applied to	(0.903)	(1.001)	(0.905)	(0.998)	(0.899)	(0.998)
One additional application	0.020	0.566	0.022	0.576	0.018	0.556
	(0.142)	(0.496)	(0.148)	(0.494)	(0.134)	(0.497)

Notes: Means are weighted to reflect sampling procedures at public universities. Means and standard deviations from Dale and Krueger's (2002) sample are reported when available. Dale and Krueger's (2002) sample includes both men and women and includes only full-year full-time workers. Our sample includes men and women, although we largely focus on women in our analysis, and also includes part-time or non-workers.

Continued: Summary Statistics of the College and Beyond Survey: Matched vs. Unmatched Applicants

	All		Women		Men	
	Unmatched applicants	Matched applicants	Unmatched applicants	Matched applicants	Unmatched applicants	Matched applicants
Two additional applications	0.124	0.337	0.112	0.342	0.136	0.331
	(0.330)	(0.473)	(0.316)	(0.475)	(0.343)	(0.471)
Three additional applications	0.170	0.092	0.157	0.080	0.183	0.104
	(0.375)	(0.289)	(0.364)	(0.272)	(0.387)	(0.306)
Four additional applications	0.049	0.005	0.042	0.001	0.056	0.008
	(0.215)	(0.069)	(0.200)	(0.038)	(0.230)	(0.090)
Undergraduate percentile rank in class	50.549	52.247	52.766	54.682	48.264	49.743
	(27.947)	(28.140)	(27.118)	(27.079)	(28.604)	(28.981)
Attained advanced degree	0.495	0.545	0.465	0.508	0.527	0.582
	(0.500)	(0.498)	(0.499)	(0.500)	(0.499)	(0.493)
Graduated from college	0.825	0.855	0.823	0.850	0.827	0.859
	(0.380)	(0.352)	(0.381)	(0.357)	(0.379)	(0.348)
Public college	0.466	0.337	0.478	0.346	0.452	0.328
	(0.499)	(0.473)	(0.500)	(0.476)	(0.498)	(0.470)
Private college	0.381	0.501	0.318	0.426	0.446	0.578
	(0.486)	(0.500)	(0.466)	(0.494)	(0.497)	(0.494)
Liberal arts college	0.154	0.162	0.204	0.228	0.102	0.093
	(0.361)	(0.368)	(0.403)	(0.420)	(0.302)	(0.291)
Married	0.755	0.759	0.741	0.742	0.769	0.777
	(0.430)	(0.428)	(0.438)	(0.438)	(0.421)	(0.416)
Never married	0.130	0.129	0.133	0.132	0.126	0.125
	(0.336)	(0.335)	(0.340)	(0.339)	(0.332)	(0.331)
Divorced	0.060	0.053	0.068	0.061	0.053	0.044
	(0.238)	(0.224)	(0.251)	(0.240)	(0.223)	(0.205)
Years of education	16.869	17.132	16.742	16.954	17.002	17.315
	(2.367)	(2.302)	(2.291)	(2.243)	(2.435)	(2.347)
1(positive earnings)	0.911	0.911	0.833	0.831	0.992	0.992
	(0.284)	(0.285)	(0.373)	(0.374)	(0.090)	(0.087)
Spousal years of education	16.703	16.902	17.058	17.282	16.351	16.527
	(2.295)	(2.276)	(2.403)	(2.387)	(2.124)	(2.094)
Spousal advanced degree	0.472	0.506	0.553	0.595	0.391	0.418
	(0.499)	(0.500)	(0.497)	(0.491)	(0.488)	(0.493)
1(positive spousal earnings)	0.701	0.695	0.925	0.923	0.480	0.471
	(0.458)	(0.460)	(0.263)	(0.267)	(0.500)	(0.499)
Log(spousal earnings)	7.628	7.600	10.233	10.268	5.081	5.005
	(5.044)	(5.100)	(3.015)	(3.069)	(5.321)	(5.338)
N	19,658	8,012	9,917	4,049	9,738	3,963

Notes: Means are weighted to reflect sampling procedures at public universities. Means and standard deviations from Dale and Krueger's (2002) sample are reported when available. Dale and Krueger's (2002) sample includes both men and women and includes only full-year full-time workers. Our sample includes men and women, although we largely focus on women in our analysis, and also includes part-time or non-workers.

