

**CDEP-CGEG WORKING PAPER SERIES**

CDEP-CGEG WP No. 68

**The Effect of Fertility on Mothers' Labor  
Supply over the Last Two Centuries**

Daniel Aaronson, Rajeev Dehejia, Andrew Jordan,  
Cristian Pop-Eleches, Cyrus Samii, and Karl Schulze

December 2018

# **The Effect of Fertility on Mothers' Labor Supply over the Last Two Centuries**

**Daniel Aaronson, Rajeev Dehejia, Andrew Jordan, Cristian Pop-Eleches, Cyrus Samii, and**

**Karl Schulze**

**February 2017**

## **Abstract**

This paper documents the evolving impact of childbearing on the work activity of mothers between 1787 and 2014. It is based on a compiled data set of 429 censuses and surveys, representing 101 countries and 46.9 million mothers, using the International and U.S. IPUMS, the North Atlantic Population Project, and the Demographic and Health Surveys. Using twin births (Rosenzweig and Wolpin 1980) and same gendered children (Angrist and Evans 1998) as instrumental variables, we show three main findings: (1) the effect of fertility on labor supply is small and often indistinguishable from zero at low levels of income and large and negative at higher levels of income; (2) these effects are remarkably consistent both across time looking at the historical time series of currently developed countries and at a contemporary cross section of developing countries; and (3) the results are robust to other instrument variation, different demographic and educational groups, rescaling to account for changes in the base level of labor force participation, and a variety of specification and data decisions. We show that the negative gradient in female labor supply is consistent with a standard labor-leisure model augmented to include a taste for children. In particular, our results appear to be driven by a declining substitution effect to increasing wages that arises from changes in the sectoral and occupational structure of female jobs into formal non-agricultural wage employment as countries develop.

---

Aaronson: Federal Reserve Bank of Chicago ([daaronson@frbchi.org](mailto:daaronson@frbchi.org)); Dehejia: New York University ([rajeev@dehejia.net](mailto:rajeev@dehejia.net)); Jordan: University of Chicago; Pop-Eleches: Columbia University ([cp2124@columbia.edu](mailto:cp2124@columbia.edu)); Samii: New York University ([cgs2083@nyu.edu](mailto:cds2083@nyu.edu)); Schulze: Federal Reserve Bank of Chicago ([kschulze@frbchi.org](mailto:kschulze@frbchi.org)). We thank seminar participants at various universities and conferences and especially Quamrul Ashraf for invaluable suggestions. The views expressed in this paper are not necessarily those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

## **I. Introduction**

The relationship between fertility and female labor supply has been widely studied in economics. For example, the link between family size and mother's work decisions has been used to explain household time allocation and the evolution of women's labor supply, particularly among rapidly growing countries in the second half of the 20<sup>th</sup> century (Carlinger, Robinson, and Tomes 1980; Angrist and Evans 1998; Del Boca, Pasqua, and Pronzata 2005; Cristia 2008; Bruijns 2014; and Hupkau and Leturcq 2016). Moreover, development economists have related the fertility-work relationship to the demographic transition and studied its implications for economic growth (Bloom, Canning, and Sevilla 2001). Yet despite the centrality of these issues in the social sciences, there is no unified evidence on how this relationship has evolved over time and with the process of economic development.

Our contribution is to provide such evidence that spans not only a plausible cross-section of countries at various stages of development but also historical examples from currently developed countries going back to the late 18<sup>th</sup> century. To provide consistent estimates over time and space, we use two common instrumental variables strategies: (i) twin births introduced by Rosenzweig and Wolpin (1980) and applied repeatedly since (e.g., Bronars and Grogger 1994; Black, Devereux, and Salvanes 2005; and Caceres-Delpiano 2006) and (ii) the gender composition of the first two children introduced by Angrist and Evans (1998). We implement these estimators using four large databases of censuses and surveys: the Integrated Public Use Micro Sample (IPUMS) International (200 country-years), the IPUMS U.S. (15 U.S. censuses and 4 Puerto Rican censuses), the North Atlantic Population Project (18 country-years), and the Demographic and Health Surveys (192 low-income country-years). Together, the data covers

429 country-years, and 46.9 million mothers, stretching from 1787 to 2014 and, consequently, a large span of economic development.

A natural starting point in thinking about the fertility-labor supply relationship is Angrist and Evans (1998). Based on U.S. IPUMS data from 1980 and 1990, Angrist and Evans find a negative effect of fertility on female labor supply using both gender mix and twin births as instruments for subsequent children, a result also established by Bronars and Grogger (1994).<sup>1</sup> Alternative instruments that rely on childless mothers undergoing infertility treatments in the U.S. and Denmark (Cristia 2008 and Lundborg, Plug, and Rasmussen 2016) or natural experiments like the introduction of birth control pills (Bailey 2013) or changes in abortion legislation (Bloom et al. 2009 and Angrist and Evans 1996) similarly conclude that children have a negative effect on their mother's labor supply or earnings. This instrument-invariant robustness is particularly notable since each IV uses a somewhat different subpopulation of compliers to estimate a local average treatment effect. That the results are consistent suggests fairly wide external validity (Angrist, Lavy, and Schlosser 2010 and Bisbee et al. 2015).

However, we show that the negative relationship between fertility and mother's work behavior holds only for countries at a later stage of economic development. At a lower level of income, including the U.S. and Western European countries prior to WWII, there is no relationship between fertility and mother's labor supply. The lack of an impact at low levels of development is in-line with Agüero and Marks' (2008, 2011) study of childless mothers undergoing infertility treatments in 32 developing countries and Godefroy's (2016) analysis of changes to women's legal rights in Nigeria. Strikingly, combining U.S historical data with data

---

<sup>1</sup> For discussions of the validity of various fertility instruments, see for example Rosenzweig and Wolpin (2000), Hoekstra et al. (2007), Angrist, Lavy, and Schlosser (2010), and Bhalotra and Clarke (2016). Clarke (2016) provides a useful summary of the empirical literature.

from a broad set of contemporary developing countries, we find that the negative gradient of the fertility-labor supply effect with respect to economic development is remarkably consistent across time and space. That is, women in the U.S. at the turn of the 20<sup>th</sup> century make the same labor supply decision in response to additional children as women in developing countries today. Moreover, we show that the negative gradient is exceedingly robust to a wide range of data issues, including using alternative instruments, changing the sample specification, conditioning on covariates, using alternative measures of mother's labor supply, and rescaling the estimates to account for varying rates of labor force participation over time.

The empirical regularities we describe are consistent with a standard labor-leisure model augmented to include a taste for children. As wages increase during the process of development, households face an increased time cost of fertility but also experience increased income. With a standard constant elasticity of substitution utility function, the former effect dominates as countries develop, creating a negative gradient.

Indeed, in exploring the mechanism behind our result, we document that the income effect from rising wages is invariant to economic development but the substitution effect falls from zero to negative and becomes economically important as real GDP per capita increases. We show that the patterns in the data cannot be explained by factors that might impact the cost of childcare as societies become wealthier, such as the declining presence of family members in or near the household or increased access to education. We instead argue that the declining substitution effect arises from changes in the sectoral and occupational structure of female jobs, as in Schultz (1991). In particular, as economies evolve, women's labor market opportunities transition from agricultural and self-employment jobs to urban wage work. Such wage work tends to be less compatible with raising children and causes some movement out of the labor

force (e.g. Jaffe and Azumi 1960; McCabe and Rosenzweig 1976; Kupinsky 1977; Goldin 1990; Galor and Weil 1996; Edwards and Field-Hendrey 2002; and Szulga 2013). We further show that the negative gradient is steeper among those in non-professional occupations that might have less opportunity to find alternative childcare.

Our key result – of a negative gradient to the fertility-labor supply relationship, ranging from roughly zero at low levels of development to significantly negative at higher income levels – has important implications both for understanding the historical evolution of women’s labor supply and the relationship between demographic transition and the process of economic development. As Goldin (1990) has documented in her comprehensive study of women’s work in the 20<sup>th</sup> century, women’s labor supply follows a U-shape over the process of economic growth, first declining before eventually increasing. Our results suggest that declining fertility rates as incomes increase may have contributed to the upswing in women’s labor supply in much of the developed world during the second half of the century. The results also motivate recent studies of the role of family policies (Olivetti and Petrolgolo 2017) and childcare costs (Del Boca 2015; Herbst 2015; and Kubota 2016) on female labor supply. At the other end of the economic development spectrum, our results suggest that the demographic transition to smaller families probably does not have immediate implications for women’s labor supply and growth. This in turn reinforces the idea suggested in the demographic transition literature (e.g. Bloom, Canning, and Sevilla 2001) that family planning policies are unlikely to enhance growth through a labor supply channel (although such policies could still be desirable for other reasons).

Our paper is organized as follows. We begin by sketching a model highlighting the key mechanism driving fertility’s impact on labor supply. Section III explains our empirical strategy, followed in section IV by a description of the data. Section V presents our findings, along with a

series of robustness checks. Section VI analyzes potential channels for our results, and section VII briefly concludes.

## II. Sketch of a Model

We believe many of the empirical patterns in the data can be explained within a standard labor-leisure model. In particular, consider a constant elasticity of substitution (CES) utility function defined over consumption  $c$ , leisure  $d$ , and fertility  $n$ :

$$(1) \quad U(c, d, n) = \left[ \gamma(c + c_0)^\rho + \alpha d^\rho + \beta \left(\frac{n}{N}\right)^\rho \right]^{1/\rho}$$

where  $c_0 < 0$  is subsistence consumption and utility from fertility is relative to  $N$ , the potential reproductive capacity. Equation (1) is a CES variant of the model used by Bloom et al. (2009). Total time (normalized to 1) is allocated between leisure  $d$ , childcare  $bn$  (where  $b$  is the time cost per child), labor  $l$ , and non-market household work  $\varepsilon$ :

$$(2) \quad 1 = l + d + bn + \varepsilon$$

Assuming households do not save, consumption is derived directly from earned income:

$$(3) \quad c = wl.$$

Substituting equations (2) and (3) into (1), we obtain the household utility function:

$$(4) \quad V(l, n) = \left[ \gamma(wl + c_0)^\rho + \alpha(1 - l - bn - \varepsilon)^\rho + \beta \left(\frac{n}{N}\right)^\rho \right]^{1/\rho}.$$

The first order conditions are:

$$(5) \quad \frac{\partial V}{\partial l} = \frac{1}{\rho} v^{\left(\frac{1}{\rho}-1\right)} [\rho \gamma w (wl + c_0)^{\rho-1} - \alpha \rho (1 - l - bn - \varepsilon)^{\rho-1}] = 0$$

$$\partial V / \partial n = \frac{1}{\rho} v^{\left(\frac{1}{\rho}-1\right)} [-\alpha \rho b (1-l-bn-\varepsilon)^{\rho-1} + \beta \rho N^{-\rho} n^{\rho-1}] = 0$$

where  $v \equiv \left[ \gamma (wl + c_0)^\rho + \alpha (1-l-bn-\varepsilon)^\rho + \beta \left( \frac{n}{N} \right)^\rho \right]$ . Re-arranging yields:

(6)

$$l = \frac{(\alpha^\theta - \alpha^\theta \varepsilon - w^\theta \gamma^\theta c_0) - \alpha^\theta b n}{w^{\theta+1} \gamma^\theta + \alpha^\theta}$$

$$n = \frac{\alpha^\theta b^\theta (1 - \varepsilon - l)}{\beta^\theta N^{-\rho\theta} + \alpha^\theta b^{\theta+1}},$$

where  $\theta \equiv 1/(\rho - 1)$ . Note that in the solution:

(7)

$$\frac{\partial l}{\partial n} = - \frac{\alpha^\theta b}{w^{\theta+1} \gamma^\theta + \alpha^\theta} < 0$$

and  $\partial^2 l / \partial n \partial w < 0$  if  $\rho \in (0,1)$  or the elasticity of substitution is between  $(0,\infty)$ . In other words, the model predicts that the effect of fertility on labor supply is negative and becomes more negative as the wage increases. As the wage increases, the agent experiences both a substitution and income effect. The former arises because an increase in the wage causes the price of leisure and the time-cost of children to also increase, leading to a substitution into labor and out of children. Higher wages also increase income, which moves households away from labor and toward children. When the elasticity of substitution is positive, the substitution effects tends to dominate, increasing the responsiveness of labor to fertility as the wage goes up.



In a small number of low-income countries, including pre-WWI U.S., we show a positive labor supply response may be empirically relevant. While this framework predicts a negative labor supply response to children, such a positive response is possible with a simple extension of the model. Suppose there is a consumption (e.g., food) cost to children so  $c = wl - kn$ , and for simplicity set  $c_0$  and  $\varepsilon$  to zero. The first-order condition with respect to labor, with rearrangement, now becomes:

(8)

$$l = \frac{\alpha^\theta + n(w^\theta \gamma^\theta k - \alpha^\theta b)}{w^{\theta+1} \gamma^\theta + \alpha^\theta}.$$

In this case  $\partial l / \partial n > 0$  is consistent with  $w > \alpha^\theta b / \gamma^\theta k$ . An increase in fertility implies an increased time cost but also reduces consumption, making increased labor more valuable. With a sufficiently high wage, the last effect can dominate leading to increased labor. In this case,  $\partial^2 l / \partial n \partial w < 0$  without further assumptions, so we would continue to expect a negative gradient of the fertility-labor relationship with respect to the wage.<sup>2</sup>

### III. Empirical Strategy

Our empirical analysis adopts the standard approach of exploiting twin births and gender composition as sources of exogenous variation in the number of children to identify the causal effect of an additional child on the labor force activity of women (e.g. Rosenzweig and Wolpin 1980; Bronars and Grogger 1994; Angrist and Evans 1998; and Black, Devereux, and Salvanes 2005). In particular, for twin births, consider a first stage regression of the form:

---

<sup>2</sup> That is,  $\text{sgn}(\partial^2 l / \partial n \partial w) = \text{sgn}(-\gamma^\theta k \gamma w^\theta + \theta k w^{-1} \alpha^\theta + (\theta + 1) \alpha^\theta) = -1$  if  $\rho \in (0,1)$ .

$$(9) \quad z_{ijt} = \gamma S_{ijt} + \rho w_{ijt} + \pi_{jt} + \mu_{ijt}$$

where  $z_{ijt}$  is an indicator of whether mother  $i$  in country  $j$  at time  $t$  had a third child, the instrument  $S_{ijt}$  is an indicator for whether the second (and third) child are the same age (twins),  $w_{ijt}$  is a vector of demographic characteristics that typically include the current age of the mother, her age at first birth, and indicators for the gender of the first two children, and  $\pi_{jt}$  are country-year fixed effects.  $\gamma$  measures the empirical proportion of mothers with at least two children who would not have had a third child in the absence of a multiple second birth.

The local average treatment effect (LATE) among mothers with multiple children is identified from a second stage regression:

$$(10) \quad y_{ijt} = \beta z_{ijt} + \alpha w_{ijt} + \theta_{jt} + \varepsilon_{ijt}$$

where  $y_{ijt}$  is a measure of labor supply for mother  $i$  in country  $j$  at time  $t$  and  $\beta$  is the IV estimate of the pooled labor supply response to the birth of twins for women with at least one prior child.<sup>3</sup> We condition on one prior child, as in Angrist and Evans (1998), to provide a family-size-consistent comparison so that both the same-gender and twins IV study the effect of a family growing from two to three children.

While twins are a widely-used source of variation for studying childbearing on mother's labor supply, it is by no means the only strategy in the literature. Perhaps the leading alternative exploits preferences for mixed gender families (Angrist and Evans 1998). In particular, Angrist and Evans estimate a first-stage regression like equation (9) but, for  $S_{ijt}$ , substitute twin births for an indicator of whether the first two children of woman  $i$  are of the same gender (boy-boy or girl-girl). Again, the sample is restricted to women with at least two children and  $\gamma$  measures the

---

<sup>3</sup> The reported estimates of  $\beta$  are weighted by the household weights supplied by the various surveys or censuses, normalized by the number of mothers in the final regression sample.

likelihood that a mother with two same gendered children is likely to have additional children relative to a mother with a boy and a girl.

Both twins and same gender children have been criticized as valid instruments on the grounds of omitted variables biases. Twin births may be more likely among healthier and wealthier mothers, and can consequently vary over time and across geographic location (see e.g. Rosenzweig and Wolpin 2000; Hoekstra et al. 2007; Bhalotra and Clarke 2016; and Clarke 2016). While the same gender instrument has proven quite robust for the U.S. and other developed countries (Butikofer 2011), there are many reasons to be cautious in samples of developing countries (Schultz 2008). Among other factors, there is the concern that same-gender siblings may be less costly to raise, leading to a violation of the exclusion restriction (Rosenzweig and Zhang 2009). More directly, households may practice either sex selection or selective neglect of children based on gender (e.g. Ebenstein 2010 and Jayachandran and Pande 2015).

We adopt the broad view of Angrist, Lavy, and Schlosser (2010) that the sources of variation used in various IV strategies are different and, therefore, so are the biases. As such, each IV provides a specification check of the other. In this spirit, we also provide a series of LATE estimates that show a) twin results at alternative family parities, b) twins results of the same gender versus mixed gender,<sup>4</sup> c) findings from a third instrument introduced by Klemp and Weisdorf (2016), which relies on exogenous variation in the timing of first births, and d) directly employ the methodology in Angrist, Lavy, and Schlosser (2010) that combines multiple IV

---

<sup>4</sup> Monozygotic (MZ) twinning is believed to be less susceptible to environmental factors. Hoekstra et al. (2007) provides an excellent survey of the medical literature. Since we cannot identify MZ versus dizygotic (DZ) twins in our data, we take advantage of the fact that MZ twins are always the same gender, whereas DZ twins share genes like other non-twin siblings and therefore are 50 percent likely to be the same gender.

estimates.<sup>5</sup> Additionally, to the extent possible given our data, we show how our results vary when we control for education and health measures such as height and body mass index that have been highlighted as key determinants of twin births (Bhalotra and Clarke 2016).

The literature analyzes a number of measures of  $y_{ijt}$ , including whether the mother worked, the number of hours worked, and the labor income earned. These measures are sometimes defined over the previous year or at the time of the survey. In order to include as wide a variety of data across time and countries as possible, we typically focus on the labor force participation (LFP) of mothers at the time of a census or survey. When LFP is unavailable, especially in some of the pre-WWII censuses, we derive LFP based on whether the woman has a stated occupation. We show that changes in the outcome variable do not change the results.

In concordance with much of the literature, our standard sample contains women aged 21 to 35 with at least two children, all of whom are 17 or younger. We exclude mothers who gave birth before age 15 and families where a child's age or gender is imputed. Furthermore, we drop mothers with an imputed age, who live in group quarters, or whose first child is a multiple birth.<sup>6</sup> It is worth emphasizing that the restrictions on mother's (21-35) and child's (under 18) age may further allay concerns about missing children that have already moved out of the household. We also experiment with even younger mother and child age cut-offs, which additionally provides some inference about difference in the labor supply response to younger and older offspring. Further sample statistics, as well as results when these restrictions are loosened, are provided in the Appendix.

---

<sup>5</sup> In particular, we combine all possible instrument variations: twins, two-boy, and two-girl.

<sup>6</sup> These restrictions depart from Angrist and Evans (1998). The final restriction takes care of rare cases of triplets.

We present our results stratified by time, country, level of development, or some combination. The prototypical plot stratifies countries-years into seven real GDP per capita bins (in 1990 U.S. dollars): under \$2,500, \$2,500-5,000, \$5,000-7,500, \$7,500-10,000, \$10,000-15,000, \$15,000-20,000, and over \$20,000. To be concrete, in this example, all country-years where real GDP per capita are, say, under \$2,500 in 1990 U.S. dollars are pooled together for the purpose of estimating equations (9) and (10). Similarly, countries with real GDP per capita between \$2,500 and \$5,000, and so on, are also pooled together for estimation. The plots report estimates of  $\gamma$  and  $\beta$ , and their associated 95 percent confidence interval based on country-year clustered standard errors, for each bin.

#### **IV. Data**

We estimate the statistical model using four large databases of country censuses and surveys.

##### **a. U.S. Census, 1860-2010**

The U.S. is the only country for which we have consistent historical microdata over a long stretch of time.<sup>7</sup> We use the 1 percent samples from the 1860, 1870, 1950, and 1970 censuses; the 5 percent samples from the 1900, 1960, 1980, 1990, and 2000 censuses; the 2010 American Community Survey (ACS) 5-year sample, which combines the 1 percent ACS samples for 2008 to 2012; and the 100 percent population counts from the 1880, 1910, 1920, 1930, and 1940 censuses.<sup>8</sup> Besides additional precision, the full count censuses allow us to stratify the

---

<sup>7</sup> We have a sporadic time-series for Canada, UK, Ireland, and France as well.

<sup>8</sup> For information on the IPUMS samples, see Steven Ruggles, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek, *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database], Minneapolis: University of Minnesota, 2010. The 100 percent counts were generously provided to us by the University of Minnesota Population Center via the data collection efforts of ancestry.com. Those files have been cleaned and harmonized by IPUMS. The 1890 U.S. census is unavailable and U.S. censuses prior to 1860 do not contain labor force information for women.

sample by geography (e.g. states) to potentially take advantage of more detailed cross-sectional variation.

IPUMS harmonizes the U.S. census samples to provide comparable coding of variables over time. However, there are unavoidable changes to some of our key measures. Perhaps most important, the 1940 census shifted our labor supply measure from an indicator of reporting any “gainful occupation” to the modern labor force definition of working or looking for work in a specific reference week. Fortunately, there does not appear to be a measurable difference between these definitions. When both measures are available in 1940, their state-level correlation is 0.95. The 1940 census is also the first to introduce years of completed schooling and earnings; therefore, when we show results invoking education or earnings, we exclude U.S. data prior to 1940.

While the 1880, 1920, 1930, and 1940 full count censuses are fully harmonized with the IPUMS samples, the 1910 full count census is not yet. For our purposes, the most important feature missing from the unharmonized data is linkages matching children to mothers. Accordingly, we create these linkages ourselves using the IPUMS rules.<sup>9</sup>

For Puerto Rico, we use the 5 percent census samples from 1980, 1990, and 2000. As in the U.S., we also include the 2010 Puerto Rico Community Survey 5-year sample, which combines the 1 percent samples for 2008 to 2012. Prior censuses are either missing labor force data or reliable information about real GDP per capita.

---

<sup>9</sup> These rules are available upon request. Of note, relationship linkages did not exist at the time Angrist and Evans (1998) was written. Nevertheless, we can nearly replicate their results using our linkage rules or the modern relationship variables, suggesting their absence in the 1910 census is not a significant issue.

## **b. IPUMS International Censuses, 1960-2011**

IPUMS harmonizes censuses from around the world, yielding measures of our key variables that are roughly comparable across countries and time. We use data from 200 of the 263 country-year censuses between 1960 and 2011 that are posted at the IPUMS-I website.<sup>10</sup> Censuses are excluded if mother-child links or labor force status is unavailable (57 censuses)<sup>11</sup> or age is defined by ranges rather than single-years (6 censuses).<sup>12</sup>

## **c. North Atlantic Population Project (NAPP), 1787-1911**

The North Atlantic Population Project (NAPP) provides 18 censuses from Canada, Denmark, Germany<sup>13</sup>, Great Britain, Norway, and Sweden between 1787 and 1911. As with IPUMS data, these samples are made available by the Minnesota Population Center.<sup>14</sup> For most samples, NAPP generates family interrelationship linkages. However, in a few cases (Canada for 1852, 1871, and 1881 and Germany in 1819), such linkages are not available. In those cases, we use similar rules developed to link mothers and children in the U.S. full count census.<sup>15</sup> Also, consistent with the pre-1940 U.S. censuses, labor force activity is based on whether women

---

<sup>10</sup> This information is as of June 28, 2016, when we downloaded the data. The tabulations of available countries exclude the U.S. and Puerto Rico.

<sup>11</sup> This unfortunately affects some censuses from Canada and the U.K. Similar to the U.S., the international linking variables use relationships, age, marital status, fertility, and proximity in the household to create mother-child links. Sobek and Kennedy (2009) compute that these linking variables have a 98 percent match rate with direct reports of family relationships.

<sup>12</sup> The 1971 to 2006 Irish censuses use ages ranges for adults but not for children younger than 20. Therefore, twins are identifiable and we do not exclude this data.

<sup>13</sup> The NAPP 1819 German data is from the small state of Mecklenburg-Schwerin, rather than the whole region of Germany. However, we refer to it as Germany for expositional purposes.

<sup>14</sup> For additional information about the NAPP samples, see Minnesota Population Center (2015), **North Atlantic Population Project: Complete Count Microdata**, Version 2.2 [Machine-readable database], Minneapolis: Minnesota Population Center.

<sup>15</sup> Details are available upon request.

report an occupation rather than the modern definition of working or seeking work within a specific reference period, and education is unavailable.<sup>16</sup>

#### **d. Demographic and Health Surveys (DHS), 1990-2015**

To collect additional observations from low-income environments, we supplement the censuses with the Demographic and Health Surveys (DHS), a series of nationally representative surveys of mothers and children in developing countries.<sup>17</sup> There have been six waves of DHS surveys, beginning in the mid-1980s and running roughly every 5 years thereafter. From the initial set of 254 country-year surveys, we exclude samples missing age of mother, marital status of mother, current work status, whether the mother works for cash, and birth history. These restrictions exclude the first wave of the DHS, when the surveys were in a relatively nascent state. Further excluding countries without valid real GDP per capita data (see below), we are left with 192 country-years, encompassing 692,923 mothers.

As a detailed survey rather than a census, the DHS includes a number of questions that are especially valuable for testing the robustness of our census results. The detailed health information allows us to control for characteristics that may be related to a mother's likelihood of twinning (Bhalotra and Clarke 2016). Moreover, the DHS indicates whether the children in a mother's birth history are in fact twins. To keep the DHS results comparable to the censuses,<sup>18</sup> our baseline estimates identify twins based on the same census year-of-birth criterion. Yet we still make use of the twin information to test how accurate our coding of census twins might be.

---

<sup>16</sup> In the NAPP, the occupation definitions are based on the variables *occgb*, *occhisco*, and *occ50us*. Note that the NAPP occupation classifications are different than those used in the U.S. censuses, with the exception of the occupational coding used for Canada in 1911.

<sup>17</sup> For additional information about the DHS files see ICF International (2015). The data is based on extracts from DHS Individual Recode files. See <http://dhsprogram.com/Data/>.

<sup>18</sup> In addition, we only consider living children who reside with the mother to keep the DHS comparable with our census samples.



Indeed, Appendix Figure A1 shows the high degree of correspondence between twinning rates when we define twins using “real” multiple births and those imputed for children sharing the same birth-year.

The DHS has a number of labor force variables but none that directly compare to those in the censuses. We chose to use an indicator of whether the mother is currently working since it is most correlated with the IPUMS labor force measures (see Appendix Figure A2).

#### **e. Real GDP per Capita**

Real GDP per capita (in US\$1990) is collected from the Maddison Project.<sup>19</sup> To reduce measurement error, we smooth each GDP series by a seven year moving average that includes three trailing and three leading years around the census/survey. We are able to match 429 country-years to the Maddison data.<sup>20</sup> This leaves a total of 46,892,809 mothers aged 21 to 35 with at least two children who are present in our baseline estimates.

When we split the 1930 and 1940 full population U.S. censuses into the 48 states and DC, we bin those samples by state-specific 1929 or 1940 income-per-capita.<sup>21</sup> The income data are converted into 1990 dollars using the Consumer Price Index.

#### **f. Summary Statistics**

---

<sup>19</sup> See <http://www.ggdnet.net/maddison/maddison-project/home.htm>.

<sup>20</sup> In a few minor cases, we were not able to match a country to a specific year but still left the census in our sample because we did not believe it would have impacted their placement in a real GDP per capita bin. In particular, the censuses of Denmark in 1787 and 1801 are matched to real GDP per capita data for Denmark in 1820 and Norway in 1801 is matched to data for Norway in 1820. Excluding these country-years has no impact on our results. More importantly, the Maddison data ends in 2010 and therefore censuses or surveys thereafter are assigned their most recently available real GDP per capita data.

<sup>21</sup> [http://www2.census.gov/library/publications/1975/compendia/hist\\_stats\\_colonial-1970/hist\\_stats\\_colonial-1970p1-chF.pdf](http://www2.census.gov/library/publications/1975/compendia/hist_stats_colonial-1970/hist_stats_colonial-1970p1-chF.pdf).

Table 1 provides descriptive statistics separately for the U.S. and non-U.S. samples as well as by real GDP per capita bins. Not surprisingly given the large number of country-years from the DHS, 210 out of 429 samples are from countries outside the U.S. with an income below \$2,500 (in 1990 U.S. dollars). Nevertheless, most GDP per capita bins have a large number of mothers for both the U.S. and non-U.S. samples. Summary statistics for a number of individual variables, including share in the labor force, number of children, mother's age at survey, 2nd child is multiple birth, are also presented in Table 1. Appendix Table A1 additionally provides descriptive statistics for all country-year datasets.

## **V. Results**

### **a. OLS Estimates**

We begin with a discussion of OLS estimates from regressions of the labor supply indicator on the indicator for a third child and the controls described above. These results do not have a clear causal interpretation, but they are useful for establishing some of the key patterns in the data. In Figure 1, we plot the coefficients for the U.S., the non-U.S. countries, and the combined world sample (labeled "All"). Country-year observations are binned into seven ranges of real GDP per capita, as reported on the x-axis (e.g. \$0-2,500, \$2,500-5,000, etc.). The results for the U.S. and non-U.S. samples corresponding to Figure 1 are shown in Table 2.<sup>22</sup>

All three plots exhibit a similar pattern. At low levels of real GDP per capita, the OLS estimate of the effect of children on mother's labor supply is negative and statistically significant at the 5 percent level but small in magnitude (e.g. -0.021 (0.004) in the lowest GDP bin). As real GDP per capita increases, the effect becomes more negative, ultimately flattening out between -

---

<sup>22</sup> In this and subsequent figures, we present 95 percent confidence interval bands based on standard errors clustered at the country-year level.

0.15 and -0.25 beyond \$15,000 for all three samples. Appendix Figure A3 shows similar evidence for four countries – Canada, France, Ireland, and the U.K. – for which we have census data at various points of their development cycle.<sup>23</sup> In these cases, we see a similar, albeit noisier, negative gradient as these economies grow.

In Figure 2, we plot the U.S.-only results over time.<sup>24</sup> The OLS estimates starts out negative, albeit relatively small (e.g. -0.011 (0.004) in 1860 and -0.013 (0.0004) in 1910), decrease from 1910 to 1980, at which point the magnitude is -0.177 (0.001), and flattens thereafter. Note that due to the sample size, 95 percent confidence intervals are provided but not visible at the scale of the figure.

Figure 3 plots the OLS estimates by real GDP per capita separately by time periods (pre-1900, 1900-1950, 1950-1989, and 1990+). Years prior to 1950 combine U.S. census and NAPP data. Years thereafter include all four of our databases. The same general pattern appears *within* time periods.<sup>25</sup> The effect of fertility on labor supply tends to be small at low levels of GDP per capita but increases as GDP per capita rises.

## **b. Twins IV**

The left panel of Figure 4 shows the first-stage effect of a twin birth on our fertility measure, the probability of having three or more children. For the U.S., non-U.S., and combined world samples, there is a notable positive and concave pattern, with the first-stage increasing with higher real GDP per capita up to \$15,000 or so and flattening thereafter. Note that the

---

<sup>23</sup> There are four Canadian censuses from 1871-1911, four British censuses from 1851-1991, eight Irish censuses from 1971-2011, and eight French censuses from 1962-2011. Pre-WWII microdata are not available for the Irish and French censuses. We also do not have access to British or Canadian census microdata around WWII and cannot identify households in some of the more recent IPUMS samples.

<sup>24</sup> Blue circles represent IPUMS samples and red diamonds represent full population counts. We take the high degree of correspondence between the estimates in years with both as validation of our implementation of mother linkages in the full count data.

<sup>25</sup> Relative to Figure 1, we had to combine some real GDP per capita bins because of small sample sizes within these tight time windows.

regression specification controls for the mother's age, but does not, indeed cannot, control for the number of children or target fertility. Therefore, the positive gradient over real GDP per capita reflects the negative impact of income on target fertility and hence the heightened impact of a twin birth on continued fertility relative to a non-twin birth.<sup>26</sup> Regardless, in all cases, the instrument easily passes all the normal statistical thresholds of first-stage relevance, including among countries with low real GDP per capita and high fertility rates.

The right panel of Figure 4 plots  $\beta$ , the instrumental variables effect of fertility on mother's labor supply.<sup>27</sup> In the world sample,  $\beta$  is mostly statistically indistinguishable from zero among countries with real GDP per capita of \$7,500 or less. Thereafter, it begins to decline and eventually flattens out between -0.05 and -0.10 at real GDP per capita at around \$15,000 and higher.<sup>28</sup> The results for the U.S. and non-U.S. samples are similar in that there is a notable negative gradient with respect to real GDP per capita. For example, above \$20,000, the U.S. estimate is -0.070 (0.008) while the non-U.S. estimate is -0.104 (0.002).

In Figure 5, we show the U.S. twin results by census decade. The pattern is broadly similar to the previous figure. The magnitude of the first stage is increasing over time, and the second-stage IV results exhibit a pronounced negative gradient, particularly post-WWII.<sup>29</sup> The same pattern appears within time periods (Figure 6), and notably begins to decline prior to 1990, well before modern fertility treatments, like IVF, became widely used in wealthy countries.

---

<sup>26</sup> The first stage coefficient,  $\gamma$ , is  $E\{z=1|S=1,w\} - E\{z=1|S=0,w\}$ . Mechanically,  $E\{z=1|S=1,w\}=1$  because of the definition of twins. This means that if, for example,  $\gamma=0.6$ , then  $E\{z=1|S=0,w\}=0.4$ , implying that 40 percent of mothers would have a third child if their second child is a singleton. The increasing coefficient over real GDP per capita means having a third child after a singleton second child is declining with development.

<sup>27</sup> The point estimates and standard errors from Figure 4 are also shown in Table 2.

<sup>28</sup> By comparison, Angrist and Evans (1998) report a twins IV estimate of -0.087 for the 1980 U.S. census.

<sup>29</sup> In our binned samples, we only include the U.S. full population for 1880 and 1910 to 1940. However, we display the single-year estimates from the IPUMS random samples for these years in Figures 2 and 5. We take the high degree of correspondence between the 1910 IPUMS and full population estimates as validation of our implementation of mother linkages.

Finally, the pattern appears across data sets (Appendix Figure A4) and geographic regions of the world (Appendix Figure A5), including four other developed countries in which we have longer time-series but highlighted by the U.K. and Ireland, where we can estimate  $\beta$  at both low- and high-income periods in their history (Appendix Figure A6).

### **c. Are There Positive Labor Supply Effects Among the Lowest Income Countries?**

One surprising finding is that at low real GDP per capita levels, we sometimes estimate a positive labor supply response to childbearing. That is particularly evident in the pre-WWI U.S. estimates displayed in Figure 5, although there are periodically positive but not statistically significant effects for some of the low-income post-1990 countries as well.<sup>30</sup> The U.S. positive results are not statistically significant different from zero for the early census samples (1860, 1870), but they are for the full population counts of 1880 and 1910.

While these positive results are not artifacts in the statistical sense, it is worth noting that the underlying rates of labor force participation for U.S. women are very low at this time in history (e.g. 6.2 and 11.8 percent for 1880 and 1910 mothers, respectively). As such, a positive effect could reflect that low income mothers are more likely to work after having children, for example because subsistence food and shelter are necessary, whereas childcare might be cheaply available. Section II discusses a simple extension to our theoretical model, the introduction of a consumption cost to children, which implies the potential for a positive labor supply response to additional children. Such a framework may be especially relevant for the subpopulation of compliers for the local average treatment effect – that is, mothers induced to have children who would not have otherwise.

---

<sup>30</sup> See Appendix Figures A5. On the regional figure, the estimates tend to be not statistically, nor economically, different from zero at low income levels, with the exception of a single pooled sample from Asia, which is positive and significant. The Asian sample between \$5,000-7,500 consists of 462,553 observations from 14 country-years. The pooled result for these 14 samples is almost completely driven by Turkey in 1990 and 2000, which have IV estimates of 0.200 (0.023) and 0.150 (0.017) and make up 163,770 and 180,069 observations, respectively.

