Culture: An Empirical Investigation of Beliefs, Work, and Fertility

A Verification and Reproduction of Fernández and Fogli (*American Economic Journal: Macroeconomics*, 2009)

Victor Gay*

Journal of Comments and Replications in Economics, Volume 2, 2023-2, DOI: *10.18718/81781.28*

JEL: J13, J16, J22, J24, Z13
Keywords: Culture, Female Labor Force Participation, Fertility, Replication

Data Availability: The data and Stata code to reproduce the results of this replication can be downloaded at JCRE's data archive (DOI: 10.15456/j1.2022252.0530923662).

Please Cite As: Gay, Victor (2023). Culture: An Empirical Investigation of Beliefs, Work, and Fertility: A Verification and Reproduction of Fernández and Fogli (2009). *Journal of Comments and Replications in Economics*, Vol 2(2023-2). DOI: 10.18718/81781.28

Abstract

In this article, I perform a verification and a reproduction of the main results in Fernández and Fogli (2009), which estimates the role of culture in explaining the labor and fertility decisions of second generation immigrant women to the United States in 1970. While I am able to verify Fernández and Fogli's (2009) main results as well as their robustness relative to both labor and fertility decisions, I am unable to reproduce them relative to labor decisions in alternative samples drawn from the same underlying population.

*Department of Social and Behavioral Sciences, Toulouse School of Economics and Institute for Advanced Study in Toulouse, University of Toulouse Capitole, Toulouse, France, victor.gay@tse-fr.eu.
Acknowledgements: I gratefully acknowledge funding from the ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program).

Received September 09, 2022; Revised September 02, 2023; Accepted September 04, 2023; Published October 16, 2023.

1 Introduction

Despite dramatic improvements over the past half-century, women's economic conditions relative to men's still exhibit profound disparities across countries, in particular regarding their labor market involvement (Olivetti and Petrongolo, 2016; Klasen, 2019). To account for such persistent gender gaps, economists have increasingly appealed to cultural explanations to complement more traditional economic ones (Fernández, 2007; Giuliano, 2021)[1]. When doing so, a common empirical strategy to distinguish the role of cultural factors consists in using an "epidemiological approach" (Fernández, 2011)[2]. This strategy identifies cultural effects by comparing behaviors among individuals with different cultural origins but who are embedded within the same institutional environment, thereby facing similar external incentives when making decisions. To proxy for culture, this approach uses past aggregate outcomes of individuals' places of origin, as only the cultural component of these variables should exhibit some explanatory power in this setting. Fernández and Fogli (2009, American Economic Journal: Macroeconomics, henceforth FF) was among the first studies to systematically apply this method[3]. Therein, the role of culture is highlighted in explaining the labor and fertility decisions of second generation immigrant women in the United States in 1970.

FF's empirical approach has been highly influential: this article is credited with 509 citations as of September 2023[4]. This represents more than twice as many citations to any article published in the same (inaugural) issue of AEJ: Macroeconomics or in inaugural issues of all three other AEJ journals. FF's impact per this citations metric also fares well compared to articles published in the concurrent issue of the American Economic Review, as only 3 of its 22 articles have received more citations than FF to date (Table A.1)[5]. Moreover, FF has dramatically influenced empirical methods in cultural economics: among its 509 citations, 138 studies have applied FF's empirical approach in a variety of contexts in order to provide a cultural explanation to variations in economic and demographic behaviors (Table A.2)[6].

Despite its status of seminal study, there has been no attempt to replicate FF's results[7]. This article fills this gap by conducting a replication of FF's main result, i.e., that women from countries with traditionally higher rates of female labor force participation (FLFP) are more likely to be in the

[1]Traditional economic explanations that account for gender gaps in labor force participation include, among others, economic development and the structural transformation along with changes in female education and fertility (Goldin, 1990; Gaddis and Klasen, 2014; Ngai and Petrongolo, 2017), medical progress in maternal health (Albanesi and Olivetti, 2016), declining prices of labor-saving consumer durable goods (Greenwood, Seshadri and Yorukoglu, 2005), and the increasing availability of oral contraceptives (Goldin and Katz, 2002; Bailey, 2006).

[2]Here, culture is understood as "those customary beliefs and values that ethnic, religious, and social groups transmit fairly unchanged from generation to generation" (Guiso, Sapienza and Zingales, 2006, p. 23, cited in Giuliano, 2021).

[3]Earlier uses of comparable methods include Blau (1992), Carroll, Rhee and Rhee (1994), Antecol (2000), and Fortin (2005). The NBER Working Paper version of FF was published in 2005.

[4]This citation count is based on data from Clarivate Web of Science—Google Scholar credits FF with 1,696 citations as of September 2023. Further including journal articles citing FF's NBER Working paper version increases the Clarivate Web of Science citation count by an additional 34 citations.

