BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Gender complementarities in the labor market

by Giacomo De Giorgi, Marco Paccagnella and Michele Pellizzari

June 2013

Number 183

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional papers)

Gender complementarities in the labor market

by Giacomo De Giorgi, Marco Paccagnella and Michele Pellizzari

Number 183 – June 2013

The series Occasional Papers *presents studies and documents on issues pertaining to the institutional tasks of the Bank of Italy and the Eurosystem. The* Occasional Papers *appear alongside the* Working Papers *series which are specifically aimed at providing original contributions to economic research.*

The Occasional Papers *include studies conducted within the Bank of Italy, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

The series is available online at www.bancaditalia.it .

ISSN 1972-6627 (print)
ISSN 1972-6643 (online)

Printed by the Printing and Publishing Division of the Bank of Italy

GENDER COMPLEMENTARITIES IN THE LABOR MARKET

by Giacomo De Giorgi*, Marco Paccagnella† and Michele Pellizzari‡

Abstract

In this paper we estimate the short-run elasticity of substitution between male and female workers, using data on employment and wages from Italian provinces from 1993-2006. We adopt a production function approach similar to that of Card and Lemieux (2001a) and Acemoglu, Autor and Lyle (2004). Our identification strategy relies on a natural experiment. In 2000 the Italian parliament passed a law to abolish compulsory military service; the reform was implemented through a gradual reduction in the number of draftees, and compulsory drafting was definitively terminated in 2004. We use data on the (planned) maximum number of draftees at the national level (as stated in the annual budgetary law), interacted with sex-ratios at birth at the provincial level, as instruments for (relative) female labor supply. Our results suggest that young males and females are imperfect substitutes, with an elasticity of substitution ranging between 1.1 and 1.6. Our results have implications for the evaluation of policies aimed at increasing female labor market participation, suggesting that they do not necessarily displace male employment.

JEL Classification: J21, J31, J16.
Keywords: gender complementarities, elasticity of substitution, employment, wages.

Contents

---

* Stanford University, BREAD, CEPR and NBER

† Bank of Italy and fRDB

‡ OECD, University of Geneva and IZA

1 Introduction*

Italy has amongst the lowest levels of female participation in the labor market across European countries. In 2010, women's employment rate was as low as 46.1%, 21.6 percentage points below the corresponding rate for men, with ample differences within the country (from 30.5% in the South to 54.8% in the Centre-North; Banca d'Italia, 2011). Gender wage gaps are, on the other hand, lower than in other OECD countries, at around 6%; after controlling for selection, Olivetti and Petrongolo (2008) estimate median wage gaps to be between 10% and 20% during the 1994-2001 period (still below the US, where the estimated gap is above 30%); more recently, Zizza (2012) estimated gender wage gaps at around 13% in 2008. There is ample consensus, both among policymakers and academics, that female participation is ”too low” and that effort must be exerted to increase employment rates.[1] Many different solutions have been proposed to achieve such goal, from gender-based taxation (Alesina, Ichino, and Karabarbounis, 2011) to increased provision of childcare facilities (Del Boca, Locatelli, and Vuri, 2005; Del Boca and Vuri, 2007; Arpino, Pronzato, and Tavares, 2010; Brilli, Del Boca, and Pronzato, 2011) and the extension of parental leaves (Pronzato, 2009; Kluve and Tamm, forthcoming), from gender quotas (Beaman, Chattopadhyay, Duflo, Pande, and Topalova, 2009; De Paola, Scoppa, and Lombardo, 2010) to targeted training programs. These

*A previous version of this paper was presented at the conference “Women and the Italian Economy”, organized by the Bank of Italy, held in Rome on March 7^{th}, 2012. The views expressed therein are those of the authors and do not necessarily reflect those of the Bank of Italy or the OECD. We would like to thank, without implications, Anna Laura Mancini, Emiliano Rustichelli and Ugo Troiano for their comments, as well as seminar participants at Bocconi University. Davide Crapis provided able research assistance. Part of this project was completed while Giacomo De Giorgi and Marco Paccagnella were visiting UC Berkeley, which they wish to thank for the hospitality.

[1] In 2001 the Lisbon Agenda set the explicit goal of a 60% female employment rate to be reached by 2010. More recently, the new Europe 2020 Strategy redefined this goal, aiming to an overall employment rate of 75%. Obviously, the achievement of this new target will require a large increase in female participation

policies are often justified both on equity (the current allocation being deemed ”unfair”) and on efficiency grounds; according to this last argument, either gender diversity in the workplace would be productivity-enhancing, and/or the specific skills possesed by women are sub-optimally represented in the economy.

There is actually ample evidence (often experimental) that men and women posses different skills (which may be complementary). For instance, men were documented to perform better in competitive environments (Gneezy, Niederle, and Rustichini, 2003; Gneezy and Rustichini, 2004; Paserman, 2010), while women appear to be less able to foster cooperation among men (Gagliarducci and Paserman, 2012), more altruistic (Eckel and Grossman, 1998; Andreoni and Vesterlund, 2001) and more risk-averse (Borghans, Golsteyn, Heckman, and Meijers, 2009). In a recent article surveying a number of experimental studies, Croson and Gneezy (2009) identify robust differences in risk preferences, social preferences and competitive preferences.[2] An obvious question to ask is thus whether men and women are actually complements or substitutes in the production process, and how large the elasticity of substitution between these two types of ”labor inputs” is.

Starting with the seminal contribution by Katz and Murphy (1992), the literature has mainly focused on estimating the elastiticity of substitution between skilled and unskilled workers (Angrist, 1995; Card and Lemieux, 2001a; Leuven, Oosterbeek, and van Ophem, 2004; Ciccone and Peri, 2005; Caselli and Coleman, 2006) or between natives and immigrants workers (Card, 2001; Borjas, 2003; Ottaviano and Peri, 2006; Borjas, Grogger, and Hanson, 2008; Borjas, 2009; D'Amuri, Ottaviano, and Peri, 2010; Barone and Mocetti, 2011; Ottaviano and Peri, 2012).

To the best of our knowledge, Acemoglu, Autor, and Lyle (2004) is the only recent

[2] Although Abrevaya and Hamermesh (2012) find no significant differences in actual behavior.

paper that provide estimates of the long-run elasticity of substitution between male and female workers, exploiting the US World War II mobilization as a ”natural experiment” that drew a large number of women into the labor market. In their paper, cross-state variation in mobilization rates is used as an instrument for (relative) female labor supply. One drawback of this approach is that *actual* mobilization rates are used, which may be correlated with local labor market conditions; this cast some doubts on the validity of the instrument. We take into account the possible endogeneity of (relative) female labor supply by instrumenting it with sex ratios at birth and by exploiting a similar natural experiment, i.e. the abolition of compulsory military service (passed by the Italian Parliament in 2000). The abolition was not sudden, and was implemented through a gradual reduction in the yearly intake of draftees. From the yearly budgetary law we derive information on the maximum number of draftees the military *planned* to enroll in each year (at the national level). These two variables (interacted with each other) should provide a valid instrument, to the extent that they both generate exogenous variation in the relative supply of female workers at the year-province level. Although coefficients estimates are often characterized by large standard errors, and the instrument does not appear to be particularly strong, we do find converging evidence of imperfect substitutability between young men and women (those mainly affected by the reform), with an implied elasticity of substitution ranging between 1.1 and 1.6, depending on the specification.