To gain further insight into the low real GDP sample results, we split the U.S. 1930 and 1940 full population counts by state of residence and pool states into income-per-capita estimation bins (matching what we did with countries in previous figures). Figure 7 shows the now familiar upward sloping pattern to the first stage results by real income per capita. In the second stage, we see that the effect of fertility on labor supply is in general statistically indistinguishable from zero at low income levels in 1930 and 1940 and overlaps with the low-income post-1990 non-U.S. results (shown in the green line). But we also find a small positive effect from the lowest income states in 1930, seemingly corroborating the positive estimates from a lower income U.S. pre-WWI.<sup>31</sup>

#### **d. Same Gender IV**

Next, we discuss results, displayed in Figure 8, that use the same gender instrument.<sup>32</sup> Like the twins IV, we estimate a positive gradient to the first stage with respect to real GDP per capita, although the interpretation of this pattern is different than for twins. In particular, the same-gender first-stage picks up the increased probability that a mother opts to have more than two children based on the gender mix of her children (rather than picking up the proportion of mothers with incremental fertility when the twin instrument is zero, i.e., for non-twin births). Most importantly, we again see a negative gradient on the second stage IV estimates, from a close-to-zero effect among low GDP countries to a negative and statistically significant effect at higher real GDP per capita that flattens at around \$15,000. Again, the negative estimates appear in the U.S. post-WWII (Appendix Figure 7).

---

<sup>31</sup> For the 1930 census, the states in that lowest bin (\$2,000-3,000) are: Alabama, Arkansas, Georgia, Mississippi, North Carolina, North Dakota, New Mexico, South Carolina, and Tennessee.

<sup>32</sup> The point estimates and standard errors from Figure 8 are shown in Table 2.

Our main intention is to highlight the similar shapes of the labor supply effect across the development cycle, despite using instruments that exploit difference sources of variation.<sup>33</sup> Indeed, not surprisingly, when we combine all possible instrument variation into a singled pooled estimator, as in Angrist, Lavy, and Schlosser (2010), our weighted average twin and same gender IV results also, unsurprisingly, shows the same strong negative gradient. That said, the magnitude of the same gender IV result is larger than the twin IV result at the high GDP per capita bins.<sup>34</sup> Since this is a local average treatment effect, this disparity suggests a greater effect of fertility on labor supply for those women induced to have an incremental child based either on son preference or the taste for a gender mix compared to those encouraged to higher fertility by a twin birth.

#### **e. Hours**

The results thus far are reported for the extensive labor force participation margin. Figure 9 plots twin IV results for the number of hours worked per week among women that are working (the intensive margin). We include all country-years that contain a measure of hours worked, which unfortunately limits us to 38 censuses.<sup>35</sup> Nevertheless, we again find no evidence of a labor supply response among mothers in low-income countries and a negative response of about 0.8 hours per week among mothers in higher-income countries. As a benchmark, employed mothers work, on average, just under 33 hours per week in countries with real GDP

---

<sup>33</sup> Like the twins estimates, we also find additional systematic evidence of a positive fertility-labor supply effect at low levels of income, which are statistically significant for the 1910, 1930, and 1940 U.S. censuses (see Appendix Figure A7).

<sup>34</sup> For example, at the \$20,000 and above bin, the twin estimate is -0.070 (0.008) for the U.S. sample and -0.104 (0.002) for the non-U.S. sample. By comparison, the same gender estimates are -0.121 (0.008) for the U.S. sample and -0.174 (0.020) for the non-U.S. sample.

<sup>35</sup> We can use eight U.S. censuses (1940-2010) and 30 censuses from other countries. The DHS and NAPP do not contain information about hours worked per week. When hours are reported as a range, we use the center of the interval.

per capita above \$20,000, suggesting a roughly 2½ percent average decline in hours as a result of an additional child, conditional on working.

## **f. Robustness**

### **f.1 Omitted Variables and Alternative Sources of Identification**

As noted above, twin and same gender instruments are susceptible to omitted variables biases. These biases are likely to differ across instrument, suggesting that the twins and same gender IV estimates can be specification checks of each other (Angrist, Lavy, and Schlosser 2010). However, in this subsection, we push this idea further by providing three other sets of estimates that exploit alternative sources of instrument variation or control for observable characteristics that are known to explain variation in the treatment.

First, we examine a third instrument for fertility – the time that elapses between the parents’ marriage and the couple’s first birth (“time to first birth” or TFB) – introduced by Klemp and Weisdorf (2016).<sup>36</sup> A long line of research in demography and medicine (e.g. Bongaarts 1975) uses birth spacing, not necessarily limited to first births, as an indicator of fecundity. Some of these papers also document the extent to which spacing is unrelated to many observable characteristics.<sup>37</sup> Klemp and Weisdorf argue that TFB is especially hard to predict based on observable characteristics outside of parent age and consequently is a valid indicator of ultimate family size. Unfortunately, TFB requires marriage and birth dates which are generally only available in the DHS. Therefore, we cannot replicate the negative gradient across the development cycle. However, we find the TFB IV estimates are near zero, and perhaps even

---

<sup>36</sup> Unfortunately, our data does not allow us to systematically study other instruments used in the literature, such as the use of infertility treatments (Cristia 2008; Aguero and Marks 2011; Lundborg, Plug, and Rasmussen 2016), changes in access to birth control (Bailey 2013), or other policy changes (Bloom et al. 2009; Godefroy 2016).

<sup>37</sup> For evidence otherwise, see Feng and Quanhe (1996), Basso, Juul, and Olsen (2000), and Juul, Karmaus, and Olsen (1999).



slightly positive, at low real GDP per capita levels, and similar to twin IV and same gender estimates at the same GDP level.<sup>38</sup>

Second, our baseline twin birth estimates condition on families with one child and compare those who then have a twin birth to those who have a singleton birth. Again following Angrist, Lavy, and Schlosser (2010), we condition on different family size parities to capture variation from different sets of mothers. For example, one might expect that mothers with a large number of previous children would be less likely to adjust their labor supply in response to unexpected incremental fertility (for example, because of low incremental childcare costs for higher births). Indeed, as shown in Figure 10, we observe a stronger first stage effect for the sample that conditions on more children, especially at higher income levels. In the second stage, we see a notably, although not always statistically significantly, more negative effect in high-income countries for women starting with one child. However, the pattern of results is similar regardless of how many children are in the household when the twins are born. In all family size circumstances (up to three initial children), we continue to find no effect among low income countries and an increasingly larger negative effect among higher income countries, flattening out around \$20,000 per capita.

Finally, it has been noted by many researchers, most recently Bhalotra and Clarke (2016), that mothers of twins are positively selected by health and wealth.<sup>39</sup> We provide additional two pieces of evidence that this selection process is not driving the negative labor supply gradient. First, when we control for the observable characteristics that have been highlighted by Bhalotra and Clarke (2016), such as mother's education, medical care availability, and mother's health,

---

<sup>38</sup> The TFB IV estimates using the DHS data are: 0.031 (0.018), 0.047 (0.015), and 0.044 (0.014) for the \$0-2,500, \$2,500-5,000, and \$5,000-10,000 GDP per capita bins, respectively.

<sup>39</sup> Related, Rosenzweig and Zhang (2009) argue twins are less costly to raise than two singleton births spaced apart. While we cannot fully address this concern, we can restrict the analysis to mothers with close birth-spacing. Appendix Figure A8 shows that this restriction has little effect on the gradient.

our results are statistically identical to the baseline estimates without these controls.<sup>40</sup> Second, a strand of the medical literature argues that there is a different process of selection into monozygotic and dizygotic twins (see e.g. Hoekstra et al. 2007). The proportion of dizygotic twins is affected by environmental and genetic factors of the type discussed by Bhalotra and Clarke (2016). By contrast, the proportion of monozygotic twins appears to be relatively constant over time (and thus the development cycle) and therefore unaffected by their omitted variables bias concern. Of course, we cannot identify monozygotic and dizygotic twins in our data but we can exploit the fact that monozygotic twins are always same gender, whereas dizygotic twins are an equal mix of same and opposite gender (like non-twin siblings).<sup>41</sup> In Figure 11, we report that results are statistically indistinguishable across same and opposite gender twins, lending additional credence to the view that our results are not driven by omitted variable bias with respect to twinning.

## **f.2 Alternative Development Benchmarks**

The labor supply patterns we have documented thus far are based on an economy's real GDP per capita. The key model prediction, however, is based on the substitution and income effects arising from changes to a woman's wage. Unfortunately, data limitations make it difficult to show full world results stratified by female (or overall) wages. Instead, we examine two alternative exercises.

---

<sup>40</sup> Appendix Figure A9 plots the results with and without mother's education covariates using all available censuses and the DHS. Health measures are available only in the DHS. We are able to roughly replicate Bhalotra and Clarke's association between twinning and doctor availability, nurse availability, prenatal care availability, mother's height, mother's BMI (underweight and obese dummies), and infant mortality prior to birth. When we specifically control for these measures, our labor supply IV estimates are identical to the baseline for the <\$2,500 bin and only slightly larger but statistically and economically indistinguishable for the \$2,500-\$5,000 bin (-0.006 (0.031) versus 0.001 (0.032)) and \$5,000 and over bin (-0.075 (0.042) versus -0.068 (0.044)).

<sup>41</sup> The rate of monozygotic twinning is approximately 4 per 1000 births and is constant across various subgroups (Hoekstra et al. 2007). Under the standard assumption that dizygotic twins have a 50 percent chance of being the same gender, approximately 43 to 59 percent of same-gender twins are monozygotic across the various GDP bins. Notably, the proportion of monozygotic twins will be highest in low-GDP countries, where Bhalotra and Clarke (2016) find the potential for the omitted variable bias is greatest.

First, we use the 1940 to 2010 U.S. censuses to compute average female real wage rates by state and census year.<sup>42</sup> Analogous to the real GDP per capita bins used in prior figures, state-years are stratified into four real hourly wage bins, ranging from \$0 to \$6 up to \$12 to \$19, based on the average wage in the state at that time. Similar to the GDP per capita results, we find no labor supply effect at the lowest two real wage levels and larger negative effects as the real hourly wage rises (Figure 12). Second, we report IV results from the U.S. and non-U.S. samples stratified by the average education level of women aged 21 to 35 (Figure 13).<sup>43</sup> We again find no effect at low education levels (below 9 years) but decreasing negative effects thereafter.

### **f.3 Other Data Issues**

There are several variable definition decisions that could conceivably be problematic. One of these decisions was to label siblings born in the same year as twins since few censuses record multiple births or the birth month/quarter.<sup>44</sup> This classification raises the risk that two births in the same calendar year could be successive rather than twins (so-called Irish twins). Fortunately, some of our data provide quarter or month of birth or direct measures of multiple births, allowing us to compare twins based on more precise birth dates with our baseline year-of-birth twins. The black line in Figure 14 represents our baseline reported in earlier figures. The blue line uses a subset of countries with quarter or month of birth. By and large, we see a very similar negative gradient despite a notably smaller sample of country-years. To make the

---

<sup>42</sup> There is no wage data prior to 1940. For all persons aged 18 to 64, we calculate the average hourly wage rate as annual earned income divided by weeks worked times hours worked per week. The age range overlaps with the cohort of mothers used in our baseline sample but we do not condition on gender or motherhood. The results are robust to using the average wage rate of men or women only as well. Wages are inflation adjusted using the consumer price index to 1990 dollars. We top- and bottom-coded the wage at the 1<sup>st</sup> and 99<sup>th</sup> percentiles in each census prior to taking means.

<sup>43</sup> Again, data availability limits our analysis to 1940 and later. We also exclude 30 country-years where years of education are not provided. By 1940, U.S. women in their twenties and thirties had, on average, at least 9 years of education. Consequently, the U.S. is included only in the two highest education bins (9-12 and 12+ years).

comparison cleaner, we also re-estimated the baseline year-of-birth estimates with the sample of countries that provide quarter or month of birth (red line). Although the pattern gets noisier with smaller sample sizes, the twin estimates based on year-of-birth appear to be, if anything, biased away from zero in low-income countries.

A second measurement issue relates to our labor supply outcome. As mentioned earlier, our historical results (1930 and earlier) must use occupation-based labor force participation, coding women as in the labor force if they report an occupation. Post-1940, we use the modern definition based on whether the person is working or searching for work at the time of the survey. We show that this change in definition has no impact on our results in two ways. First, we use the 1940 U.S. census to compare across-state results using both measures. We find a 0.95 cross-state correlation between the two LFP measures and a 0.82 correlation of the state-level IV results (Appendix Figure A11). Second, Figure 15 shows that our baseline trend across real GDP per capita is essentially invariant to the use of an occupation-based LFP. The patterns are also similar if we use employment (versus in the labor force) or worked over the prior year (rather than at the time of the census or survey).

More broadly, our findings are robust to a number of other reasonable tweaks to our specification, variable definitions, and sample selection, such as excluding country-year fixed effects<sup>44</sup> and alternative ways to specify the mother's age and age at first birth covariates, as well as parsing the sample by age, age at first birth, education, and marital status of the mother.<sup>46</sup>

While we find consistently larger negative effects among single (relative to married) and

---

<sup>44</sup> Another measurement concern relates to non-biological children. Child-parent biological relationships are not consistently available across censuses. However, when we have that information, we find that restricting our sample to biological children has little impact on the results (see Appendix Figure A10).

<sup>45</sup> In the absence of fixed effects, the fertility effect is positive at some low income levels. However, that result is driven by some outliers in our data (e.g., Nicaragua in 1998 with an estimate of 0.385 (0.040) based on 3,733 mothers).

<sup>46</sup> These figures can be found in Appendix Figures A12 to A16.

younger (relative to older) mothers, especially in countries with higher GDP per capita, those cases still exhibit the same negative gradient across development. Moreover, there is no statistical or economic difference across gender and mother's education at any level of GDP per capita.

Lastly, we also used the methods proposed by Angrist and Fernandez-Val (2010) and Bisbee et al. (2015) to calibrate our IV estimates to a common LATE, namely the LATE for compliers in the U.S. in 1980.<sup>47</sup> That refinement has no impact on the results (Appendix Figure A17).

## VI. Channels

This section explores some of the potential mechanisms that account for the remarkably robust negative income gradient of mother's labor supply response to children.<sup>48</sup>

### a. Accounting for Base Rates of Labor Force Participation

One possibility is that the negative gradient is simply a function of the base rate of labor force participation. With respect to our theoretical model, a lower base rate of labor force participation would imply more corner ( $l = 0$ ) cases, for which there is no scope for a negative fertility effect on labor supply. This mechanically limits the scale of any average causal effect of fertility. We can account for this possibility by rescaling estimates to the relevant base rate. The logic of this rescaling is based on the assumption that effects tend to be monotonic in the population under study. That is, write the average effect in population  $s$  as,

$$(11) \quad \beta_s = E_s[Y_1 - Y_0],$$

---

<sup>47</sup> This calculation targets the estimation to the covariate distribution of compliers in the U.S. in 1980.

<sup>48</sup> As the main area of interest is the causal labor supply effect of children and the strength of the instruments are apparent, we stop reporting the first-stage estimates. For brevity, we concentrate solely on the twin estimates.

where  $Y_1$  and  $Y_0$  are potential labor outcomes (with support  $\{0,1\}$ ) under the condition of three or more children and less than three children, respectively. Effect monotonicity implies  $Y_1 \leq Y_0$ , which also means

$$(12) \quad E_s[Y_1 - Y_0 | Y_0 = 0] = 0.$$

This further implies that

$$(13) \quad \beta_s = E_s[Y_1 - Y_0 | Y_0 = 1] E_s[Y_0],$$

in which case the average effect of having three or more children *among those for which there can be an effect* is given by

$$(13) \quad \beta_s^r = E_s[Y_1 - Y_0 | Y_0 = 1] = \frac{\beta_s}{E_s[Y_0]}.$$

Comparing trends in  $\beta_s$  versus  $\beta_s^r$  allows us to assess the influence of base participation rates. For example, suppose we have two populations,  $s$  and  $s'$ , perhaps corresponding to the same country but at different points in time, or perhaps corresponding to two different countries at different levels of development. Then, if  $\beta_s < \beta_{s'}$ , but  $\beta_s^r = \beta_{s'}^r$  and  $E_s[Y_0] < E_{s'}[Y_0]$ , we could infer that the effect of fertility *among those for whom an effect is possible* is constant, but that an increase in the base rate of participation from population  $s$  to  $s'$  leads to a stronger average effect when taking all women in the populations into account. Such a pattern of evidence would suggest no fundamental change in the way fertility tends to affect labor supply. If, however, we see that  $\beta_s^r < \beta_{s'}^r$ , then this would suggest that the negative gradient in the average effect is not simply a function of changes in the base rate.

This rescaling only recovers a meaningful effect in populations for which the monotonicity assumption is reasonable. We have already noted that for some of our country-year samples we estimate positive fertility effects, in which case this rescaling strategy would not

be applicable. The analysis here is applied only to the country-year samples for which we observe either negative or (statistically indistinguishable from) zero fertility effects.

Given that we are estimating complier LATEs via IV, the populations indexed by  $s$  correspond to the compliers in our various country years. As such, the relevant base rate,  $E_s[Y_0]$ , corresponds to the labor force participation rate among compliers with instrument values equal to 0. We compute these complier-specific rates using the IV approach of Angrist, Pathak, and Walters (2013).<sup>49</sup>

Figure 16 shows the results of applying the rescaling to the baseline estimates. For the US, the rescaling results in a substantial flattening past \$7,500 per capita. For the non-US populations, the rescaled estimates are consistent (taking into account the uncertainty in the estimates) with a flattening after \$10,000 per capita. However, a negative gradient is still evident over lower levels of income. This indicates that the decline in the labor supply effect of an additional child is not solely driven by increases in the base rate of mother's LFP and motivates further analysis into the channel driving the negative gradient, particular over income levels under \$10,000 per capita. The analyses below examine results both with and without the base-rate rescaling.

#### **b. Changes to the Income and Substitution Effect Across Stages of Development**

We believe the remaining negative gradient is due to a declining substitution effect, in combination with an unchanging income effect resulting from increasing wages for women during the process of economic development.

We identify the substitution effect primarily through changes in job opportunities. This exercise is motivated by previous work that documents a U-shape of female employment with

---

<sup>49</sup> Specifically, we stack the two-stage estimation used in Angrist, Pathak, and Walters to calculate the complier-control mean with our baseline two-stage least squares regression to get the covariance between the base rate and the labor supply effect.

development in the U.S. (Goldin, 1990) and across countries (Schultz 1991 and Mammen and Paxson 2000). Schultz (1991) shows that the U-shape is not observed within sector. Rather, it is explained by changes in the sectoral composition of the female labor force. In particular, women are less likely to participate in unpaid family work (mostly in agriculture) and self-employment and more likely to be paid a wage in the formal sector in the later stages of the development process. In addition, we have reason to believe that the changes in the types of jobs that women have over time might become less compatible with raising children. For example, in rural agricultural societies, women can work on family farms while simultaneously taking care of children but the transition to formal urban wage employment is less compatible with providing care at home (Jaffe and Azumi 1960; McCabe and Rosenzweig 1976; Kupinsky 1977; Goldin 1990; Galor and Weil 1996; Edwards and Field-Hendrey 2002; and Szulga 2013).

Given that consistent information on occupations and sectors across our many samples is limited, we rely on two coarse indicators of job type that can be consistently measured in almost all of our data. First, we try to capture the distinction between urban/rural and formal/informal occupations by comparing women who work for a wage versus those who work but are unpaid. These results, unscaled (left) and scaled (right), are presented in Figure 17. We find consistent evidence that the changing relationship between fertility and labor supply is driven by women who work for wages. The response from women who are working but not for wages is small and statistically indistinguishable at different levels of real GDP per capita. Note again, that since these are rescaled estimates the gradient – or lack thereof – is driven not by changes in aggregate levels of labor force participation at different levels of GDP per capita, but by changes in the labor-childbearing tradeoff at the individual level.



A second proxy of sectoral shifts is whether women work in the agricultural or non-agricultural sectors (Figure 18). Although the scaled results presented in the right plot are unfortunately noisy for agricultural labor, the labor supply response of women in non-agricultural sectors becomes clearly more negative as real GDP per capita rises.

In settings where nearly all labor is in the formal sector, it becomes especially hard to identify differences in the effects of women's labor market opportunities. In light of the inherent selection problem around observed wages, the fertility response literature has long used a woman's education to proxy for the type of jobs and wages available to her. While Gronau (1986) documents several results finding education is correlated with fertility response, this correlation appears to reverse once Angrist and Evans (1998) apply instrumental variables. We find no strong heterogeneity by education (Appendix Figure A15). We also observe in Figure 19 that fertility has almost no differential effect across the development cycle on female labor supply to professional occupations, despite the fact that these occupations tend to have higher wages.<sup>50</sup> Instead, the changing gradient seems to be driven entirely by women who work in non-professional occupations, suggesting either that education and professional status are poor proxies for the substitution effect, or that the opportunity differences they capture are small in comparison to the sectoral shifts out of agricultural and non-wage work.

By contrast, we believe the income effect of rising wages is likely small and invariant to the stage of development.<sup>51</sup> We show this in two ways. First, we look at the husband's labor supply response to children using the same twin IV estimator. A long literature, tracing back to

---

<sup>50</sup> Professional occupations are defined somewhat differently across data sources. For the U.S., we use the 1950 occupation codes. For IPUMS-I, we use the International Standard Classification of Occupations (ISCO) occupation codes. For the NAPP, we use the Historical ISCO codes, except for 1911 Canada where we use 1950 U.S. occupation codes. We dropped the 1851 and 1881 U.K. censuses due to difficulty convincingly identifying professionals. For the DHS, we use their occupation codes.

<sup>51</sup> Henceforward, we will present the unscaled IV estimates since changes in the base rate of mothers' work are less of a concern.

classic models of fertility such as Becker (1960) and Willis (1973), argues that an increase to the husband's wage increases the demand for having children, possibly because men spend less time rearing children. That is, the income effect is dominant. In Figure 20, we return to the unscaled estimates and show that the husband's labor supply response is economically indistinguishable from zero and invariant to the level of real GDP per capita.

Second, we use the 1940 to 2010 U.S. censuses, which contain hourly wages of husbands, to measure the differential labor supply response of women throughout the hourly wage distribution of their spouse. One version of those results, where mothers are stratified into three real wage groups of their husbands (\$0-10, \$10-\$16 and above \$16 measured in 1990 dollars), is displayed in Figure 21.<sup>52</sup> Generally, we do not find a differential response, again suggestive that the income effect is unlikely to be a driver of the negative gradient in the labor supply response to children over the development cycle.

### **c. Changes to Family Structure**

Next, we explore an alternative explanation, which we refer to as a family structure channel. According to this view, the increased time cost of looking after children in our model is driven by changes in family composition that come with development, such as fewer opportunities for non-parent members of the household to help with childcare or household chores. Indeed, a large body of work has analyzed the decline of intergenerational co-residence over time (e.g. Ruggles 1994). We explore this potential channel by using the information available in our data to locate the presence of a grandparent or, more generally, any non-mother female older than 15 in the household. These results are presented in Figures 22 and 23.

---

<sup>52</sup> Figure 21 is an extension of Figure 12, where the states are grouped into bins by the average wage of all 18-64 year olds and mothers are separated within bins by their spouse's wage.

Strikingly, we find little difference in the results by co-residence of grandparents or other non-mother adult females.

As further corroborating evidence, we find that the labor supply responses of U.S. mothers who still reside in their birth state, and thus are more likely to live near their own parents or other family members, are not different from U.S. mothers who migrate away from their state of birth (Figure 24). Consequently, changes in the family structure, and their implication on the price of access to childcare, do not appear to drive our main results.

Finally, the structure of the nuclear family also has potential implications for childcare costs. Older children in the household may reduce the incremental cost of childcare, reducing the need for mothers to adjust their labor supply after an incremental birth. Moreover, more eligible workers, whether they are older children or husbands, may also offset some of the labor supply consequences of motherhood. On the other hand, having young children may increase childcare costs. Figures 25 and 26 stratify the samples based on the age of the oldest and youngest child (ages 0-6, 7-12, or 13-17), respectively. Again, in every case, we see a near-zero labor supply response to children at low levels of real GDP per capita, a declining labor supply elasticity at real GDP per capita around \$7,000 - \$15,000, and a large, relatively constant negative effect above \$15,000.

#### **d. Changes to Schooling Opportunities for Children**

A related possibility is that development changes access to school. Given the rise of universal schooling around the world, children spend more time in school and less time at home during their childhood, and this could reduce the need for childcare. First, we explore whether access to schooling influences the labor supply responses of mothers using the 1930 and 1940 U.S. full population counts. We split the U.S. state-years by the share of school-age kids in

school and by the stringency of compulsory schooling laws.<sup>53</sup> We find no differential mother labor supply response when we stratify states by the earliest age of required schooling or by the total number of mandated schooling years (Appendix Figures A18 and A19). Second, we find no difference in country-years that have a higher than 80 percent school attendance rate for children aged 5 to 18 and those who do not.<sup>54</sup> We also find no differential response when we stratify U.S. states by the earliest age of required schooling or by the total number of mandated schooling years. Therefore, we conclude that there is little evidence that access to schooling can explain the labor supply patterns we document.

## **VII. Conclusion**

In her magisterial monograph of the evolution of women's work in the United States, Goldin (1990) documents and dissects the now-classic U-shaped evolution of women's labor supply over the 20<sup>th</sup> century. At the same time, she notes the paucity of historical causal evidence on the link between fertility and labor supply. A parallel literature in development economics has investigated the implications of evolving patterns of fertility in developing countries on economic growth (and implicitly labor supply). While there have been notable and pioneering studies on the effect of fertility on labor supply in developing countries, these studies naturally tend to focus on single countries or non-causal evidence.

Using a twin birth and same gender of the first two children as instruments for incremental fertility, this paper links these two literatures by examining causal evidence on the evolution of the response of labor supply to incremental fertility across a wide swath of countries in the world and over 200 years of history. Our paper has two robust findings. First, the effect of fertility on labor supply is small, indeed typically indistinguishable from zero, at low levels of

---

<sup>53</sup> Compulsory schooling laws are taken from Goldin and Katz (2003).

<sup>54</sup> These results are insensitive to choosing 70 or 90 percent thresholds.

income, and negative and substantially larger at higher levels of income. Second, the magnitude of these effects is remarkably consistent across the contemporary cross-section of developing countries and the historical time series (using primarily U.S. census data but also other developed-country historical samples from the NAPP), as well as across demographic and education groups.

Our results are consistent with a standard labor-leisure model. As income increases, individuals face an increased time cost of looking after children but also experience higher incomes. Our results suggest the former dominates the latter. This substitution effect seems to come from changes in the sectoral and occupational structure of female jobs, in particular the rise of non-professional, non-agricultural wage work that flourishes with development.

We see three implications of our results. First, in thinking about the U-shaped pattern of labor force participation that has been widely documented in the economic history literature, our results suggest that decreases in fertility play a significant causal role in explaining part of this effect. As fertility rates have declined over the latter half of the 20<sup>th</sup> century, the responsiveness of labor supply to fertility has increased, contributing to increases in female labor force participation. Second, among developing countries, our results however suggest that changes in fertility tend not to have a large impact on labor force participation, arguing against fertility-reduction policies specifically motivated by women's labor force participation and its contribution to growth. Third, at least when it comes to fertility and labor supply, our results point to a remarkable consistency between historical and contemporary developing country data, suggesting that each of these disciplines has important insights for the other.

## Bibliography

- Agüero, Jorge and Mindy Marks, 2008, "Motherhood and Female Labor Force Participation: Evidence from Infertility Shocks," *American Economic Review*, 98 (2), 500–504.
- Agüero, Jorge and Mindy Marks, 2011, "Motherhood and Female Labor Supply in the Developing World Evidence from Infertility Shocks," *Journal of Human Resources*, 46 (4), 800–826.
- Angrist, Joshua and William Evans, 1996, "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," Working Paper, NBER.
- Angrist, Joshua and William Evans, 1998, "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review*, 88 (3), 450–477.
- Angrist, Joshua and Ivan Fernandez-Val, 2010, "ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework," Working Paper, NBER.
- Angrist, Joshua, Victor Lavy, and Analia Schlosser, 2010, "Multiple Experiments for the Causal Link between the Quantity and Quality of Children," *Journal of Labor Economics* 28(4), 773–824.
- Angrist, Joshua, Parag Pathak, and Christopher Walters, 2013, "Explaining Charter School Effectiveness," *American Economic Journal: Applied Economics*, 5(4): 1-27.
- Bailey, Martha, 2013, "Fifty Years of Family Planning: New Evidence on the Effects of Increasing Access to Contraception," *Brookings Papers on Economic Activity*, 46(1), p. 341-409.
- Basso, Olga, Svend Juul, and Jorn Olsen, 2000, "Time to pregnancy as a correlate of fecundity: differential persistence in trying to become pregnant as a source of bias," *International Journal of Epidemiology* 29(5), 856-861,
- Becker, Gary, 1960, "An economic analysis of fertility", in: *Demographic and Economic Change in Developed Countries*, Universities-National Bureau of Economic Research Conference Series 11, NBER, Princeton, NJ, pp. 209-231.
- Bisbee, James, Rajeev Dehejia, Cristian Pop-Eleches, and Cyrus Samii, 2015, "Local Instruments, Global Extrapolation: External Validity of the Labor Supply-Fertility Local Average Treatment Effect," forthcoming *Journal of Labor Economics*.
- Bhalotra, Sonia and Damian Clarke, 2016, "The Twin Instrument," Working Paper, IZA.
- Black, Sandra, P. J. Devereux and K. G. Salvanes, 2005, "The More the Merrier? The Effect of Family Composition on Children's Education," *Quarterly Journal of Economics*, 120(2), 2005, pp.669-700.
- Bloom, David, David Canning, and Jaypee Sevilla, 2001, "Economic Growth and Demographic Transition," Working Paper, NBER.

- Bloom, David, David Canning, Günther Fink, and Jocelyn Finlay, 2007, "Fertility, Female Labor Force Participation, and the Demographic Dividend," *Journal of Economic Growth*, 14(2), 79-101.
- Bongaarts, John, 1975, "A Method for the Estimation of Fecundability," *Demography* 12, 645-660.
- Bronars, Stephen, and Jeffrey Grogger, 1994, "The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment," *American Economic Review*, 84(5), 1141-1156.
- Bruijns, Sanne, 2014, "Children and Their Parents' Labor Supply: Evidence from 2012 Dutch Data," Working Paper, University of Amsterdam.
- Butikofer, Aline, 2011, "Sibling Sex Composition and the Cost of Children," Working Paper.
- Caceres-Delpiano, Julio, 2006, "The Impact of Family Size on Investment in Child Quality," *Journal of Human Resources* 41(4), p. 738-754.
- Carlinger, Geoffrey, Christopher Robinson, and Nigel Tomes, 1980, "Female Labor Supply and Fertility in Canada," *Canadian Journal of Economics* 13(1): 46-64.
- Clarke, Damian, 2016, "Fertility and Causality," Working Paper, University of Oxford, Centre for the Study of African Economies.
- Cristia, Julian, 2008, "The Effect of a First Child on Female Labor Supply: Evidence from Women Seeking Fertility Services," *Journal of Human Resources*, 43 (3), 487-510.
- Del Boca, Daniela, 2015, "Child Care Arrangements and Labor Supply," Working Paper, Inter-American Development Bank.
- Del Boca, Daniela, Silvia Pasqua, and Chiara Pronzato, 2005, "Fertility and Employment in Italy, France, and the UK," *Labour* 19(1): 51-77.
- Ebenstein, Avraham, 2010, "The "Missing Girls" of China and the Unintended Consequences of the One Child Policy," *Journal of Human Resources*, 45(1): 87-115.
- Edwards, Linda, and Elizabeth Field-Hendry, 2002, "Home-Based Work and Women's Labor Force Decisions," *Journal of Labor Economics* 20(1): 170-200.
- Feng, Wang and Yang Quanhe, 1996, "Age at Marriage and the First Birth Interval: The Emerging Change in Sexual Behavior Among Young Couples in China," *Population and Development Review* 22(2), 299-320.
- Galor, Oded, 2012, "The demographic transition: causes and consequences," *Cliometrica*, 6(1), 1-28.
- Galor, Oded and David Weil, 1996, "The Gender Gap, Fertility, and Growth," *American Economic Review* 86(3): 374-387.

- Godefroy, Raphael, 2016, "How Women's Rights Affect Fertility: Evidence from Nigeria," Working Paper.
- Goldin, Claudia, and Lawrence Katz, 2003, "Mass Secondary Schooling and the State," Working Paper, NBER.
- Goldin, Claudia, 1990, *Understanding the Gender Gap: An Economic History of American Women*, New York: Oxford University Press.
- Gronau, Reuben. "Home Production -- A Survey", in Orley Ashenfelter and Richard Layard, eds. *"Handbook of Labor Economics"*, Vol 1. Amsterdam: North-Holland, 1986, pp. 273-304.
- Herbst, Chris, 2015, "The Rising Cost of Child Care in the United States: A Reassessment of the Evidence," Working Paper, Arizona State University.
- Hoekstra, Chantal, Zhen Zhen Zhao, Cornelius Lambalk, Gonneke Willemsen, Nicholas Martin, Dorret Boomsma, and Grant Montgomery, 2007, "Dizygotic Twinning," *Human Reproduction Update*, p. 1-11.
- Hupkau, Claudia, and Marion Leturcq, 2016, "Fertility and Labor Supply: New Evidence from the UK," Working Paper, Centre for Economic Performance, London School of Economics.
- ICF International. Demographic and Health Surveys (various) [Datasets], 2015, Calverton, Maryland: ICF International [Distributor].
- Jaffe, A. J. and K. Azumi, 1960, 'The birth rate and cottage industries in under-developed countries', *Economic Development and Cultural Change* 9, 52-63.
- Jayachandran, Seema, and Rohini Pande, 2015, "Why Are Indian Children So Short?" Working Paper, NBER.
- Juul, S., W. Karmaus, J. Olsen, 1999, "Regional differences in waiting time to pregnancy: pregnancy-based surveys from Denmark, France, Germany, Italy and Sweden," *Human Reproduction* 14, 1250-1254.
- Klemp, Marc and Jacob Weisdorf, 2016, "Fecundity, Fertility and the Formation of Human Capital", Working Paper, University of Warwick.
- Kubota, So, 2016, "Child Care Costs and Stagnating Female Labor Force Participation in the US," Working Paper, Princeton University.
- Kupinsky, S, 1977, *The Fertility of Working Women: A Synthesis of International Research*. New York: Praeger Publishers.
- Laughlin, Lynda, 2013, "Who's Minding the Kids? Child Care Arrangements: Spring 2011," U.S. Census Bureau, Report P70-135.



Lundborg, Petter, Erik Plug, and Astrid Wurtz Rasmussen, 2016, "Can Women Have Children and a Career? IV Evidence from IVF Treatments," forthcoming, *American Economic Review*.

The Maddison-Project, <http://www.ggdcc.net/maddison/maddison-project/home.htm>, 2013 version.

Mammen, Kristin and Christina Paxson, 2000, "Women's Work and Economic Development," *Journal of Economic Perspectives*, 14(4), 141-164.

McCabe, James and Mark R. Rosenzweig, 1976, "Female Labor Force Participation, Occupational Choice, and Fertility in Developing Countries," *Journal of Development Economics* 3, 141-160.

Minnesota Population Center, 2015, North Atlantic Population Project: Complete Count Microdata, Version 2.2 [Machine-readable database], Minneapolis: Minnesota Population Center.

Olivetti, Claudia, and Barbara Petrongolo, 2017, "The Economic Consequences of Family Policies: Lessons from a Century of Legislation in High-Income Countries," *Journal of Economic Perspectives*, 31(1), 205-230.

Rosenzweig, Mark, and Kenneth Wolpin, 1980, "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, 48(1), 227-240.

Rosenzweig, Mark, and Kenneth Wolpin, 2000, "Natural 'Natural Experiments' in Economics," *Journal of Economic Literature*, 38(4), 827-874.

Rosenzweig, Mark, and Junsen Zhang, 2009, "Do Population Control Policies Induce More Human Capital Investment? Twin, Birth Weight and China's 'One-Child' Policy," *The Review of Economic Studies*, 76(3), 1149-1174.

Ruggles, Steven, 1994, "The Transformation of American Family Structure," *American Historical Review*, 99, 103-128.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek, 2010, Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database], Minneapolis: University of Minnesota.

Schultz, Theodore, 1991, "International Differences in Labor Force Participation in Families and Firms," Working Paper, Yale University.