Bibliography

Alm, James, J. Sebastian Leguizamon, and Susane Leguizamon. 2014. “Revisiting the Income Tax Effects of Legalizing Same-Sex Marriages.” *Journal of Policy Analysis and Management* 33 (2): 263–289.

Alm, James, and Leslie A. Whittington. 1995a. “Does the income tax affect marital decisions?” *National Tax Journal* 48 (4): 565–572.

———. 1995b. “Income taxes and the marriage decision.” *Applied Economics* 27 (1): 25–31.

———. 1999. “For Love or Money? The Impact of Income Taxes on Marriage.” *Economica* 66 (263): 297–316.

Antecol, Heather. 2011. “Chapter 2 the Opt-Out Revolution: Recent Trends in Female Labor Supply.” In *Research in labor economics,* 45–83.

Antecol, Heather, and Michael D. Steinberger. 2013. “Labor Supply Differences Between Married Heterosexual Women and Partnered Lesbians: A Semi-Parametric Deconomposition Approach.” *Economic Inquiry* 51 (1): 783–805.

Badgett, MV Lee. 1995. “The wage effects of sexual orientation discrimination.” *Industrial and Labor Relations Review* 48 (4): 726–739.

Bailey, Martha J, Brad Hershbein, and Amalia R Miller. 2012. "The Opt-In Revolution? Contraception and the Gender Gap in Wages." *American Economic Journal: Applied Economics* 4 (3): 225–54.

Baldwin, Alex, Michael Allgrunn, and Raymond Ring. 2011. "Does the Male-Female Partition Still Apply to Household Labor Supply?" *International Journal of Applied Economics* 8 (1): 46–54.

Bastian, Jacob, and Katherine Michelmore. 2016. "The Intergenerational Impact of the Earned Income Tax Credit on Education and Employment Outcomes." *Working Paper.*

Becker, Gary. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy* 81 (4): 813–846.

———. 1974. "A Theory of Marriage: Part II." *Journal of Political Economy* 82 (2): 511–526.

Belkin, Lisa. 2003. "The Opt-Out Revolution." The New York Times Magazine. Accessed March 16, 2018. `http://www.nytimes.com/2003/10/26/magazine/the-opt-out-revolution.html`.

Bertrand, Marianne, Claudia Goldin, and Lawrence F Katz. 2010. "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors." *American Economic Journal: Applied Economics* 2 (3): 228–55.

Bitler, Marianne P., Jonah B. Gelbach, Hilary Williamson Hoynes, and Madeline Zavodny. 2004. "The Impact of Welfare Reform on Marriage and Divorce." *Demography* 41 (2): 213–236.

Black, Dan A, and Jeffrey A Smith. 2004. "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching." *Journal of Econometrics* 121 (1-2): 99–124.

———. 2006. "Estimating the Returns to College Quality with Multiple Proxies for Quality." *Journal of Labor Economics* 24 (3): 701–728.

Black, Dan, Gary Gates, Seth Sanders, and Lowell Taylor. 2007. "The Measurement of Same-Sex Unmarried Partner Couples in the 2000 U.S. Census." *California Center for Population Research On-Line Working Paper Series.*

Blau, Francine D, and Lawrence M Kahn. 1997. "Swimming Upstream: Trends in the Gender Wage Differential in the 1980s." *Journal of labor Economics* 15 (1, Part 1): 1–42.

———. 2006. "The US Gender Pay Gap in the 1990s: Slowing Convergence." *Industrial and Labor Relations Review* 60 (1): 45–66.

———. 2007. "Changes in the labor supply behavior of married women: 1980–2000." *Journal of Labor Economics* 25 (3): 393–438.

Bowen, William G, and Derek Bok. 1998. *The Shape of the River.* Princeton, NJ: Princeton University Press.

Brewer, Dominic J, Eric R Eide, and Ronald G Ehrenberg. 1999. "Does it Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings." *Journal of Human Resources:* 104–123.

Brien, Michael J, Lee A Lillard, and Steven Stern. 2006. "Cohabitation, marriage, and divorce in a model of match quality." *International Economic Review* 47 (2): 451–494.