[5]Considering annual rather than total citations, it is clear that starting from 2016 FF has been in the same category as these top-three AER articles in terms of citations (Figure B.1).

[6]Though widely used, FF's epidemiological approach is not without criticism, as selection into migration might bias the cultural effects identified through this method (Beblo, Görges and Markowsky, 2020a; 2020b).

[7]The data and reproduction code were not made available by the authors. The webpage of the article (https://www.aeaweb.org/articles?id=10.1257/mac.1.1.146) and of both authors (https://sites.google.com/site/raquelfernandezsite and https://sites.google.com/site/alessandrafoglisite) were accessed in September 2023.

labor force and that those from countries with traditionally lower total fertility rates (TFR) have relatively less children. More specifically, I replicate results in FF's Table 2, which reports coefficients from estimating the following equation on a sample of second-generation immigrant women to the United States aged 30 to 40 in 1970:

$$Z_{isj} = \beta_0 + \boldsymbol{\beta}_1' \mathbf{X}_i + \beta_2 \tilde{Z}_j + f_s + \varepsilon_{isj}, \tag{1}$$

where Z_{isj} is the work or fertility decision of woman i residing in the Standard Metropolitan Statistical Area (SMSA) s and of ancestry j[8]. $\mathbf{X}_i$ includes a set of individual characteristics, f_s is a set of SMSA of residence fixed effects, and $\tilde{Z}_j$, a proxy for culture—past values of FLFP or TFR for i's country of ancestry j. Standard errors are clustered at the country-of-ancestry level.

Throughout this replication attempt, I strictly follow Clemens's (2017, p. 327) definition of the nature of a replication test:

> A *replication* test estimates parameters drawn from the same sampling distribution as those in the original study. [...] A replication test can take two forms: A *verification* test means ensuring that the exact statistical analysis reported in the original paper gives materially the same results reported in the paper, either using the original data set or remeasuring with identical methods the same traits of the same sample of subjects. [...] A *reproduction* test means resampling precisely the same population but otherwise using identical methods to the original study.

I first attempt to *verify* estimates reported in FF's Table 2 by constructing the same regression sample and estimating Equation 1 based on the guidelines provided in FF's original article. Then, I attempt to *reproduce* estimates reported in FF's Table 2 by resampling precisely the same population but otherwise using identical methods. In particular, while FF's analysis relies on the Metro sample of the US census of 1970, I use two alternative samples that are drawn from the same underlying population: the State and the Neighborhood samples of the US census of 1970. Overall, while I am able to verify estimates reported in FF's Table 2 as well as their robustness relative to both labor and fertility outcomes, I am unable to reproduce these estimates relative the labor outcome in alternative samples drawn from the same underlying population.

In the reminder of this article, I describe the construction of the dataset used in this replication exercise (Section 2) then perform the verification (Section 3) and reproduction (Section 5) tests of the estimates reported in FF's Table 2.

2 Data

In this section, I describe the procedures I implement to reconstruct the regression sample of FF's Table 2 (Section 2.1), its analysis variables (Section 2.2), and its cultural proxy variables (Section 2.3). I further describe the construction of alternative samples drawn from the same underlying population (Section 2.1). To assess the accuracy of my procedures relative to FF's original

[8] FF's Table 2 (p. 157) is reproduced from the original article in Table A.3.

dataset, I leverage summary statistics at the country and individual levels reported in FF's Tables 1 and A.1[9].

2.1 Regression Sample

The 1970 Metro Sample of the US Census FF's Table 2 uses the 1 percent 1970 Form 2 Metro Sample of the US Census, which can be retrieved from IPUMS USA (Ruggles et al., 2021)[10]. It is a 1-in-100 random sample drawn from the 15 percent random sample of the population that was given Form 2 and in which the smallest identifiable geographic units are SMSAs (Bureau of the Census, 1972, p. 194–195).

Sample selection procedures The regression sample of the analysis in FF's Table 2 includes married women aged 30 to 40 residing in non-farming households, who hold non-agricultural occupations, were born in an identified US state, and whose fathers' were born in an identified country outside the United States[11]. Respondents with fathers born in Russia, centrally planned economies, or in countries with less than 15 respondents are excluded[12].

Country-level summary statistics reported in FF's Table 1 provide the opportunity to assess the accuracy of my sample selection procedures. The original and verification samples are nearly identical: while the original regression sample contains 6,774 observations, the verification sample contains 6,768 observations (Table A.8). The 6 missing observations are from Italy (4), Germany (1), and the Philippines (1). I was unable to find the reason for these missing observations.