Schultz, Theodore, 2008, "Population Policies, Fertility, Women's Human Capital, and Child Quality," In *Handbook of Development Economics, Volume Four*, ed. Paul Schultz T., Strauss John A., 3249-3303. Amsterdam: Elsevier Science B.V.

Sobek, Matthew and Sheela Kennedy, 2009, "The Development of Family Interrelationship Measures for International Census Data," Working Paper, University of Minnesota Population Center.

Staiger, Douglas and James Stock, 1997, "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65, 557-586.

Szulga, Radek, 2013, "A dynamic model of female labor force participation rate and human capital investment," Working Paper, SSRN.

United States Census Bureau, 1970, "National Income and Wealth," Series F 297-314, p. 243, retrieved from:

[http://www2.census.gov/library/publications/1975/compendia/hist\\_stats\\_colonial-1970/hist\\_stats\\_colonial-1970p1-chF.pdf](http://www2.census.gov/library/publications/1975/compendia/hist_stats_colonial-1970/hist_stats_colonial-1970p1-chF.pdf)

Willis, Robert, 1973, "A New Approach to the Economic Theory of Fertility Behavior," *Journal of Political Economy* 81, S14-S64.

Figure 1 - OLS, by Real GDP/Capita

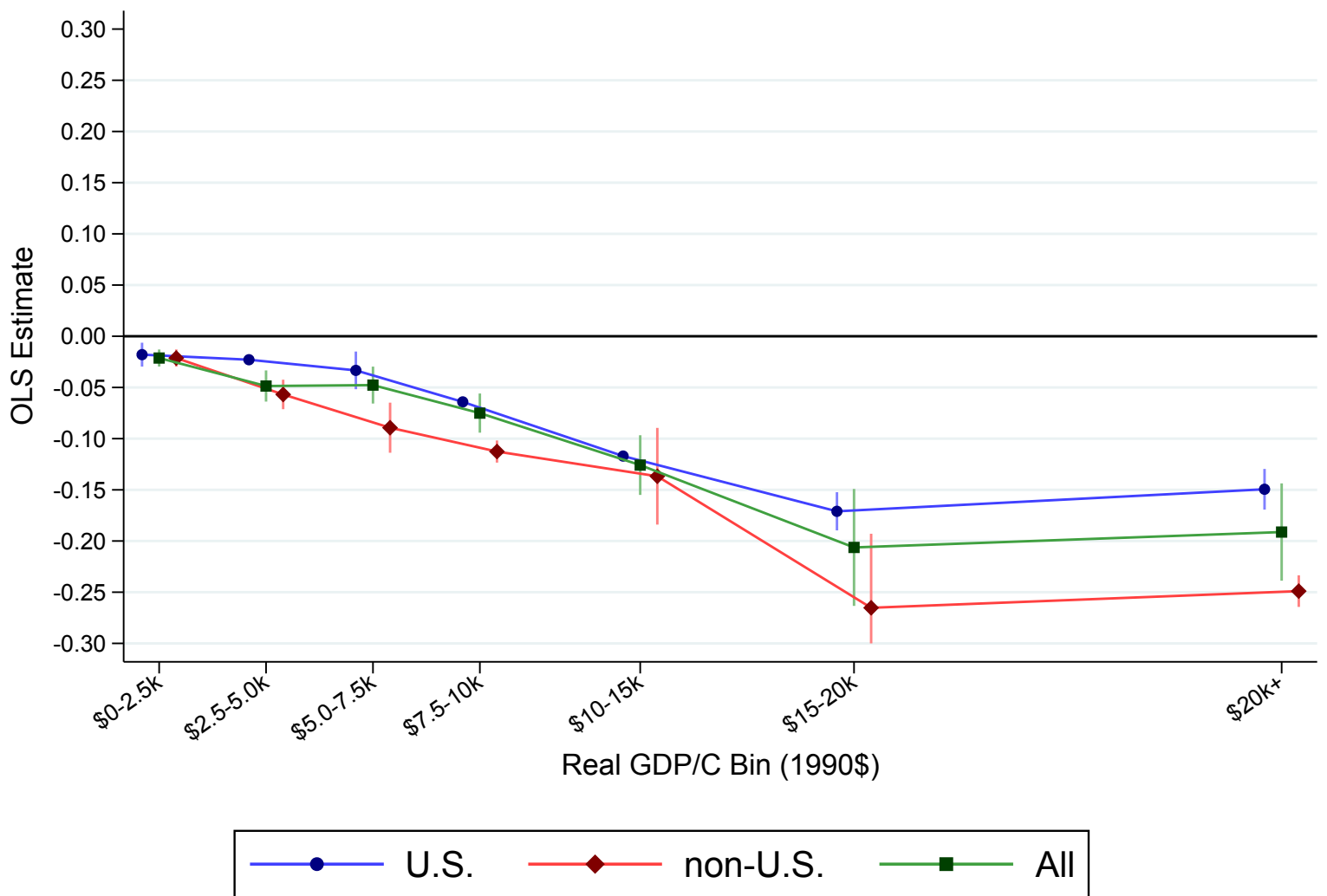


Figure 2 - OLS, U.S. by Time

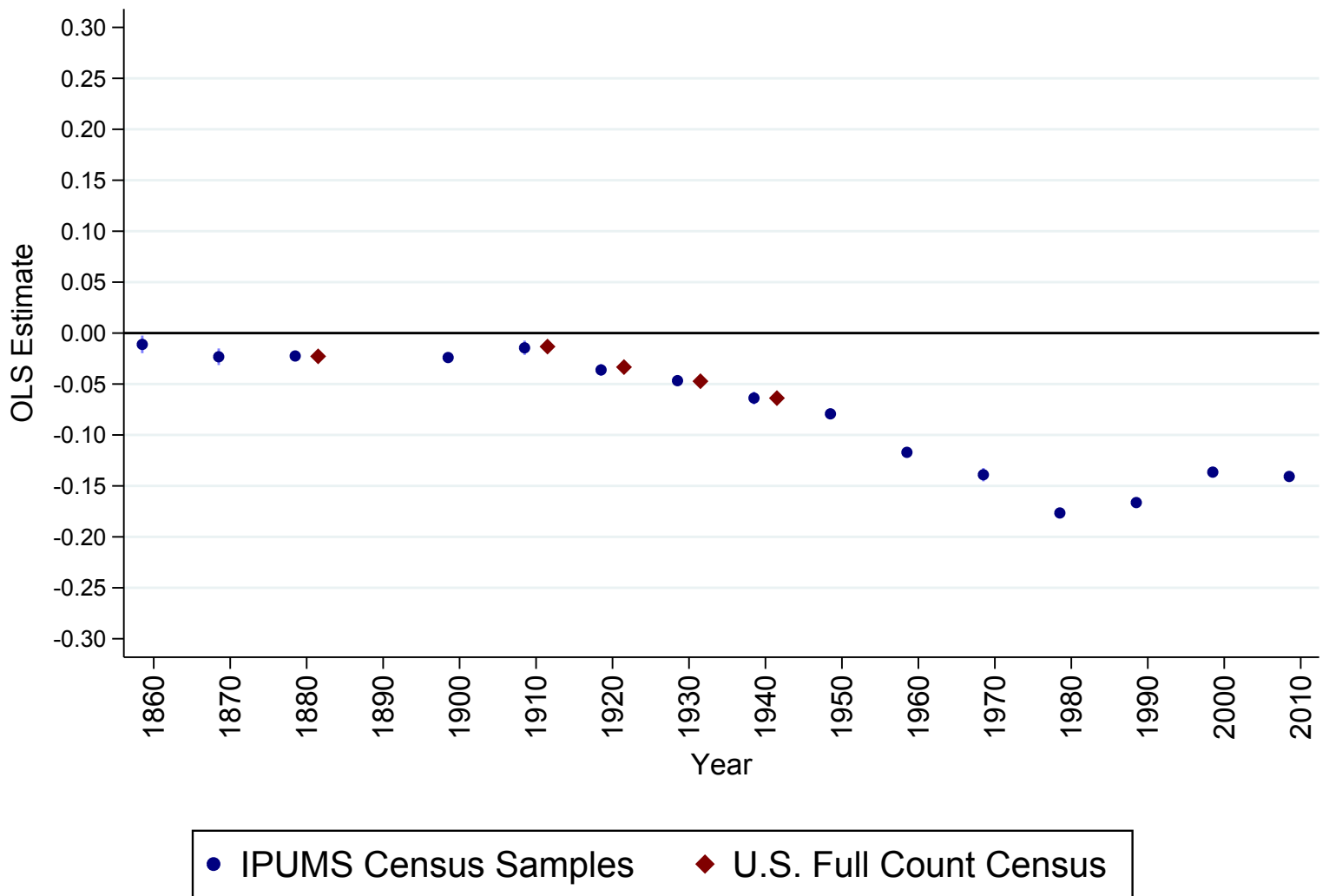


Figure 3 - OLS, by Time and Real GDP/Capita Bin

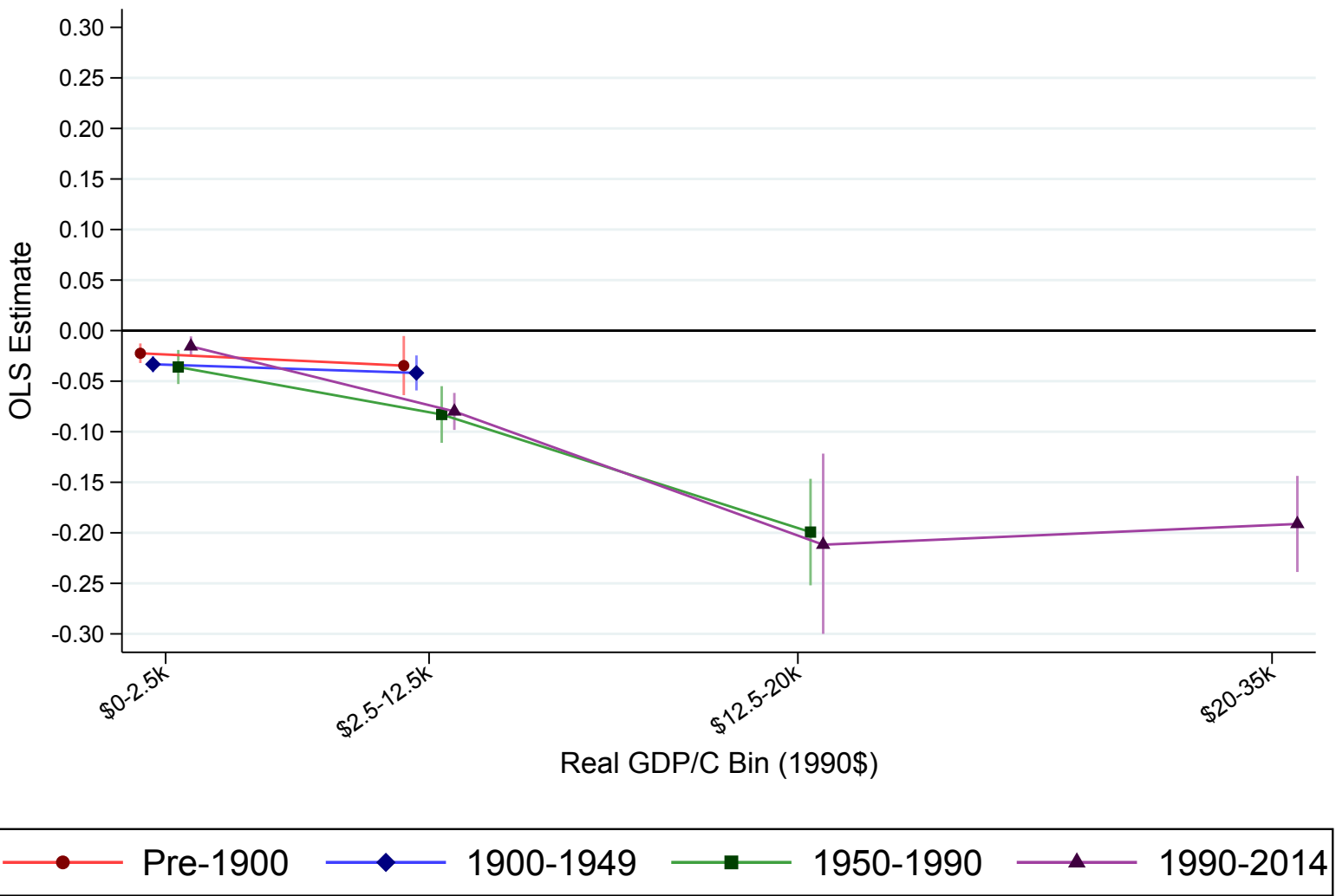


Figure 4 - Twin IV, by Real GDP/Capita