Bruze, Gustaf. 2015. "Male and Female Marriage Returns to Schooling." *International Economic Review* 56 (1): 207–234.

Buchmueller, Thomas, and Christopher S Carpenter. 2010. "Disparities in health insurance coverage, access, and outcomes for individuals in same-sex versus different-sex relationships, 2000–2007." *American Journal of Public Health* 100 (3): 489–495.

Card, David, and Lara D Shore-Sheppard. 2004. "Using discontinuous eligibility rules to identify the effects of the federal medicaid expansions on low-income children." *Review of Economics and Statistics* 86 (3): 752–766.

Carpenter, Christopher S. 2007. "Revisiting the income penalty for behaviorally gay men: Evidence from NHANES III." *Labour economics* 14 (1): 25–34.

Chetty, Raj. 2009. "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods." *Annual Review of Economics* 1 (1): 451–488.

———. 2012. "Bounds on Elasticites with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica* 80 (3): 969–1018.

Chetty, Raj, John N Friedman, and Emmanuel Saez. 2013. "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *The American Economic Review* 103 (7): 2683–2721.

Chetty, Raj, and Emmanuel Saez. 2013. "Teaching the tax code: Earnings responses to an experiment with EITC recipients." *American Economic Journal: Applied Economics* 5 (1): 1–31.

Chiappori, Pierre-Andre. 1992. "Collective Labor Supply and Welfare." *Journal of Political Economy* 100 (3): 437–467.

Chiappori, Pierre-Andre, Bernard Fortin, and Guy Lacroix. 2002. "Marriage market, divorce legislation, and household labor supply." *Journal of Political Economy* 110 (1): 37–72.

Cox, Amanda. 2015. "Tax Day: Are You Receiving a Marriage Penalty or Bonus?" New York Times. Accessed August 24, 2017. `https://www.nytimes.com/interactive/2015/04/16/upshot/marriage-penalty-couples-income.html?mcubz=0`.

Crossley, Thomas F., and Sung-Hee Jeon. 2007. "Joint Taxation and the Labor Supply of Married Women: Evidence from the Canadian Tax Reform of 1988." *Fiscal Studies* 28 (3): 343–365.

Dahl, Gordon B, and Lance Lochner. 2012. “The impact of family income on child achievement: Evidence from the earned income tax credit.” *The American Economic Review* 102 (5): 1927–1956.

Dale, Stacy Berg, and Alan B Krueger. 2002. “Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables.” *The Quarterly Journal of Economics* 117 (4): 1491–1527.

———. 2014. “Estimating the Effects of College Characteristics Over the Career Using Administrative Earnings Data.” *Journal of Human Resources* 49 (2): 323–358.

DeSilver, Drew. 2013. “How many same-sex marriages in the U.S.? At least 71,165, probably more.” Pew Research Center. Accessed April 4, 2017. `http://www.pewresearch.org/fact-tank/2013/06/26/how-many-same-sex-marriages-in-the-u-s-at-least-71165-probably-more/`.

Dickert-Conlin, Stacy. 1999. “Taxes and Transfers: Their Effect on the Decision to End a Marriage.” *Journal of Public Economics* 73 (2): 217–240.

Dickert-Conlin, Stacy, and Scott Houser. 2002. “EITC and Marriage.” *National Tax Journal* 55 (1): 25–40.

Eissa, Nada, and Hilary Williamson Hoynes. 2000. “Tax and transfer policy, and family formation: Marriage and cohabitation.” *Unpublished Manuscript. University of California, Berkeley, CA.*

———. 2003. “Good News for Low Income Families? Tax-transfer Schemes and Marriage.” *CESifo Venice Summer Institute 2003; Workshop on Taxation and the Family* (July).

———. 2004. “Taxes and the labor market participation of married couples: the earned income tax credit.” *Journal of Public Economics* 88 (9): 1931–1958.

Eissa, Nada, and Jeffrey B. Liebman. 1996. "Labor Supply Response to the Earned Income Tax Credit." *The Quarterly Journal of Economics* 111 (2): 605–637.

Ellwood, David T. 2000. "The Impact of the Earned Income Tax Credit and Social Policy Reforms on Work, Marriage, and Living Arrangements." *National Tax Journal* 53 (4): 1063–1106.