The rest of the paper is organized as follows. In section 2 we present the theoretical framework at the basis of our empirical application. In section 3 we present the data used in this study and we discuss the validity of our identification strategy. In section 4 we present our empirical results. Section 5 concludes.

2 Theoretical framework

The starting point of our analysis is a standard Cobb-Douglas production function, featuring constant returns to scale and heterogeneus labor inputs:

$$Y_{jt} = A_{jt} K_{jt}^{\alpha} L_{jt}^{1-\alpha} \tag{1}$$

where labor input L_{jt} is a constant elasticity of substitution (CES) aggregate of male and female workers (Acemoglu, Autor, and Lyle, 2004):

$$L_{jt} = [(B(f)_{jt} F_{jt})^{\rho} + (B(m)_{jt} M_{jt})^{\rho}]^{\frac{1}{\rho}} \tag{2}$$

In equations 1 and 2, Y_{jt} and K_{jt} are, respectively, total output and capital stock in province j at time t, M_{jt} and F_{jt} represent the supply of, respectively, male and female workers, A is a neutral productivity term and the Bs are productivity parameters allowed to be gender-dependent and normalized so that they sum to one. The elasticity of substitution between male and female workers is defined as the percentage change in the relative demand for female (male) workers following a percentage change in the relative price of male (female) workers, and can be expressed as

$$\sigma \equiv \frac{1}{1-\rho}, \qquad \rho \in (-\infty, 1)$$

Male and female workers are gross substitutes if $\sigma > 1$ ($\rho > 0$), and gross complements if $\sigma < 1$ ($\rho < 0$).

Now let assume a competitive labor market, where inputs are paid their marginal product. The first-order conditions for the firm's problem equate the derivatives of (1) with respect to male and female labor to male and female wages, respectively:

$$\frac{\partial Y_{jt}}{\partial F_{jt}} = w_{jt}^{F} = (1-\alpha)\, K_{jt}^{\alpha} A_{jt} B(f)_{jt}^{\rho} F_{jt}^{(\rho-1)} L_{jt}^{\frac{1-\alpha-\rho}{\rho}} \tag{3}$$

$$\frac{\partial Y_{jt}}{\partial M_{jt}} = w_{jt}^{M} = (1-\alpha)\, K_{jt}^{\alpha} A_{jt} B(m)_{jt}^{\rho} M_{jt}^{(\rho-1)} L_{jt}^{\frac{1-\alpha-\rho}{\rho}} \tag{4}$$

Dividing equation 3 by equation 4 and taking logarithms we obtain:

$$\ln\left(\frac{w_{jt}^{F}}{w_{jt}^{M}}\right) = \rho \ln\left(\frac{B(f)_{jt}}{B(m)_{jt}}\right) + (\rho-1)\ln\left(\frac{F_{jt}}{M_{jt}}\right)$$

which can be more conveniently expressed as:

$$\ln\left(\frac{w_{jt}^{F}}{w_{jt}^{M}}\right) = \frac{\sigma-1}{\sigma} \ln\left(\frac{B(f)_{jt}}{B(m)_{jt}}\right) - \frac{1}{\sigma}\ln\left(\frac{F_{jt}}{M_{jt}}\right) \tag{5}$$

Equation 5 can be directly estimated using data on average male and female wages and average male and female labor supply by province-year. Year and province fixed effects would control for unobserved productivity differences $B(f)_{jt}/B(m)_{jt}$[3].

One of the reasons behind the popularity of the CES framework is its extreme flexibility. Workers can differ in many different dimensions other than gender, and the model can be easily adapted to accomodate such further complications. For the purpose of the present study, a relevant dimension to look at is workers' age, since our identification strategy will rely on the reform of compulsory military service, that arguably affected only young workers.[4] Following Card and Lemieux (2001a) and Ottaviano and Peri

[3]We will return to this specification in section 4

[4]Before the reform, young men were called to serve in the army when they turned 18. Enrollement could be postponed for a few years for various reasons (the most common of which was enrollement in higher education). We will return to these issues in section 3.

(2012), we allow for a nested CES production function, in which the two groups of male and female workers are nested into age groups[5]. This model assumes a constant elasticity of substitution between any pair of age groups, allowing for a different parameter characterizing substitutability across gender. Now, omitting time and province subscripts for notational convenience, the labor aggregate L can be redefined as

$$L = \left[\sum_{x=1}^{X} \theta_x L_x^{\rho_X} \right]^{\frac{1}{\rho_X}} \tag{6}$$

where L_x is the number of workers in age class x, θ_x is the relative productivity parameter of that group and ρ_X is a parameter such that $\rho \equiv 1 - 1/\sigma_X$, where σ_X is the elasticity of substitution between any two age class. Gender differences are then easily nested in this framework by letting

$$L_x = [(B(f)_{jt} F_{jt})^{\rho} + (B(m)_{jt} M_{jt})^{\rho}]^{\frac{1}{\rho}} \tag{7}$$

as in equation 2. Following the same steps taken for the basic non-nested model, it is easy to derive an expression equivalent to equation 5, with a further subscript x indexing age classes.

3 Data and institutional details

The main data source for our study are microdata from the archives of the National Institute for Social Security (INPS). INPS has released an archive containing information on all individuals born on the 10th of March, June, September and December that have opened at least one position at INPS. Data are currently available for the 1985-2004

[5]In our framework, age can be thought of as a proxy of labor market experience.

period. For 2005-06 we use aggregate data (at the year-province level) made available on the INPS website. We collected informations on wages and paid weeks, from which we derive measures of weekly wages and of weeks worked. For part-timers, weeks worked are adjusted to make them ”full-time equivalent”. For each year and each italian province[6] we compute average relative (female/male) wages and relative labor supply, defined as the ratio between the total number of weeks worked by females and the total number of weeks worked by males. In table 1 we present basic descriptive statistics for our main variables.[7]

The raw gender wage gap is around 13% for the sample of all workers; it decreases substantially for younger workers, being null for people in the 15-19 age class. The gap in labor supply is much higher, especially when we look at weeks and hours worked. The gap is smaller for employment levels, probably due to the fact that part-time jobs are relatively more common among women.