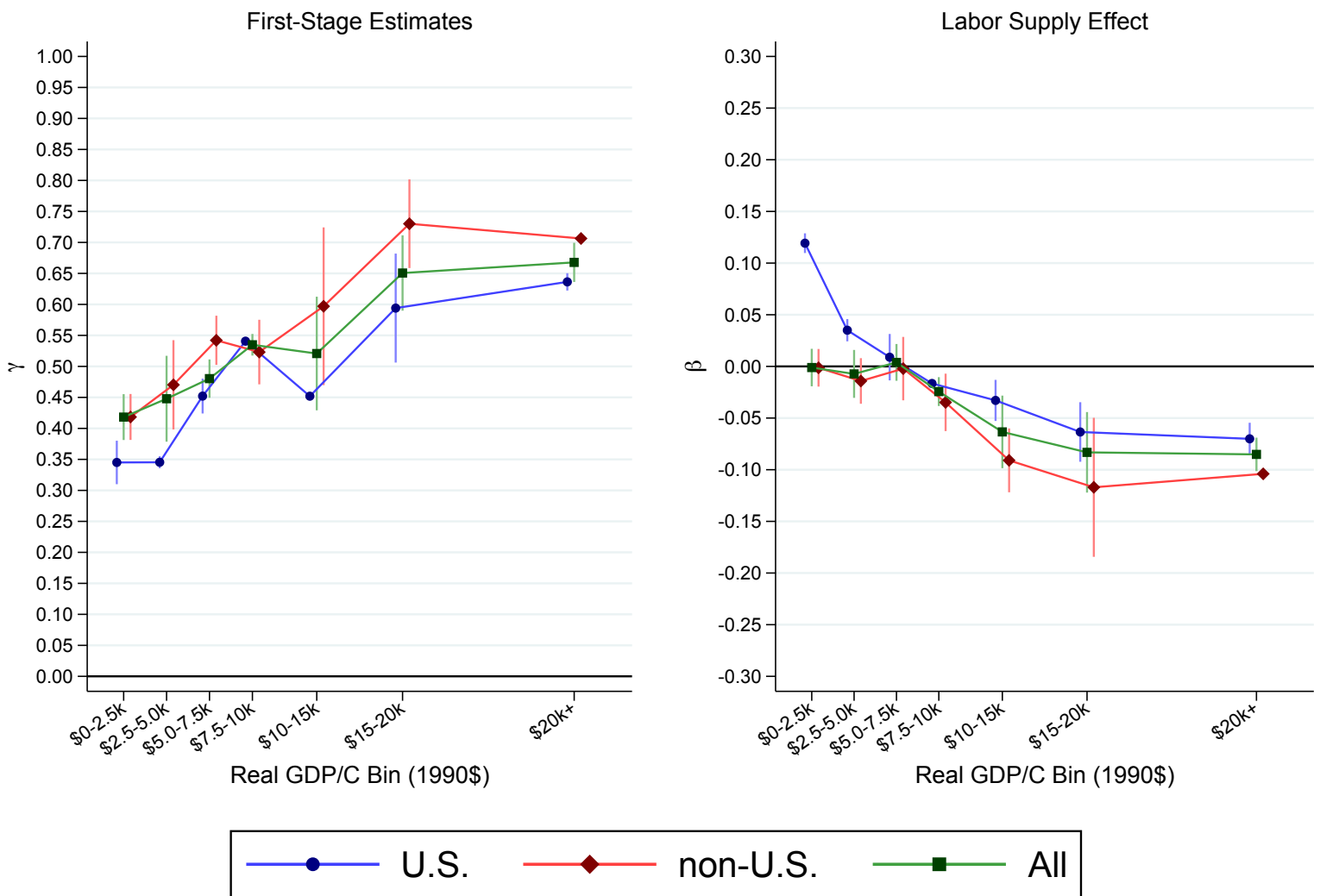


Figure 5 - Twin IV, U.S. by Time

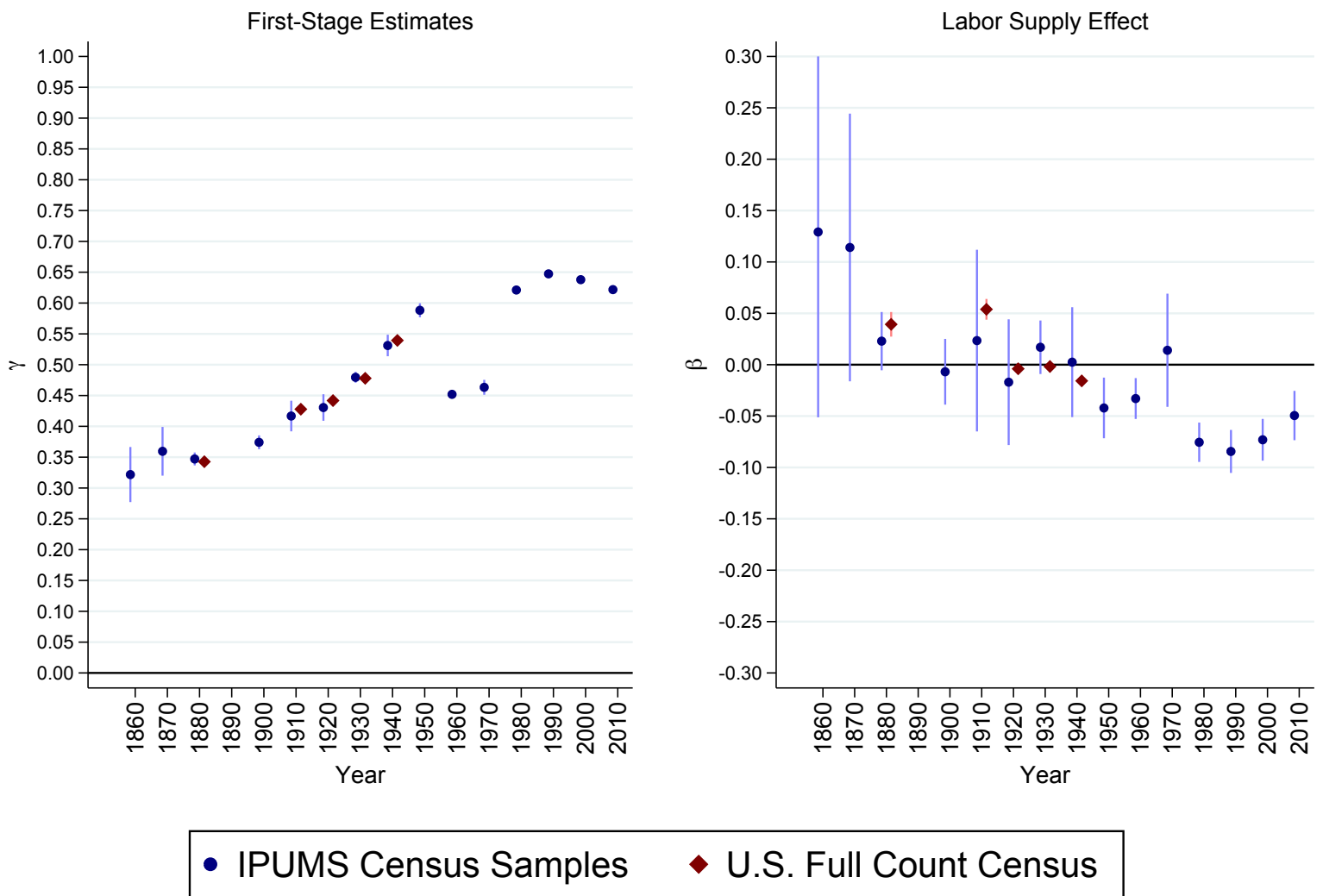


Figure 6 - Twin IV, by Time and Real GDP/Capita Bin

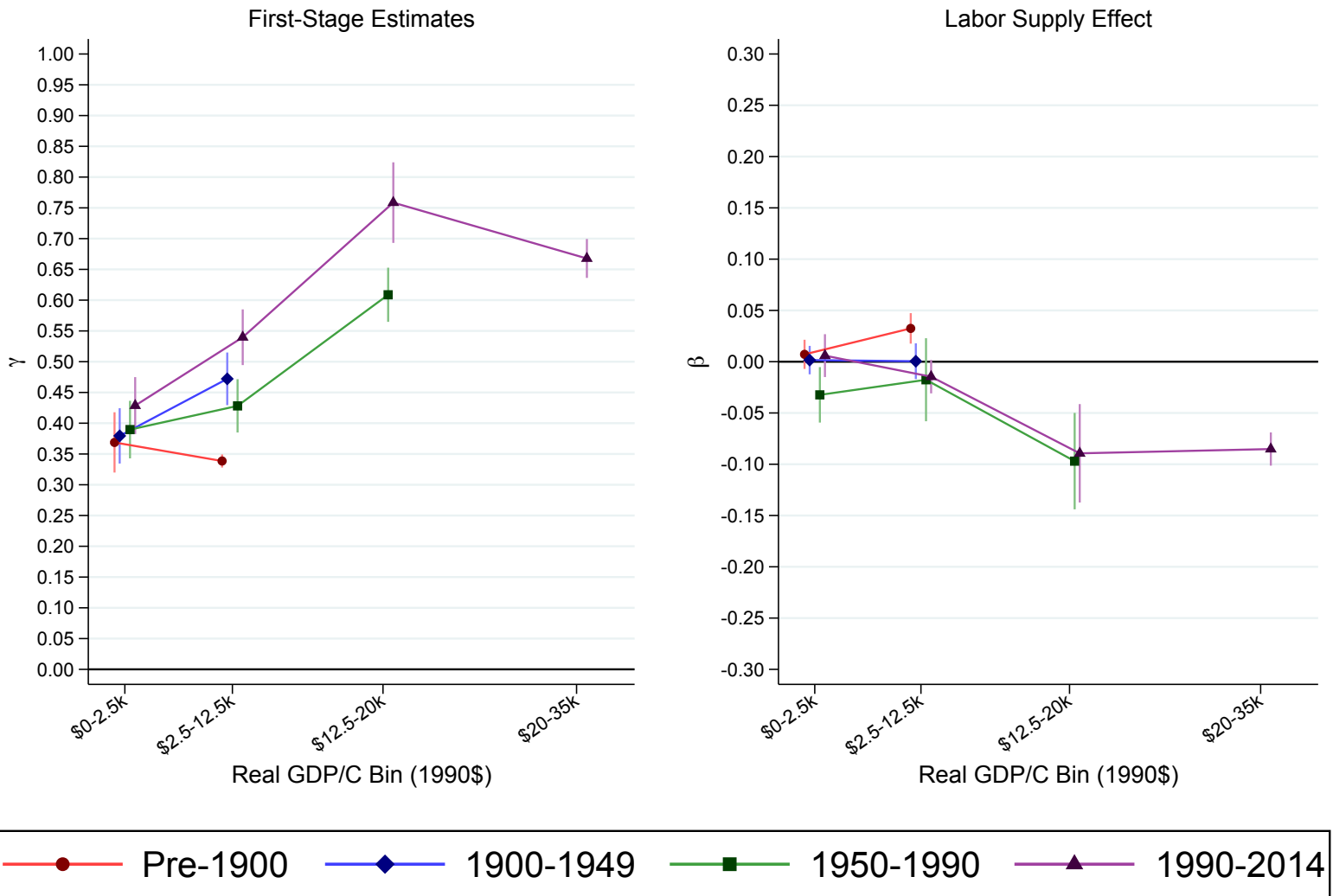




Figure 7 - U.S. States in 1930 and 1940, Twin IV

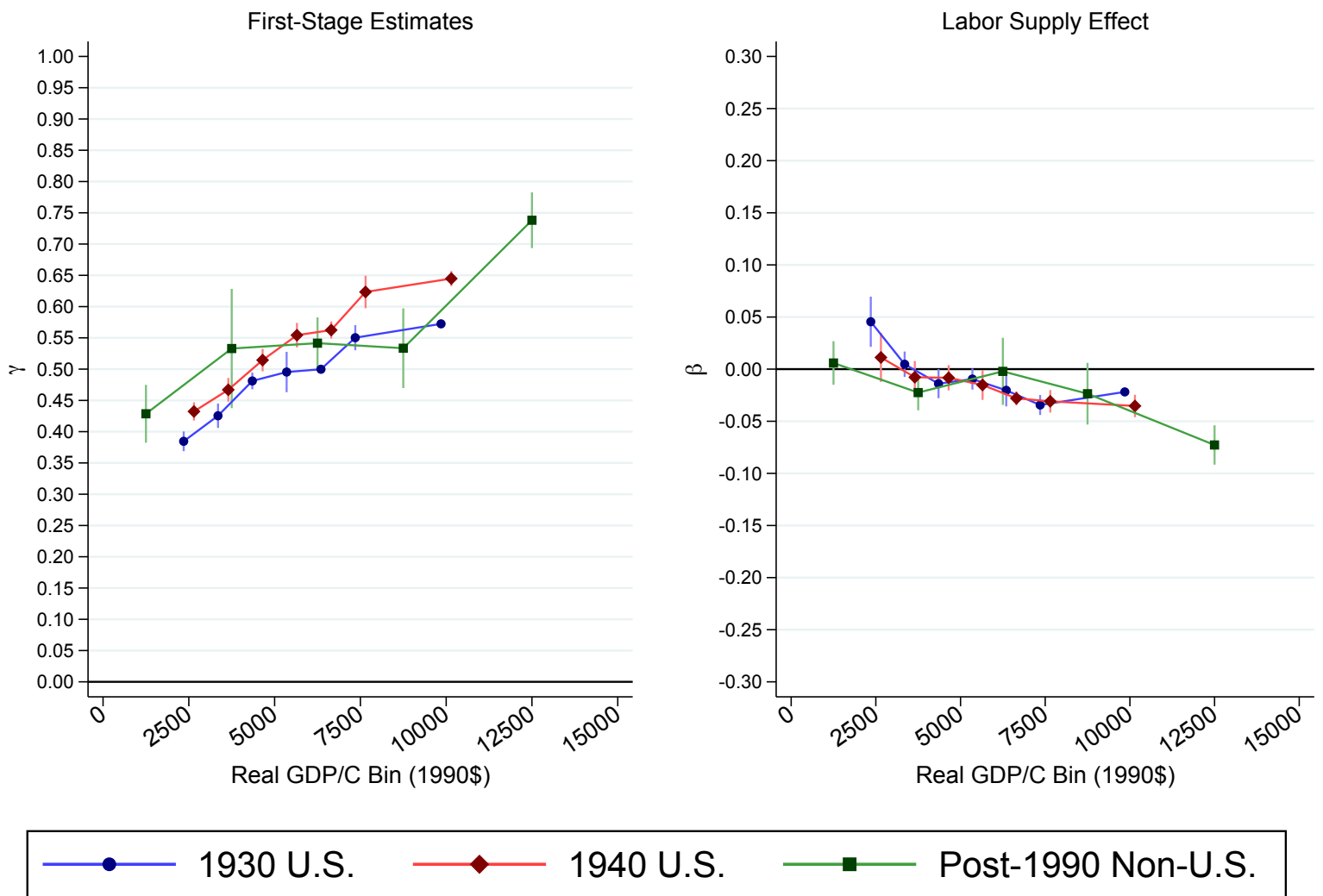


Figure 8 - Same Gender IV, by Real GDP/Capita

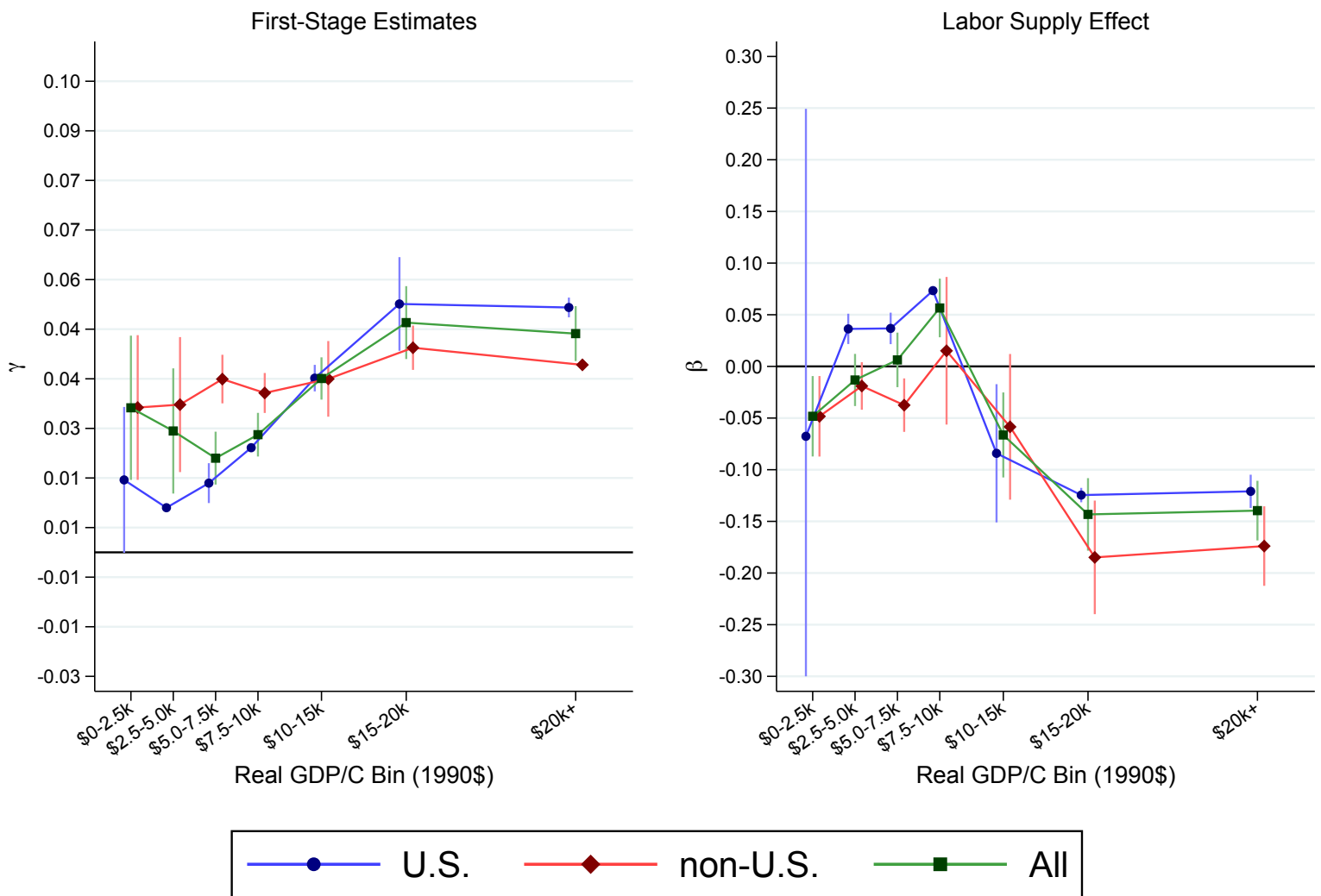


Figure 9 - Hours Conditional on Working, Twin IV

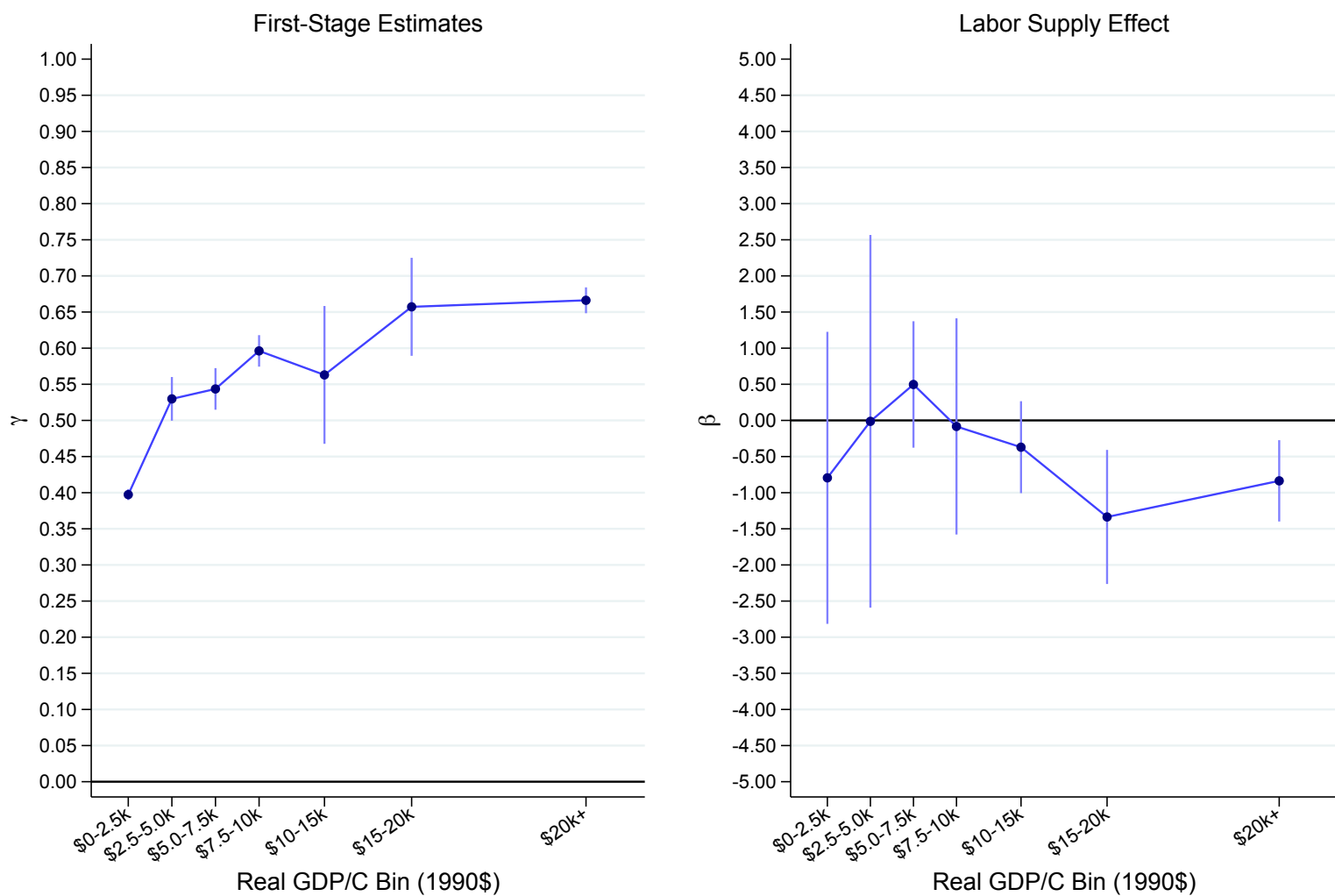


Figure 10 - Different Child Parities, Twin IV

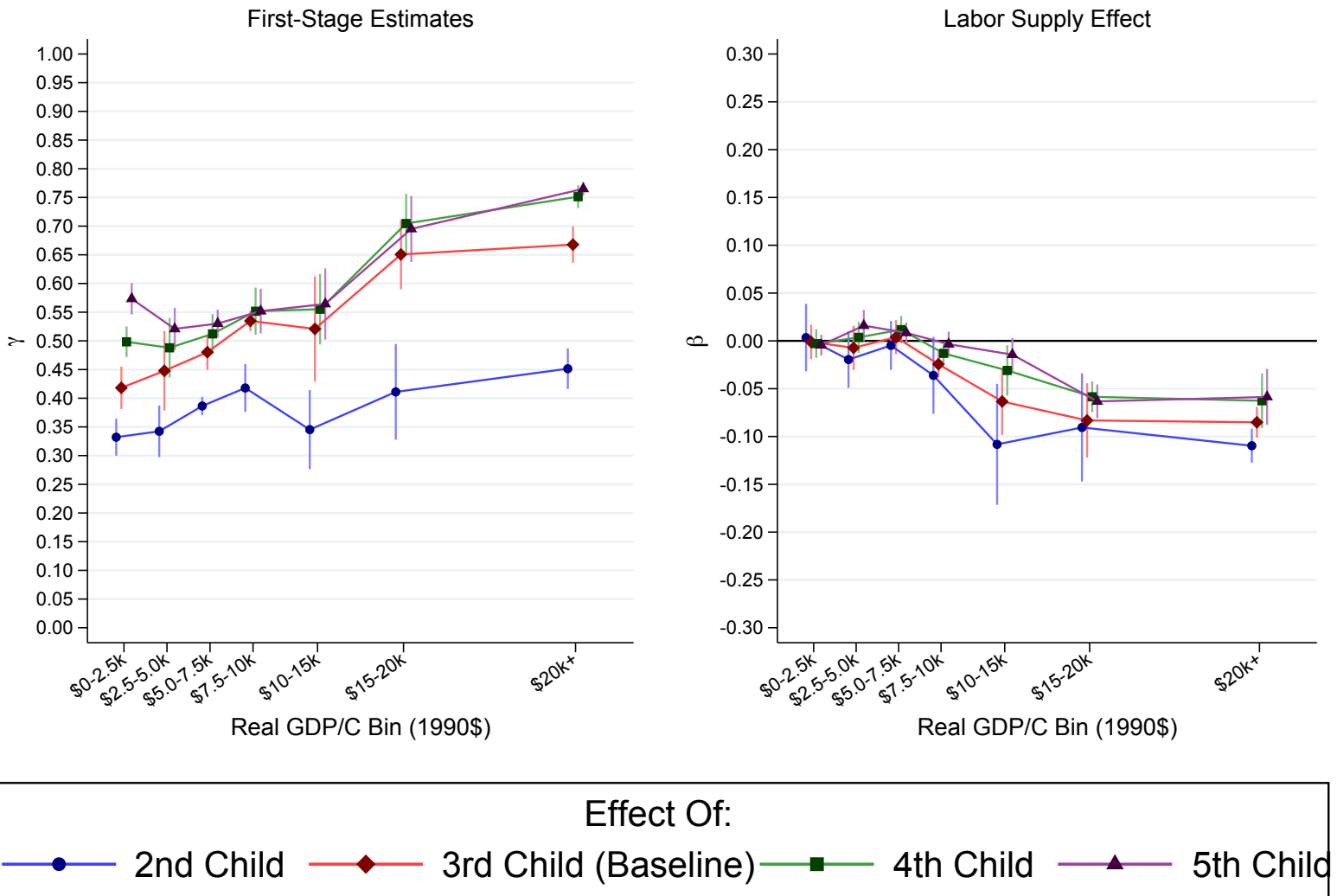


Figure 11 - By Gender of Twins

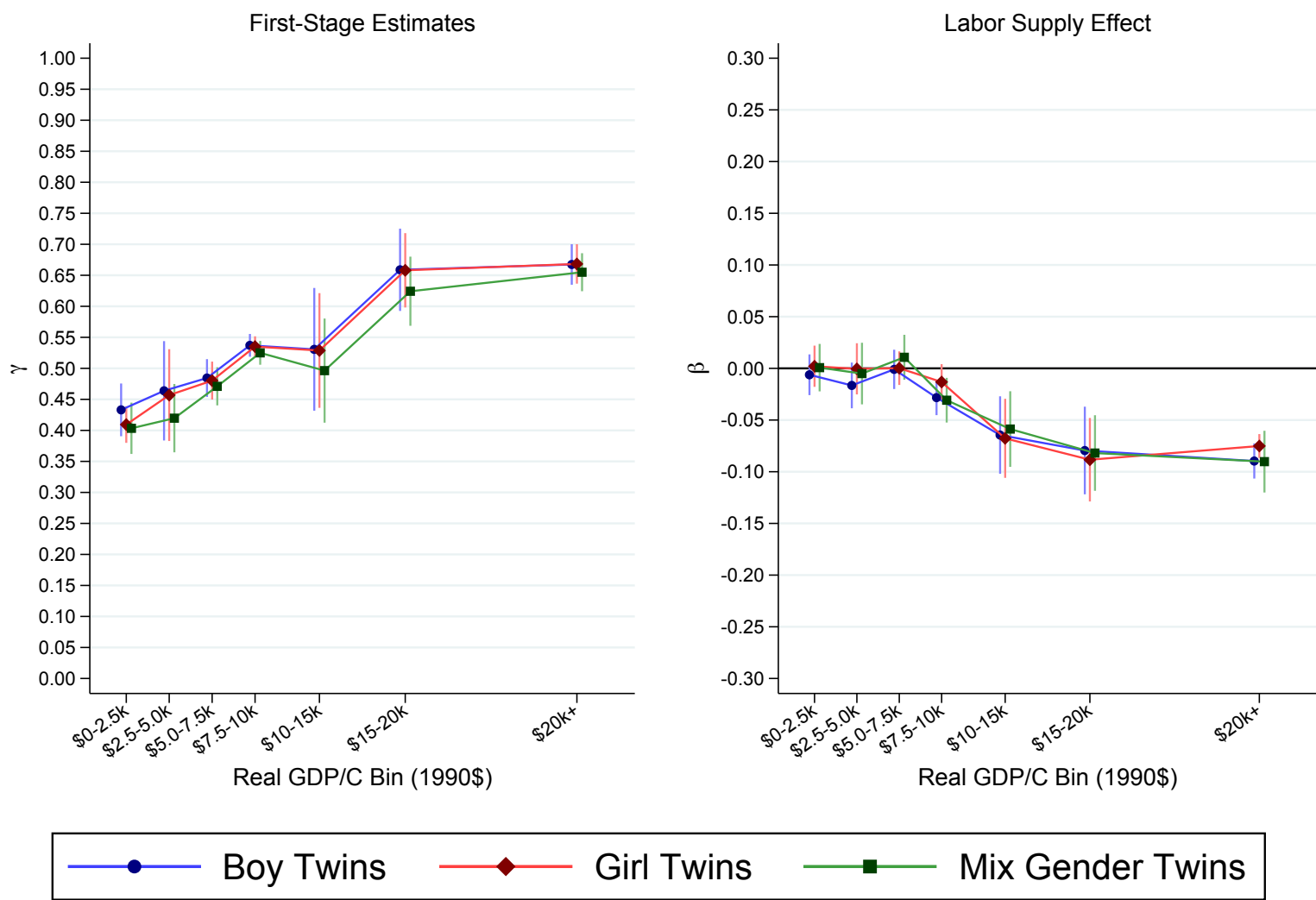


Figure 12 - Alternative Development Benchmark  
by U.S. State Mean Hourly Wage, 1940-2010, Twin IV

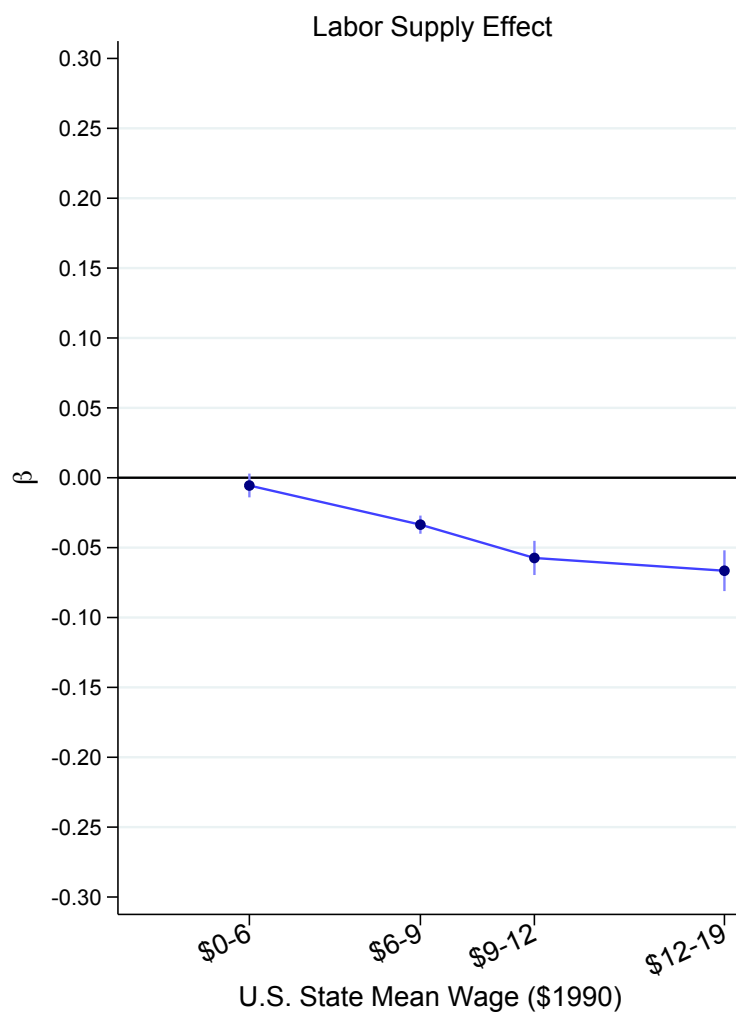
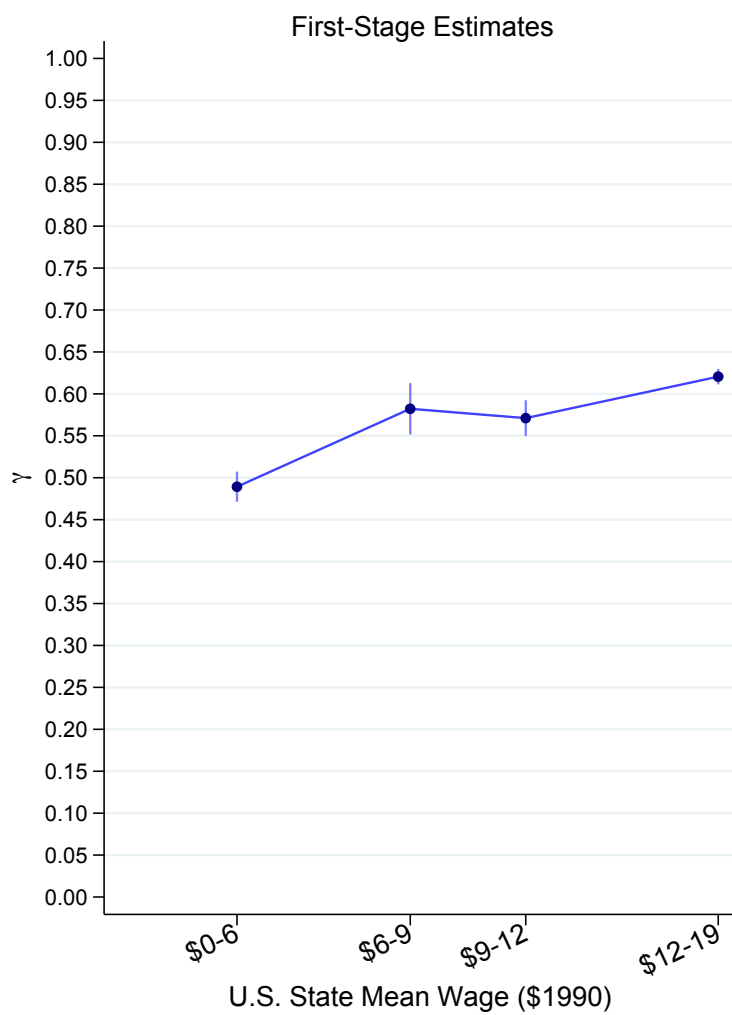


Figure 13 - Alternate Development, by Benchmark by Female Education, Twin IV

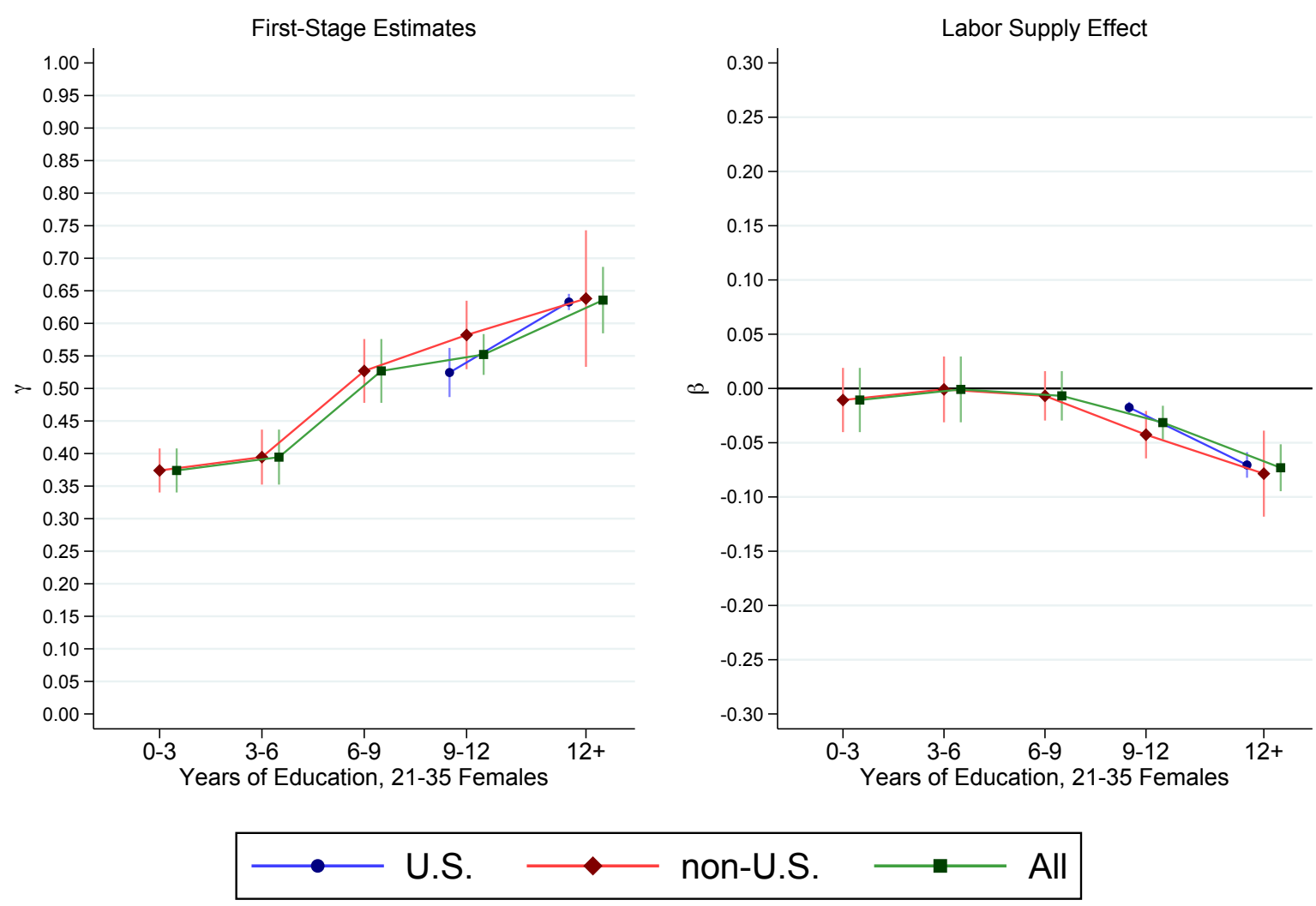


Figure 14 - By Definition of Twins

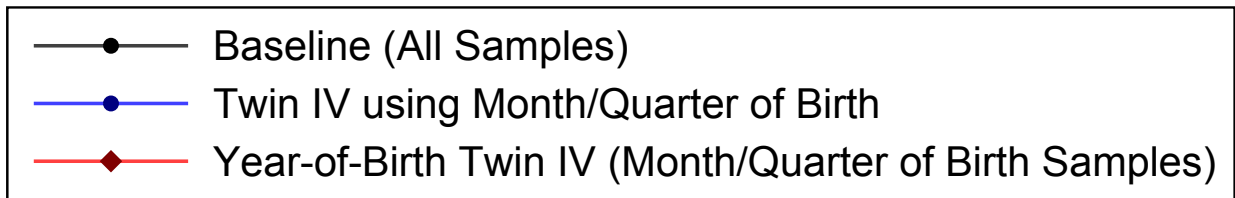
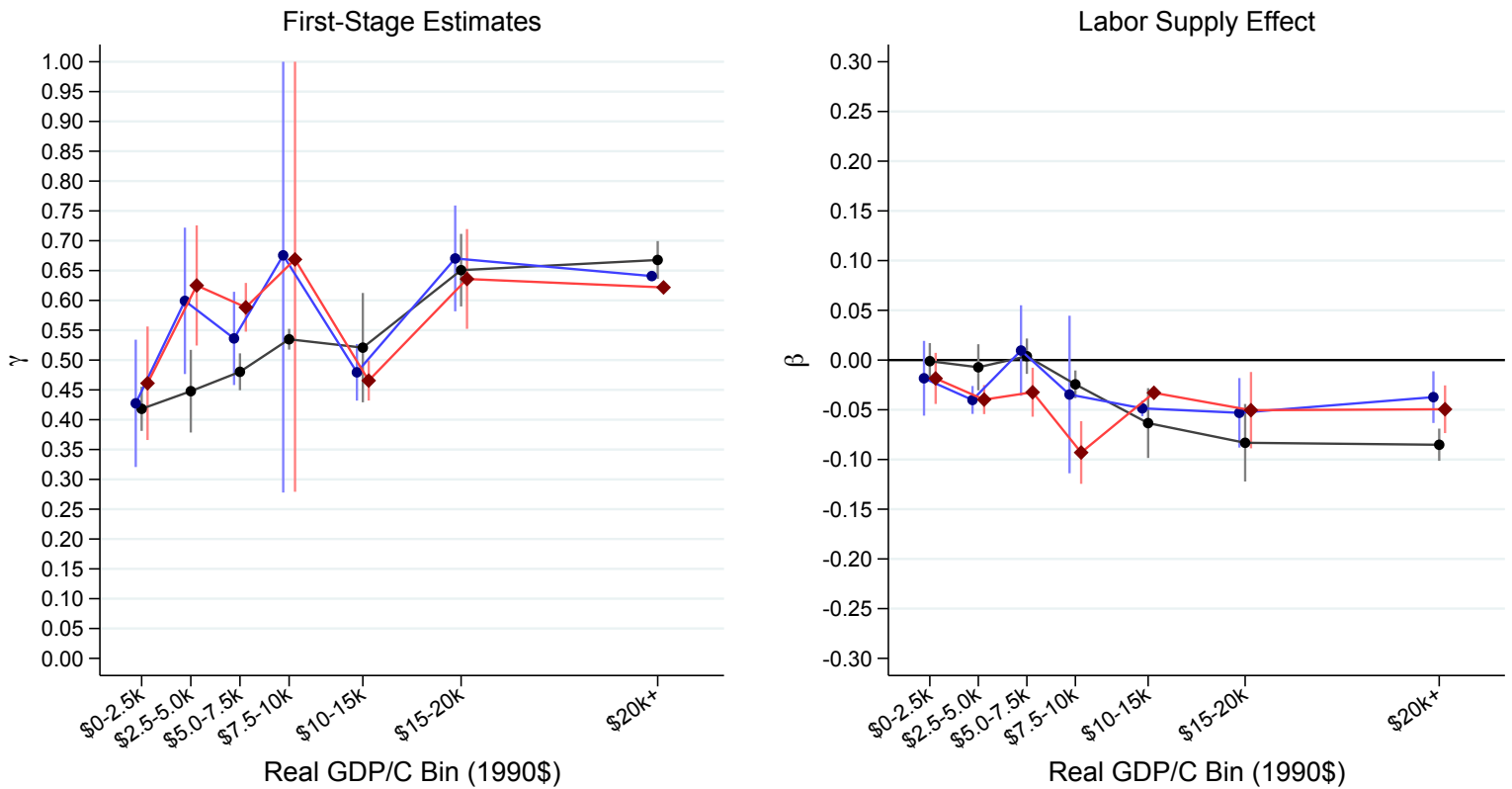




Figure 15 - By Alternative Labor Supply Measures, Twin IV

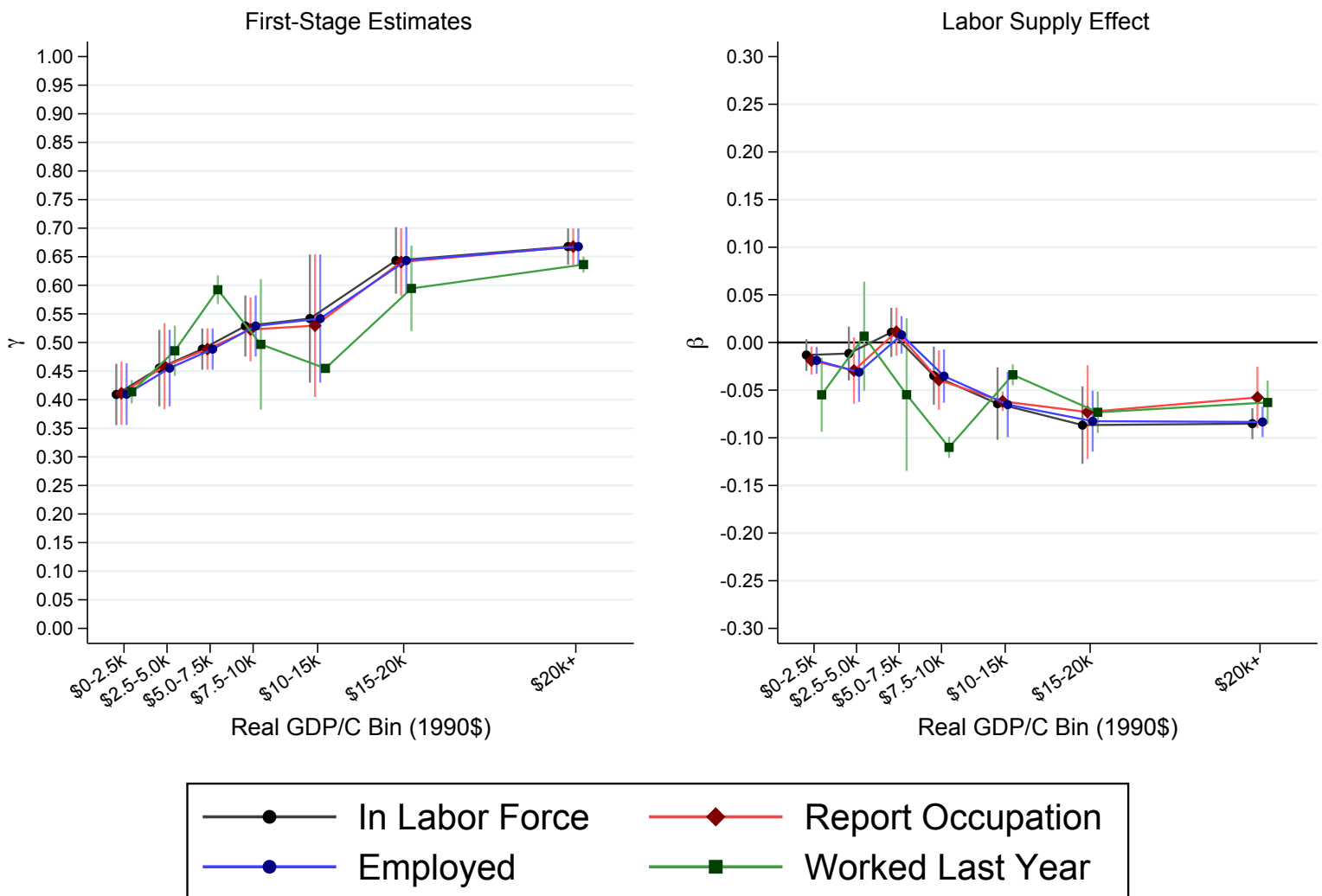


Figure 16 - Twin IV, by Real GDP/Capita  
Rescaled by Complier-Control Outcome Mean

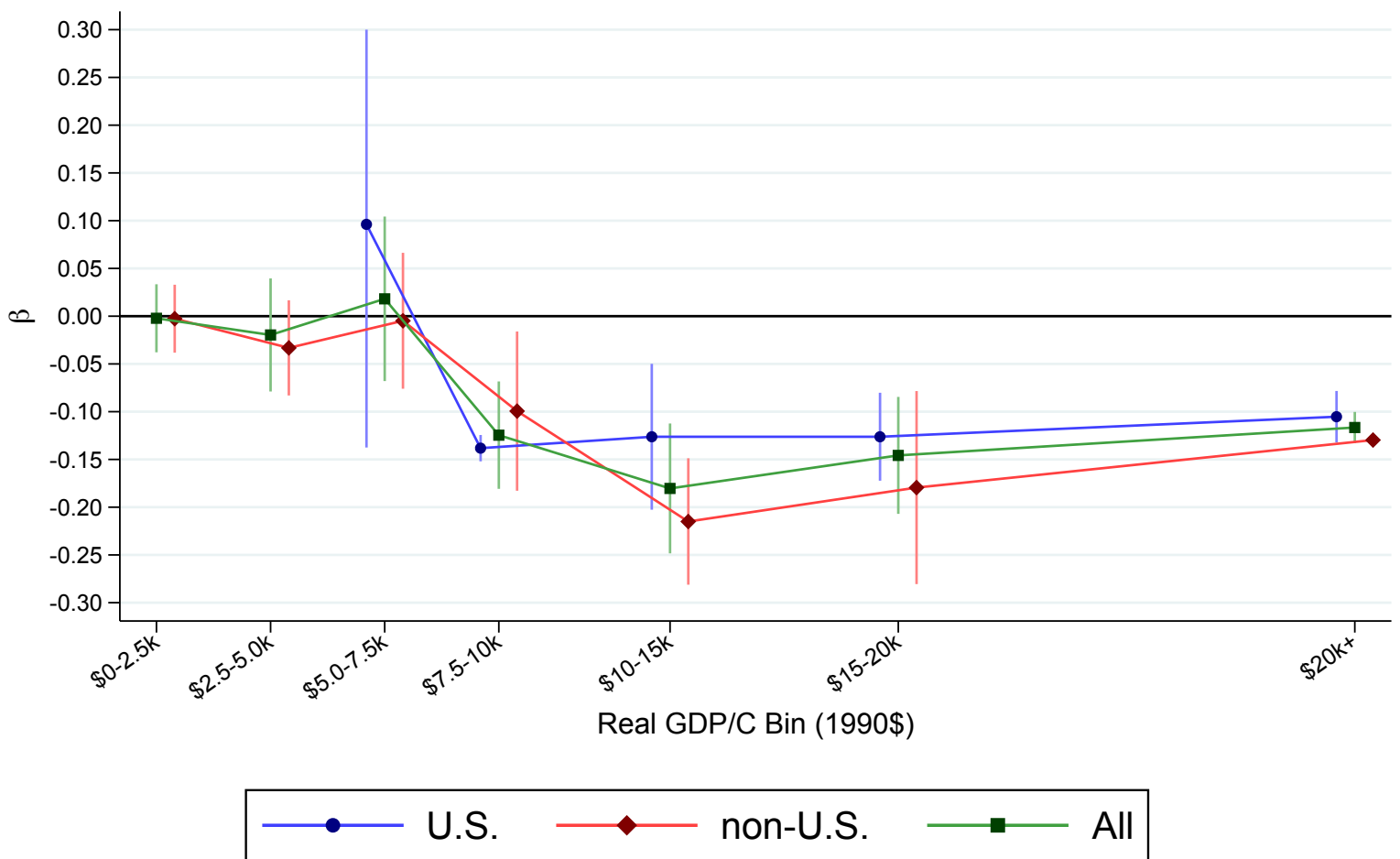
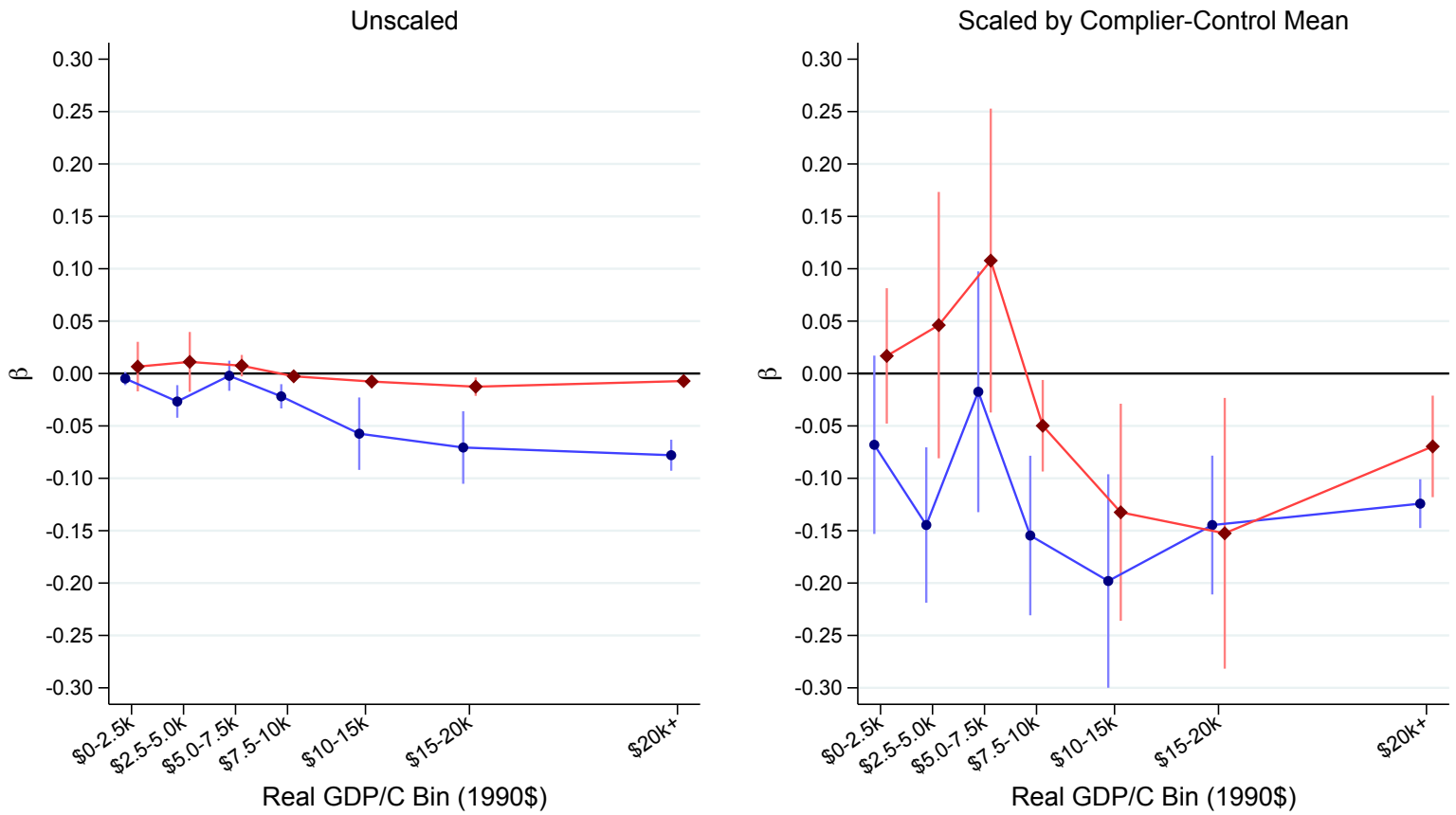


Figure 17 - By Class of Worker, Twin IV



Outcome:

—●— Works for Wages      —◆— Non-wage Worker

Figure 18 - By Agricultural Occupation of Worker, Twin IV

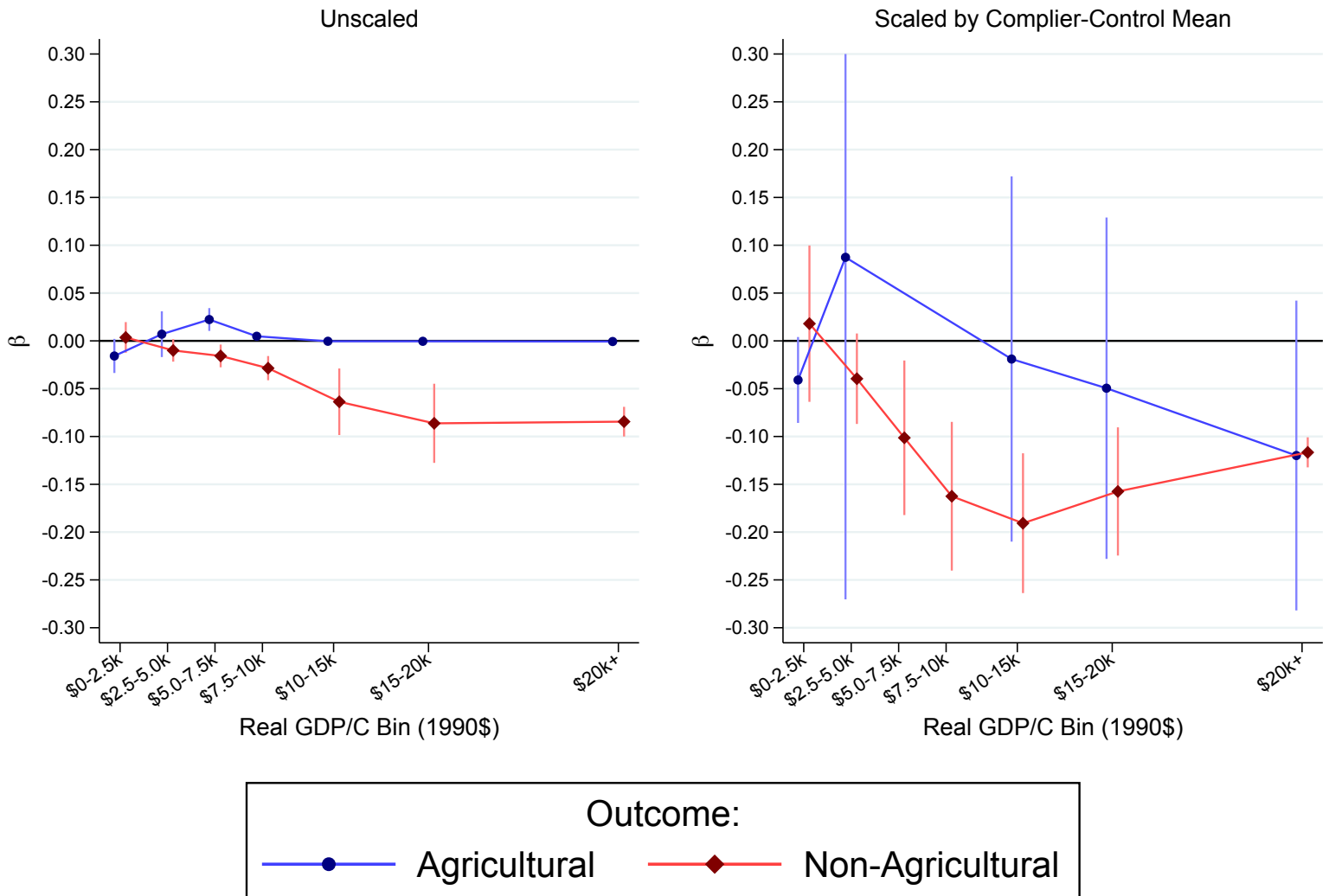


Figure 19 - By Professional Occupation of Worker, Twin IV

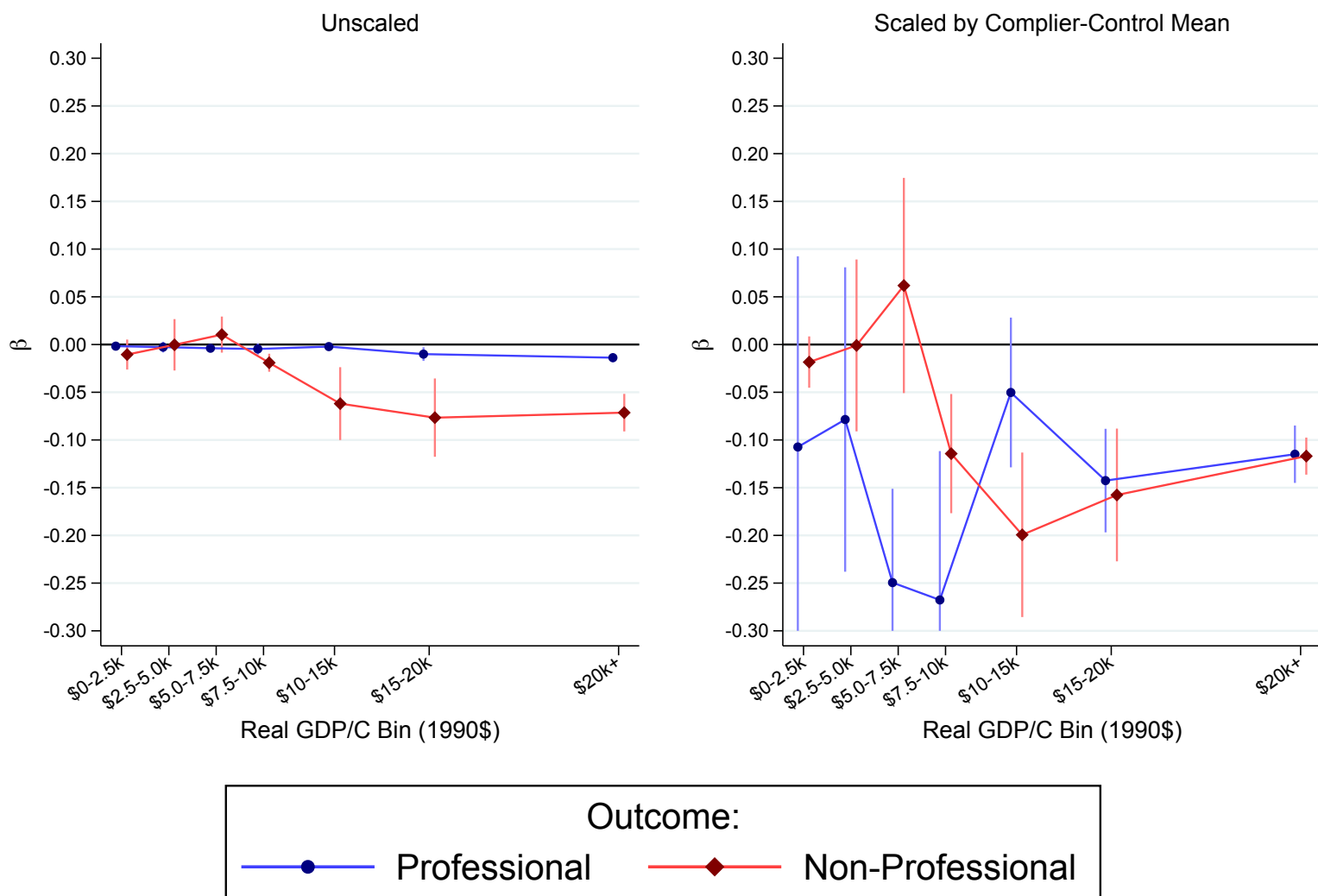


Figure 20 - By Husband and Wife, Twin IV

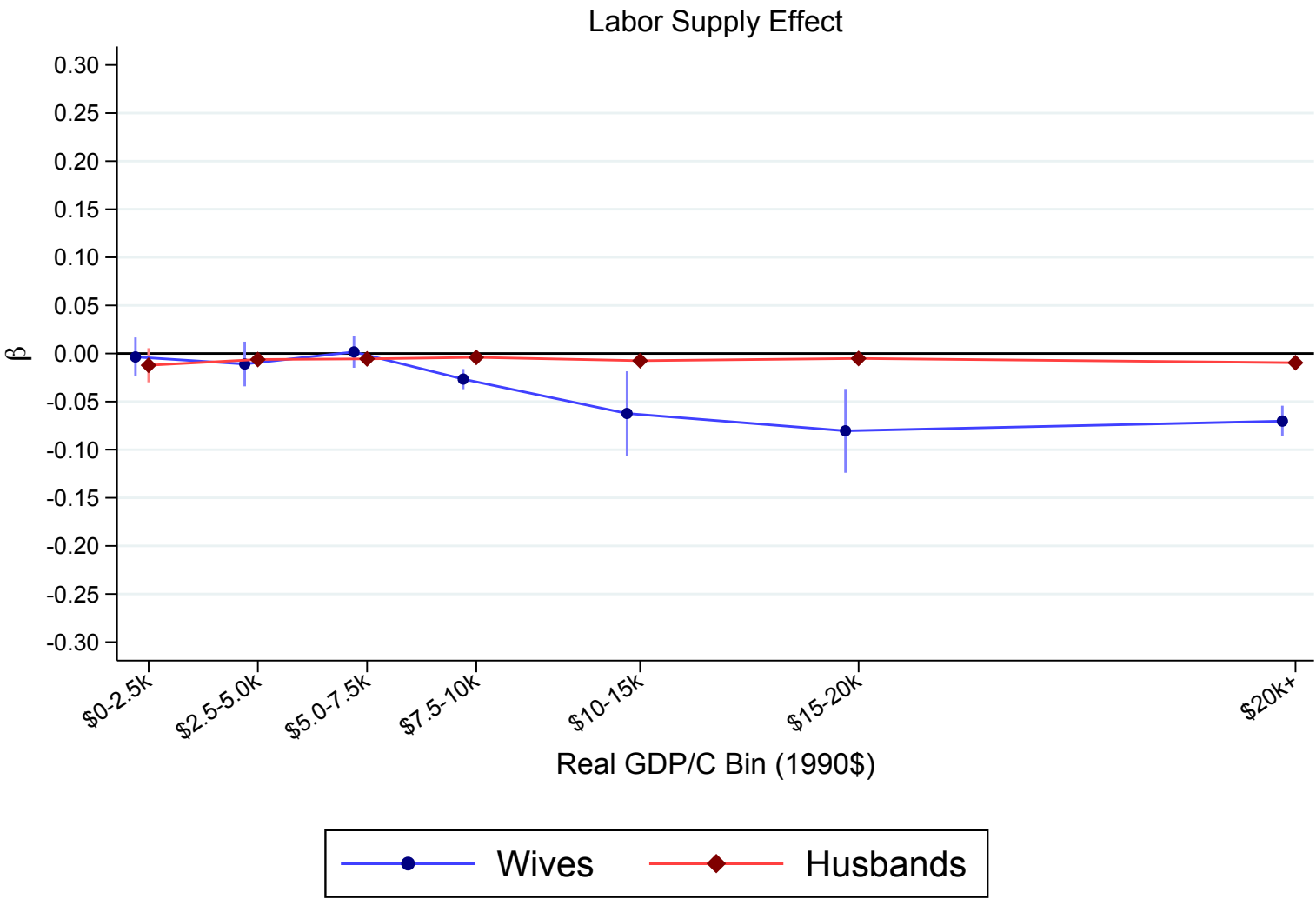


Figure 21 - By State and Husband Wage, U.S. 1940-2010, Twin IV

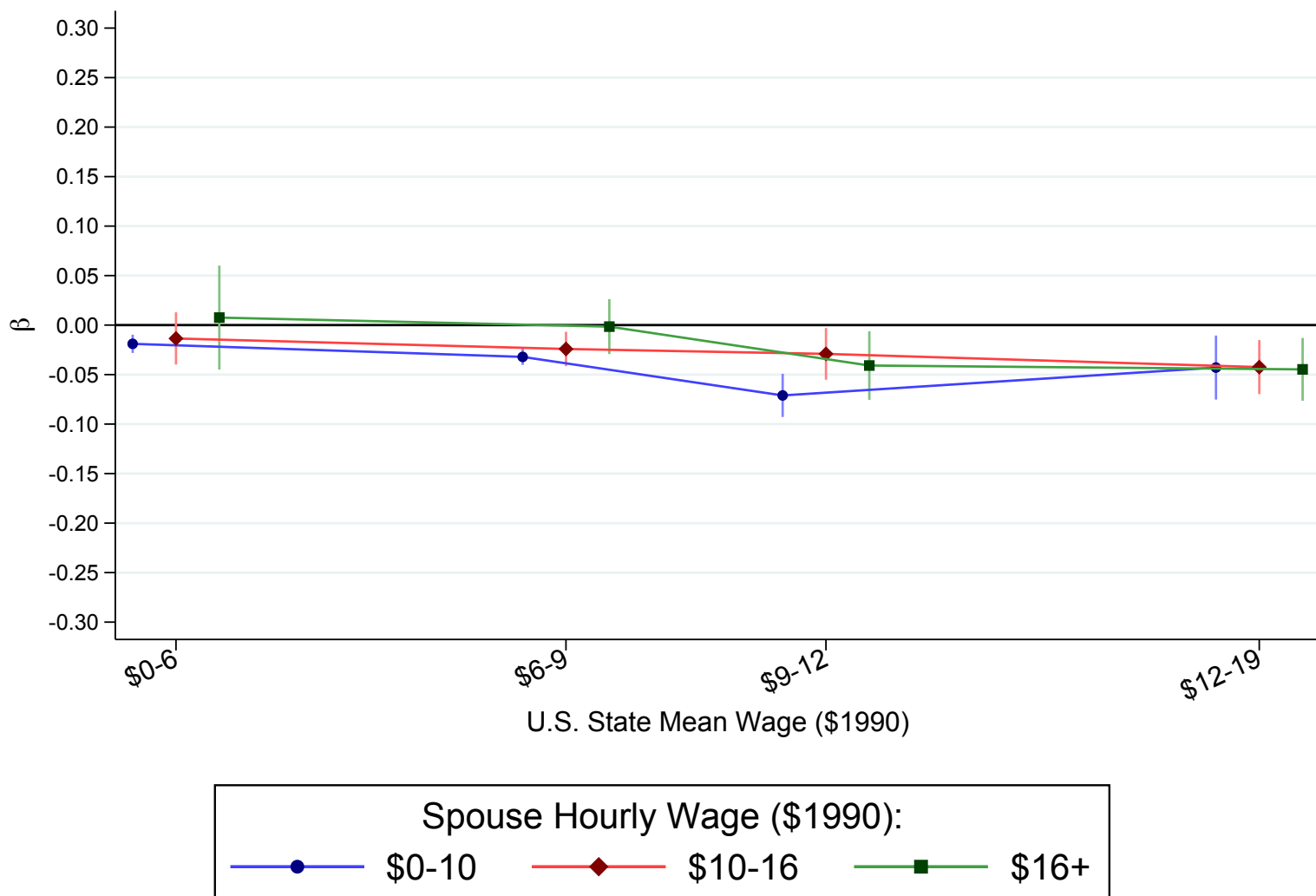


Figure 22 - By Presence of Grandparent, Twin IV

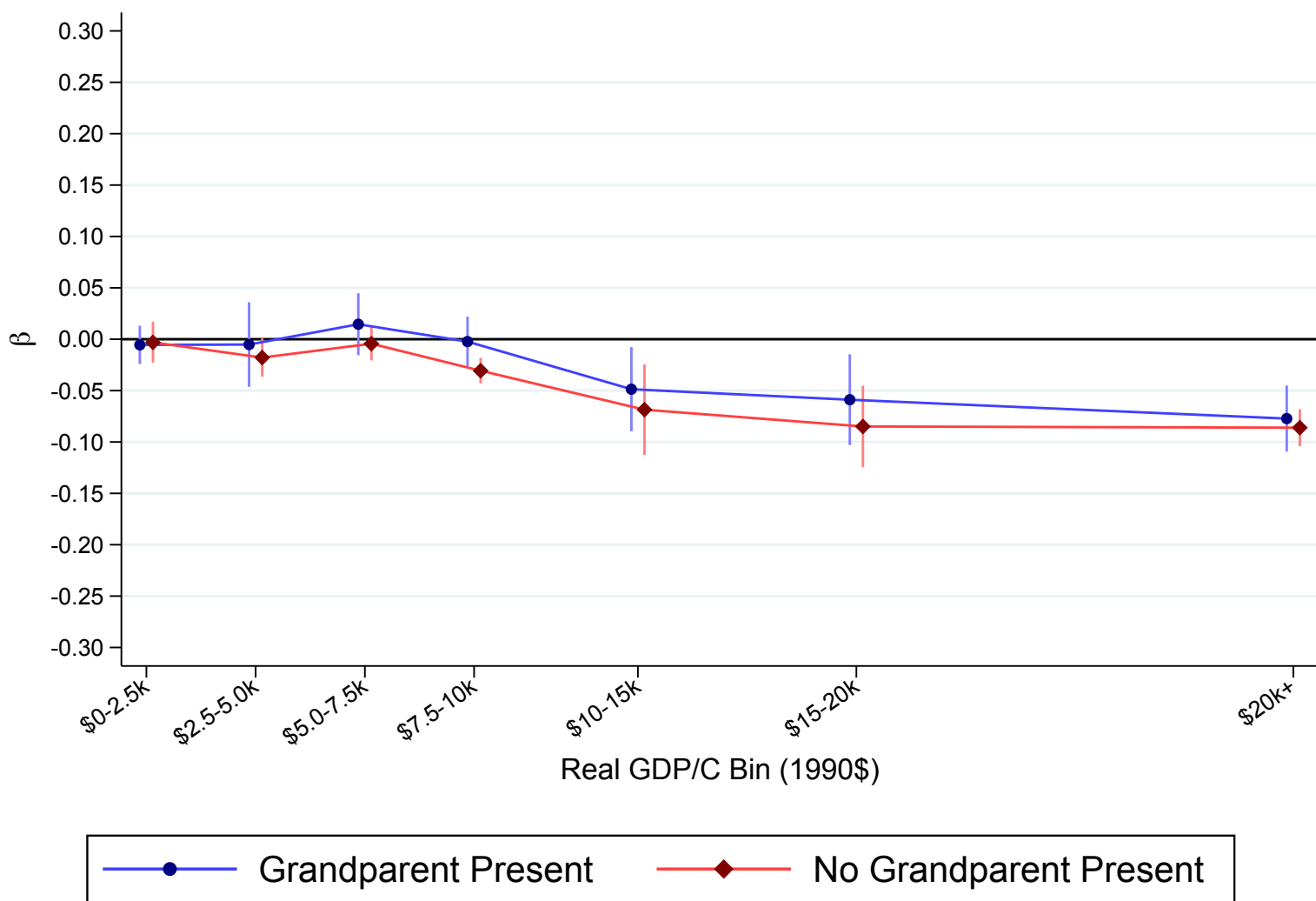




Figure 23 - By Number of non-Mother Adult Women in Household, Twin IV

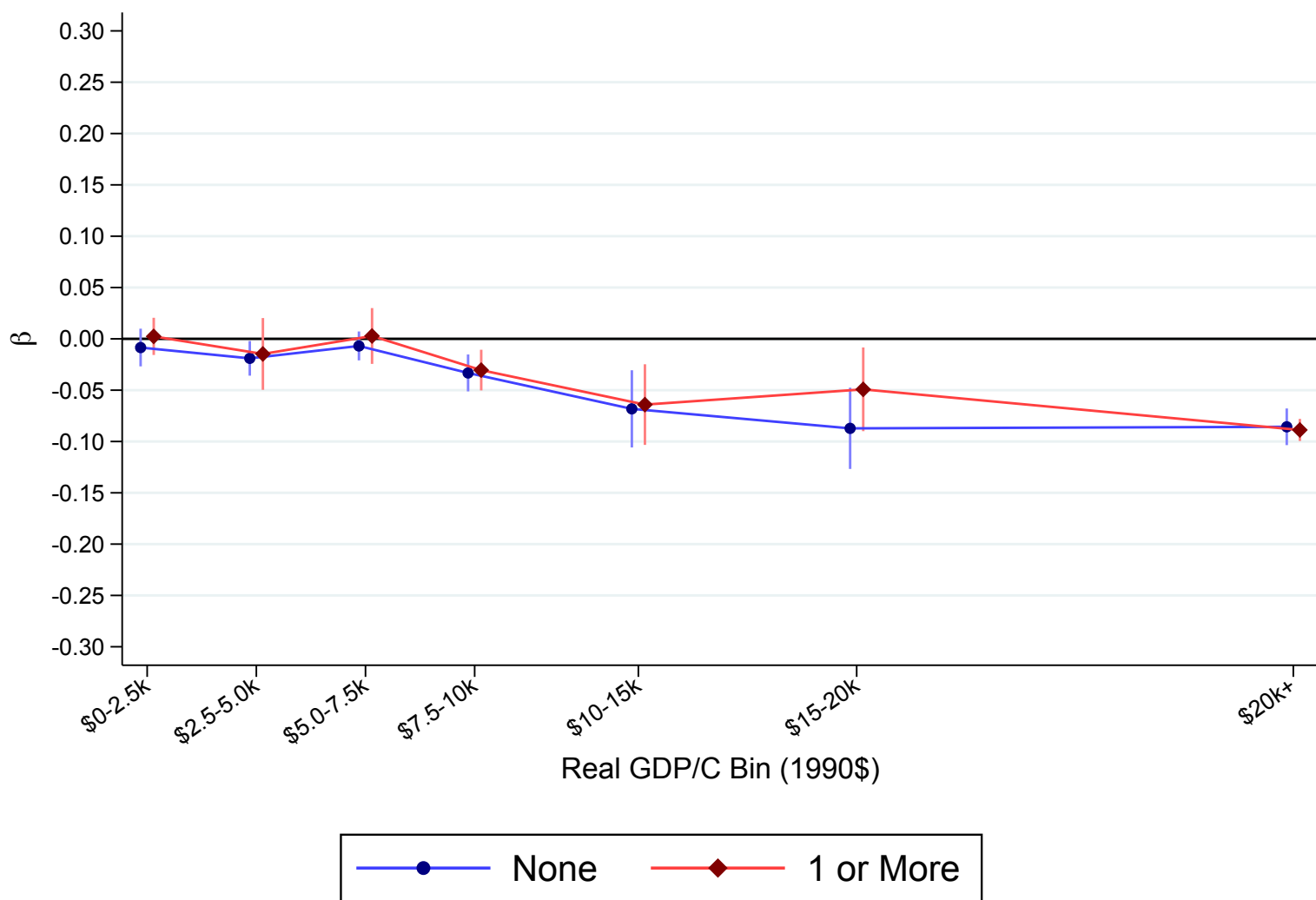


Figure 24 - By Mother's State of Birth, U.S., Twin IV

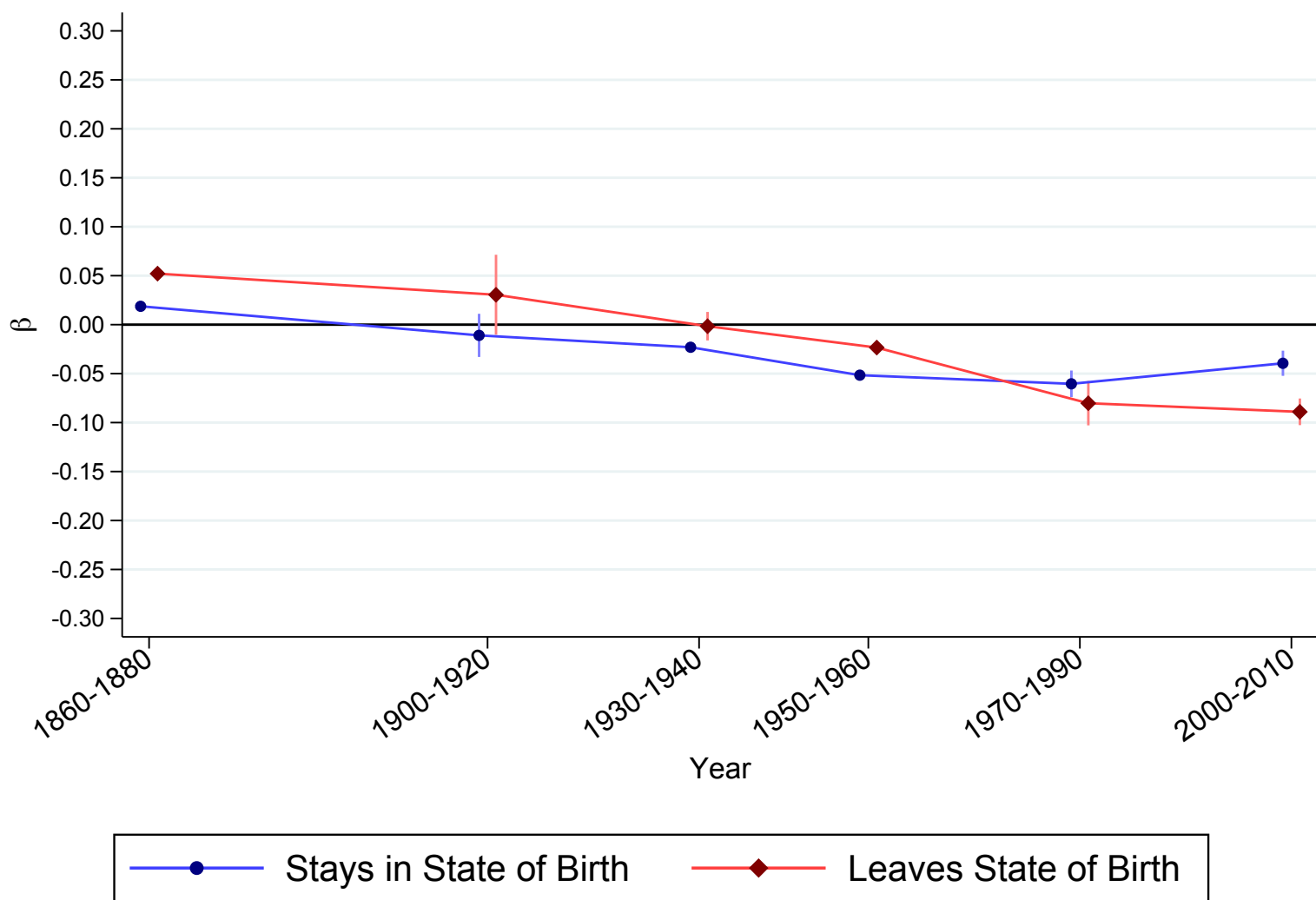


Figure 25 - By Age of Oldest Child, Twin IV

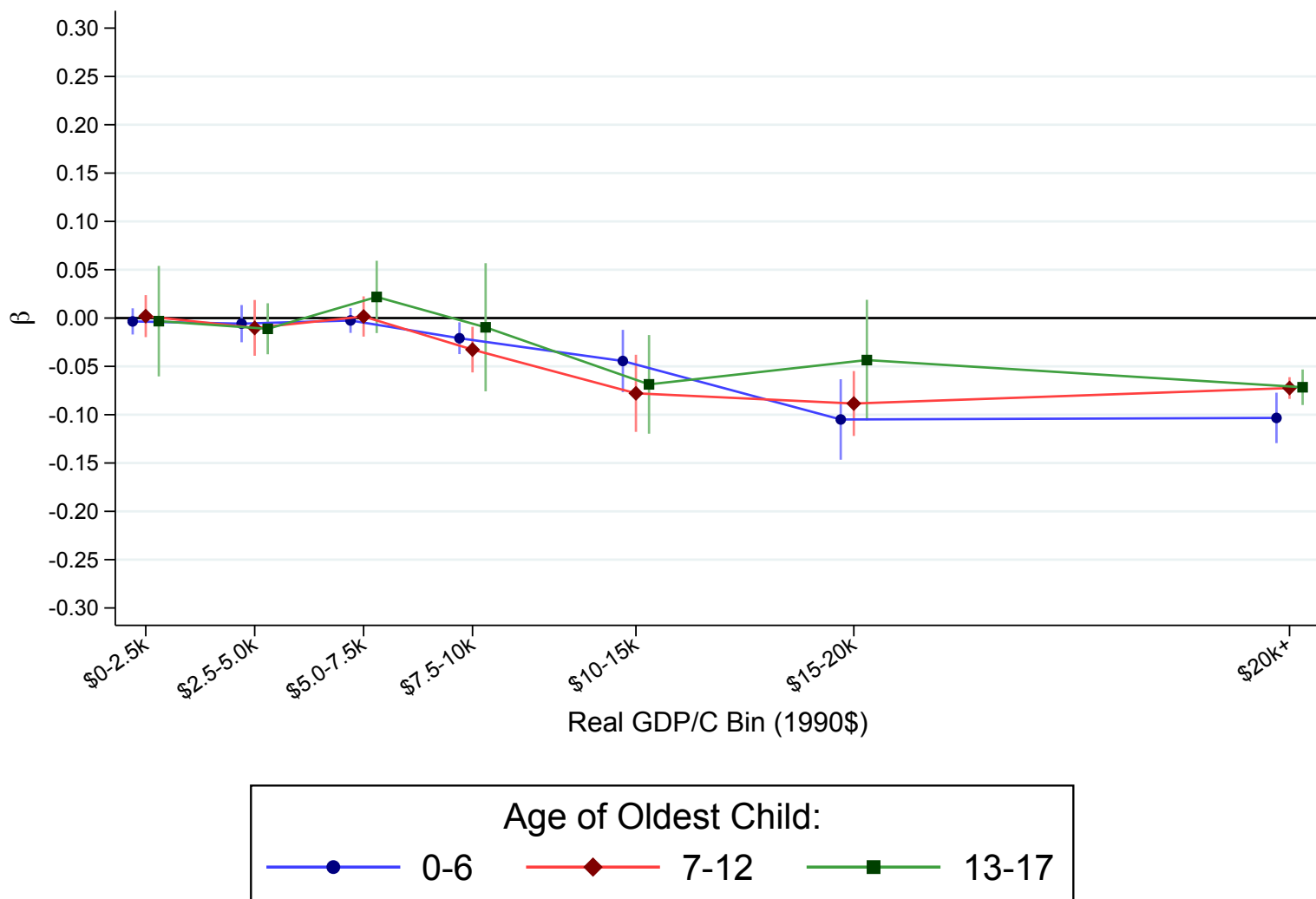


Figure 26 - By Age of Youngest Child, Twin IV

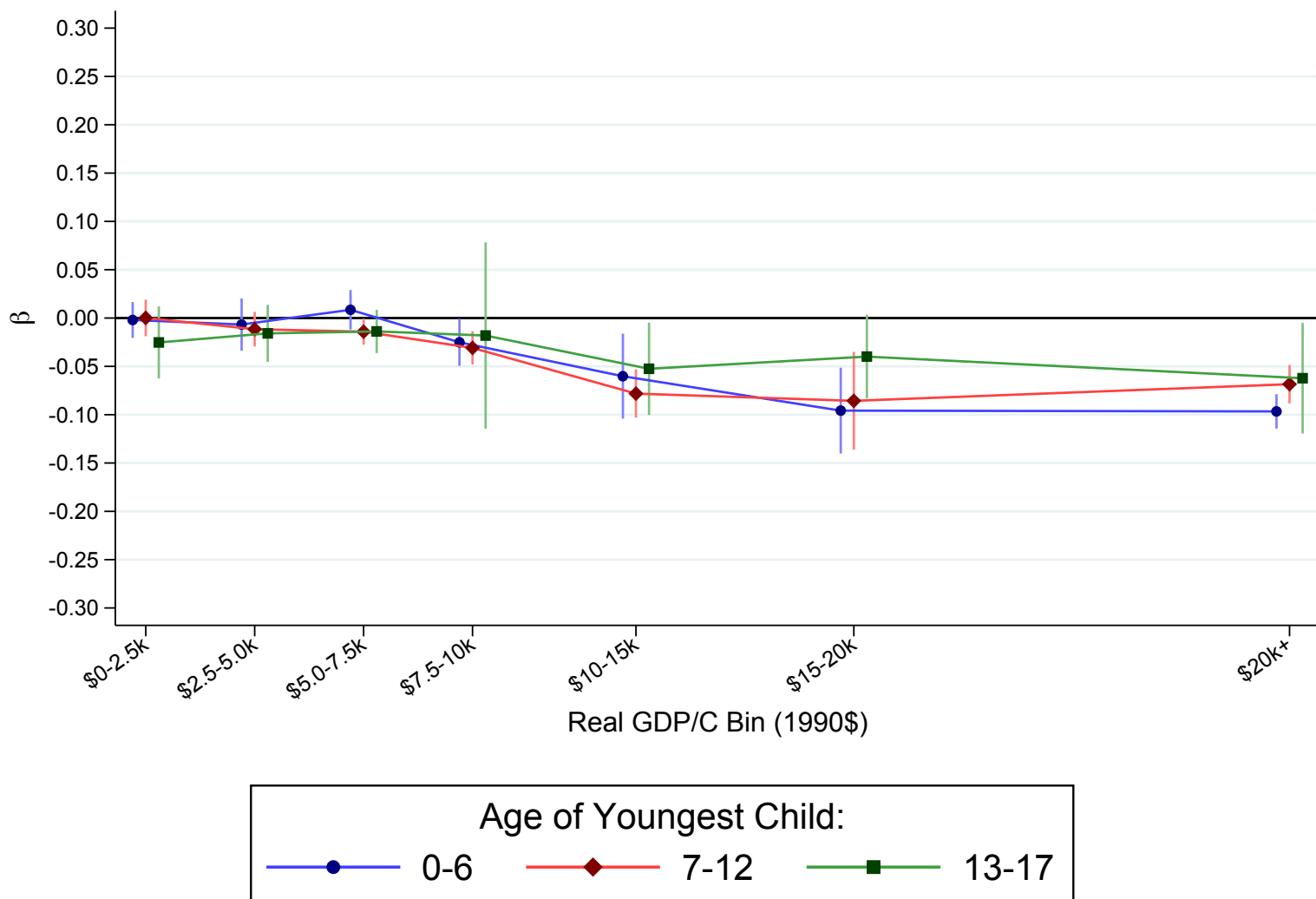


Table 1 - Sample Summary Statistics by GDP Group												
US												
GDP	N. Mothers	N. Samples	In Labor Force	3 or More Children	2nd Child is Multiple Birth	Number of Children (in Household)	Mother's Age at Survey	Mother's Age at First Birth	First Child is Boy	Second Child is Boy	Age of First Child	Age of Second Child
0-2,500	32,531	2	5.1%	62.5%	0.74%	3.27	29.02	21.04	50.8%	50.7%	7.99	5.18
2,500-5,000	2,557,639	2	6.2%	63.9%	0.68%	3.32	28.99	20.95	50.7%	50.8%	8.04	5.31
5,000-7,500	12,959,066	3	9.2%	55.5%	0.86%	3.09	29.30	21.13	50.6%	50.6%	8.17	5.46
7,500-10,000	4,706,116	2	10.6%	47.0%	0.87%	2.88	29.48	20.94	50.8%	50.8%	8.54	5.66
10,000-15,000	470,378	1	22.8%	55.1%	1.70%	2.99	29.30	21.40	51.0%	50.8%	7.90	5.28
15,000-20,000	598,515	2	46.8%	39.0%	1.29%	2.58	29.68	21.07	51.2%	51.0%	8.61	5.68
20,000-35,000	1,312,550	3	62.9%	36.6%	1.46%	2.50	30.28	21.85	51.1%	50.9%	8.42	5.17
Non-US												
0-2,500	9,006,590	210	43.9%	56.7%	1.22%	3.04	29.07	20.67	50.6%	51.5%	8.39	5.42
2,500-5,000	7,529,849	99	36.0%	50.9%	1.05%	2.97	29.82	21.19	51.1%	51.0%	8.63	5.50
5,000-7,500	4,132,620	51	37.2%	46.3%	1.22%	2.77	29.41	20.45	50.9%	50.8%	8.96	5.79
7,500-10,000	1,474,565	17	32.4%	46.5%	1.36%	2.76	29.46	20.74	51.1%	50.7%	8.72	5.58
10,000-15,000	592,503	18	36.4%	37.1%	1.21%	2.63	29.97	21.64	51.4%	51.1%	8.33	5.28
15,000-20,000	454,756	11	58.8%	29.3%	1.18%	2.39	30.84	22.79	51.5%	51.3%	8.05	4.78
20,000-35,000	1,065,131	8	73.8%	29.0%	1.43%	2.38	31.23	24.00	51.2%	51.1%	7.23	3.99

Table 2 - Baseline Estimates by GDP Group

GDP	N. Mothers	N. Samples	LFP	N. Mothers	N. Samples	LFP	OLS		Twin IV				Same-Sex IV			
									non-US		non-US		non-US		non-US	
							US	non-US	US	2S	FS	2S	FS	2S	FS	2S
0-2,500	32,531	2	5.1%	9,006,590	210	43.9%	-0.018 (0.006)	-0.021 (0.004)	0.345 (0.018)	0.119 (0.005)	0.418 (0.019)	-0.001 (0.009)	0.015 (0.007)	-0.068 (0.162)	0.029 (0.007)	-0.048 (0.020)
2,500-5,000	2,557,639	2	6.2%	7,529,849	99	36.0%	-0.023 (0.000)	-0.057 (0.005)	0.345 (0.005)	0.035 (0.005)	0.470 (0.037)	-0.014 (0.011)	0.009 (0.000)	0.036 (0.007)	0.030 (0.007)	-0.019 (0.012)
5,000-7,500	12,959,066	3	9.2%	4,132,620	51	37.2%	-0.033 (0.009)	-0.089 (0.012)	0.452 (0.014)	0.009 (0.011)	0.542 (0.020)	-0.002 (0.016)	0.014 (0.002)	0.037 (0.008)	0.035 (0.002)	-0.038 (0.013)
7,500-10,000	4,706,116	2	10.6%	1,474,565	17	32.4%	-0.064 (0.001)	-0.113 (0.006)	0.541 (0.002)	0.017 (0.001)	0.523 (0.027)	-0.035 (0.014)	0.021 (0.000)	0.073 (0.002)	0.032 (0.002)	0.015 (0.036)
10,000-15,000	470,378	1	22.8%	592,503	18	36.4%	-0.117 (0.001)	-0.137 (0.024)	0.452 (0.002)	0.033 (0.010)	0.597 (0.065)	-0.091 (0.016)	0.035 (0.001)	-0.084 (0.034)	0.035 (0.004)	-0.059 (0.036)
15,000-20,000	598,515	2	46.8%	454,756	11	58.8%	-0.171 (0.010)	-0.265 (0.037)	0.594 (0.045)	0.063 (0.015)	0.730 (0.037)	-0.117 (0.034)	0.050 (0.005)	-0.125 (0.004)	0.041 (0.002)	-0.185 (0.028)
20,000-35,000	1,312,550	3	62.9%	1,065,131	8	73.8%	-0.149 (0.010)	-0.249 (0.008)	0.636 (0.007)	0.070 (0.008)	0.706 (0.003)	-0.104 (0.002)	0.049 (0.001)	-0.121 (0.008)	0.038 (0.001)	-0.174 (0.020)