Feenberg, Daniel Richard, and Elizabeth Coutts. 1993. "An Introduction to the TAXSIM Model." *Journal of Policy Analysis and Management* 12 (1): 189–194.

Feldstein, Martin. 1999. "Tax avoidance and the deadweight loss of the income tax." *The Review of Economics and Statistics* 81 (4): 674–680.

Fetter, Daniel K., and Lee M. Lockwood. 2017. "Government Old-Age Support and Labor Supply: Evidence from the Old Age Assistance Program." *Working Paper.*

Fisher, Robin, Geof Gee, and Adam Looney. 2016. "Joint Filing by Same-Sex Couples after Windsor: Characteristics of Married Filers in 2013 and 2014." *Office of Tax Analysis Working Paper 108.*

Gao, Huasheng, and Wei Zhang. 2016. "Employment Nondiscrimination Acts and Corporate Innovation." *Management Science.*

Gates, Gary J., and Michael D. Steinberger. 2010. "Same-Sex Unmarried Partner Couples in the American Community Survey: The Role of Misreporting, Miscoding and Misallocation." *Working Paper.*

Gayle, George-Levi, and Andrew Shephard. 2016. "Optimal Taxation, Marriage, Home Production, and Family Labour Supply." *Working Paper.*

Ge, Suqin. 2011. "Women's College Decisions: How Much Does Marriage Matter?" *Journal of Labor Economics* 29 (4): 773–818.

Gemici, Ahu, and Steve Laufer. 2014. "Marriage and cohabitation." *Working Paper. New York University.*

Geruso, Michael, and Heather Royer. 2018. "The Impact of Education on Family Formation: Quasi-Experimental Evidence from the UK."

Goldin, Claudia. 2006. "The Quiet Revolution that Transformed Women's Employment, Education, and Family." *American Economic Review: Papers and Proceedings (Richard T. Ely Lecture)* 96 (2): 1–21.

Goldin, Claudia, and Lawrence F Katz. 2008. "Transitions: Career and Family Life Cycles of the Educational Elite." *American Economic Review* 98 (2): 363–69.

Gonzales, Gilbert, and Lynn A Blewett. 2014. "National and state-specific health insurance disparities for adults in same-sex relationships." *American Journal of Public Health* 104 (2): e95–e104.

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos. 2014. "Marry Your Like: Assortative Mating and Income Inequality." *American Economic Review, Papers and Proceedings* 104 (5): 348–53.

Gruber, Jon, and Emmanuel Saez. 2002. "The elasticity of taxable income: evidence and implications." *Journal of Public Economics* 84 (1): 1–32.

Heim, Bradley T. 2007. "The incredible shrinking elasticities married female labor supply, 1978–2002." *Journal of Human Resources* 42 (4): 881–918.

———. 2009. "Structural estimation of family labor supply with taxes estimating a continuous hours model using a direct utility specification." *Journal of Human Resources* 44 (2): 350–385.

Herbst, Chris M. 2011. "The Impact of the Earned Income Tax Credit on Marriage and Divorce: Evidence from Flow Data." *Population Research and Policy Review* 30 (1): 101–128.

Herr, Jane Leber, and Catherine D Wolfram. 2012. "Work Environment and Opt-Out Rates at Motherhood Across High-Education Career Paths." *Industrial and Labor Relations Review* 65 (4): 928–950.

Hersch, Joni. 2013. "Opting out among women with elite education." *Review of Economics of the Household* 11 (4): 469–506.

Hoynes, Hilary W, and Ankur J Patel. 2015. "Effective policy for reducing inequality? The earned income tax credit and the distribution of income." *National Bureau of Economic Research Working Paper No. 21340.*

Hoynes, Hilary, Doug Miller, and David Simon. 2015. "Income, the earned income tax credit, and infant health." *American Economic Journal: Economic Policy* 7 (1): 172–211.

Hryshko, Dmytro, Chinhui Juhn, and Kristin McCue. 2017. "Trends in earnings inequality and earnings instability among U.S. couples: How important is assortative matching?" *Labour Economics* 48:168–182.

Immervoll, Herwig, Henrik Jacobsen Klevin, Claus Thustrup Kreiner, and Emmanuel Saez. 2007. "Welfare Reform in European Countries: A Microsimulation Analysis." *The Economic Journal* 117 (516): 1–44.