2.2 Analysis Variables

Outcome variables FF's Table 2 reports coefficients from estimating Equation 1 on two outcomes: the number of hours worked in the previous week and the number of children ever born to a woman. While the number of children is precisely reported in the original census data for up to 12 children, hours worked are reported in intervals. FF computes a measure of time worked by assigning the midpoint of each of these intervals (Table A.9)[13].

A comparison with country-level summary statistics reported in FF's Table 1 reveals that both outcome variables in the original and verification samples display identical means by country of origin, except for countries for which observations are missing compared to the original sample—though differences are marginal (Table A.11). Turning to individual-level summary statistics reported in FF's Table A.1 similarly shows little differences across the original and verification samples (Table A.12).

[9]These tables are reproduced from the original article in Tables A.4 and A.5.

[10]The specification of the data extract from IPUMS USA is detailed in Table A.6.

[11]Sample selection procedures are thoroughly described in FF's pages 152–154 and footnotes 18, 19, 22, and 23.

[12]The specific sample selection procedures applied to generate the regression sample of FF's Table 2 are detailed in Table A.7.

[13]The highest category for the hours worked variable is `60+ hours`. For this category, I assign the value `66`, as it is the maximum value reported in FF's Table A1. I also construct a measure of weeks worked last year in the same way to make comparisons of summary statistics across the original and verification samples.

Control variables Regressions in FF's Table 2 control for a set of individual characteristics: all specifications include respondents' age and age squared, some include their educational attainment, and the "full specification" further controls for their husbands' age, educational attainment, and total income. While FF applies no transformation to census data for respondents' age and their husbands' total income besides expressing it in tens of thousands of dollars, four indicator variables are constructed to measure educational attainment: below high school (omitted from regressions), high school degree, some college, and at least a college degree (Table A.10). The same transformations are applied to capture husbands' educational attainment. Moreover, FF creates ten age-range indicators to control for husbands' age. FF provides no indication regarding the size of these ranges, so I create ten equally-sized bins of 8.6 years from 14 to 100—the minimum and maximum of husbands' ages in the data.

As with outcome variables, comparing individual-level summary statistics of control variables across the original and verification samples reveals little differences (Table A.12).

2.3 Cultural Proxy Variables

To proxy for culture, FF uses past aggregate outcomes of respondents' countries of ancestry. Because the census does not provide the country of birth of respondents' mothers when both their parents were born outside the United States, FF uses their fathers' birthplace to assign country-of-ancestry culture.

Female labor force participation in 1950 To capture the cultural determinants of women's working behavior, FF uses country-of-ancestry FLFP in 1950 from the International Labour Organization (ILO). In particular, notes below FF's Table 1 specify that this variable is from "ILO, Economically Active Population, 1950–2010, (Geneva, 1997)" along with the following bibliographical reference: "International Labour Office. 1988. Current International Recommendations on Labour Statistics. Geneva: International Labour Organization."

These references are not entirely accurate. Going back to the original source, FF's data for FLFP in 1950 are from Table 4 of ILO's *Economically Active Population, 1950–2010, Vol. I, Asia* (1996, p. 39–203), *Vol. III, Latin America and the Caribbean* (1997a, p. 27–131), and *Vol. IV, Northern America - Europe - Oceania* (1997b, p. 41–211), entitled "Population and Economically Active Population by Sex and Age Group, 1950–2010." Moreover, while FF claims to be using "the rate of the economically active population for women over 10 years of age," a close comparison of the original ILO data to those in FF's Table 1 reveals that FF is actually using FLFP rates calculated relative to the total female population and not relative to the population of women over 10 years old.

Comparing both approaches reveals important differences (Table A.13). On average, FLFP rates relative to the total female population are 6 percentage points lower than those relative to the population of women over 10 years old, with differences ranging from 2 to 19 percentage points. FLFP rates relative to the population of women over 10 years old further exhibits more dispersion as its standard deviation across countries is 15 percentage points, while it is limited to 12 percentage points for FLFP rates relative to the total female population.

In the verification and reproduction tests, I consider the FLFP rate relative to the population of women over 10 years old as the appropriate measure since it is the measure claimed to be used by FF. I will however show how this inaccuracy affects the results when using FF's original FLFP values as per FF's Table 1.

Total fertility rate in 1950 To capture the cultural determinants of women's fertility, FF uses country-of-ancestry TFR in 1950 from the United Nations (UN). In particular, notes below FF's Table 1 specify that this variable is from the "United Nations *Demographic Yearbook* 1997, Historical supplement table 4" with no further indication in the bibliography.

The appropriate reference is UN's *Demographic Yearbook* 1997, Historical Supplement (1999), Table 4 (entitled "Selected Derived Measures of Natality: 1948–1997"), column *Total Fertility Rate*. However, it is unclear which years FF selected, as TFR data for 1950 is not available for all countries of ancestry present in the regression sample—the closest year for which TFR data is available across all countries is 1953[14].