## Appendix

Figure A1 - Comparison of Twinning Rates in DHS

Corr: .87

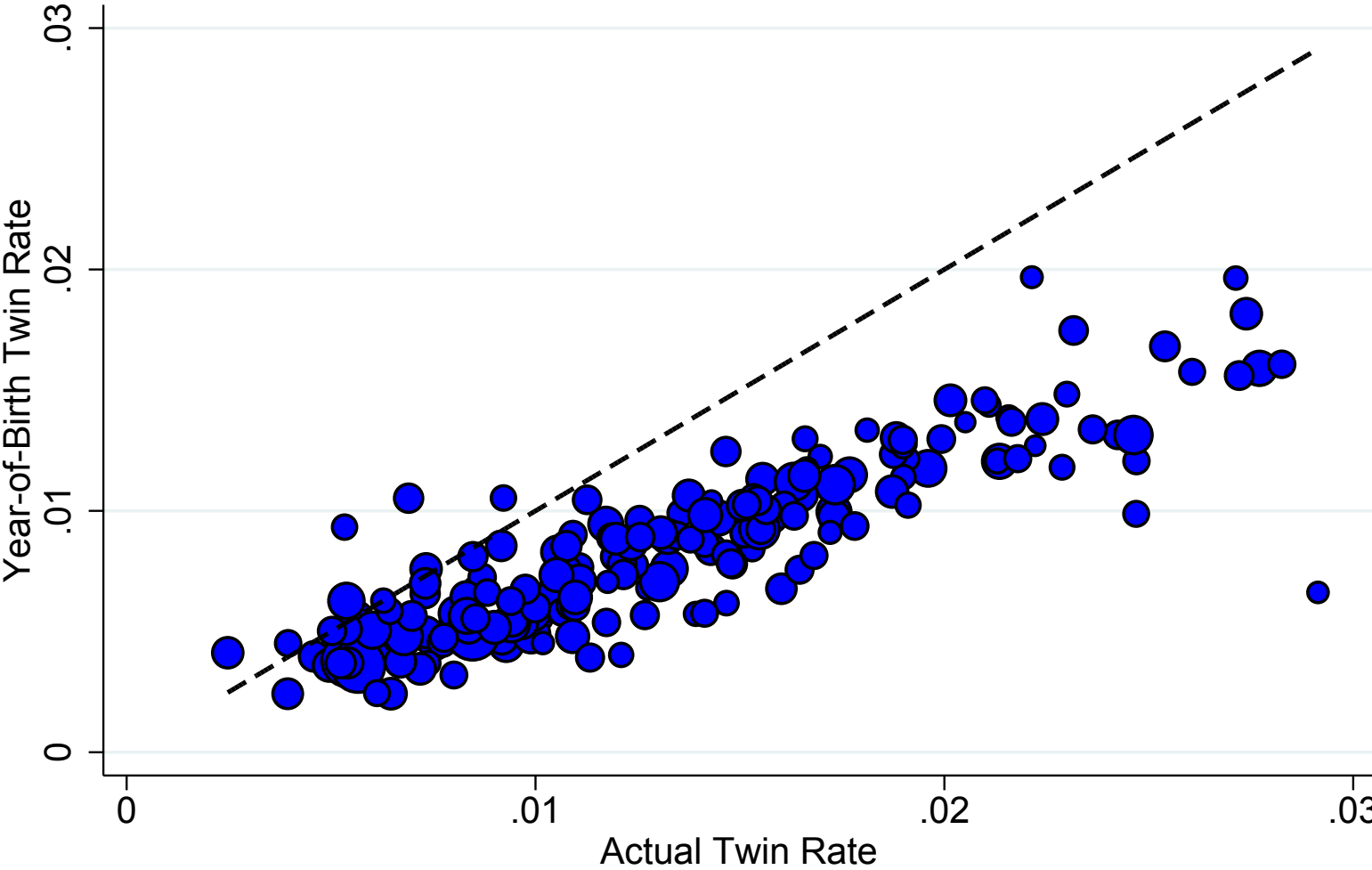
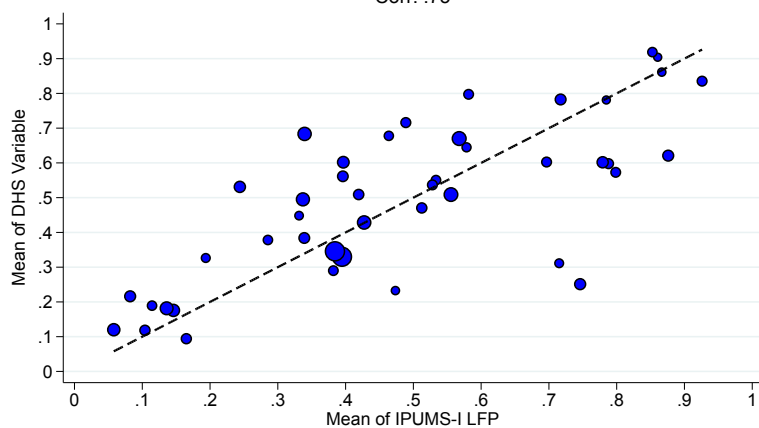


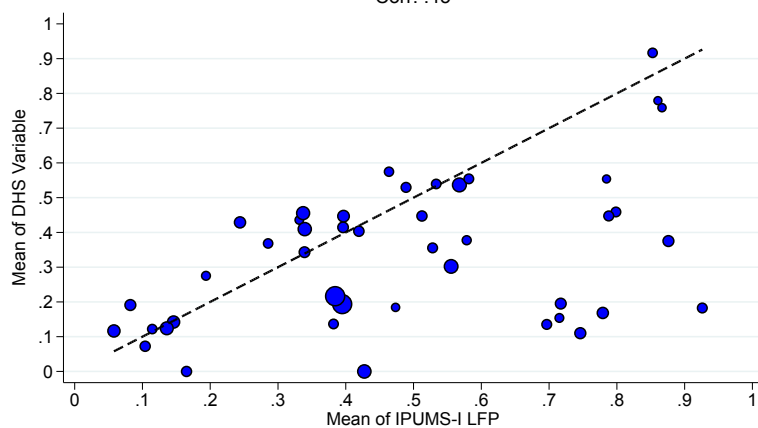


Figure A2 - Comparison of DHS Work Measures with IPUMS LFP  
Observation = Country-Year

Any Current Work  
Corr: .75



Any Current Work for Cash  
Corr: .45



Any Current Work for Cash away from Home  
Corr: -.3

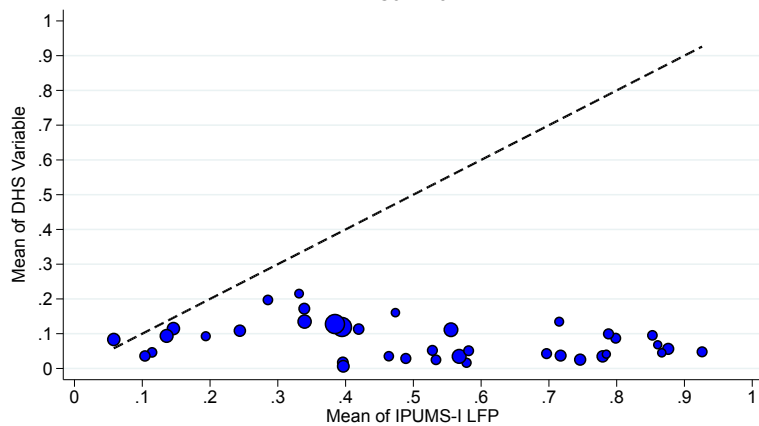


Figure A3 - OLS, By Country and Real GDP/Capita

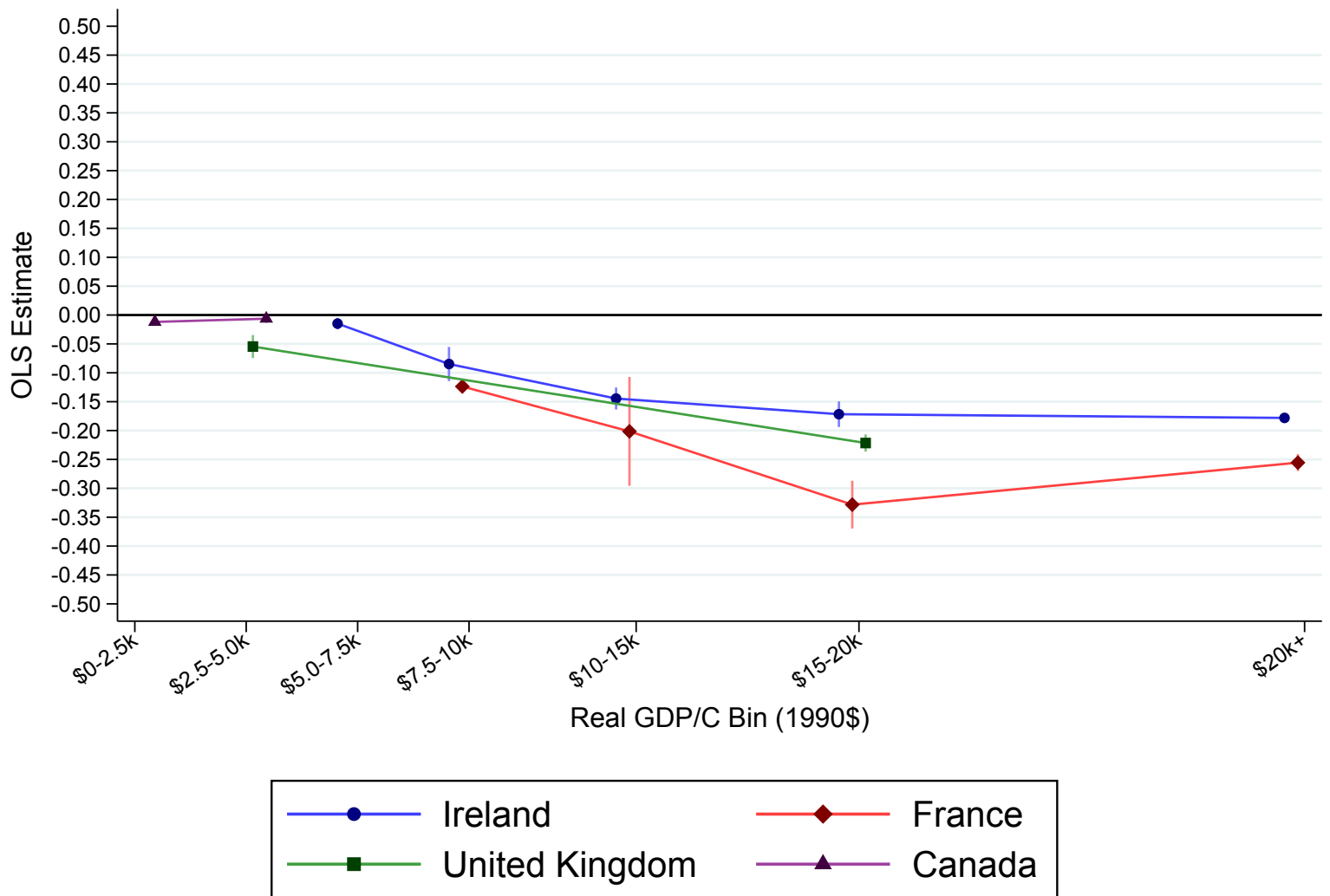


Figure A4 - By Data Source, Twin IV

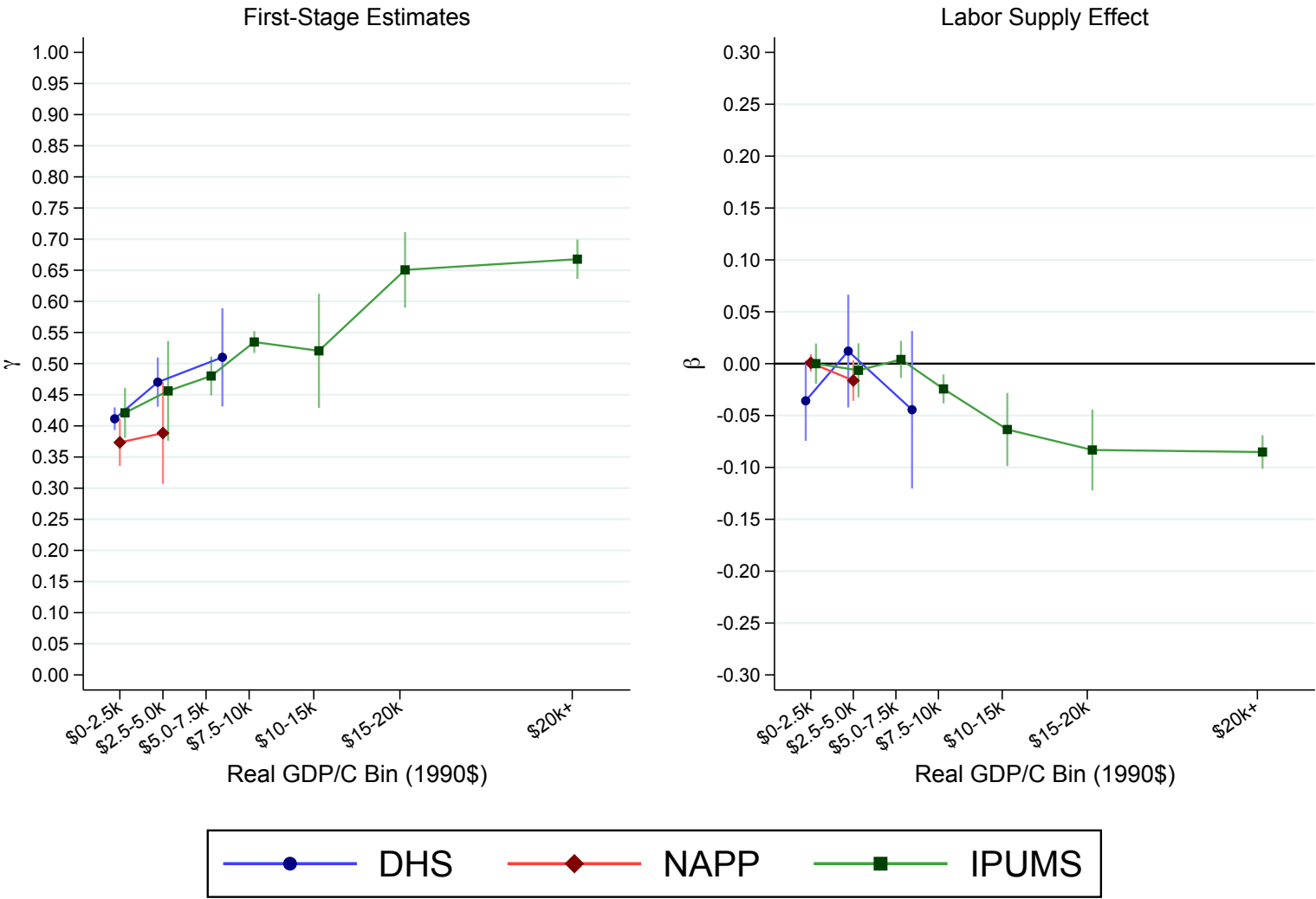


Figure A5 - By Region, Twin IV

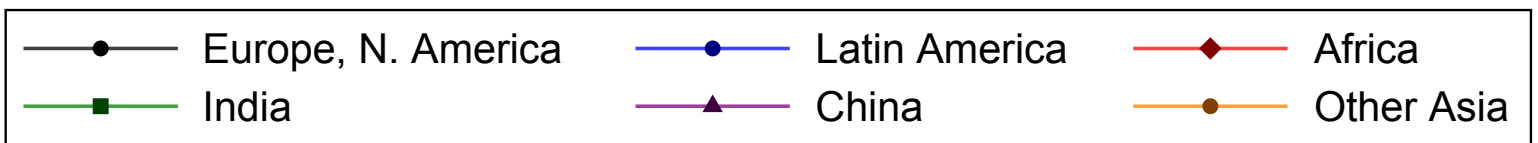
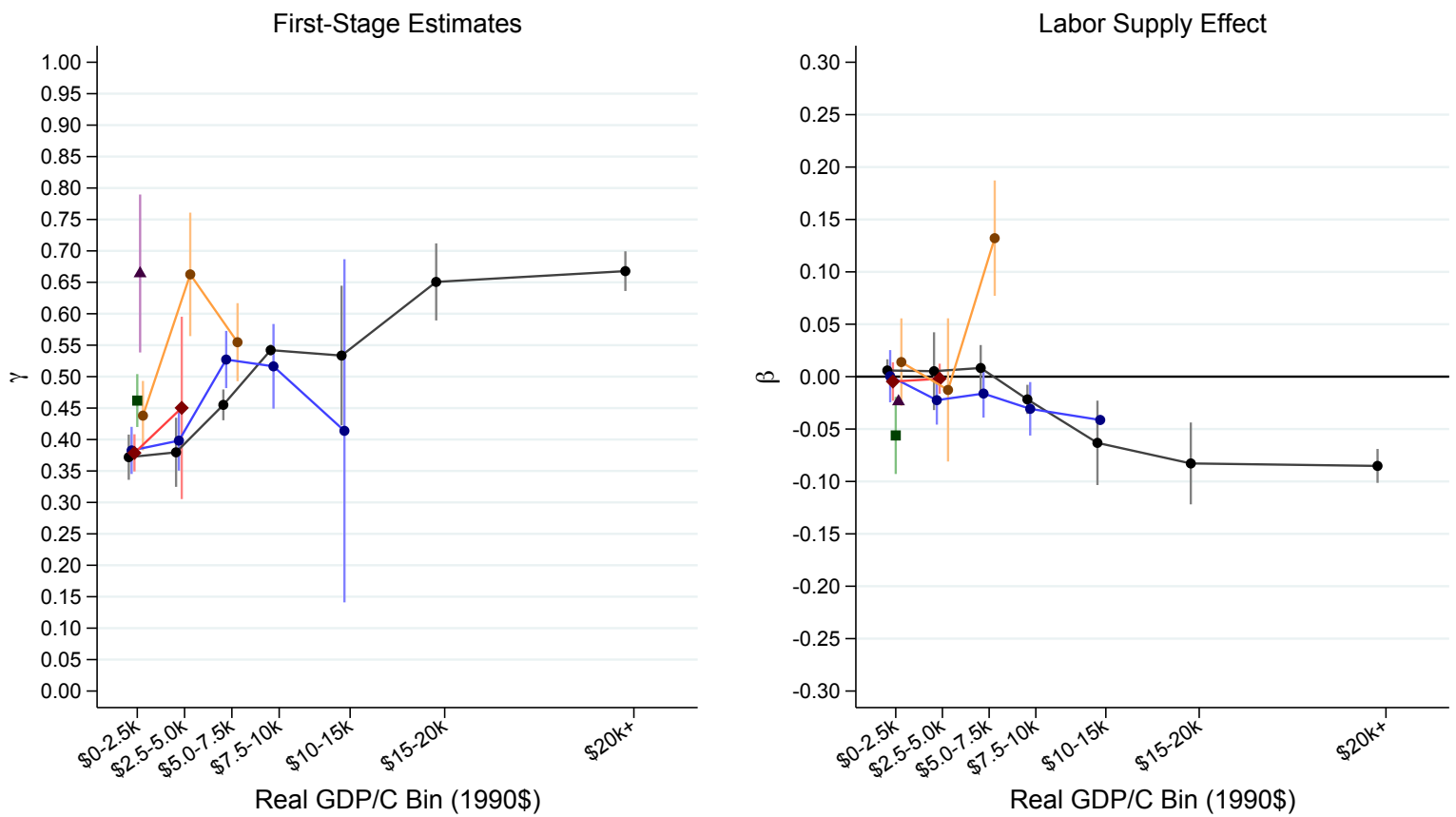


Figure A6 - Twin IV, By Country and Real GDP/Capita

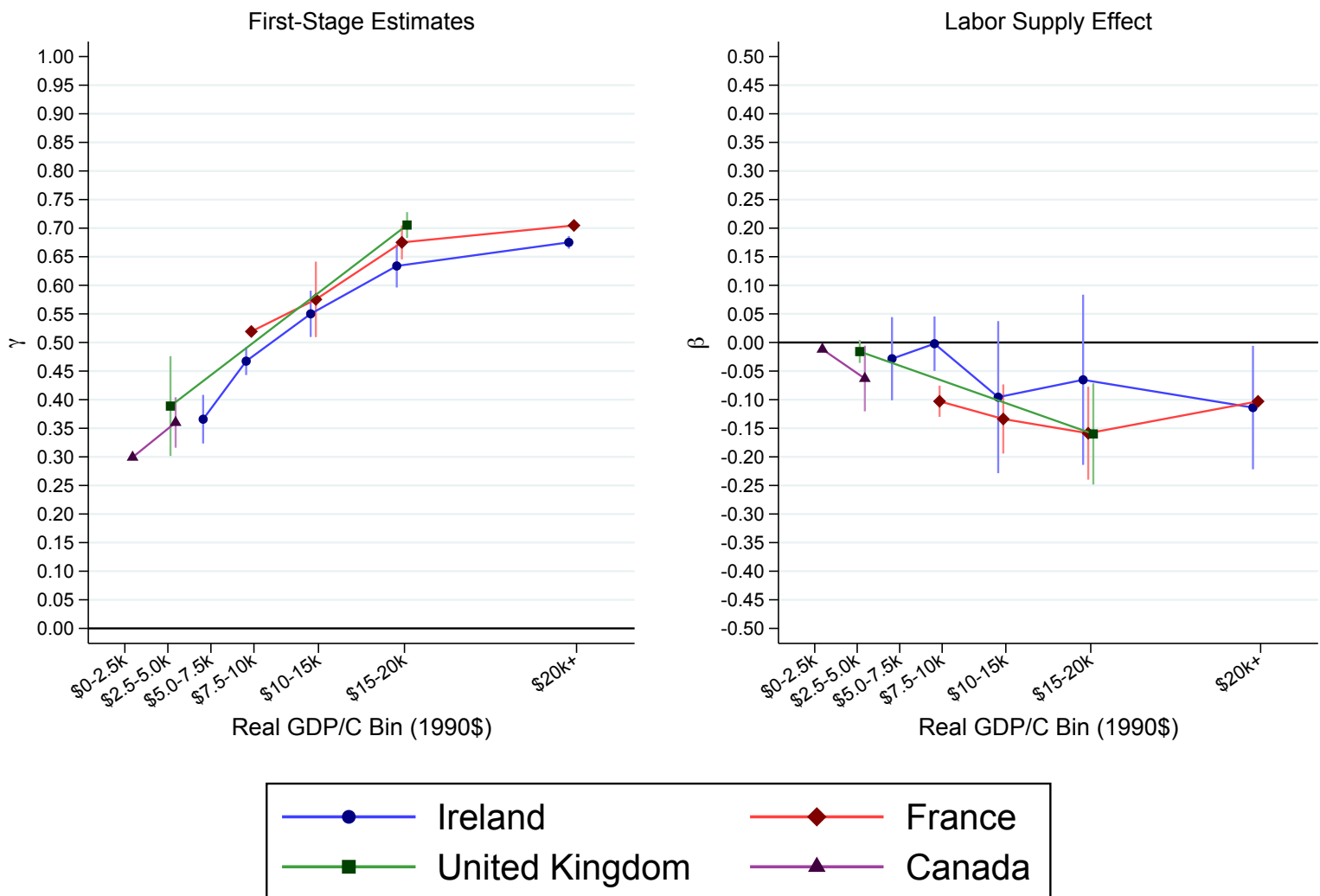


Figure A7 - Same Gender IV, U.S. by Time

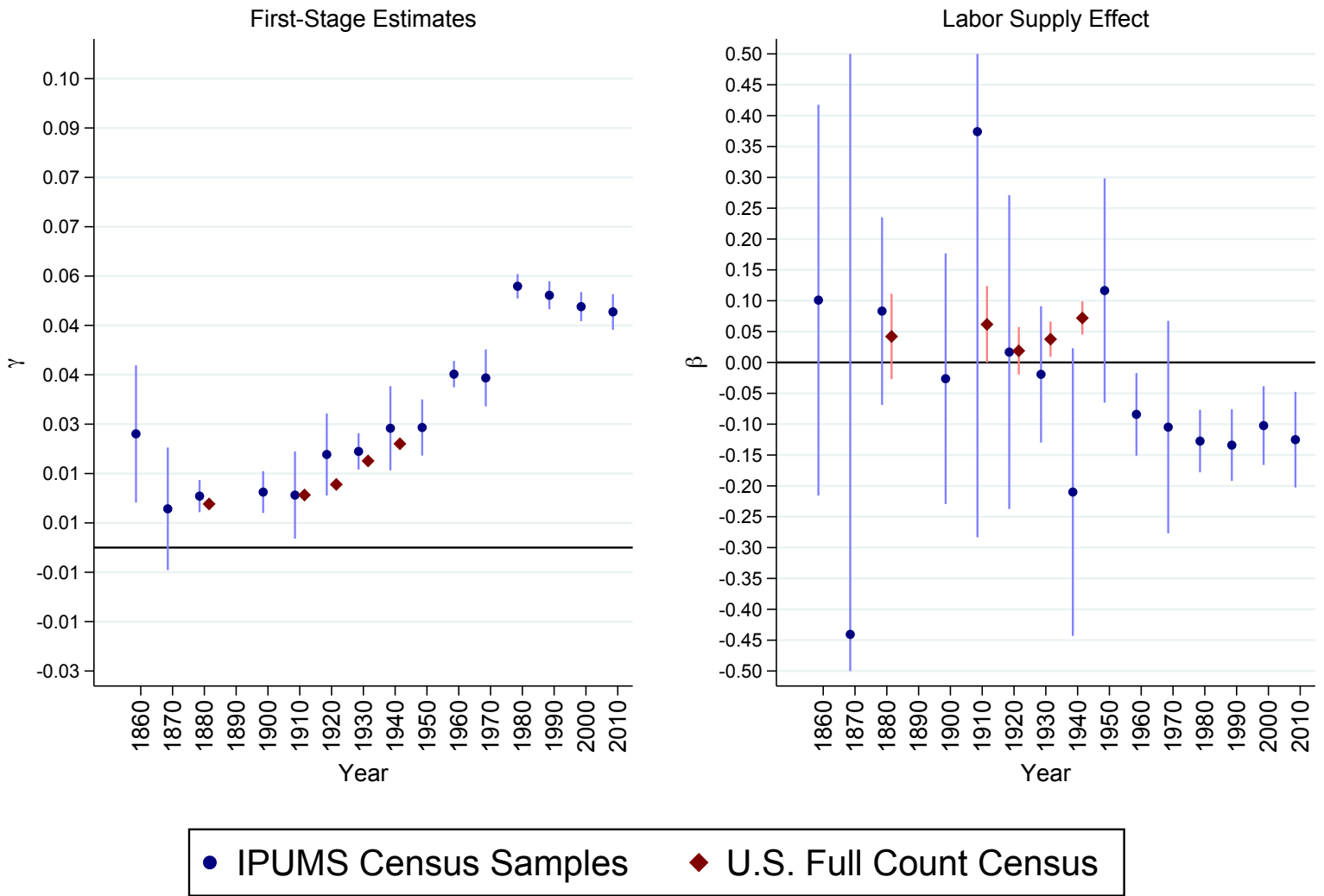
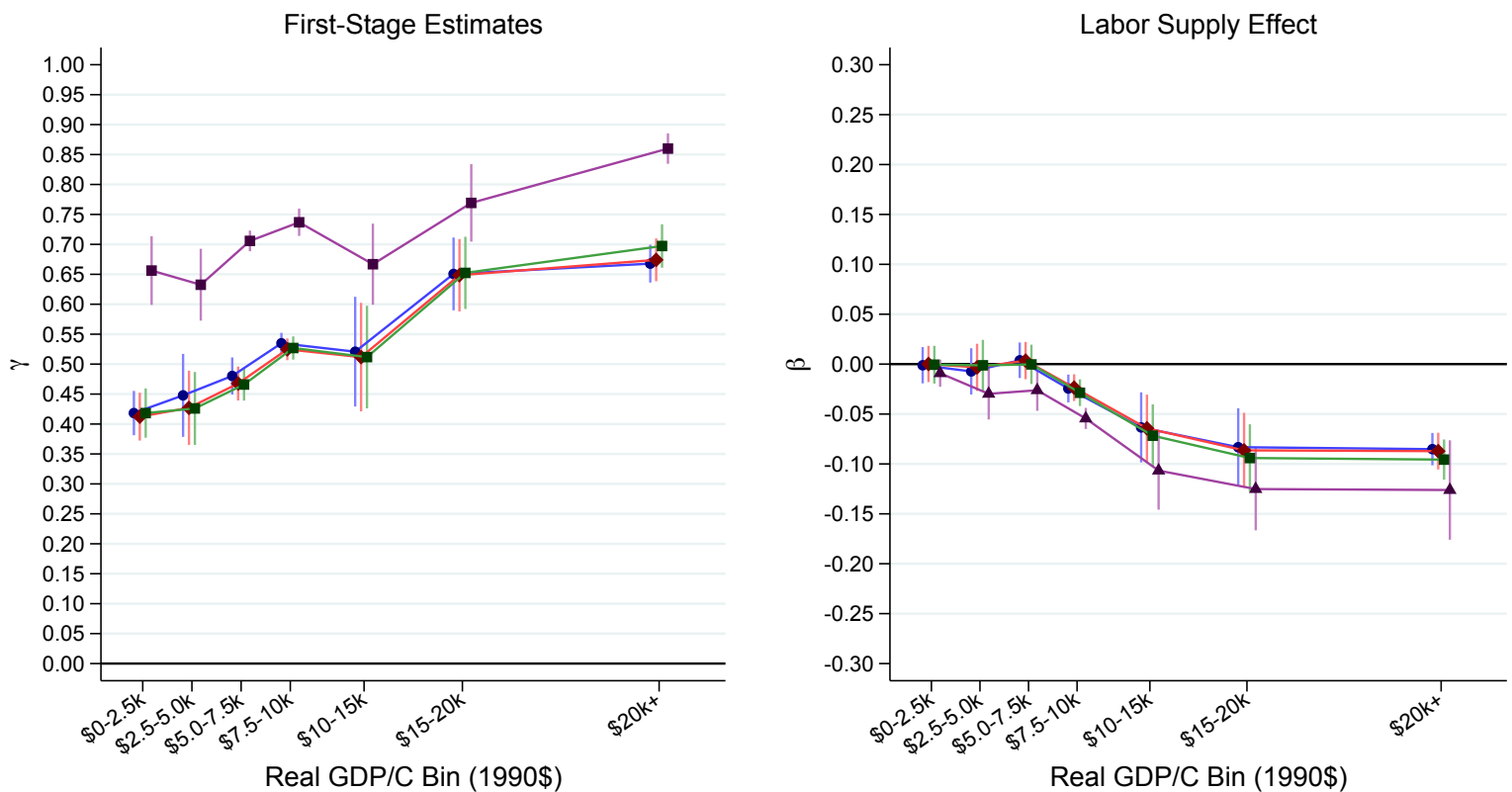


Figure A8 - Spacing of Births, Twin IV



Largest Difference in Ages of First Two/Three Children:

- |   |  |
|---|--|
| <span style="color: blue;">—●—</span> No Restriction (Baseline) | <span style="color: red;">—◆—</span> 5 Years   |
| <span style="color: green;">—■—</span> 3 Years                  | <span style="color: purple;">—▲—</span> 1 Year |

Figure A9 - Robustness to Education, Twin IV

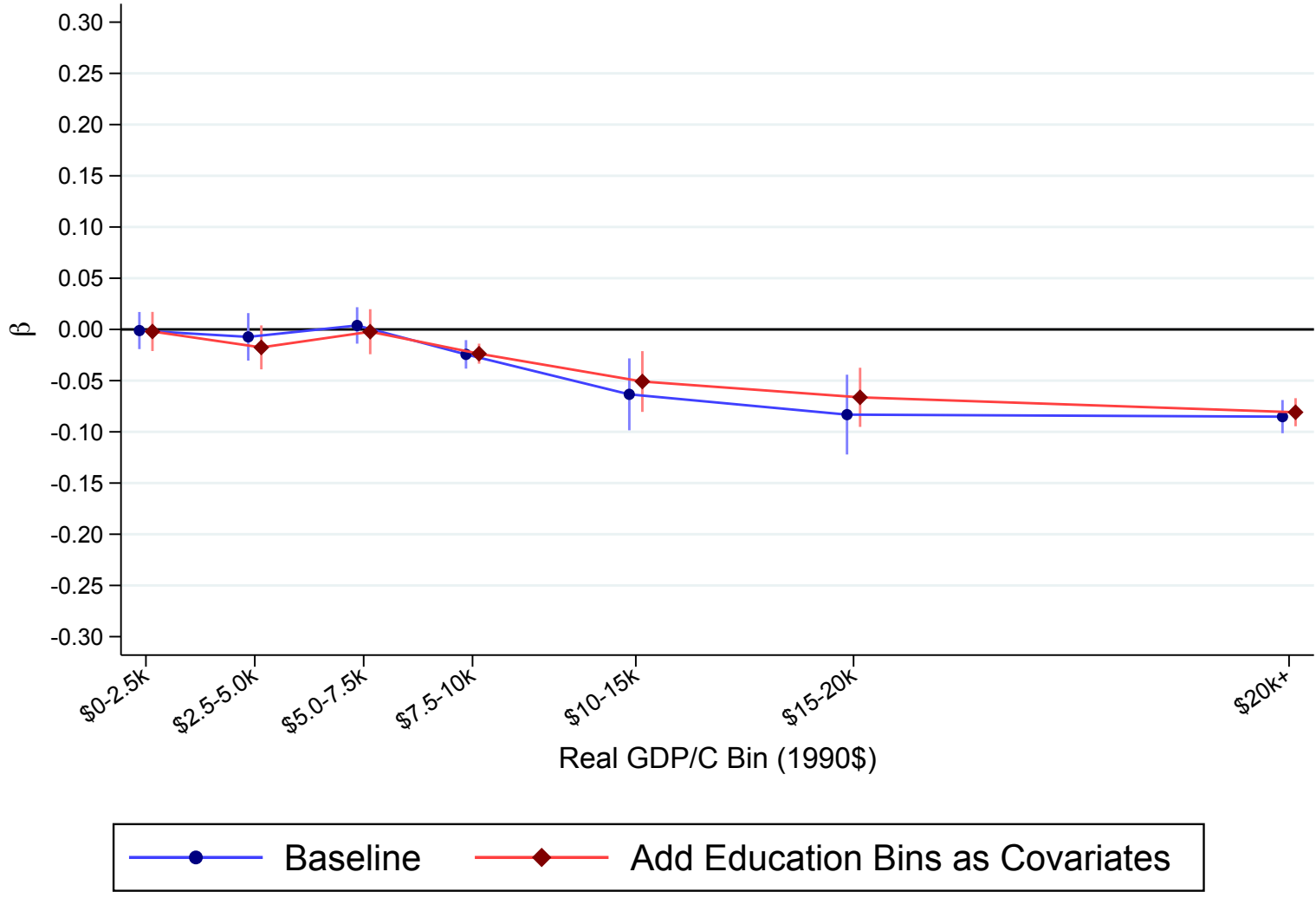




Figure A10 - Robustness to Non-Biological Children  
Country-Years with Information on Number of Children Ever Born

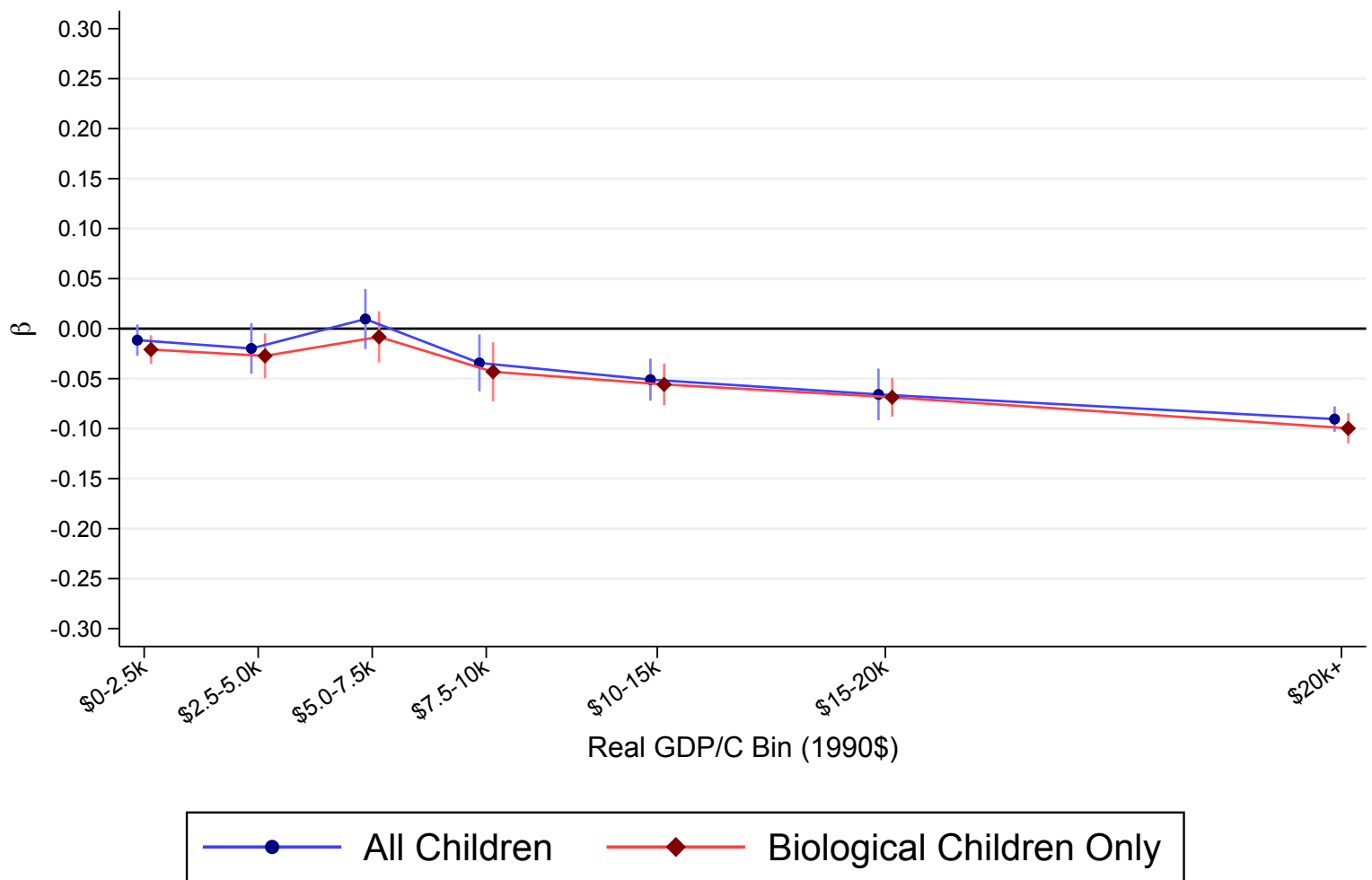


Figure A11 - Alternative Measures of Labor Force Participation by State  
U.S. 1940 Full Census