3.1 Instrumental variables

Given that (relative) labor supply is likely to be correlated with shifts in relative labor demand, an instrumental-variable strategy is preferable to OLS when estimating equation 5. To construct instruments for relative labor supply we use two sources. The first is a natural experiment, i.e. the abolition of compulsory military service that took place during the 2000’s. Military service in Italy used to be compulsory for all males

[6]There are currently 110 provinces in Italy; however, some of them were created in recent years, which is why in this paper we work with 94 provinces, aggregating up some of the newly establised provinces.

[7]Alternative measures of labor supply could be computed using microdata from the National Statistical Institute (Istat) Labor force surveys (LFS), namely relative employment levels and relative hours worked. However, they would not be perfectly consistent with wage data, so we will stick using weeks as our main labor supply variable. Results using employment levels and hours are presented in the Appendix.

Table 1: Descriptive statistics

	Full sample	15-19 years old	20-24 years old
Avg. w^F/w^M	0.877	1.041	0.971
(between SD)	(0.120)	(0.218)	(0.106)
(within SD)	(0.325)	(0.734)	(0.319)
Avg. F/M - weeks worked	0.514	0.506	0.701
(between SD)	(0.126)	(0.251)	(0.147)
(within SD)	(0.050)	(0.544)	(0.166)
Avg. F/M - employment	0.656	0.640	0.756
(between SD)	(0.150)	(0.212)	(0.195)
(within SD)	(0.073)	(0.517)	(0.202)
Avg. F/M - hours worked	0.540	0.607	0.692
(between SD)	(0.132)	(0.190)	(0.181)
(within SD)	(0.056)	(0.492)	(0.198)

Data on (weekly) wages and weeks worked are from INPS archives; data on employment and hours worked are from the Labor Force Survey.
Between standard deviation across 94 provinces; within standard deviation across 14 years (1992-2006).

as they turned 18 years old.[8] Enrollement could be waived under exceptional circumstances: the most common cases for a waiver were physical deficiencies, or the need to work in order to sustain the family (eg. in the case of orphans). Deferments were more common, and were related to continuing education: individuals enrolled in tertiary education could defer the service, provided they had a regular academic record.

[8]People were also allowed to opt out of military service and serve in civilian organizations such as NGO's.

Once people turned 26, no further postponement was allowed. The length of service used to vary according to the sector of the military (army, navy or air force), but it was gradually reduced until convergence to a period of 12 months for all sectors. In 2000 a reform set the framework for the gradual abolition of compulsory military service and for a transition toward a regime of professional and volunteer soldiers. The end of compulsory service was set to 2007 (meaning that the cohort of individuals born in 1985 would be the first to be totally exempted from service), the length of service was further reduced to 10 months and a gradual reduction in the size of the military was decided. As a consequence, during the transition phase, less and less individuals were called to service. In 2004 a new law anticipated the end of the transition period, and after 2005 nobody was called to serve. In the yearly budgetary law (so called *Legge Finanziaria*) the competent Minister defined the maximum number of graduate individuals that were to be called in the next year.[9] In figure 1 we plot the yearly number of (planned) draftees; the clear downward trend is consistent with the transition toward the new system based on professionals and volunteers.

Our instrument is based on the idea that the variation in figure 1 has generated exogenous variation in (relative) labor supply: as less and less young men were called to serve, they had the possibility to enter the labor market.[10] In figure 2 we plot the evolution of relative labor supply (defined as the log of the ratio of weeks worked by females over weeks worked by males) for three age classes: 15-19 years old, 20-24 years old, and the entire (active) population.

[9] *Graduati di leva aiuto specialisti*. These were people with a high-school degree that allowed them to specialize in some tasks during the service. This is a subset of the total number of individuals called to serve, but should be a good proxy for total enrollement.

[10] The abolition of compulsory military service could have many effects, not only on labor market participation, but also on educational choices (Card and Lemieux, 2001b; Di Pietro, forthcoming; Cipollone and Rosolia, 2007). We abstract from these issues and only look (empirically) at the relationship between the number of people called to serve and relative labor supply.

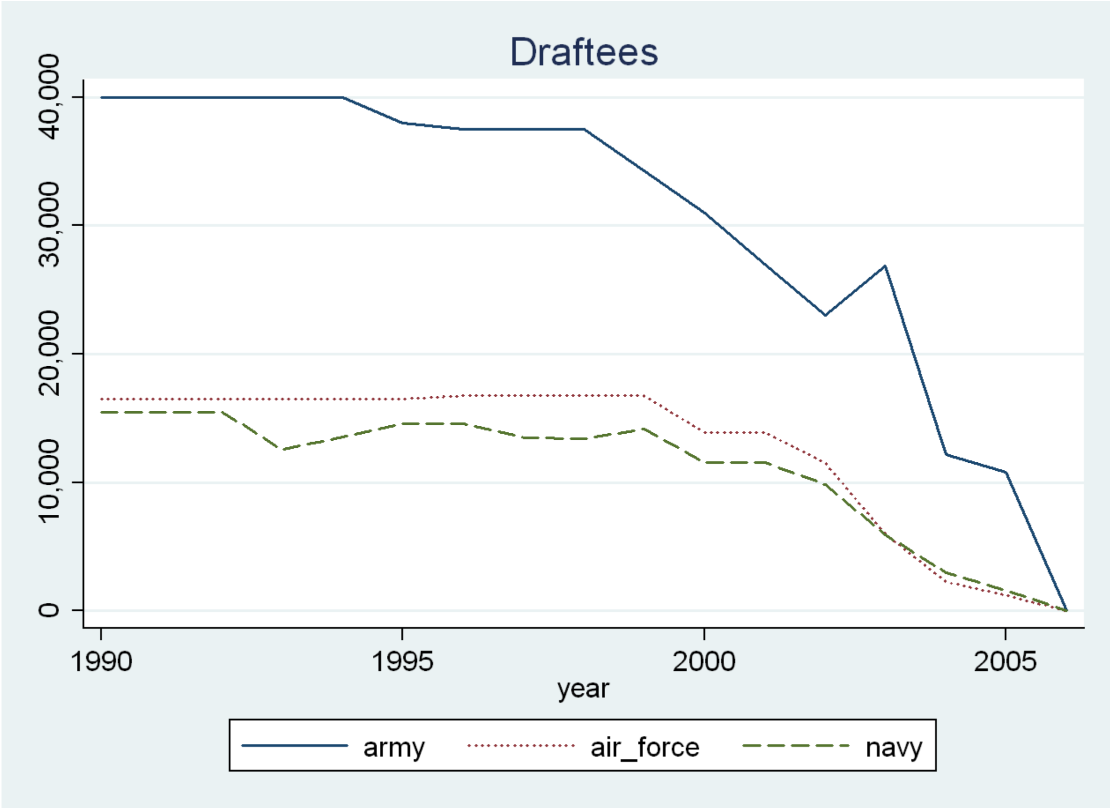

Figure 1: Maximum number of draftees

Figure 2 nicely illustrates how our instrument is supposed to work: as the number of draftees decrease over time, so does relative labor supply of young female workers, as more men are pushed into the labor force. Not surprisingly, we do not see any effect of this kind for aggregate labor supply (the dashed line in figure 2). Notice that we don't want to use as instrument the number of people who *actually* served in the military (or in NGO's), since that number is also likely to be endogenous to local labor market conditions (in particular due to the norms that regulated deferment of service). The *planned* number of draftees at the national level should be a better instrument, with the drawback that it does not display geographical variation across regions or provinces.[11]