As a result, the 1950 TFR values (or the closest year to 1950) only corresponds to the values in FF's Table 1 for 6 out of 25 countries. Because the 1953 TFR values are available for all countries in the regression sample, I use these values in the verification attempt. This also ensures that the measure is defined for the same year. Reassuringly, absolute differences between the 1953 TFR values and FF's original values remain moderate, as their means are 3.48 for the former and 3.66 for the latter, with a country-wise average absolute difference of 0.20 (Table A.14). The difference is nonetheless substantial for the Philippines, since FF reports a value of 7.29 while the original value corresponds to 3.14. I will show how these differences affect the results when using FF's original TFR values as per FF's Table 1.

2.4 Alternative Samples from the 1970 US Census

Two alternative 1-in-100 samples were drawn from the same 15 percent sample of the population as the 1 percent 1970 Form 2 Metro Sample used in FF's Table 2: the 1 percent 1970 Form 2 State Sample and the 1 percent 1970 Form 2 Neighborhood Sample. The selection process of these samples was such that they are mutually exclusive and are representative of the same underlying population (Bureau of the Census, 1972, p. 197–198). They can therefore be used independently as well as combined for a reproduction test.

Sample selection and variables transformation procedures applied to these samples are identical to those applied to the Metro sample. The inspection of country and individual-level summary statistics reveals that characteristics of observations across these regression samples are nearly identical (Table A.15). Comparing the distributions of the main variables of the analysis along their CDFs directly similarly confirms that they are not different (Figures B.2 and B.3)[15].

[14] It also remains unclear how the values for Germany are computed, since it was then split between the FRG and the GDR. I take the average TFR value of both countries in 1955, since this is the first year for which TFR data is available for the GDR.

[15] More specifically, I compare the distribution of the treatment (FLFP in 1950 and TFR in 1953) and outcome (hours worked and number of children) variables of interest pair-wise across the Metro, State, and Neighborhood samples based on Goldman and Kaplan's (2018) procedure and implemented through Kaplan's (2019) `distcomp` Stata command. These

3 Verification

Baseline results In this section, I verify the results reported in FF's Table 2 by estimating Equation 1 on the sample of second-generation immigrant women to the United States described in Section 2. The baseline verification test uses as cultural proxy variables the FLFP rates of women over 10 years old in 1950 and the TFR in 1953. I report the original FF estimates of interest $\widehat{\beta}_2$ in Panel A of Table 1 along with the verification estimates in Panel B. In Panel C, I report estimates when using the cultural proxy values reported in FF's Table 1[16].

Verification estimates are relatively close to those reported in FF's Table 2. In FF's preferred specification (Columns 2 and 5), the verification coefficient on FLFP is 0.059 (s.e. of 0.014) relative to an original coefficient of 0.072 (s.e. of 0.015). The verification coefficient on TFR is 0.228 (s.e. of 0.040) relative to an original coefficient of 0.219 (s.e. of 0.041). Verification coefficients that use FF cultural proxy variables in Panel C are nearly identical to those in FF's Table 2. This suggests that the discrepancies between the verification and original FF estimates are entirely driven by cultural proxy variables. Overall, the verification of FF's Table 2 is successful.

Robustness FF claims that these results are robust to changes in sample criteria and alternative variables as cultural proxies, though the article does not provide statistical output to support this assertion. In particular, FF claims that these results are robust to including all women independently of their marital status, including Russia or independently excluding China, Mexico, and Italy, and changing the sample to women that aged 40–50. FF further claims that these results are robust to using the following alternative cultural proxy variables: the percentage of the workforce that is female in 1960, the labor force participation of women aged 30–34 in 1950, and 1960 FLFP and TFR values.

To assess the robustness of results in FF's Table 2, I run FF's preferred specification for both outcomes of interest using the above alternative sample selection criteria and cultural proxy variables. Results are reported in Tables 2 and 3. Verification estimates are robust to all alternative cultural proxies and sample selection criteria, except when restricting the sample to women aged 40–50 for the hours worked outcome, and when excluding Mexico for the children outcome. Nevertheless, verification estimates can be considered as generally robust.

4 Reproduction

In this section, I reproduce the results reported in FF's Table 2 across alternative samples drawn from the same underlying population: the State and the Neighborhood samples of the 1970 US census. Because the Metro, State, and Neighborhood samples are mutually exclusive, I further combine them to create a Pooled sample of the 1970 US census, representing a 3-in-100 sample.

These samples differ along one dimension: while the smallest identifiable geographic units in the Metro sample are SMSAs, those in the State sample are states and those in the Neighborhood

tests all fail to reject the null that these distributions are different at the 1, 5, or 10 percent levels.