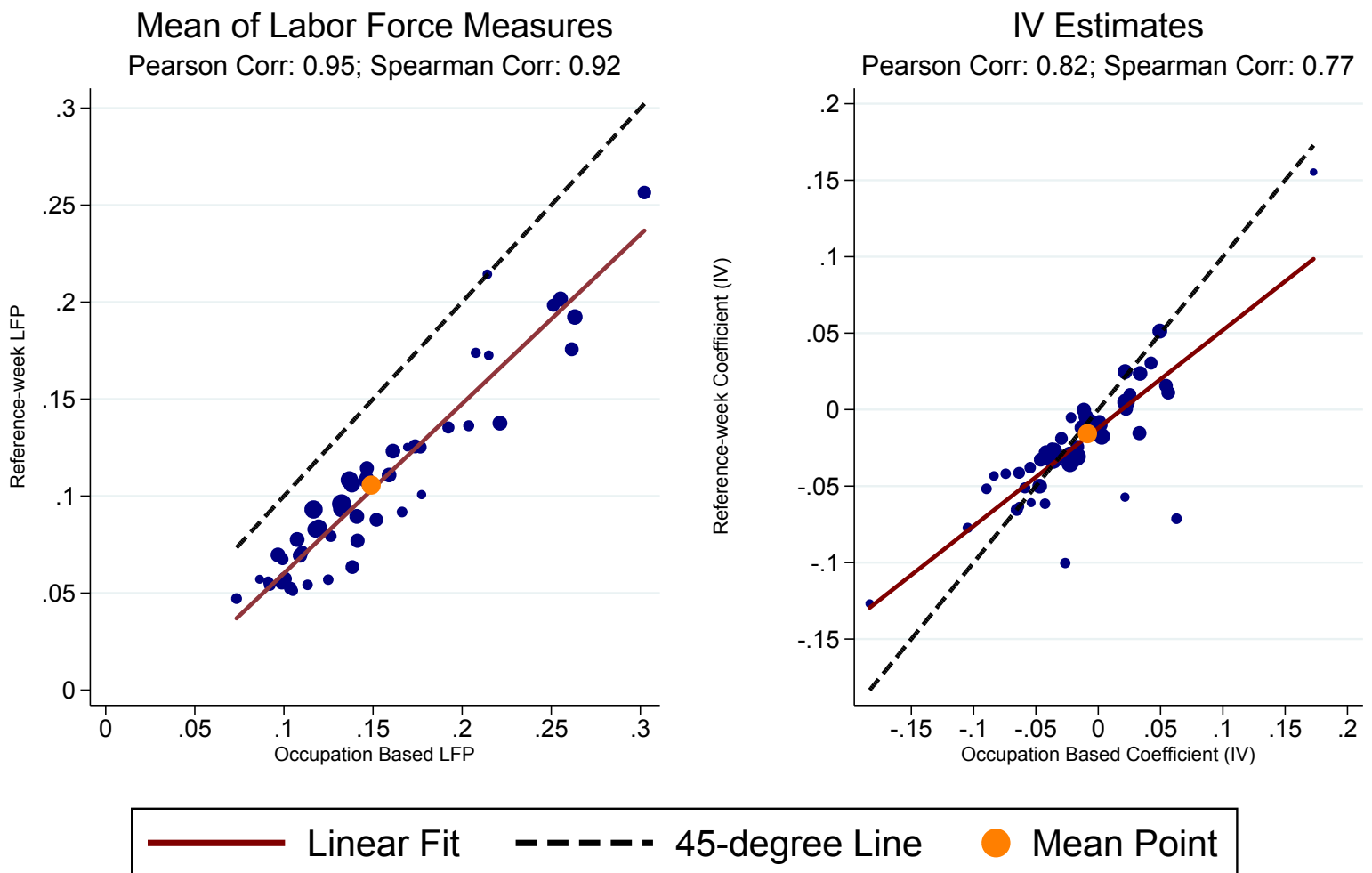
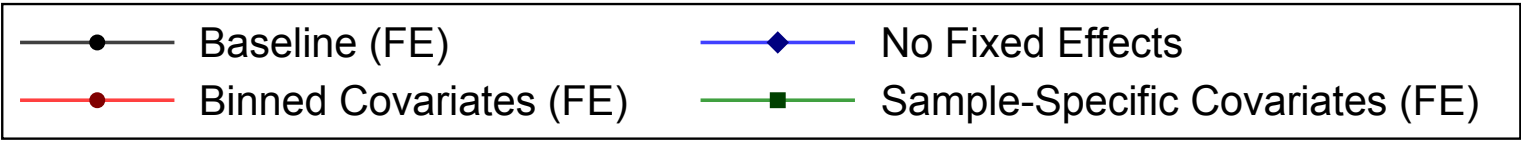
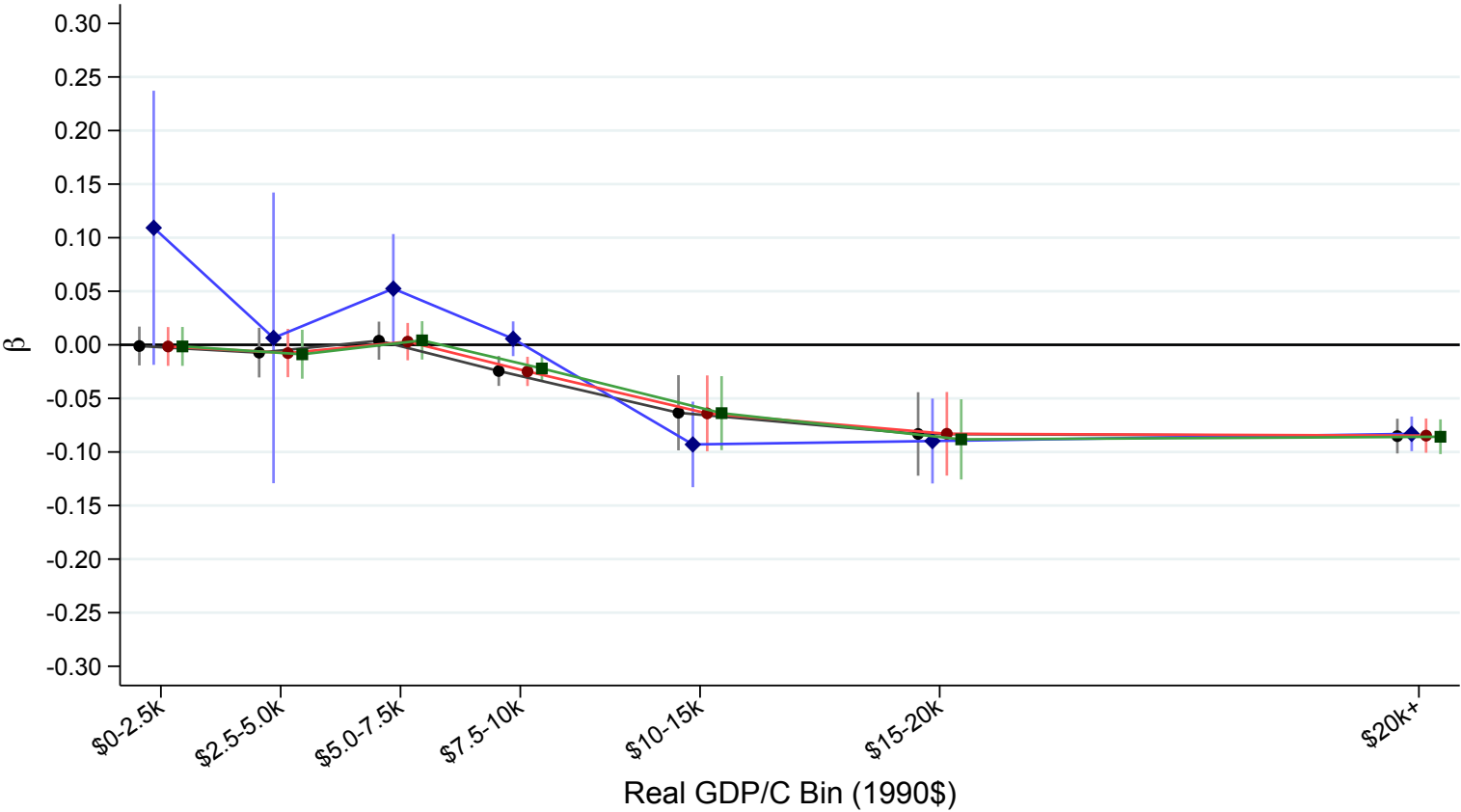


Figure A12 - Robustness to Specification, Twin IV



Figuer A13 - By Age of Mother, Twin IV

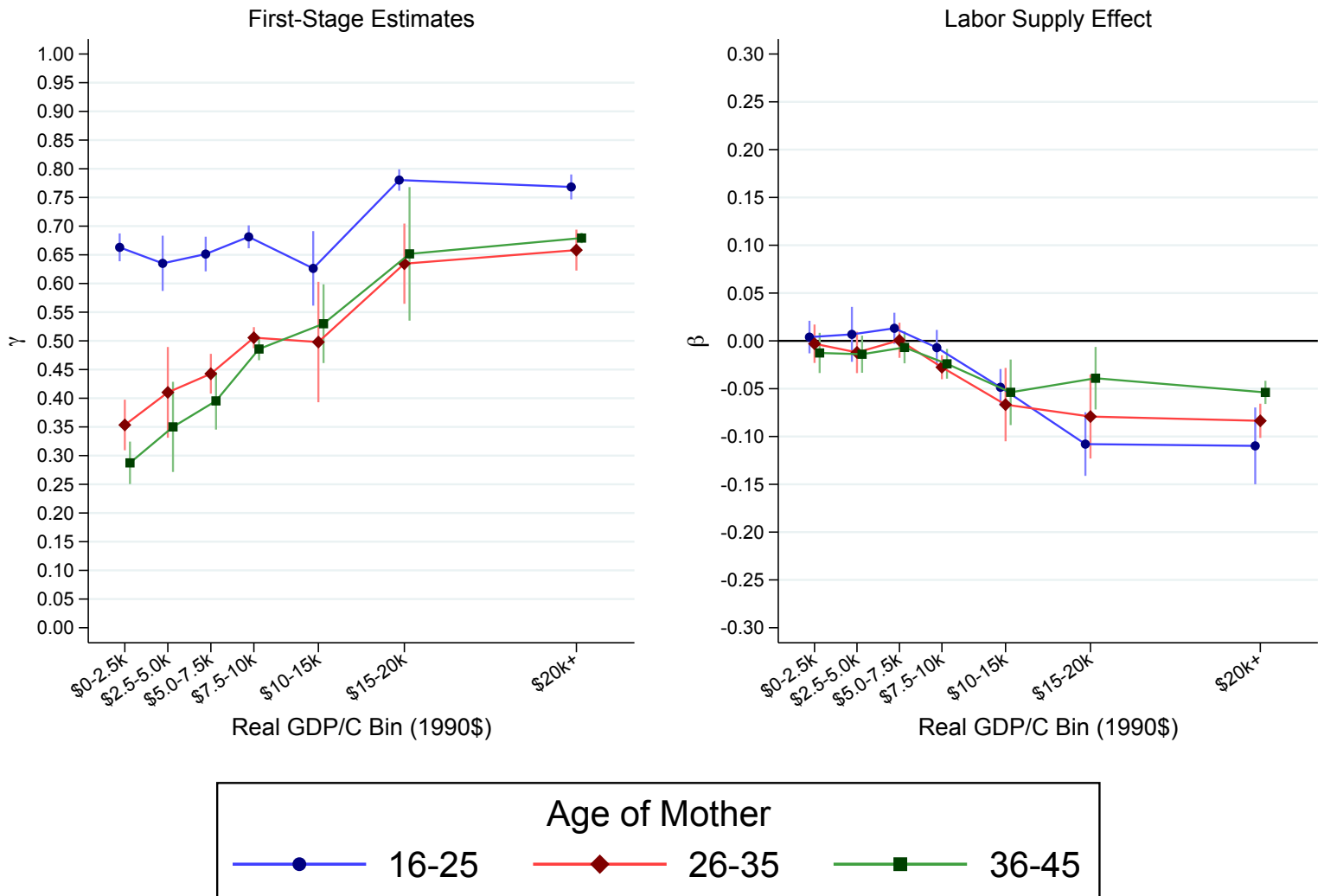


Figure A14 - By Age of Mother at First Birth, Twin IV

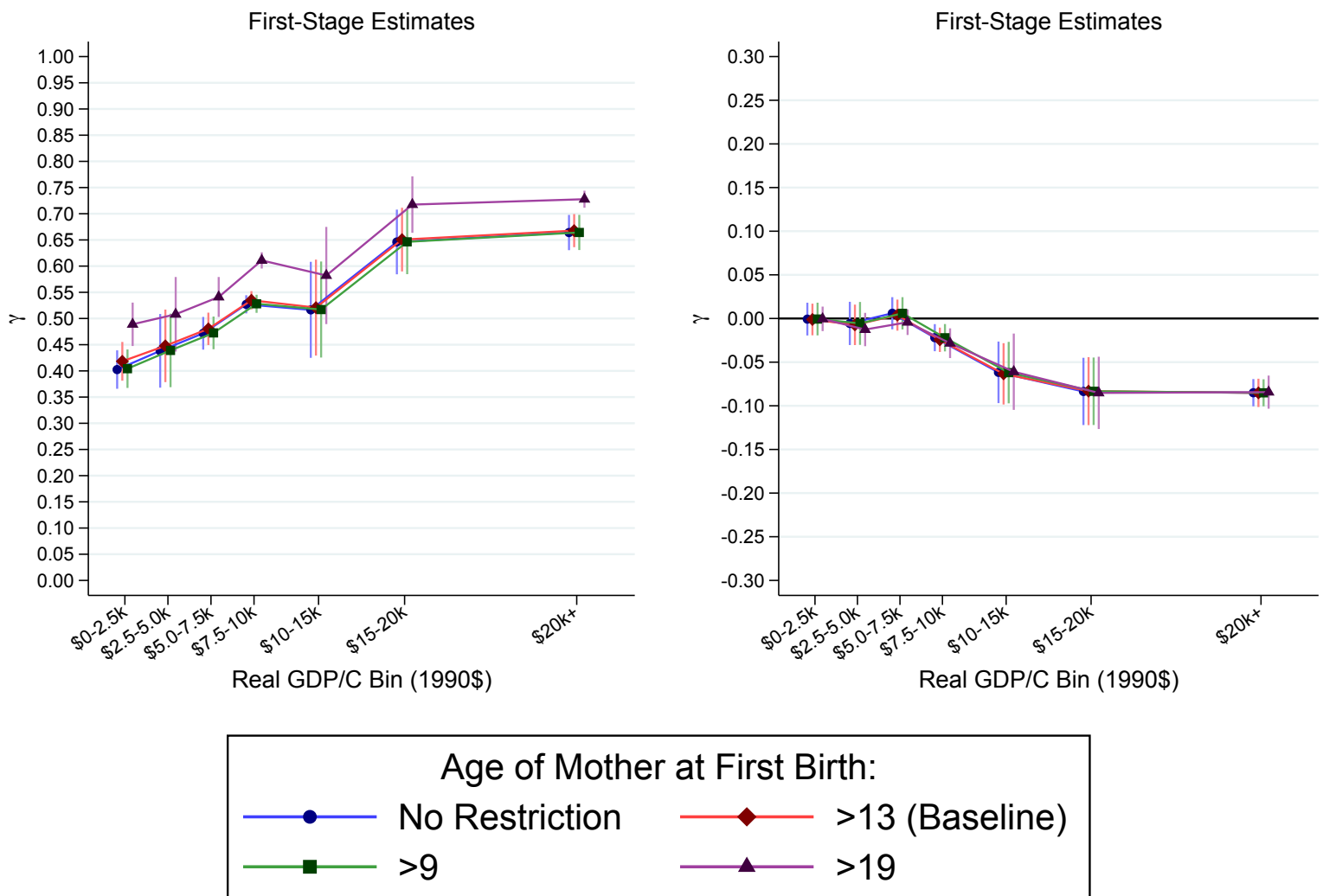


Figure A15 - By Mother's Education, Twin IV

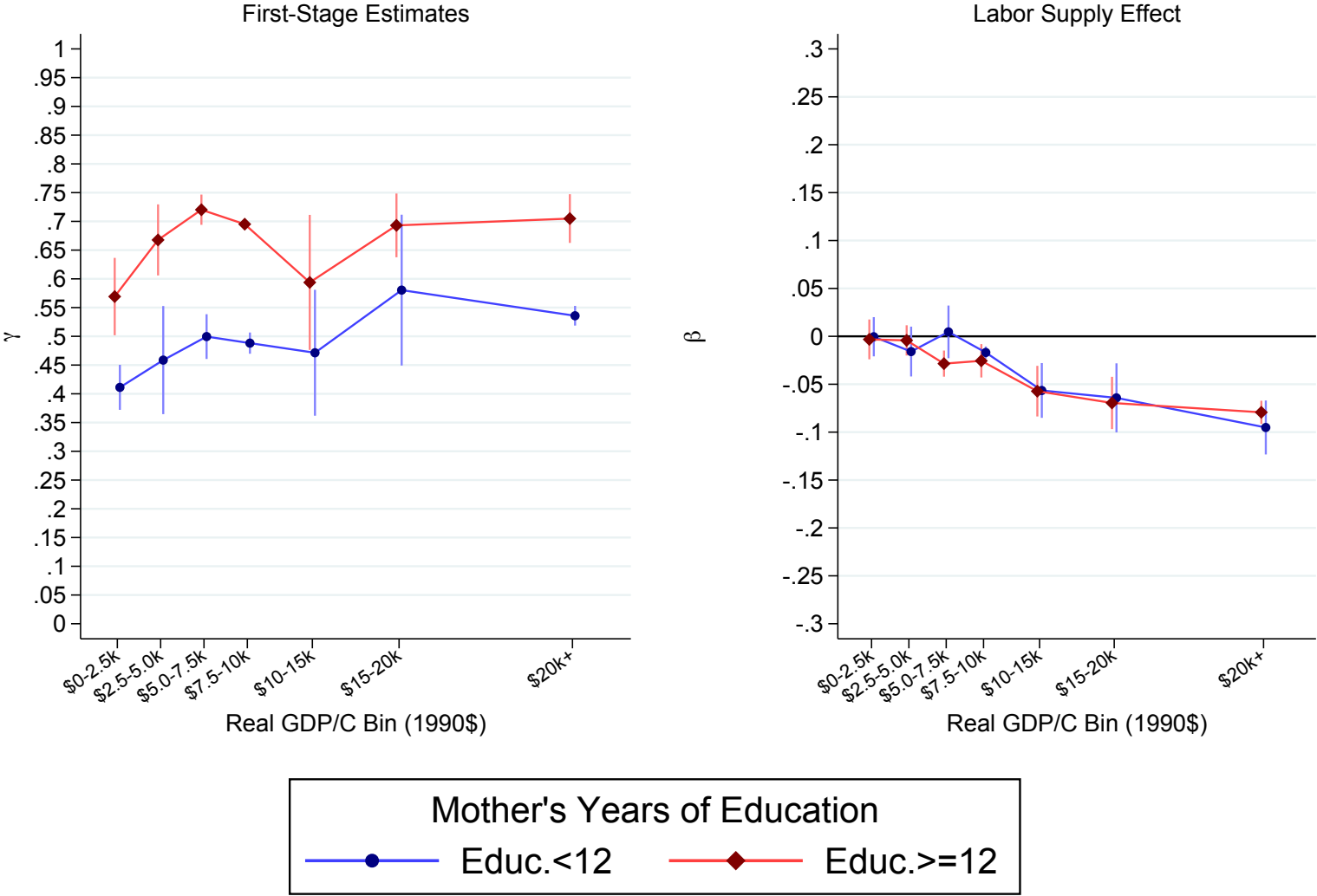


Figure A16 - By Marital Status, Twin IV

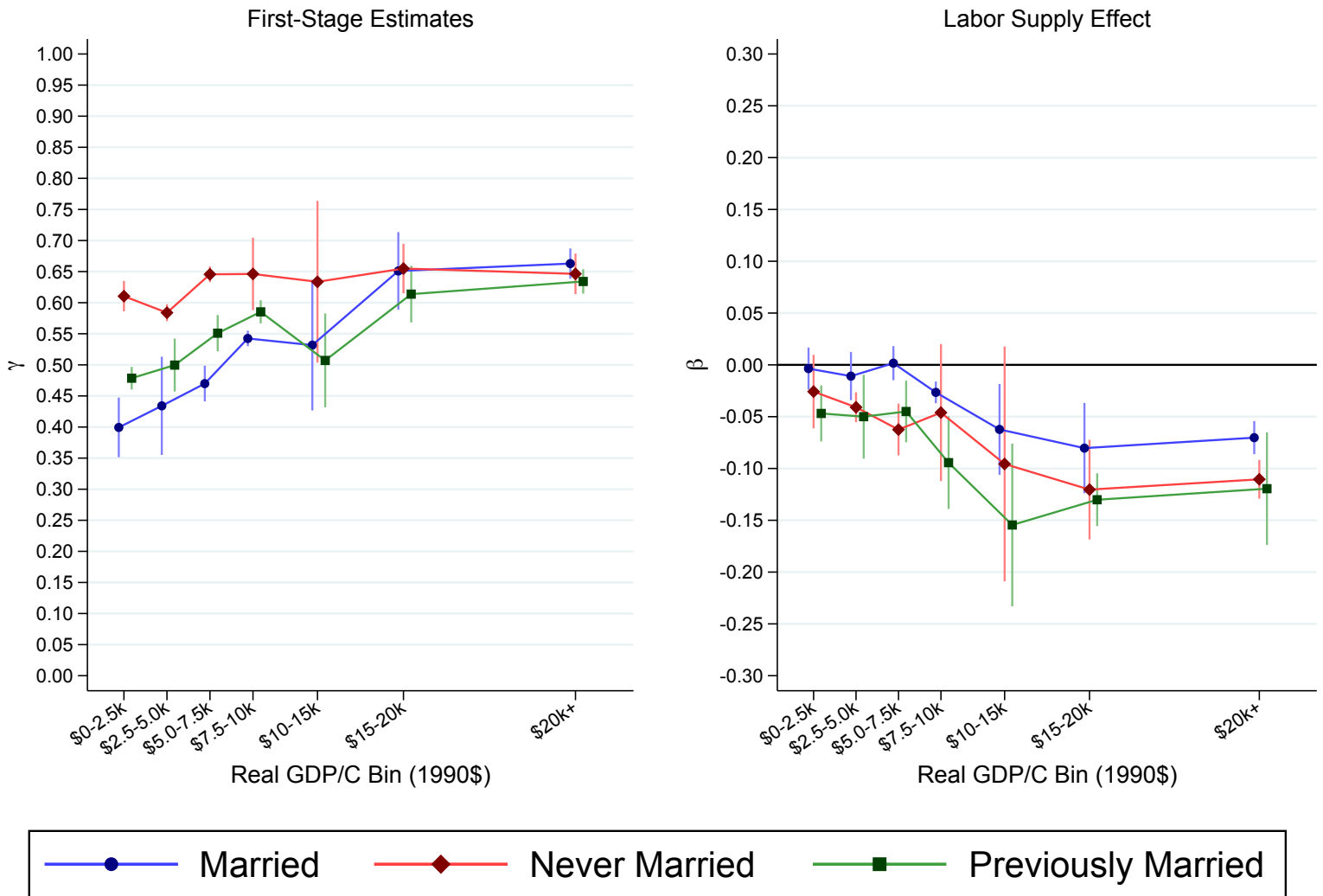


Figure A17 - Reweight Covariates to U.S. 1980 Compliers, Twin IV

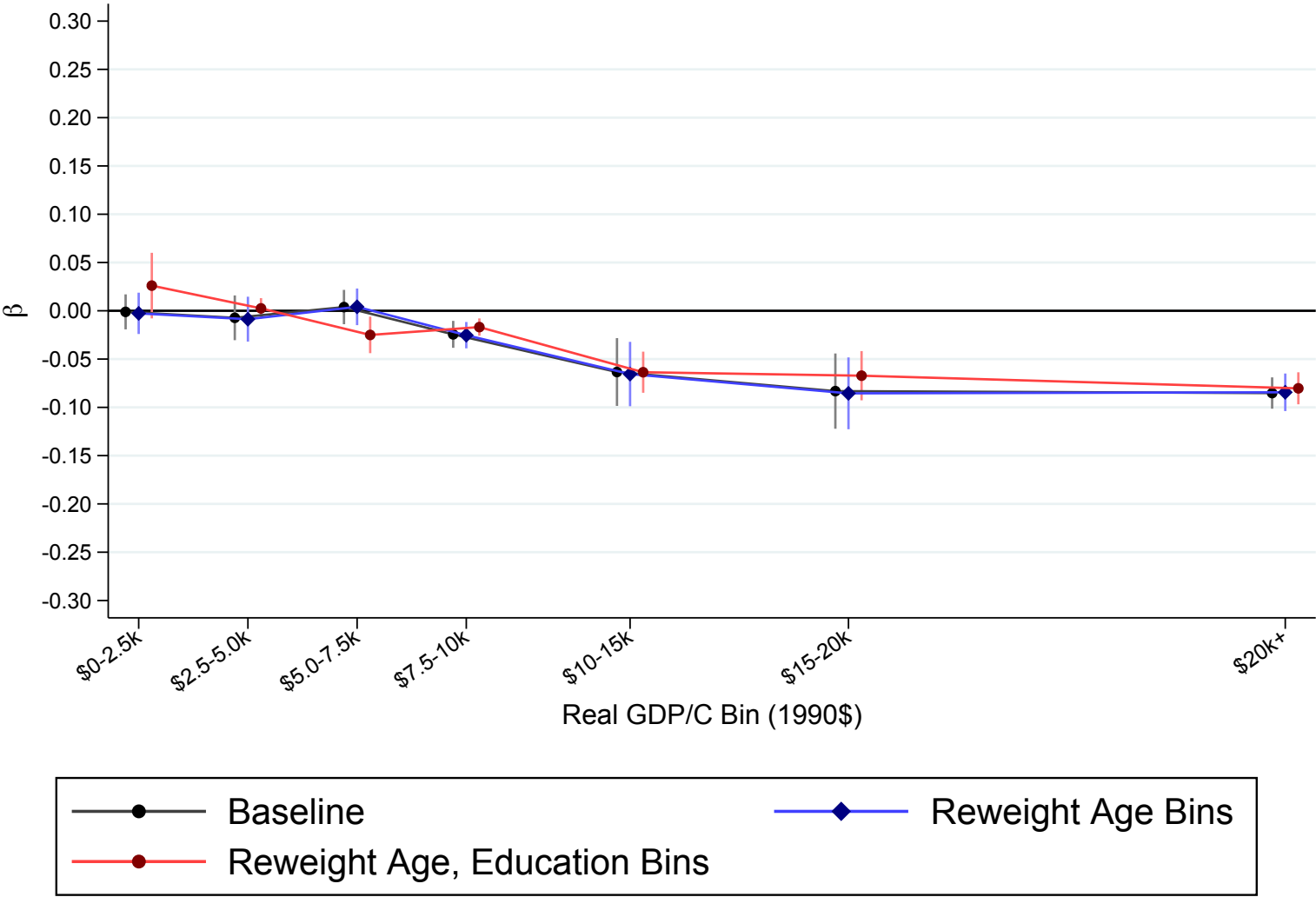
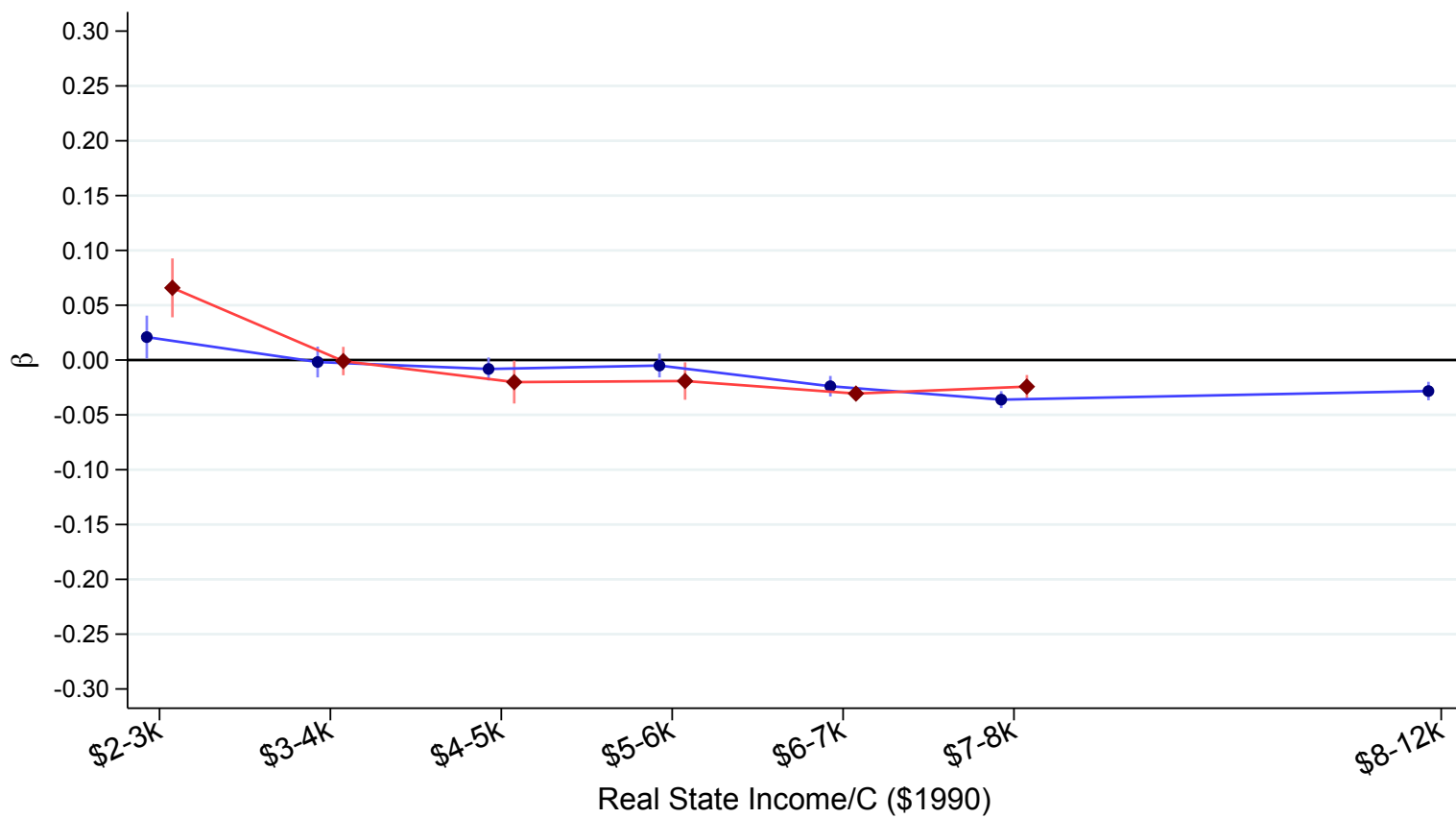




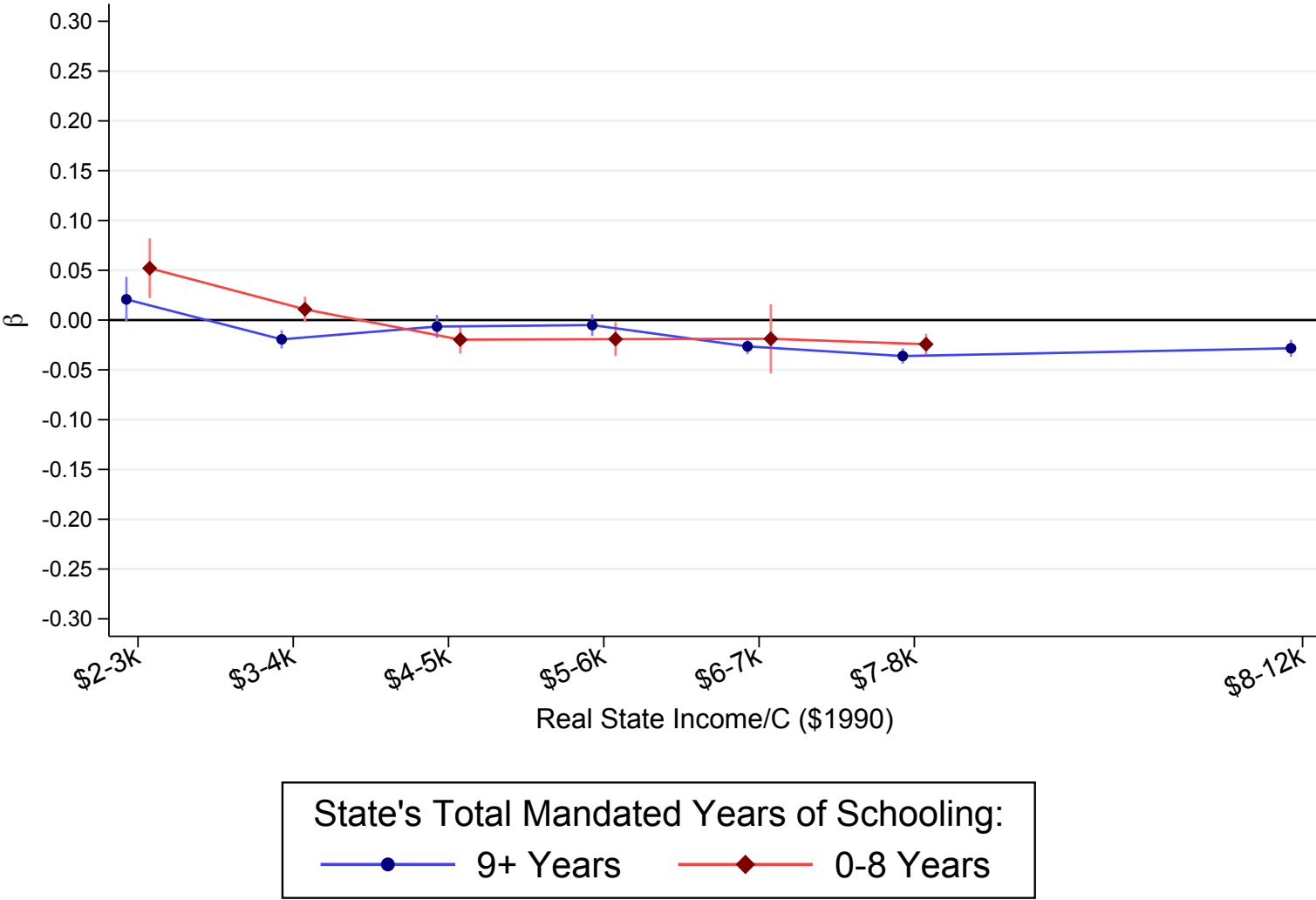
Figure A18 - By US Schooling Age of Entry, 1930-1940, Twin IV



State's Required Age of Entry:

—●— 6 or 7 Entry      —◆— 8, 9, or No Law

Figure A19 - By US Schooling Mandated Years, 1930-1940, Twin IV









USA	1870	IPUMS-USA	18,167	0.20%	\$2,497	5.2%	61.5%	0.81%	29.1	21.0			(0.004)	<b>-0.023</b>	<b>0.360</b>	0.114	0.008	-0.441
Uganda	1991	IPUMS-I: Africa	84,404	0.93%	\$584	72.7%	66.1%	1.81%	28.0	19.7	X		(0.004)	(0.020)	(0.066)	(0.005)	(0.539)	
Uganda	1995	DHS	2,144	0.02%	\$654	65.0%	67.4%	0.75%	28.1	19.9	X		(0.004)	(0.007)	(0.040)	(0.003)	(4.401)	
Uganda	2000	DHS	2,236	0.02%	\$780	79.8%	69.3%	0.55%	27.8	19.7	X	X	0.045	<b>0.331</b>	-0.219	-0.019	1.322	
Uganda	2002	DHS											(0.030)	(0.062)	(0.387)	(0.019)	(1.726)	
Uganda	2002	IPUMS-I: Africa	136,380	1.51%	\$835	58.2%	69.8%	2.55%	28.0	19.5	X	X	0.012	<b>0.388</b>	-0.077	-0.007	-1.956	
Uganda	2006	DHS	2,685	0.03%	\$989	88.9%	68.5%	0.97%	28.3	19.9	X	X	(0.026)	(0.055)	(0.318)	(0.018)	(5.519)	
Uganda	2011	DHS	2,593	0.03%	\$1,158	75.8%	66.1%	1.13%	28.2	20.0	X	X	<b>-0.017</b>	<b>0.277</b>	0.023	0.000	-11.084	
Vietnam	1989	IPUMS-I: Asia	166,529	1.84%	\$1,009	87.9%	55.4%	1.06%	29.4	21.8	X	X	(0.003)	(0.004)	(0.030)	(0.002)	(147.754)	
Vietnam	1997	DHS	1,910	0.02%	\$1,560	92.0%	43.1%	0.57%	30.1	21.6	X	X	0.025	<b>0.301</b>	-0.117	<b>0.039</b>	-0.078	
Vietnam	1999	IPUMS-I: Asia	133,016	1.47%	\$1,739	85.3%	37.6%	0.61%	30.0	21.4	X	X	(0.017)	(0.059)	(0.262)	(0.016)	(0.331)	
Vietnam	2002	DHS	1,634	0.02%	\$2,039	93.1%	28.5%	0.22%	30.3	21.2	X	X	-0.037	<b>0.348</b>	-0.143	0.013	2.092	
Yemen	1991	DHS	1,505	0.02%	\$2,380	12.1%	78.6%	0.87%	28.7	20.1	X	X	(0.025)	(0.048)	(0.267)	(0.018)	(3.278)	
Zambia	1990	IPUMS-I: Africa	33,408	0.37%	\$772	27.8%	69.4%	2.25%	28.3	19.1	X		(0.002)	(0.008)	(0.027)	(0.003)	(0.067)	
Zambia	1992	DHS	1,963	0.02%	\$730	59.1%	65.7%	0.70%	28.2	19.5	X		0.004	<b>0.635</b>	-0.120	<b>0.087</b>	-0.070	
Zambia	1996	DHS	2,302	0.03%	\$635	53.2%	62.9%	0.78%	28.1	19.7	X	X	(0.015)	(0.082)	(0.203)	(0.022)	(0.153)	
Zambia	2001	DHS	2,288	0.03%	\$616	60.8%	62.1%	0.58%	28.1	19.6	X	X	0.003	<b>0.620</b>	<b>-0.079</b>	<b>0.068</b>	-0.029	
Zambia	2007	DHS	2,267	0.03%	\$716	52.6%	64.5%	1.41%	28.3	19.9	X	X	(0.003)	(0.011)	(0.029)	(0.003)	(0.036)	
Zambia	2010	IPUMS-I: Africa	78,308	0.87%	\$795	52.8%	66.8%	1.78%	28.3	19.4	X		-0.030	<b>0.628</b>	<b>0.063</b>	<b>0.129</b>	-0.164	
Zambia	2013	DHS	5,091	0.06%	\$795	57.0%	65.0%	0.99%	28.6	19.8	X	X	(0.015)	(0.094)	(0.021)	(0.023)	(0.106)	
Zimbabwe	1994	DHS	1,467	0.02%	\$1,341	59.8%	59.1%	1.13%	28.9	20.1	X	X	-0.031	<b>0.200</b>	-0.278	0.004	1.325	
Zimbabwe	1999	DHS	1,240	0.01%	\$1,311	57.1%	47.5%	0.31%	28.7	20.2	X	X	(0.024)	(0.057)	(0.303)	(0.019)	(7.594)	
Zimbabwe	2005	DHS	2,135	0.02%	\$872	37.4%	44.6%	1.00%	28.8	20.2	X	X	<b>-0.043</b>	<b>0.298</b>	0.007	-0.004	-0.432	
Zimbabwe	2010	DHS	2,246	0.02%	\$750	38.1%	40.6%	1.00%	28.9	20.3	X	X	(0.006)	(0.009)	(0.056)	(0.004)	(1.189)	
													<b>-0.089</b>	<b>0.288</b>	-0.246	0.007	2.960	
													(0.029)	(0.067)	(0.425)	(0.018)	(9.012)	
													<b>-0.093</b>	<b>0.534</b>	-0.051	-0.014	0.227	
													(0.026)	(0.046)	(0.227)	(0.018)	(1.601)	
													<b>-0.054</b>	<b>0.451</b>	-0.011	-0.001	12.981	
													(0.027)	(0.066)	(0.311)	(0.018)	(309.046)	
													-0.048	<b>0.394</b>	0.056	-0.002	-0.931	
													(0.028)	(0.052)	(0.253)	(0.018)	(11.688)	
													<b>0.017</b>	<b>0.321</b>	0.068	-0.002	1.337	
													(0.004)	(0.006)	(0.042)	(0.003)	(2.727)	
													0.041	<b>0.432</b>	-0.006	-0.002	6.480	
													(0.021)	(0.045)	(0.183)	(0.013)	(49.240)	
													-0.058	<b>0.422</b>	-0.547	0.035	0.622	
													(0.034)	(0.074)	(0.383)	(0.022)	(0.875)	
													-0.011	<b>0.561</b>	0.260	0.022	1.779	
													(0.038)	(0.141)	(0.354)	(0.025)	(2.443)	
													<b>-0.120</b>	<b>0.631</b>	0.009	-0.007	-5.259	
													(0.027)	(0.064)	(0.184)	(0.022)	(16.944)	
													<b>-0.117</b>	<b>0.600</b>	-0.244	0.006	-1.728	
													(0.026)	(0.060)	(0.167)	(0.019)	(6.744)	



Morocco	1994	IPUMS-I: Africa	60,890	0.60%	\$2,626	11.4%	66.0%	1.43%	29.5	20.8	X	X	(0.024)	(0.082)	(0.496)	(0.018)	(0.678)
Morocco	2003	DHS	2,718	0.03%	\$3,167	11.8%	53.2%	0.58%	29.6	21.0	X	X	<b>-0.058</b> (0.004)	<b>0.332</b> (0.009)	0.034 (0.034)	<b>0.021</b> (0.003)	0.129 (0.123)
Morocco	2004	IPUMS-I: Africa	60,390	0.60%	\$3,286	10.4%	53.0%	1.23%	29.6	21.0	X	X	<b>-0.037</b> (0.017)	<b>0.541</b> (0.067)	0.188 (0.200)	<b>0.045</b> (0.017)	-0.175 (0.288)
Mozambique	2011	DHS	3,843	0.04%	\$2,613	41.6%	63.0%	1.18%	28.5	19.9	X	X	<b>-0.015</b> (0.003)	<b>0.487</b> (0.010)	-0.009 (0.023)	<b>0.034</b> (0.003)	0.034 (0.074)
Namibia	1992	DHS	988	0.01%	\$3,335	35.6%	54.2%	0.93%	29.0	21.1	X	X	-0.044 (0.023)	<b>0.425</b> (0.034)	-0.162 (0.164)	-0.008 (0.016)	0.133 (2.278)
Namibia	2000	DHS	1,108	0.01%	\$3,652	38.5%	44.2%	1.85%	29.6	21.3	X	X	<b>-0.085</b> (0.036)	<b>0.435</b> (0.070)	<b>-0.669</b> (0.253)	0.025 (0.029)	-0.792 (1.465)
Namibia	2006	DHS	1,413	0.01%	\$4,277	49.0%	38.3%	1.26%	29.3	21.2	X	X	<b>-0.208</b> (0.041)	<b>0.505</b> (0.058)	-0.035 (0.305)	0.004 (0.034)	-8.258 (62.033)
Nicaragua	1971	PUMS-I: America	10,485	0.10%	\$2,906	17.2%	74.5%	0.77%	28.3	19.5	X		<b>-0.134</b> (0.036)	<b>0.503</b> (0.055)	-0.132 (0.255)	0.044 (0.027)	-0.848 (0.840)
Panama	1970	PUMS-I: America	8,373	0.08%	\$3,828	23.6%	72.1%	1.28%	28.4	20.0	X		<b>-0.112</b> (0.010)	<b>0.284</b> (0.023)	-0.052 (0.145)	<b>0.025</b> (0.008)	-0.169 (0.292)
Panama	1980	PUMS-I: America	10,736	0.11%	\$4,850	30.4%	62.7%	1.45%	28.9	20.1	X		<b>-0.152</b> (0.012)	<b>0.298</b> (0.019)	0.205 (0.150)	0.008 (0.009)	0.279 (1.302)
Panama	1990	PUMS-I: America	12,549	0.12%	\$4,818	29.9%	55.3%	1.44%	29.0	20.2	X		<b>-0.145</b> (0.011)	<b>0.397</b> (0.019)	0.026 (0.094)	0.012 (0.008)	-0.279 (0.757)
Paraguay	1982	PUMS-I: America	15,623	0.15%	\$3,193	15.3%	63.2%	0.97%	28.5	20.6	X		<b>-0.153</b> (0.009)	<b>0.465</b> (0.016)	0.033 (0.074)	<b>0.046</b> (0.008)	0.152 (0.180)
Paraguay	1990	DHS	1,519	0.02%	\$3,226	34.2%	60.1%	1.02%	29.1	20.8	X	X	<b>-0.112</b> (0.007)	<b>0.394</b> (0.018)	0.100 (0.080)	<b>0.020</b> (0.007)	0.209 (0.310)
Paraguay	1992	PUMS-I: America	22,777	0.23%	\$3,274	19.4%	61.6%	0.97%	29.0	20.6	X		<b>-0.162</b> (0.030)	<b>0.422</b> (0.073)	0.146 (0.305)	0.029 (0.023)	-0.574 (0.909)
Paraguay	2002	PUMS-I: America	24,926	0.25%	\$2,997	36.8%	56.3%	1.03%	29.3	20.3	X		<b>-0.127</b> (0.006)	<b>0.396</b> (0.015)	-0.010 (0.065)	<b>0.031</b> (0.006)	0.181 (0.177)
Peru	1991	DHS	3,929	0.04%	\$3,196	52.8%	56.9%	0.59%	29.1	20.6	X	X	<b>-0.141</b> (0.007)	<b>0.447</b> (0.014)	0.050 (0.068)	<b>0.025</b> (0.006)	-0.089 (0.242)
Peru	1993	PUMS-I: America	113,466	1.12%	\$3,220	24.4%	55.1%	0.92%	29.2	20.6	X		<b>-0.043</b> (0.020)	<b>0.439</b> (0.053)	<b>-0.525</b> (0.268)	-0.002 (0.015)	11.967 (98.079)
Peru	1996	DHS	7,325	0.07%	\$3,531	51.3%	55.4%	0.40%	29.3	20.5	X		<b>-0.092</b> (0.003)	<b>0.449</b> (0.007)	0.014 (0.030)	<b>0.023</b> (0.003)	0.037 (0.111)
Peru	2000	DHS	6,371	0.06%	\$3,766	57.2%	49.4%	0.48%	29.6	20.6	X	X	<b>-0.093</b> (0.017)	<b>0.496</b> (0.039)	-0.051 (0.218)	<b>0.031</b> (0.013)	-0.357 (0.493)
Peru	2007	PUMS-I: America	115,601	1.15%	\$4,923	34.0%	41.4%	0.94%	29.6	20.4	X		<b>-0.052</b> (0.018)	<b>0.504</b> (0.042)	0.012 (0.186)	0.011 (0.014)	-0.243 (1.506)
Peru	2007	DHS	7,867	0.08%	\$4,923	67.1%	44.3%	0.64%	29.9	20.4	X	X	<b>-0.095</b> (0.003)	<b>0.591</b> (0.006)	-0.024 (0.024)	<b>0.026</b> (0.003)	0.068 (0.107)
Philippines	2008	DHS	2,717	0.03%	\$2,863	42.3%	54.0%	0.67%	29.5	21.4	X	X	-0.006 (0.016)	<b>0.609</b> (0.044)	0.097 (0.138)	<b>0.029</b> (0.014)	0.252 (0.528)
Philippines	2013	DHS	3,014	0.03%	\$3,024	42.1%	48.7%	0.54%	29.6	21.3	X	X	<b>-0.096</b> (0.023)	<b>0.452</b> (0.054)	-0.152 (0.224)	0.011 (0.018)	1.877 (3.688)
Romania	1992	IPUMS-I: Europe	100,657	1.00%	\$3,191	74.0%	34.4%	0.89%	29.8	20.8	X	X	<b>-0.113</b> (0.020)	<b>0.501</b> (0.066)	0.262 (0.244)	<b>0.036</b> (0.017)	0.119 (0.521)
Romania	2002	IPUMS-I: Europe	71,737	0.71%	\$3,456	54.3%	22.5%	0.87%	30.2	20.7	X	X	<b>-0.192</b> (0.003)	<b>0.655</b> (0.006)	<b>-0.065</b> (0.023)	<b>0.034</b> (0.003)	-0.075 (0.079)
South Africa	1996	IPUMS-I: Africa	133,590	1.32%	\$3,700	68.2%	47.4%	2.34%	29.7	20.3	X		<b>-0.168</b> (0.004)	<b>0.786</b> (0.005)	<b>-0.080</b> (0.025)	<b>0.035</b> (0.003)	0.083 (0.107)
South Africa	2001	IPUMS-I: Africa	136,950	1.36%	\$4,005	72.3%	43.2%	2.44%	29.8	20.3	X		<b>-0.095</b> (0.003)	<b>0.515</b> (0.004)	-0.014 (0.016)	<b>0.022</b> (0.002)	-0.093 (0.117)
South Africa	2007	IPUMS-I: Africa	33,071	0.33%	\$4,783	82.3%	39.6%	2.47%	29.7	20.4	X		<b>-0.093</b> (0.003)	<b>0.559</b> (0.004)	0.005 (0.014)	<b>0.019</b> (0.002)	0.221 (0.133)
South Africa	1998	DHS	2,067	0.02%	\$3,812	32.7%	40.0%	0.85%	29.8	20.5	X	X	<b>-0.063</b> (0.005)	<b>0.614</b> (0.007)	-0.035 (0.024)	<b>0.016</b> (0.005)	0.096 (0.275)
Sudan	2008	IPUMS-I: Africa	289,810	2.87%	\$3,021	24.7%	72.0%	1.47%	28.9	20.1	X		<b>-0.134</b> (0.027)	<b>0.578</b> (0.042)	-0.149 (0.232)	0.011 (0.022)	1.676 (4.276)
Swaziland	2006	DHS	851	0.01%	\$2,967	42.8%	50.2%	0.68%	28.8	20.1	X	X	<b>-0.009</b> (0.003)	<b>0.288</b> (0.005)	0.012 (0.035)	<b>0.018</b> (0.002)	0.095 (0.138)
Turkey	1985	IPUMS-I: Asia	150,756	1.49%	\$4,578	39.2%	57.4%	1.39%	29.0	20.1	X		<b>-0.088</b> (0.040)	<b>0.508</b> (0.105)	-0.030 (0.440)	-0.005 (0.032)	3.471 (22.011)
USA	1880	US Full Count	2,391,227	23.70%	\$3,032	6.2%	64.1%	0.66%	29.0	20.9			<b>0.103</b> (0.003)	<b>0.462</b> (0.005)	<b>0.293</b> (0.024)	<b>0.051</b> (0.002)	-0.045 (0.050)
USA	1900	IPUMS-USA	166,412	1.65%	\$4,161	6.1%	61.3%	0.89%	29.2	21.2		X	<b>-0.023</b> (0.000)	<b>0.343</b> (0.002)	<b>0.039</b> (0.006)	<b>0.009</b> (0.001)	0.042 (0.035)
Ukraine	2007	DHS	755	0.01%	\$4,487	72.0%	12.8%	0.57%	30.8	20.7	X	X	<b>-0.024</b> (0.001)	<b>0.374</b> (0.006)	-0.007 (0.016)	<b>0.011</b> (0.002)	-0.026 (0.104)
Uruguay	1963	PUMS-I: America	9,974	0.10%	\$4,909	17.4%	44.4%	1.02%	29.6	21.8	X		<b>-0.155</b> (0.055)	<b>0.855</b> (0.031)	-0.018 (0.221)	<b>0.052</b> (0.026)	-0.437 (0.662)
Uzbekistan	1996	DHS	1,275	0.01%	\$3,223	46.2%	55.9%	0.78%	28.9	21.2	X	X	<b>-0.062</b> (0.008)	<b>0.582</b> (0.018)	0.054 (0.068)	<b>0.039</b> (0.009)	0.152 (0.200)
Vietnam	2009	IPUMS-I: Asia	745,767	7.39%	\$3,063	87.2%	20.7%	0.61%	30.2	21.4	X	X	<b>-0.146</b> (0.036)	<b>0.388</b> (0.052)	0.070 (0.506)	<b>0.081</b> (0.025)	0.306 (0.392)
Yemen	2013	DHS	6,699	0.07%	\$3,165	9.4%	69.9%	0.63%	28.6	19.8		X	<b>-0.014</b> (0.001)	<b>0.790</b> (0.003)	<b>-0.066</b> (0.010)	<b>0.084</b> (0.001)	<b>-0.041</b> (0.012)
													-0.018 (0.010)	<b>0.205</b> (0.058)	0.015 (0.267)	<b>0.024</b> (0.012)	-0.578 (0.434)





\$7,500-10,000 GDP/C Bin																	
Country	Year (#Samples)	Source	N	Percent of Pooled	Mean GDP/C	In Labor Force	3 or More Children	2nd Child is Multiple Birth	Mother's Age at Survey	Mother's Age at First Birth	Education?	Month/Quarter of Birth?	OLS	FS, Twin IV	Z5, Twin IV	FS, Same-Sex	Z5, Same-Sex
Pooled	19		6,180,681		\$8,004	15.8%	46.9%	0.98%	29.5	20.9			-0.075 (0.010)	0.535 (0.009)	-0.024 (0.007)	0.024 (0.002)	0.057 (0.014)
Argentina	1980	IPUMS-I: Americas	135,408	2.19%	\$7,826	20.4%	47.7%	1.38%	29.3	21.8	X		-0.079 (0.003)	0.530 (0.006)	-0.050 (0.021)	0.043 (0.003)	0.103 (0.068)
Argentina	2001	IPUMS-I: Americas	150,620	2.44%	\$8,049	49.1%	50.0%	1.22%	29.4	20.6	X		-0.116 (0.003)	0.509 (0.005)	-0.055 (0.023)	0.023 (0.002)	-0.133 (0.110)
Armenia	2005	DHS	1,315	0.02%	\$8,617	21.5%	25.5%	0.89%	29.4	20.6	X	X	-0.062 (0.035)	0.844 (0.076)	-0.207 (0.060)	0.110 (0.028)	0.136 (0.262)
Costa Rica	2011	IPUMS-I: Americas	17,905	0.29%	\$7,997	34.9%	34.1%	0.99%	29.6	19.9	X		-0.096 (0.008)	0.656 (0.014)	0.009 (0.055)	0.033 (0.007)	0.056 (0.212)
France	1962	IPUMS-I: Europe	92,331	1.49%	\$8,073	20.3%	49.3%	2.68%	30.1	22.2	X		-0.124 (0.003)	0.519 (0.004)	-0.103 (0.014)	0.026 (0.003)	-0.173 (0.100)
Greece	1981	IPUMS-I: Europe	45,467	0.74%	\$8,897	21.3%	24.0%	1.19%	29.7	22.0	X	X	-0.024 (0.005)	0.761 (0.005)	-0.011 (0.023)	0.063 (0.004)	-0.046 (0.060)
Ireland	1981	IPUMS-I: Europe	13,484	0.22%	\$8,641	8.9%	53.2%	1.28%	29.4	22.8	X		-0.070 (0.006)	0.456 (0.017)	0.031 (0.051)	0.040 (0.008)	0.126 (0.128)
Ireland	1986	IPUMS-I: Europe	12,809	0.21%	\$9,597	16.7%	50.6%	1.12%	29.6	22.7			-0.100 (0.007)	0.481 (0.020)	-0.038 (0.062)	0.058 (0.008)	-0.105 (0.112)
Malaysia	2000	IPUMS-I: Asia	20,415	0.33%	\$7,759	34.1%	57.9%	1.66%	30.2	22.4	X		-0.080 (0.008)	0.461 (0.014)	0.209 (0.056)	0.028 (0.006)	-0.680 (0.264)
Mexico	2010	IPUMS-I: Americas	644,685	10.43%	\$7,716	33.7%	43.4%	0.94%	29.5	20.3	X		-0.111 (0.003)	0.582 (0.006)	-0.004 (0.020)	0.030 (0.002)	0.083 (0.089)
Portugal	1981	IPUMS-I: Europe	19,031	0.31%	\$7,979	46.3%	29.0%	1.02%	29.9	22.1	X		-0.141 (0.008)	0.703 (0.011)	-0.045 (0.051)	0.043 (0.006)	0.252 (0.174)
Puerto Rico	1980	IPUMS-PR	8,246	0.13%	\$7,918	35.1%	51.7%	1.84%	29.3	21.0	X	X	-0.167 (0.011)	0.465 (0.018)	-0.062 (0.082)	0.048 (0.010)	-0.191 (0.216)
USA	1940	US Full Count	4,602,622	74.47%	\$7,942	10.6%	47.1%	0.86%	29.5	20.9	X		-0.064 (0.000)	0.539 (0.001)	-0.016 (0.003)	0.021 (0.000)	0.072 (0.014)
USA	1950	IPUMS-USA	103,494	1.67%	\$9,643	14.0%	43.1%	1.02%	29.3	21.7	X		-0.079 (0.002)	0.588 (0.006)	-0.042 (0.015)	0.024 (0.003)	0.117 (0.093)
Uruguay	1996	IPUMS-I: Americas	11,642	0.19%	\$8,086	54.8%	39.9%	1.22%	29.9	21.3	X		-0.117 (0.010)	0.584 (0.017)	-0.019 (0.071)	0.029 (0.008)	-0.195 (0.311)
Uruguay	2006	IPUMS-I: Americas	9,121	0.15%	\$9,084	62.8%	41.0%	1.24%	30.0	20.6	X		-0.148 (0.013)	0.563 (0.028)	-0.076 (0.100)	0.027 (0.011)	-0.306 (0.459)
Venezuela	1981	IPUMS-I: Americas	80,451	1.30%	\$9,827	26.1%	60.9%	2.36%	28.6	20.4	X		-0.134 (0.004)	0.380 (0.005)	-0.012 (0.026)	0.029 (0.003)	0.062 (0.106)
Venezuela	1990	IPUMS-I: Americas	98,117	1.59%	\$8,785	32.1%	56.0%	2.35%	29.1	20.3	X		-0.152 (0.004)	0.427 (0.005)	-0.075 (0.026)	0.030 (0.003)	-0.157 (0.113)
Venezuela	2001	IPUMS-I: Americas	113,518	1.84%	\$8,138	33.5%	49.5%	1.45%	29.3	20.1	X		-0.132 (0.003)	0.518 (0.005)	-0.043 (0.022)	0.035 (0.003)	0.064 (0.081)

\$10,000-15,000 GDP/C Bin																	
Country	Year (#Samples)	Source	N	Percent of Pooled	Mean GDP/C	In Labor Force	3 or More Children	2nd Child is Multiple Birth	Mother's Age at Survey	Mother's Age at First Birth	Education?	Month/Quarter of Birth?	OLS	FS, Twin IV	Z5, Twin IV	FS, Same-Sex	Z5, Same-Sex
Pooled	19		1,062,881		\$11,484	30.4%	45.1%	1.42%	29.7	21.5			-0.126 (0.015)	0.521 (0.047)	-0.063 (0.018)	0.035 (0.002)	-0.066 (0.021)
Armenia	2010	DHS	1,178	0.11%	\$10,215	22.5%	19.9%	0.73%	29.6	21.2	X	X	-0.061 (0.038)	0.810 (0.040)	0.072 (0.207)	0.125 (0.025)	-0.233 (0.235)
Armenia	2011	IPUMS-I: Asia	15,059	1.42%	\$10,215	47.4%	22.7%	0.93%	29.7	21.4	X	X	-0.012 (0.010)	0.794 (0.013)	-0.111 (0.051)	0.107 (0.006)	-0.088 (0.076)
Austria	1971	IPUMS-I: Europe	30,982	2.91%	\$10,195	34.3%	40.9%	1.04%	29.4	21.6	X		-0.076 (0.006)	0.594 (0.010)	-0.055 (0.044)	0.026 (0.005)	-0.060 (0.210)
Austria	1981	IPUMS-I: Europe	27,991	2.63%	\$13,779	43.6%	29.9%	1.00%	29.8	21.2	X		-0.102 (0.007)	0.697 (0.008)	-0.143 (0.041)	0.042 (0.005)	-0.253 (0.140)
Chile	2002	PUMS-I: America	56,760	5.34%	\$10,777	31.4%	31.1%	0.94%	30.4	20.8	X		-0.081 (0.004)	0.688 (0.007)	-0.044 (0.028)	0.026 (0.004)	-0.187 (0.149)
France	1968	IPUMS-I: Europe	95,250	8.96%	\$10,432	24.5%	46.6%	1.05%	30.0	22.3	X		-0.153 (0.003)	0.539 (0.006)	-0.084 (0.024)	0.033 (0.003)	-0.104 (0.082)
France	1975	IPUMS-I: Europe	103,331	9.72%	\$13,254	36.9%	38.9%	1.13%	29.4	21.8	X		-0.249 (0.003)	0.607 (0.006)	-0.172 (0.021)	0.026 (0.003)	0.088 (0.120)
Greece	1991	IPUMS-I: Europe	40,657	3.83%	\$10,062	37.0%	21.8%	1.22%	30.3	21.5	X	X	-0.080 (0.006)	0.781 (0.005)	-0.054 (0.027)	0.059 (0.004)	-0.035 (0.081)
Greece	2001	IPUMS-I: Europe	28,882	2.72%	\$12,660	51.6%	20.4%	1.13%	31.1	22.5	X		-0.070 (0.007)	0.801 (0.006)	-0.086 (0.034)	0.042 (0.005)	0.038 (0.139)
Ireland	1991	IPUMS-I: Europe	10,937	1.03%	\$11,843	31.3%	45.7%	1.24%	30.0	22.7	X		-0.144 (0.010)	0.550 (0.021)	-0.096 (0.068)	0.060 (0.008)	-0.279 (0.146)
Portugal	1991	IPUMS-I: Europe	15,987	1.50%	\$10,872	63.3%	22.8%	1.15%	30.7	21.5	X		-0.184 (0.010)	0.771 (0.009)	-0.046 (0.047)	0.021 (0.006)	0.120 (0.375)
Portugal	2001	IPUMS-I: Europe	11,704	1.10%	\$13,831	74.5%	16.8%	1.13%	31.2	22.2	X		-0.144 (0.012)	0.866 (0.010)	-0.061 (0.045)	0.026 (0.007)	-0.359 (0.330)
Portugal	2011	IPUMS-I: Europe	8,445	0.79%	\$14,279	80.7%	17.2%	1.35%	31.6	22.8	X		-0.164 (0.013)	0.851 (0.011)	-0.017 (0.042)	0.025 (0.008)	-0.225 (0.331)
Puerto Rico	1990	IPUMS-PR	8,442	0.79%	\$10,477	41.7%	47.0%	1.42%	29.7	20.9	X		-0.148 (0.012)	0.509 (0.018)	-0.096 (0.089)	0.055 (0.011)	0.011 (0.204)
Puerto Rico	2000	IPUMS-PR	7,809	0.73%	\$13,881	43.1%	40.7%	1.41%	29.7	21.0	X		-0.106 (0.013)	0.561 (0.020)	-0.193 (0.084)	0.042 (0.011)	-0.458 (0.283)
Spain	1991	IPUMS-I: Europe	59,957	5.64%	\$12,030	40.0%	23.2%	1.07%	31.1	22.4	X		-0.112 (0.005)	0.768 (0.006)	-0.095 (0.024)	0.045 (0.003)	-0.051 (0.088)
USA	1960	IPUMS-USA	470,378	44.26%	\$11,380	22.8%	55.1%	1.70%	29.3	21.4	X	X	-0.117 (0.001)	0.452 (0.002)	-0.033 (0.010)	0.035 (0.001)	-0.084 (0.034)
Uruguay	2011	PUMS-I: America	10,012	0.94%	\$11,526	65.7%	36.5%	0.88%	30.1	20.5	X	X	-0.142 (0.011)	0.628 (0.020)	-0.016 (0.080)	0.026 (0.009)	-0.478 (0.380)
Venezuela	1971	PUMS-I: America	59,120	5.56%	\$10,429	16.0%	70.5%	2.28%	28.4	20.1	X		-0.083 (0.004)	0.289 (0.006)	-0.044 (0.034)	0.017 (0.003)	0.416 (0.207)

\$15,000-20,000 GDP/C Bin																	
Country	Year (#Samples)	Source	N	Percent of Pooled	Mean GDP/C	In Labor Force	3 or More Children	2nd Child is Multiple Birth	Mother's Age at Survey	Mother's Age at First Birth	Education?	Month/Quarter of Birth?	OLS	F5, Twin IV	Z5, Twin IV	F5, Same-Sex	Z5, Same-Sex
Pooled	13		1,053,271		\$17,532	52.0%	34.8%	1.24%	30.2	21.8			-0.206 (0.029)	0.651 (0.031)	-0.083 (0.020)	0.046 (0.004)	-0.143 (0.018)
Austria	1991	IPUMS-I: Europe	28,036	2.66%	\$16,956	51.4%	24.7%	0.93%	30.2	21.8	X		-0.117 (0.007)	0.763 (0.008)	-0.136 (0.040)	0.036 (0.005)	-0.232 (0.167)
France	1982	IPUMS-I: Europe	117,660	11.17%	\$15,076	52.2%	33.5%	1.08%	30.3	22.0	X		-0.339 (0.003)	0.663 (0.005)	-0.212 (0.020)	0.041 (0.003)	-0.243 (0.068)
France	1990	IPUMS-I: Europe	91,261	8.66%	\$17,309	64.1%	34.1%	1.04%	30.7	22.4	X		-0.357 (0.003)	0.656 (0.006)	-0.207 (0.025)	0.042 (0.003)	-0.160 (0.072)
France	1999	IPUMS-I: Europe	86,473	8.21%	\$19,690	68.1%	29.6%	1.24%	31.3	23.5	X		-0.279 (0.004)	0.706 (0.006)	-0.061 (0.020)	0.039 (0.003)	-0.203 (0.076)
Great Britain	1991	IPUMS-I: Europe	20,003	1.90%	\$16,403	46.2%	32.1%	1.11%	30.3	22.5			-0.222 (0.008)	0.705 (0.012)	-0.160 (0.045)	0.079 (0.006)	-0.232 (0.086)
Ireland	1996	IPUMS-I: Europe	9,165	0.87%	\$15,683	43.1%	39.8%	1.16%	30.2	22.9	X		-0.172 (0.011)	0.634 (0.019)	-0.065 (0.076)	0.064 (0.009)	-0.217 (0.156)
Puerto Rico	2010	IPUMS-PR	4,397	0.42%	\$15,074	57.1%	36.0%	1.39%	30.0	20.8	X	X	-0.159 (0.018)	0.635 (0.029)	-0.149 (0.106)	0.064 (0.014)	-0.070 (0.243)
Spain	2001	IPUMS-I: Europe	34,927	3.32%	\$15,874	51.2%	16.2%	2.31%	31.9	23.7	X	X	-0.066 (0.007)	0.882 (0.003)	-0.025 (0.020)	0.034 (0.004)	-0.072 (0.156)
Spain	2011	IPUMS-I: Europe	39,595	3.76%	\$16,797	87.7%	15.0%	1.05%	32.0	24.7	X	X	-0.098 (0.009)	0.864 (0.007)	-0.016 (0.027)	0.029 (0.005)	0.132 (0.171)
Switzerland	1970	IPUMS-I: Europe	11,998	1.14%	\$16,668	21.8%	35.6%	0.81%	30.2	23.2	X		-0.083 (0.008)	0.655 (0.016)	-0.075 (0.058)	0.019 (0.008)	-0.230 (0.403)
Switzerland	1980	IPUMS-I: Europe	11,241	1.07%	\$18,315	28.4%	23.1%	0.70%	30.8	23.1	X		-0.079 (0.010)	0.790 (0.011)	-0.167 (0.048)	0.042 (0.008)	-0.339 (0.202)
USA	1970	IPUMS-USA	93,241	8.85%	\$15,334	33.4%	52.5%	1.41%	29.3	20.8	X	X	-0.139 (0.003)	0.463 (0.006)	0.014 (0.028)	0.034 (0.003)	-0.105 (0.088)
USA	1980	IPUMS-USA	505,274	47.97%	\$18,487	49.3%	36.5%	1.27%	29.8	21.1	X	X	-0.177 (0.001)	0.621 (0.002)	-0.075 (0.010)	0.053 (0.001)	-0.127 (0.026)

\$20,000 and Above GDP/C Bin																	
Country	Year (#Samples)	Source	N	Percent of Pooled	Mean GDP/C	In Labor Force	3 or More Children	2nd Child is Multiple Birth	Mother's Age at Survey	Mother's Age at First Birth	Education?	Month/Quarter of Birth?	OLS	FS, Twins	2S, Twins	FS, Same-Sex	2S, Same-Sex
Pooled	11		2,377,681		\$24,420	67.8%	33.2%	1.45%	30.7	22.8			-0.191 (0.024)	0.668 (0.016)	-0.085 (0.008)	0.044 (0.003)	-0.140 (0.015)
Austria	2001	IPUMS-I: Europe	24,022	1.01%	\$20,997	72.7%	23.6%	1.00%	31.1	22.8	X		-0.127 (0.007)	0.781 (0.009)	-0.153 (0.041)	0.041 (0.005)	-0.200 (0.140)
France	2006	IPUMS-I: Europe	510,203	21.46%	\$21,540	73.3%	28.8%	1.43%	31.3	24.0	X		-0.263 (0.002)	0.707 (0.002)	-0.100 (0.008)	0.037 (0.001)	-0.210 (0.034)
France	2011	IPUMS-I: Europe	485,266	20.41%	\$21,477	76.2%	29.3%	1.46%	31.2	24.1	X		-0.248 (0.002)	0.702 (0.003)	-0.105 (0.008)	0.038 (0.001)	-0.156 (0.032)
Ireland	2002	IPUMS-I: Europe	7,664	0.32%	\$22,315	45.8%	35.4%	1.55%	30.2	23.0	X		-0.180 (0.013)	0.663 (0.018)	-0.158 (0.067)	0.037 (0.010)	-0.097 (0.300)
Ireland	2006	IPUMS-I: Europe	8,025	0.34%	\$24,076	55.6%	32.8%	1.37%	30.0	22.9	X		-0.182 (0.013)	0.681 (0.018)	0.036 (0.070)	0.047 (0.010)	-0.128 (0.231)
Ireland	2011	IPUMS-I: Europe	10,654	0.45%	\$22,013	62.0%	34.0%	1.40%	31.3	23.7	X		-0.176 (0.011)	0.680 (0.013)	-0.188 (0.059)	0.048 (0.009)	0.172 (0.200)
Switzerland	1990	IPUMS-I: Europe	10,612	0.45%	\$20,699	38.7%	26.7%	1.05%	31.0	23.8	X		-0.116 (0.011)	0.751 (0.012)	-0.022 (0.058)	0.043 (0.008)	-0.274 (0.213)
Switzerland	2000	IPUMS-I: Europe	8,685	0.37%	\$22,122	61.0%	26.1%	1.01%	31.7	24.6	X		-0.152 (0.012)	0.763 (0.016)	-0.164 (0.069)	0.043 (0.009)	0.143 (0.244)
USA	1990	IPUMS-USA	505,189	21.25%	\$22,901	60.6%	35.7%	1.28%	30.2	21.7	X		-0.166 (0.002)	0.647 (0.002)	-0.084 (0.011)	0.051 (0.001)	-0.134 (0.030)
USA	2000	IPUMS-USA	438,854	18.46%	\$28,100	62.8%	36.5%	1.58%	30.3	21.9	X		-0.136 (0.002)	0.638 (0.002)	-0.073 (0.010)	0.049 (0.002)	-0.102 (0.033)
USA	2010	IPUMS-USA	368,507	15.50%	\$30,491	66.2%	38.1%	1.57%	30.4	21.9	X	X	-0.141 (0.002)	0.622 (0.003)	-0.049 (0.012)	0.048 (0.002)	-0.125 (0.040)