[11]In spite of repeated contacts with the military administration, we were not able to get precise infor-

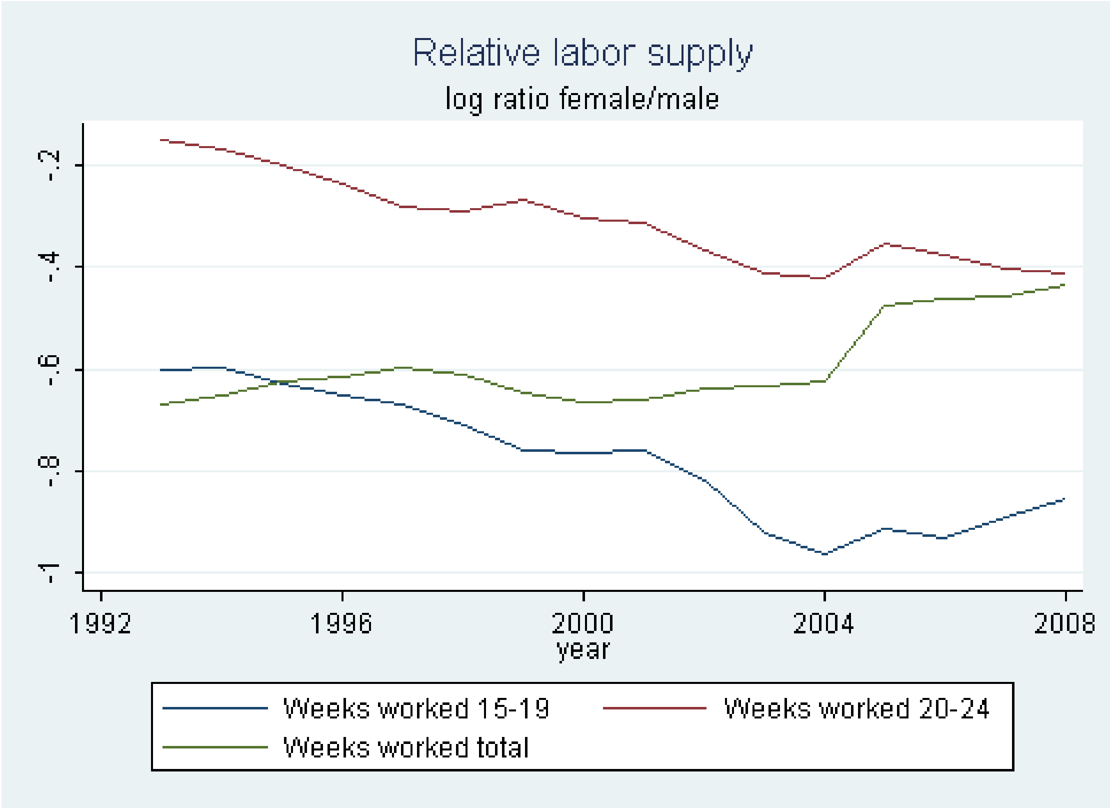

Figure 2: Relative labor supply

In order to get cross-province variation, we interact the number of draftees with sex-ratios at birth, computed using historical population data from the National Statistical Institute (Istat). We construct sex-ratios at birth for individuals in a given cohort by taking the past number of newborn boys and girls in that province. The sex-ratio at birth for the cohort of individuals aged 15-19 years old in year x in province j is thus defined as the ratio between female and male children born $x - 19$ to $x - 15$ years before in province j. In figure 3 we plot sex-ratios at birth for the age class 15-19 for a few provinces. Given that our historical population data only starts in 1970, we compute

mations concerning the criteria that were used to select draftees according to their residence. This is due to the fact that, in principle, service was a duty for everyone. Each municipality had to compile a list of all eligible individuals, randomly ordered, which was then to be used by the military administration.

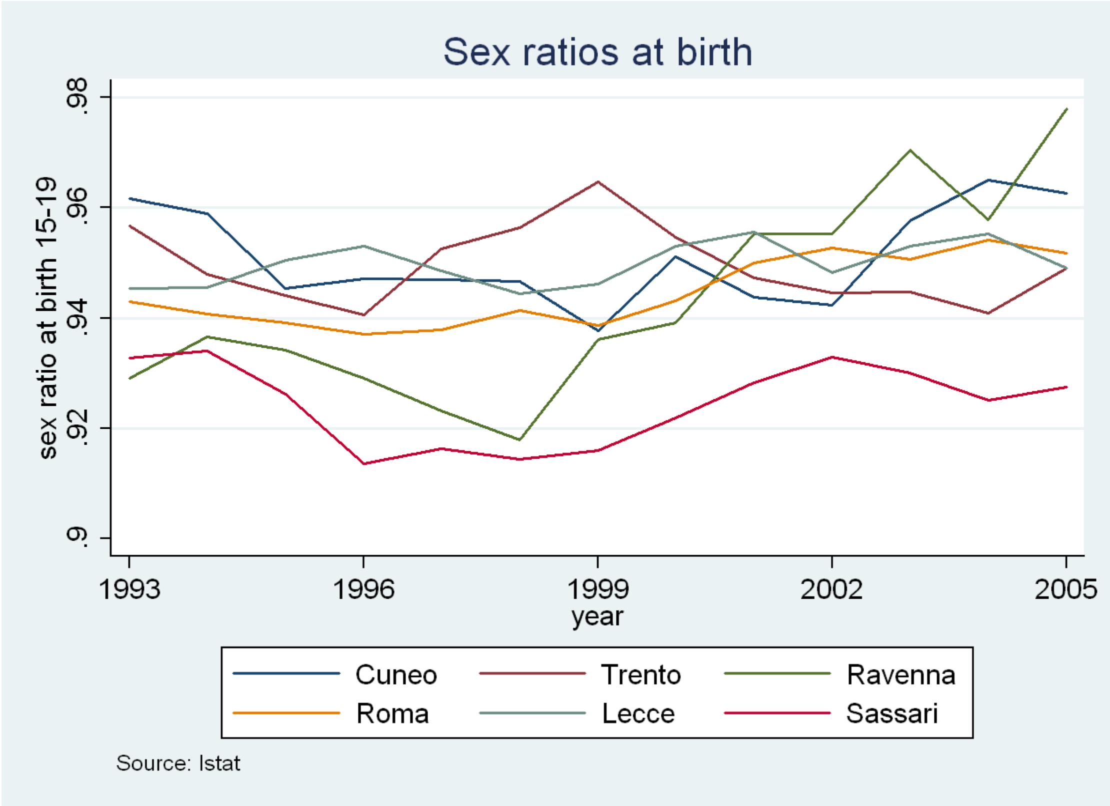

Figure 3: Sex-ratios at birth

sex-ratios at birth for two age classes (15-19 and 20-24 years old). The focus on younger cohorts is also motivated by the fact that they are the most affected by the reform of the military service. There is a good amount of variation in the data, with smaller provinces being (understandably) more erratic than larger ones. In section 4 we will therefore use, as an instrument for relative labor supply, the interaction between sex-ratios at birth and the yearly number of draftees.