[16] Tables A.16 and A.17 reproduce the entire FF's Table 2 corresponding to Panels B and C of Table 1, respectively.

Table 1: Verification of FF Table 2.

Dependent variable	Hours worked			Children		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Original FF estimates						
Female	0.047***	0.072***	0.053***			−0.010
LFP 1950	(0.012)	(0.015)	(0.016)			(0.008)
TFR 1950			−0.225**	0.250***	0.219***	0.194***
			(0.103)	(0.056)	(0.041)	(0.051)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	6,774	6,774	6,774	6,774	6,774	6,774
B. Verification estimates (verification cultural proxies)						
Female	0.040***	0.059***	0.043***			−0.009
LFP 1950	(0.010)	(0.014)	(0.013)			(0.006)
TFR 1953			−0.291**	0.266***	0.228***	0.205***
			(0.105)	(0.051)	(0.040)	(0.047)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	6,768	6,768	6,768	6,768	6,768	6,768
C. Verification estimates (FF cultural proxies)						
Female	0.047***	0.072***	0.052***			−0.010
LFP 1950	(0.012)	(0.015)	(0.015)			(0.008)
TFR 1950			−0.238**	0.250***	0.215***	0.190***
			(0.105)	(0.056)	(0.041)	(0.050)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	6,768	6,768	6,768	6,768	6,768	6,768

Notes: This table reproduces the estimates of interest of FF Table 2. Panel A reports the original FF estimates, Panel B, verification estimates when using verification cultural proxies, and Panel C, verification estimates when using FF cultural proxies according to FF Table 1. All specifications include respondents' age and age squared, their husbands' age-range indicators, and SMSA fixed effects. *Controls* include education indicators for both respondents and their husbands. Robust standard errors in parentheses account for clustering at country level.
*** Significant at the 1 percent level. ** Significant at the 5 percent level.

sample are census regions. Still, the Metro sample contains state information and all three samples contain census-region information. Therefore, the reproduction proceeds by estimating Equation 1

Table 2: Robustness of Verification Estimates to Alternative Sample Restrictions.

Sample	Baseline	All marital statuses	Include Russia	Exclude China	Exclude Mexico	Exclude Italy	Aged 40–50
	A. Dependent variable is hours worked						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female LFP 1950	0.059***	0.046***	0.057***	0.068***	0.046***	0.057***	0.028
	(0.014)	(0.012)	(0.010)	(0.013)	(0.015)	(0.016)	(0.029)
Observations	6,768	8,280	7,559	6,715	5,929	4,863	10,732
	B. Dependent variable is children						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female TFR 1953	0.228***	0.210***	0.236***	0.236***	0.129	0.195***	0.291***
	(0.040)	(0.036)	(0.038)	(0.040)	(0.091)	(0.033)	(0.070)
Observations	6,768	8,280	7,559	6,715	5,929	4,863	10,732

Notes: This table reproduces estimates from the full specification of FF Table 2 across alternative sample restrictions: the baseline sample in Column (1), including women of all marital statuses (together with marital status fixed effects but without husband controls) in Column (2), including Russia in Column (3), excluding China in Column (4), Mexico in Column (5), and Italy in Column (6), and on the sample of women aged 40–50 in Column (7). All specifications include respondents' age and age squared, their husbands' age-range indicators, SMSA fixed effects, and education indicators for both respondents and their husbands. Robust standard errors in parentheses account for clustering at country level.
*** Significant at the 1 percent level.

with alternative residence fixed effects, depending on their availability in the sample. A proper comparison of estimates obtained on all three samples can hence only be operated across specifications that use census-region fixed effects.

I report results in Table 4. First, reproduction estimates on the FF (Metro) sample using census-region instead of SMSA fixed effects renders coefficients stronger for the hours worked outcome and does not affect much those for the children outcome (Column 1). This suggests that a specification using census-region fixed effects can be reasonably used as a benchmark. When using the same specification with census-region fixed effects on the State and the Neighborhood samples (Columns 4 and 6), estimates hold for the children outcome—although they are slightly weaker—but are insignificant and closer to zero for the hours worked outcome. Pooling all three samples generates reproduction estimates that constitute a weighted average of sample-specific estimates (Column 7). Results are similar when the original FF cultural proxy variables are used instead (Table A.18).

Table 3: Robustness of Verification Estimates to Alternative Cultural Proxy Variables.