Isaac, Elliott. 2017. "Marriage, Divorce, and Tax and Transfer Policy." *Working Paper.*

Kalíšková, Klára. 2014. "Labor supply consequences of family taxation: Evidence from the Czech Republic." *Labour Economics* 30:234–244.

Kaufmann, Katja Maria Maria, Matthias Messner, and Alex Solis. 2013. "Returns to Elite Higher Education in the Marriage Market: Evidence from Chile." *Working Paper.*

Keane, Michael P. 2011. "Labor Supply and Taxes: A Survey." *Journal of Economic Literature* 49 (4): 961–1075.

Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez. 2009. "The optimal income taxation of couples." *Econometrica* 77 (2): 537–560.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2018. "Children and Gender Inequality: Evidence from Denmark."

Kreider, Rose M, and Daphne A Lofquist. 2015. "Matching survey data with administrative records to evaluate reports of same-sex married couple households." *Washington, DC: US Census Bureau.*

Lafortune, Jeanne. 2013. "Making Yourself Attractive: Pre-Marital Investments and the Returns to Education in the Marriage Market." *American Economic Journal: Applied Economics* 5 (2): 151–78.

LaLumia, Sara. 2008. "The effects of joint taxation of married couples on labor supply and non-wage income." *Journal of Public Economics* 92 (7): 1698–1719.

Lefgren, Lars, and Frank McIntyre. 2006. "The Relationship Between Women's Education and Marriage Outcomes." *Journal of Labor Economics* 24 (4): 787–830.

Li, Hongbin, Lingsheng Meng, Xinzheng Shi, and Binzhen Wu. 2012. "Does Attending Elite Colleges Pay in China?" *Journal of Comparative Economics* 40 (1): 78–88.

Light, Audrey, and Yoshiaki Omori. 2008. "Economic Incentives and Family Formation." *Working Paper.*

Lofquist, Daphne A. 2015. "Using Names to Improve Measurement of Same-sex Married Couples in the American Community Survey." *Washington, DC: US Census Bureau.*

Long, Mark C. 2008. “College Quality and Early Adult Outcomes.” *Economics of Education Review* 27 (5): 588–602.

———. 2010. “Changes in the Returns to Education and College Quality.” *Economics of Education Review* 29 (3): 338–347.

McClelland, Robert, Shannon Mok, and Kevin Pierce. 2014. “Labor Force Participation Elasticities of Women and Secondary Earners within Married Couples.” *Congressional Budget Office Working Paper 2014-06* (September).

Meyer, Bruce D., and Dan T. Rosenbaum. 2001. “Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers.” *Quarterly Journal of Economics* 116 (3): 1063–1114.

Michelmore, Katherine. 2015. “The Earned Income Tax Credit and Union Formation: The Impact of Expected Spouse Earnings.” *Working paper.*

Miller, Amalia R. 2011. “The Effects of Motherhood Timing on Career Path.” *Journal of Population Economics* 24 (3): 1071–1100.

Moffitt, Robert A. 2003. “The Temporary Assistance for Needy Families Program.” *Means-Tested Transfer Programs in the United States:* 291–364.

Moulton, Jeremy G., Alexandra Graddy-Reed, and Lauren Lanahan. 2016. “Beyond the EITC: The Effect of Reducing the Earned Income Tax Credit on Labor Force Participation.” *National Tax Journal* 69 (2): 261–284.

Neal, Derek A., and William R. Johnson. 1996. “The Role of Premarket Factors in Black-White Wage Differences.” *Journal of Political Economy* 104 (5): 869–895.

Nichols, Austin, and Jesse Rothstein. 2015. “The Earned Income Tax Credit (EITC).” *National Bureau of Economic Research Working Paper No. 21211.*

Oreffice, Sonia. 2011. "Sexual orientation and household decision making, Same-sex couples' balance of power and labor supply choices." *Labour Economics* 18 (2): 145–158.

Parker, Kim, and Renee Stepler. 2017. "As U.S. marriage rate hovers at 50%, education gap in marital status widens." Pew Research Center. Accessed March 25, 2018. http://www.pewresearch.org/fact-tank/2017/09/14/as-u-s-marriage-rate-hovers-at-50-education-gap-in-marital-status-widens/.