4 Estimates of the elasticity of substitution

The starting point of our empirical exercise is equation 5. In particular, we run the following regression:

$$\ln\left(\frac{w_{jt}^F}{w_{jt}^M}\right) = \alpha + \phi_j + \phi_t - \frac{1}{\sigma}\ln\left(\frac{F_{jt}}{M_{jt}}\right) + \epsilon_{jt} \tag{8}$$

where ϕ_j and ϕ_t are, respectively, province and year fixed effects, that control for unobserved productivity differences, relative supply is measured by (relative) weeks worked, and ϵ_{jt} is an error term. We weight each observations by the number of individuals (in the INPS microdata archive) used to construct each province-year cell. In the first row of table 2 we report results from OLS estimation of equation 8. In odd columns we use region fixed effects, while in even columns we use province fixed effects. In the third and fourth column we restrict the sample to individuals aged 15-19 years old, and in the last two columns we restrict the sample to individuals aged 20-24 years old.

The estimated coefficients for the full sample imply large values of σ. When using province fixed effects, the coefficient is not significantly different from zero, implying an infinite elasticity of substitution. Restricting the sample to individuals aged 20-24 years, the implied estimated elasticity is in the 4.5-6.6 range. We then present instrumental variables estimates. For the full-sample we use as instrument the interaction between the number of draftees and sex-ratios at birth for individuals aged 15-24. This is obviously not an ideal measure, but it is the best we can do given that population data start in 1970. It is not a surprise that, for the full sample, the instrument is extremely weak: the first-stage F-statistics is smaller than 1, and the first-stage regression coefficient (presented in table 3) has the "wrong" sign. In fact, we expect the first-stage coefficient to be positive: both higher sex-ratios at birth (which imply there should be

Table 2: Estimates of $-1/\sigma$ in the non-nested CES model

Estimation method	Full-sample[a]		15-19 years old		20-24 years old	
OLS	0.067**	0.108	0.021	0.038	0.154***	0.219***
	(0.032)	(0.104)	(0.023)	(0.024)	(0.034)	(0.040)
2SLS	0.549	-0.245	0.169	0.185	1.471	1.737
	(0.812)	(1.911)	(0.215)	(0.202)	(0.965)	(1.476)
	[0.417]	[0.897]	[0.410]	[0.362]	[0.008]	[0.011]
Fuller LIML (F=1)	0.453	-0.089	0.158	0.175	1.309*	1.470
	(0.590)	(1.070)	(0.197)	(0.189)	(0.783)	(1.070)
	[0.417]	[0.897]	[0.410]	[0.362]	[0.008]	[0.011]
Fuller LIML (F=4)	0.309	0.028	0.132	0.153	0.998**	1.039*
	(0.329)	(0.467)	(0.159)	(0.157)	(0.490)	(0.556)
	[0.417]	[0.897]	[0.410]	[0.362]	[0.008]	[0.011]
First-stage F-stat	0.51	0.34	9.06	12.74	3.14	1.73
Year FE	YES	YES	YES	YES	YES	YES
Region FE	YES	NO	YES	NO	YES	NO
Province FE	NO	YES	NO	YES	NO	YES
Observations	1,316	1,316	1,290	1,290	1,315	1,315

Clustered standard errors in parentheses. P-value of the weak-instruments robust Anderson-Rubin test for the significance of the endogenous regressor in brackets. Observations are weighted by the number of individuals in each cell. The labor supply measure is the number of weeks worked. *p<0.1,**p<0.05,***p<0.01.

[a] When the full sample is used, the instrument is the interaction between the number of draftees and sex-ratios at birth for individuals 15-24 years old.

more women relative to men) and a higher number of draftees (which imply there are less men available to work) should increase relative female labor supply.[12] Parameters estimates are very imprecise, and are never statistically significant. Results do not change when we employ estimation methods (such as Fuller-LIML) and tests (as the one proposed by Anderson-Rubin) that are more robust to weak instruments. The IV strategy is supposed to work better with younger individuals, for which we have the ”correct” sex-ratios and who are most affected by the reform of the military service. In particular, the instrument seems to work reasonably well for individuals in the 15-19 age bracket (columns 3 and 4 of tables 2 and 3). The estimated coefficients do not change much if we use province or region fixed effects, but estimates are quite imprecise and are never statistically different from zero, implying that male and female workers in that age class are perfect substitutes.

When we turn to individuals aged 20-24 years (columns 5 and 6) the instrument looses strenght, with a first-stage F-stat below 5. For this reason we estimate coefficients using Fuller-Limited Information Maximum Likelihood, with Fuller parameter set to 1 (which yields the most unbiased estimator) and 4 (to minimize the mean-squared error of the estimator; Fuller, 1977). Parameters estimated using Fuller-LIML are actually quite different from 2SLS. The Anderson-Rubin test provides weak-instrument robust inference, confirming that the estimated parameters are statistically different from zero. Taking Fuller-LIML with parameter 4 as our preferred specification would imply an elasticity of substitution close to 1 (meaning that the CES production function is actually a Cobb-Douglas). This value is much smaller than the ones proposed by Acemoglu, Autor, and Lyle (2004), who estimated an implied (long-run) elasticity of substitution

[12]The presence of a negative coefficient could be due to some mean-reversion mechanisms: provinces with high sex-ratios for the 15-24 cohort have probably lower sex-ratios for older cohorts.