Proxy	Baseline	FF proxies	FLFP or TFR 1960	Share female 1960	FLFP 1950 30–34
	A. Dependent variable is hours worked				
	(1)	(2)	(3)	(4)	(5)
Female LFP	0.059***	0.072***	0.064***	0.088***	0.045***
	(0.014)	(0.015)	(0.015)	(0.017)	(0.012)
Observations	6,768	6,768	6,768	6,768	6,768
	B. Dependent variable is children				
	(1)	(2)	(3)		
Female TFR	0.228***	0.215***	0.262***		
	(0.040)	(0.041)	(0.045)		
Observations	6,768	6,768	6,768		

Notes: This table reproduces estimates from the full specification of FF Table 2 using alternative cultural proxies: the baseline verification proxies in Column (1), the proxies from FF Table 1 in Column (2), proxies evaluated in 1960 in Column (3), the percentage of the workforce that is female in 1960 in Column (4), and the labor force participation of women aged 30–34 in 1950 in Column (5). All specifications include respondents' age and age squared, their husbands' age-range indicators, SMSA fixed effects, and education indicators for both respondents and their husbands. Robust standard errors in parentheses account for clustering at country level.

*** Significant at the 1 percent level.

To rationalize this unsuccessful reproduction on the hours worked outcome, I first inspect whether it can be explained by a different composition of the effective sample relative to respondents' countries of ancestry. In particular, I compute for each sample the share of the total residual variance by country of ancestry (Table A.19). I find little differences across samples: in all three samples, respondents from Mexico, China, Japan, and Germany are the bigger contributors to building the estimate of interest (Aronow and Samii, 2016). Then, I construct residual variance plots of residual hours worked against residual FLFP in 1950 across all samples based on the specification of Columns 1, 4, 6, and 7 of Table 4 (Figure B.4). Again, I find no clear outlier across samples. These analyses suggest that the underlying composition of the different reproduction samples cannot explain the discrepancy found in the resulting estimates.

Next, to explore whether this failure to reproduce FF's estimates for the hours worked outcome is due to "unprobable" draws of the State and Neighborhood samples, I combine the three census samples and draw 1,000 different random samples that each represent 1-in-100 samples from the

Table 4: Reproduction of FF Table 2 Across Census Samples.

	A. Dependent variable is hours worked						
1970 1% Form 2 Sample	Metro			State		Neighb.	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female LFP 1950	0.076***	0.052***	0.059***	0.031	0.019	0.027	0.045**
	(0.020)	(0.015)	(0.014)	(0.026)	(0.018)	(0.025)	(0.019)
Residence FE	Region	State	SMSA	Region	State	Region	Region
Observations	6,768	6,768	6,768	6,694	6,694	6,804	20,266

	B. Dependent variable is children						
1970 1% Form 2 Sample	Metro			State		Neighb.	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
TFR 1953	0.217***	0.218***	0.228***	0.158***	0.174***	0.165***	0.181***
	(0.040)	(0.040)	(0.040)	(0.043)	(0.043)	(0.035)	(0.038)
Residence FE	Region	State	SMSA	Region	State	Region	Region
Observations	6,768	6,768	6,768	6,694	6,694	6,804	20,266

Notes: This table reproduces estimates from the full sepcifications of FF Table 2 across census extracts: the 1970 1% Form 2 Metro sample in Columns (1)–(3), the 1970 1% Form 2 State sample in Columns (4)–(5), the 1970 1% Form 2 Neighbordhood sample in Column (6), and the combination of all three extracts in Column (7). All specifications include respondents' age and age squared, their husbands' age-range indicators, and education indicators for both respondents and their husbands. Robust standard errors in parentheses account for clustering at country level.

*** Significant at the 1 percent level. ** Significant at the 5 percent level.

1970 US census (6,700–800 observations). In a bootstrapping approach, I then run Equation 1 with census-region instead of SMSA fixed effects on each of these random samples for the hours worked outcome. I plot the resulting coefficients on the FLFP variable in Figure 1 and report summary statistics in Table A.20. They suggest that the original FF Metro sample is rather unusual compared to the State and Neighborhood samples. Indeed, estimates based on the Metro sample are in the 95th percentile of the distribution of estimates and only 53 percent of estimates are significant the 10 percent level. Results are similar when using the original FF proxy variable instead (Figure B.5)[17].

[17]The same conclusion applies when using the children outcome, although the magnitude of resulting coefficients is large enough that they are all different from zero at conventional significance levels (Figure B.6).