Plug, Erik, Dinand Webbink, and Nick Martin. 2014. "Sexual orientation, prejudice, and segregation." *Journal of Labor Economics* 32 (1): 123–159.

Romich, Jennifer L., and Thomas Weisner. 2000. "How Families View and Use the EITC: Advance Payment Versus Lump Sum Delivery." *National Tax Journal* 53 (4, Part 2): 1245–1266.

Rosen, Harvey S. 1977. "Is it time to abandon joint filing?" *National Tax Journal:* 423–428.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2017. "Integrated Public Use Microdata Series: Version 7.0 [dataset]." *University of Minnesota, Minneapolis.* https://doi.org/10.18128/D010.V7.0.

Saez, Emmanuel. 2016. "Taxing the Rich More: Preliminary Evidence from the 2013 Tax Increase." *NBER Working Paper No. 22798.*

Saez, Emmanuel, Joel Slemrod, and Seth H. Giertz. 2012. "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *Journal of Economic Literature* 50 (1): 3–50.

Schoeni, Robert F., and Rebecca M. Blank. 2003. "What Has Welfare Reform Accomplished? Impacts on Welfare Participation, Employment, Income, Poverty, and Family Structure." *PSC Research Report.*

Sekhri, Sheetal. 2014. "Prestige Matters: Value of Connections Formed in Elite Colleges." *Working Paper.*

Selin, Håkan. 2014. "The rise in female employment and the role of tax incentives. An empirical analysis of the Swedish individual tax reform of 1971." *International Tax and Public Finance* 21 (5): 894–922.

Sheran, Michelle. 2007. "The career and family choices of women: A dynamic analysis of labor force participation, schooling, marriage, and fertility decisions." *Review of Economic Dynamics* 10 (3): 367–399.

Stern, Nicholas. 1986. "On the specification of labour supply functions." In *Unemployment, Search and Labour Supply,* edited by Richard Blundell and Ian Walker, 143–189. Cambridge: Cambridge University Press.

Stevenson, Adam. 2012. "The Labor Supply and Tax Revenue Consequences of Federal Same-Sex Marriage Legalization." *National Tax Journal* 65 (4): 783–806.

Story, Louise. 2005. "Many Women at Elite College Set Career Path to Motherhood." The New York Times. Accessed March 16, 2018. `http://www.nytimes.com/2005/09/20/us/many-women-at-elite-colleges-set-career-path-to-motherhood.html`.

Tebaldi, Edinaldo, and Bruce Elmslie. 2006. "Sexual orientation and labour supply." *Applied Economics* 38 (5): 549–562.

Teitler, Julien O., Nancy E. Reichman, Lenna Nepomnyaschy, and Irwin Garfinkel. 2009. "Effects of Welfare Participation on Marriage." *Journal of Marriage and Family* 71 (4): 878–891.

United States General Accounting Office. 1992. "Earned Income Tax Credit: Advance Payment Option is Not Widely Known or Understood by the Public." GAO/GGD-92-26 (February).

U.S. Census Bureau. 2009. *History: 2000 Census of Population and Housing.* Vol. 1. U.S. Government Printing Office.

U.S. Congress. 1990. "Omnibus Budget Reconciliation Act of 1990": 408–415.

———. 2004. *Background Material and Data on the Programs within the Jurisdiction of the Committee on Ways and Means (Green Book).* Committee on Ways / Means.

U.S. Department of the Treasury, Internal Revenue Service. 2013. "Revenue Ruling 2013-17."

———. 2016. "Table 1.2. All Returns: Adjusted Gross Income, Exemptions, Deductions, and Tax Items, by Size of Adjusted Gross Income and by Marital Status, Tax Year 2014 (Filing Year 2015)." *IRS, Statistics of Income Division, Publication 1304, August 2016.* Accessed October 29, 2017. `https://www.irs.gov/pub/irs-soi/14in12ms.xls`.

Wang, Shing-Yi. 2013. "Marriage Networks, Nepotism, and Labor Market Outcomes in China." *American Economic Journal: Applied Economics* 5 (3): 91–112.

Whittington, Leslie A., and James Alm. 1997. "'Till Death or Taxes Do Us Part: The Effect of Income Taxation on Divorce." *Journal of Human Resources* 32 (2): 388–412.