Table 3: First-stage and reduced-form regressions

	PANEL A: Non-nested CES model					
	Full-sample[a]		15-19 years old		20-24 years old	
Reduced-form coefficient	0.432	0.074	-0.512	-0.612	-1.898***	-1.796**
	(0.530)	(0.568)	(0.618)	(0.668)	(0.703)	(0.693)
First-stage coefficient	-0.788	0.301	3.028***	3.312***	1.290***	1.034**
	(1.105)	(0.514)	(1.006)	(0.928)	(0.729)	(0.786)
Year FE	YES	YES	YES	YES	YES	YES
Region FE	YES	NO	YES	NO	YES	NO
Province FE	NO	YES	NO	YES	NO	YES
Angrist-Pischke F-stat	0.51	0.34	9.06	12.74	3.14	1.73

	Panel B - Nested CES model	
Reduced-form coefficient	-1.471***	-1.393***
	(0.450)	(0.490)
First-stage coefficient	1.711***	1.382***
	(0.651)	(0.656)
Year FE	YES	YES
Region FE	YES	NO
Province FE	NO	YES
Angrist-Pischke F-stat	6.91	7.80

Clustered standard errors in parentheses. Observations are weighted by the number of individuals in each cell. In panel B we also control for age-class dummies (15-19 and 20-24). The labor supply measure is the total number of weeks worked. *p<0.1,**p<0.05,***p<0.01.

[a] When the full sample is used, the instrument is the interaction between the number of draftees and sex-ratios at birth for individuals 15-24 years old.

between 2 and 3 for their entire sample, and about 5 for a subsample of individuals aged 25-34. Notice that the 2SLS estimates would suggest an even smaller elasticity of substitution (between .57 and .68, implying that male and female are actually gross complements). Given the weakness of the instrument, however, these estimates should be taken with extreme caution.

Taking into account imperfect substitutability between workers of different ages, we can use the nested CES model presented in equations 6 and 7 and estimate the following equation:

$$\ln\left(\frac{w^F_{jtx}}{w^M_{jtx}}\right) = \alpha + \phi_j + \phi_t + \phi_x - \frac{1}{\sigma}\ln\left(\frac{F_{jtx}}{M_{jtx}}\right) + \epsilon_{jtx} \tag{9}$$

constructing relative wages and relative labor supply at the province-year-age class level, and adding fixed effects for age class. Results are presented in panel B of table 3 and in table 4. In the full sample we include all seven age classes (15-19, 20-24, 25-29, 30-39, 40-49, 50-59, over 60); notice however that for this specifications we do not have available instruments. This is why we next focus on a restricted sample with only two age classes (15-19 and 20-24). Least squares estimates for the full sample are similar to the ones obtained for the non-nested model, but are estimated more precisely. The implied elasticity of substitution is about 16, and drops to 10 in the restricted sample with province fixed effects. Results change substantially when we turn to instrumental-variable estimation. Coefficients are precisely estimated, and the instruments are reasonably strong. For robustness reasons, we continue to present Fuller-LIML estimates and p-values for the Anderson-Rubin test. The implied elasticity of substitution for the sample of younger workers is in the range 1.2-1.5, closer to the values estimated by Acemoglu, Autor, and Lyle (2004) and to commonly accepted values of the elasticity of

Table 4: Estimates of $-1/\sigma$ in the nested CES model

Estimation method	Full-sample[a]		Restricted sample[b]	
OLS	0.059***	0.061***	0.069***	0.102***
	(0.015)	(0.017)	(0.020)	(0.023)
2SLS	-	-	0.860**	0.760*
			(0.422)	(0.401)
			[0.004]	[0.006]
Fuller LIML (F=1)	-	-	0.809**	0.724**
			(0.382)	(0.369)
			[0.004]	[0.006]
Fuller LIML (F=4)	-	-	0.689**	0.634**
			(0.296)	(0.296)
			[0.004]	[0.006]
First-stage F-stat	-	-	6.91	7.80
Age class FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Region FE	YES	NO	YES	NO
Province FE	NO	YES	NO	YES
Observations	9,053	9,053	2,605	2,605

Clustered standard errors in parentheses. P-value of the weak-instruments robust Anderson-Rubin test for the significance of the endogenous regressor in brackets. Observations are weighted by the number of individuals in each cell. The labor supply measure is the number of weeks worked. *p<0.1,**p<0.05,***p<0.01.

[a] In the full sample we use 7 age-class dummies: 15-19, 20-24, 25-29, 30-39, 40-49, 50-59, over 60.

[b] In the restricted sample only the first two age classes are used (15-19 and 20-24 years). In the restricted sample relative labor supply is instrumented with the interaction between the number of draftees in each year and the sex-ratio at birth for the corresponding age class.

substitution between skilled and unskilled workers.[13]

5 Conclusions

In this paper we provide estimates of the elasticity of substitution between male and female workers. We exploit a natural experiment (the abolition of compulsory military service that took place in Italy in the early 2000s) and we use sex-ratios at birth to construct instruments for relative labor supply. We focus our attention on young workers (aged 15-24), given that they are the ones most directly affected by the reform and those for whom we were able to compute sex-ratios at birth. We are not able to reject the hypothesis that men and women in the 15-19 years age class are perfect substitutes, although parameters are not estimated very precisely. On the other hand, for workers aged 20-24 we find evidence of imperfect substitutability, with an elasticity of substitution between 0.7 and 1. When we take into account imperfect substitutability between workers of different age, our estimates of the elasticity of substitution ranges between 1.2 and 1.6 for workers in the 15-24 age interval. Such estimates are smaller than what previously found in the literature (Acemoglu, Autor, and Lyle, 2004) and are close to commonly accepted values of the elasticity of substitution between high-skilled and low-skilled workers. Our results have important implications for the evaluation of policies aimed at increasing female labor force participation. The small degree of substitutability between male and female workers would imply small effects of increased

[13]As a robustness check, in the Appendix we present estimation results using alternative measures of labor supply, namely employment levels (in tables A-1 and A-2) and hours worked (in tables A-3 and A-4). Both measures are constructed using microdata from the Labor Force Survey. Overall, the results are broadly consistent with what we found using weeks worked. In the non-nested model, we still find that men and women are perfect substitutes in the 15-19 age bracket; for workers aged 20-24, the instrument appears to be too weak and we don't get meaningful parameter estimates. The nested CES model seems to work better. The implied elasticity of substitution ranges between 0.7 and 1.6 when using employment levels and between 0.8 and 1.6 when using hours worked.

female participation on male wages and employment levels; males would actually benefit from increased female participation if the true elasticity were below 1 (i.e. if male and female workers were complements). Given that the estimated elasticity of substitution decreases a lot when we focus on 20-24 years old individuals compared to 15-19 years old, it would be reasonable to expect smaller degrees of substitutability between older workers. Our parameter estimates could be fruitfully employed in the calibration of richer models that take into account the response of the capital stock or other dimensions of heterogeneity in labor inputs (e.g. race or education), with the purpose of simulating the effects of different labor supply shocks (as in D'Amuri, Ottaviano, and Peri, 2010 and Ottaviano and Peri, 2012).