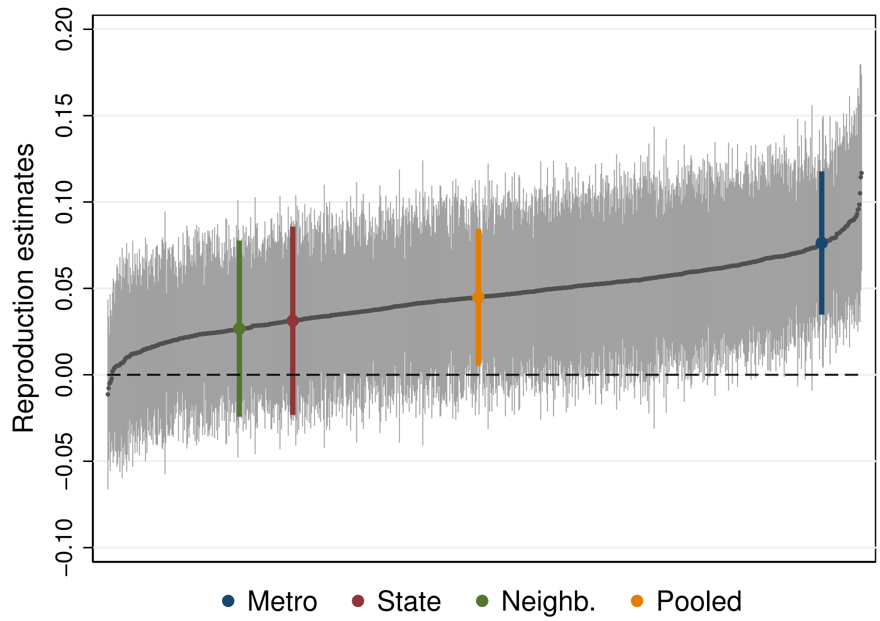

Figure 1: Reproduction Estimates of the FLFP Variable on 1,000 Random Samples. This figure plots coefficients on the FLFP verification variable from estimating Equation 1 on the hours worked outcome with census-region instead of SMSA fixed along with 95 percent confidence intervals on 1,000 different random samples representing 1-in-100 samples from the 1970 US census, from a sample pooling the Metro, State, and Neighborhood samples of the 1970 US census. It also highlights coefficients obtained when using the original Metro, State, Neighborhood, or pooled samples.

5 Conclusion

In this article, I perform a replication Fernández and Fogli's (2009) main results. While I am able to verify Fernández and Fogli's (2009) estimates and their robustness, results relative to the hours worked outcome cannot be reproduced in alternative samples drawn from the same underlying population. Extensions to other samples and meta-analytic approaches are therefore advisable to assess the validity and generalizability of the findings in this seminal study.

References

Albanesi, S., & Olivetti, C. (2016). "Gender Roles and Medical Progress." *Journal of Political Economy*, 124(3), 650–695. DOI: 10.1086/686035.

Antecol, H. (2000). "An Examination of Cross-Country Differences in the Gender Gap in Labor Force Participation Rates." *Labour Economics*, 7(4), 409–426. DOI: 10.1016/S0927-5371(00)00007-5.

Aronow, P. M., & Samii, C. (2016). "Does Regression Produce Representative Estimates of Causal Effects?" *American Journal of Political Science*, 60(1), 250–267. DOI: 10.1111/ajps.12185.

Bailey, M. J. (2006). "More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life Cycle Labor Supply." *The Quarterly Journal of Economics*, 121(1), 289–320. DOI: 10.1093/qje/121.1.289.

Beblo, M., Görges, L., & Markowsky, E. (2020a). "Gender Equality as a Confounder in the Epidemiological Approach." *Economics Bulletin*, 40(2), 1292–1299. PDF: http://www.accessecon.com/Pubs/EB/2020/Volume40/EB-20-V40-I2-P110.pdf.

Beblo, M., Görges, L., & Markowsky, E. (2020b). "Gender Matters in Language and Economic Behaviour: Can we Measure a Causal Cognition Effect of Speaking?" *Labour Economics*, 65, 101850. DOI: 10.1016/j.labeco.2020.101850.

Blau, F. D. (1992). "The Fertility of Immigrant Women: Evidence from High-Fertility Source Countries." In: George J. Borjas and Richard B. Freeman (Eds.), *Immigration and the Work Force: Economic Consequences for the United States and Source Areas*, 93–134. University of Chicago Press. URL: http://www.nber.org/chapters/c6907..

Bureau of the Census. U.S. Department of Commerce (1972). "Public Use Samples of Basic Records from the 1970 Census: Description and Technical Documentation." Washington, DC: US Government Printing Office. PDF: https://usa.ipums.org/usa/resources/codebooks/1970_PUMS_codebook.pdf.

Carroll, C. D., Rhee, B.-K., & Rhee, C. (1994). "Are There Cultural Effects on Saving? Some Cross-Sectional Evidence." *The Quarterly Journal of Economics*, 109(3), 685–699. DOI: 10.2307/2118418.

Clemens, M. A. (2017). "The Meaning of Failed Replications: A Review and Proposal." *Journal of Economic Surveys*, 31(1), 326–342. DOI: 10.1111/joes.12139.