Appendix

Table A-1: Estimates of $-1/\sigma$ in the non-nested CES model - Employment

Estimation method	Full-sample[a]		15-19 years old		20-24 years old	
OLS	0.069	0.057	0.006	0.021	0.016	0.003
	(0.047)	(0.074)	(0.018)	(0.019)	(0.029)	(0.034)
2SLS	0.315	0.094	0.137	0.158	3.285	-9.371
	(0.357)	(0.705)	(0.178)	(0.208)	(4.093)	(38.539)
	[0.417]	[0.897]	[0.437]	[0.435]	[0.008]	[0.011]
Fuller LIML (F=1)	0.301	0.090	0.129	0.147	1.816	-0.955*
	(0.336)	(0.634)	(0.167)	(0.190)	(1.353)	(0.577)
	[0.417]	[0.897]	[0.437]	[0.435]	[0.008]	[0.011]
Fuller LIML (F=4)	0.268	0.082	0.109	0.123	0.782**	-0.255**
	(0.287)	(0.487)	(0.139)	(0.151)	(0.358)	(0.109)
	[0.417]	[0.897]	[0.437]	[0.435]	[0.008]	[0.011]
First-stage F-stat	2.56	2.79	9.94	6.68	0.66	0.06
Year FE	YES	YES	YES	YES	YES	YES
Region FE	YES	NO	YES	NO	YES	NO
Province FE	NO	YES	NO	YES	NO	YES
Observations	1,316	1,316	1,242	1,242	1,315	1,315

Clustered standard errors in parentheses. P-value of the weak-instruments robust Anderson-Rubin test for the significance of the endogenous regressor in brackets. Observations are weighted by the number of individuals in each cell. The labor supply measure is the number of empèloyed individuals from the Labor force survey. *p<0.1,** p<0.05,***p<0.01.

[a] When the full sample is used, the instrument is the interaction between the number of draftees and sex-ratios at birth for individuals 15-24 years old.

Table A-2: Estimates of $-1/\sigma$ in the nested CES model - Employment

Estimation method	Full-sample[a]		Restricted sample[b]	
OLS	0.004	0.009	0.009	0.011
	(0.015)	(0.015)	(0.017)	(0.017)
2SLS	-	-	1.117**	1.448
			(0.602)	(1.060)
			[0.004]	[0.006]
Fuller LIML (F=1)	-	-	0.986**	1.118
			(0.494)	(0.673)
			[0.004]	[0.006]
Fuller LIML (F=4)	-	-	0.730**	0.667**
			(0.317)	(0.300)
			[0.004]	[0.006]
First-stage F-stat	-	-	5.23	2.21
Age class FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Region FE	YES	NO	YES	NO
Province FE	NO	YES	NO	YES
Observations	9,003	9,003	2,557	2,557

Clustered standard errors in parentheses. P-value of the weak-instruments robust Anderson-Rubin test for the significance of the endogenous regressor in brackets. Observations are weighted by the number of individuals in each cell. The labor supply measure is the number of employed individuals from the Labor force survey. *p<0.1,** p<0.05,***p<0.01.

[a] In the full sample we use 7 age-class dummies: 15-19, 20-24, 25-29, 30-39, 40-49, 50-59, over 60.

[b] In the restricted sample only the first two age classes are used (15-19 and 20-24 years). In the restricted sample relative labor supply is instrumented with the interaction between the number of draftees in each year and the sex-ratio at birth for the corresponding age class.

Table A-3: Estimates of $-1/\sigma$ in the non-nested CES model - Hours

Estimation method	Full-sample[a]		15-19 years old		20-24 years old	
OLS	0.058	0.024	-0.007	0.005	0.018	0.003
	(0.046)	(0.073)	(0.017)	(0.019)	(0.026)	(0.029)
2SLS	0.303	0.103	0.148	0.163	3.143	-9.043
	(0.345)	(0.773)	(0.180)	(0.195)	(3.821)	(37.997)
	[0.417]	[0.897]	[0.409]	[0.390]	[0.008]	[0.011]
Fuller LIML (F=1)	0.289	0.092	0.136	0.150	1.719	-0.872*
	(0.323)	(0.669)	(0.167)	(0.178)	(1.232)	(0.524)
	[0.417]	[0.897]	[0.409]	[0.390]	[0.008]	[0.011]
Fuller LIML (F=4)	0.254	0.074	0.110	0.122	0.736**	-0.231**
	(0.273)	(0.476)	(0.136)	(0.141)	(0.325)	(0.099)
	[0.417]	[0.897]	[0.409]	[0.390]	[0.008]	[0.011]
First-stage F-stat	2.23	2.29	8.50	5.94	0.67	0.06
Year FE	YES	YES	YES	YES	YES	YES
Region FE	YES	NO	YES	NO	YES	NO
Province FE	NO	YES	NO	YES	NO	YES
Observations	1,316	1,316	1,237	1,237	1,315	1,315

Clustered standard errors in parentheses. P-value of the weak-instruments robust Anderson-Rubin test for the significance of the endogenous regressor in brackets. Observations are weighted by the number of individuals in each cell. The labor supply measure is the number of hours worked from the Labor force survey. *p<0.1,** p<0.05,***p<0.01.

[a] When the full sample is used, the instrument is the interaction between the number of draftees and sex-ratios at birth for individuals 15-24 years old.

Table A-4: Estimates of $-1/\sigma$ in the nested CES model - Hours

	Full-sample[a]		Restricted sample[b]	
Estimation method				
OLS	0.001	0.002	0.001	0.002
	(0.014)	(0.014)	(0.015)	(0.015)
2SLS	-	-	1.096*	1.321
			(0.570)	(0.929)
			[0.004]	[0.005]
Fuller LIML (F=1)	-	-	0.958**	1.037*
			(0.461)	(0.606)
			[0.004]	[0.005]
Fuller LIML (F=4)	-	-	0.696**	0.630**
			(0.290)	(0.278)
			[0.004]	[0.005]
First-stage F-stat	-	-	5.43	2.35
Age class FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Region FE	YES	NO	YES	NO
Province FE	NO	YES	NO	YES
Observations	8,996	8,996	2,552	2,552

Clustered standard errors in parentheses. P-value of the weak-instruments robust Anderson-Rubin test for the significance of the endogenous regressor in brackets. Observations are weighted by the number of individuals in each cell. The labor supply measure is the number of hours worked from the Labor force survey. *p<0.1,** p<0.05,***p<0.01.

[a] In the full sample we use 7 age-class dummies: 15-19, 20-24, 25-29, 30-39, 40-49, 50-59, over 60.

[b] In the restricted sample only the first two age classes are used (15-19 and 20-24 years). In the restricted sample relative labor supply is instrumented with the interaction between the number of draftees in each year and the sex-ratio at birth for the corresponding age class.

References

ABREVAYA, J., AND D. S. HAMERMESH (2012): "Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?," Review of Economics and Statistics, 94(1), 202–207.