Fernández, R. (2007). "Women, Work, and Culture." *Journal of the European Economic Association*, 5(2–3), 305–332. DOI: 10.1162/jeea.2007.5.2-3.305.

Fernández, R. (2011). "Does Culture Matter?" In: Jess Benhabib, Alberto Bisin and Matthew O. Jackson (Eds.), *Handbook of Social Economics*, 481–510. North-Holland. DOI: 10.1016/B978-

0-444-53187-2.00011-5.

Fernández, R., & Fogli, A. (2005). "Culture: An Empirical Investigation of Beliefs, Work, and Fertility." National Bureau of Economic Research Working Paper, 11268. URL: https://www.nber.org/papers/w11268.

Fernández, R., & Fogli, A. (2009). "Culture: An Empirical Investigation of Beliefs, Work, and Fertility." *American Economic Journal: Macroeconomics*, 1(1), 146–177. DOI: 10.1257/mac.1.1.146.

Fortin, N. M. (2005). "Gender Role Attitudes and the Labour-market Outcomes of Women across OECD Countries." *Oxford Review of Economic Policy*, 21(3), 416–438. DOI: 10.1093/oxrep/gri024.

Gaddis, I., & Klasen, S. (2014). "Economic Development, Structural Change, and Women's Labor Force Participation: A Reexamination of the Feminization U Hypothesis." *Journal of Population Economics*, 27, 639–681. DOI: 10.1007/s00148-013-0488-2.

Giuliano, P. (2021). "Gender and Culture." *Oxford Review of Economic Policy*, 36(4), 944–961. DOI: 10.1093/oxrep/graa044.

Goldin, C. (1990). "Understanding the Gender Gap: An Economic History of American Women." Oxford University Press.

Goldin, C., & Katz, L. F. (2002). "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions." *Journal of Political Economy*, 110(4), 730–770. DOI: 10.1086/340778.

Goldman, M., & Kaplan, D. M. (2018). "Comparing Distributions by Multiple Testing Across Quantiles or CDF Values." *Journal of Econometrics*, 206(1), 143–166. DOI: 10.1016/j.jeconom.2018.04.003.

Greenwood, J., Seshadri, A., & Yorukoglu, M. (2005). "Engines of Liberation." *The Review of Economic Studies*, 72(1), 109–133. DOI: 10.1111/0034-6527.00326.

Guiso, L., Sapienza, P., & Zingales, L. (2006). "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives*, 20(2), 23–48. DOI: 10.1257/jep.20.2.23.

International Labour Office, Bureau of Statistics (1996). "Economically Active Population, 1950–2010, Vol. I, Asia." STAT working paper, International Labour Organization. PDF: https://www.ilo.org/public/libdoc/ilo/1996/96B09_462_e_f_s.pdf.

International Labour Office, Bureau of Statistics (1997a). "Economically Active Population, 1950–2010, Vol. III, Latin America and the Caribbean." STAT working paper, International Labour Organization. PDF: https://www.ilo.org/public/libdoc/ilo/1997/97B09_299_e_f_s.pdf.

International Labour Office, Bureau of Statistics (1997b). "Economically Active Population, 1950–2010, Vol. IV, Northern America - Europe - Oceania." STAT working paper, International Labour Organization. PDF: https://www.ilo.org/public/libdoc/ilo/1997/97B09_300_e_f_s.pdf.

Kaplan, D. M. (2019). "distcomp: Comparing Distributions." *The Stata Journal*, 19(4), 832–848. DOI: 10.1177/1536867X19893626.

Klasen, S. (2019). "What Explains Uneven Female Labor Force Participation Levels and Trends in Developing Countries?" *The World Bank Research Observer*, 34(2), 161–197. DOI: 10.1093/wbro/lkz005.

Ngai, R. L., & Petrongolo, B. (2017). "Gender Gaps and the Rise of the Service Economy." *American Economic Journal: Macroeconommics*, 9(4), 1–44. DOI: 10.1257/mac.20150253.

Olivetti, C., & Petrongolo, B. (2016). "The Evolution of Gender Gaps in Industrialized Countries." *Annual Review of Economics*, 8(1), 405–434. DOI: 10.1146/annurev-economics-080614-115329.

Ruggles, S., Flood, S., Foster, S., Goeken, R., Pacas, J., Schouweiler, M., & Sobek, M. (2021). "IPUMS USA: VERSION 11.0" [dataset]. Minneapolis, MN: IPUMS. DOI: 10.18128/D010.V11.0.

United Nations, Department of Economic and Social Affairs (1999). "1997 Demographic Yearbook, Historical Supplement." United Nations. URL: https://unstats.un.org/unsd/demographic/products/dyb/dybhist.htm.