ACEMOGLU, D., D. H. AUTOR, AND D. LYLE (2004): "Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury," Journal of Political Economy, 112(3), 497–551.

ALESINA, A., A. ICHINO, AND L. KARABARBOUNIS (2011): "Gender-Based Taxation and the Division of Family Chores," American Economic Journal: Economic Policy, 3(2), 1–40.

ANDREONI, J., AND L. VESTERLUND (2001): "Which Is the Fair Sex? Gender Differences in Altruism," The Quarterly Journal of Economics, 116(1), 293–312.

ANGRIST, J. D. (1995): "The Economic Returns to Schooling in the West Bank and Gaza Strip," American Economic Review, 85(5), 1065–87.

ARPINO, B., C. PRONZATO, AND L. P. TAVARES (2010): "All in the family: informal childcare and mothers' labour market participation," ISER Working Paper Series 2010-24, Institute for Social and Economic Research.

BANCA D'ITALIA (2011): L'Economia delle Regioni Italiane, no. 22 in Economie regionali. Banca d'Italia.

BARONE, G., AND S. MOCETTI (2011): "With a Little Help from Abroad: the Effect of Low-Skilled Immigration on the Female Labor Supply," Labour Economics, 18(5), 664–675.

BEAMAN, L., R. CHATTOPADHYAY, E. DUFLO, R. PANDE, AND P. TOPALOVA (2009): "Powerful Women: Does Exposure Reduce Bias?," The Quarterly Journal of Economics, 124(4), 1497–1540.

BORGHANS, L., B. H. H. GOLSTEYN, J. J. HECKMAN, AND H. MEIJERS (2009): "Gender Differences in Risk Aversion and Ambiguity Aversion," Journal of the European Economic Association, 7(2-3), 649–658.

BORJAS, G. J. (2003): "The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," The Quarterly Journal of Economics, 118(4), 1335–1374.

——— (2009): "The Analytics of the Wage Effect of Immigration," NBER Working Papers 14796, National Bureau of Economic Research, Inc.

BORJAS, G. J., J. GROGGER, AND G. H. HANSON (2008): "Imperfect Substitution between Immigrants and Natives: A Reappraisal," NBER Working Papers 13887, National Bureau of Economic Research, Inc.

BRILLI, Y., D. DEL BOCA, AND C. PRONZATO (2011): "Exploring the Impacts of Public Childcare on Mothers and Children in Italy: Does Rationing Play a Role?," IZA Discussion Papers 5918, Institute for the Study of Labor (IZA).

CARD, D. (2001): "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," Journal of Labor Economics, 19(1), 22–64.

CARD, D., AND T. LEMIEUX (2001a): "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis," The Quarterly Journal of Economics, 116(2), 705–746.

——— (2001b): "Going to College to Avoid the Draft: The Unintended Legacy of the Vietnam War," American Economic Review, 91(2), 97–102.

CASELLI, F., AND W. J. COLEMAN (2006): "The World Technology Frontier," American Economic Review, 96(3), 499–522.

CICCONE, A., AND G. PERI (2005): "Long-Run Substitutability Between More and Less Educated Workers: Evidence from U.S. States, 1950-1990," Review of Economics and Statistics, 87(4), 652–663.

CIPOLLONE, P., AND A. ROSOLIA (2007): "Social Interactions in High School: Lessons from an Earthquake," American Economic Review, 97(3), 948–965.

CROSON, R., AND U. GNEEZY (2009): "Gender Differences in Preferences," Journal of Economic Literature, 47(2), 448–74.

D'AMURI, F., G. I. OTTAVIANO, AND G. PERI (2010): "The Labor Market Impact of Immigration in Western Germany in the 1990s," European Economic Review, 54(4), 550–570.

DE PAOLA, M., V. SCOPPA, AND R. LOMBARDO (2010): "Can gender quotas break down negative stereotypes? Evidence from changes in electoral rules," Journal of Public Economics, 94(5-6), 344–353.

DEL BOCA, D., M. LOCATELLI, AND D. VURI (2005): "Child-Care Choices by Working Mothers: The Case of Italy," Review of Economics of the Household, 3(4), 453–477.

DEL BOCA, D., AND D. VURI (2007): "The mismatch between employment and child care in Italy: the impact of rationing," Journal of Population Economics, 20(4), 805–832.

DI PIETRO, G. (forthcoming): "Military conscription and university enrolment: evidence from Italy," Journal of Population Economics.

ECKEL, C. C., AND P. J. GROSSMAN (1998): "Are Women Less Selfish Than Men? Evidence from Dictator Experiments," The Economic Journal, 108(448), 726–35.

FULLER, W. (1977): "Some Properties of a Modification of the Limited Information Estimator," Econometrica, 45, 939–954.

GAGLIARDUCCI, S., AND M. D. PASERMAN (2012): "Gender Interactions within Hierarchies: Evidence from the Political Arena," Review of Economic Studies, 79(3), 1021–1052.

GNEEZY, U., M. NIEDERLE, AND A. RUSTICHINI (2003): "Performance in Competitive Environments: Gender Differences," The Quarterly Journal of Economics, 118(3), 1049–1074.

GNEEZY, U., AND A. RUSTICHINI (2004): "Gender and Competition at a Young Age," American Economic Review, 94(2), 377–381.

KATZ, L. F., AND K. M. MURPHY (1992): "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," The Quarterly Journal of Economics, 107(1), 35–78.

KLUVE, J., AND M. TAMM (forthcoming): "Parental Leave Regulations, Mothers' Labor Force Attachement and Fathers' Childcare Involvement: Evidence from a Natural Experiment," Journal of Population Economics.

LEUVEN, E., H. OOSTERBEEK, AND H. VAN OPHEM (2004): "Explaining International Differences in Male Skill Wage Differentials by Differences in Demand and Supply of Skill," The Economic Journal, 114(495), 466–486.

OLIVETTI, C., AND B. PETRONGOLO (2008): "Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps," Journal of Labor Economics, 26(4), 621–654.

OTTAVIANO, G. I., AND G. PERI (2006): "The Economic Value of Cultural Diversity: Evidence from US Cities," Journal of Economic Geography, 6(1), 9–44.

——— (2012): "Rethinking the Effect of Immigration on Wages," Journal of the European Economic Association, 10(1), 152–197.

PASERMAN, M. D. (2010): "Gender Differences in Performance in Competitive Environments? Evidence from Professional Tennis Players," Boston University - Department of Economics - Working Papers Series WP2010-047, Boston University - Department of Economics.

PRONZATO, C. (2009): "Return to work after childbirth: does parental leave matter in Europe?," Review of Economics of the Household, 7(4), 341–360.

ZIZZA, R. (2012): "The Gender Wage Gap in Italy," mimeo, Bank of Italy.