WORKING PAPER | ISSUE 02/2025 | 24 FEBRUARY 2025

GENDER DIVERSITY AND ECONOMIC GROWTH

JONATHAN D. OSTRY

Most macroeconomic and growth accounting models assume that male and female workers are perfectly substitutable in the aggregate production function. Whether this assumption is valid is an empirical question that this paper aims to answer by estimating the elasticity of substitution between female and male labour. We apply linear and non-linear techniques to firm-level data, cross-country sectoral data and cross-country aggregate data. We find that women and men are far from being perfect substitutes in production, a result that is consistent with much microeconomic evidence, but has not permeated to macroeconomics. The failure to account for imperfect gender substitutability has far-reaching implications. In particular, standard growth accounting exercises are likely to attribute to technological progress gains that are more properly attributable to the impact of greater gender inclusiveness in the labour force over time. Put differently, the gains from gender inclusiveness are likely to be much larger than standard economic models estimate.

Keywords: Female labour force participation, gender inclusion, aggregate production function, elasticity of substitution, growth.
JEL Classification: O47, J24.

Jonathan D. Ostry is a Non-resident fellow at Bruegel

Helpful comments from Jorge Alvarez, Shankha Chakraborty, Doug Gollin, Gita Gopinath, Duygu Güner, Narayana Kocherlakota, Christine Lagarde, Petia Topalova and seminar participants at the International Monetary Fund are gratefully acknowledged. Thanks also to Grace Li for sharing the Chinese manufacturing data, and to Zidong An and Marina Conseca Martinez for outstanding research assistance. This work benefited from the financial support of the Munk School of Global Affairs and Public Policy at the University of Toronto and the UK's Foreign, Commonwealth and Development Office. The views expressed in this paper are those of the author solely and should not be attributed to institutions with which they may be affiliated.

bruegel

Recommended citation:
Ostry, J. D. (2025) 'Gender diversity and economic growth', *Working Paper* 02/2025, Bruegel

1 Introduction

Female labour force participation (FLFP) remains stubbornly below male labour force participation (MLFP). While cultural factors, including those related to the historical role of agricultural work in the division of labour, may partly explain this legacy, technological improvements at home, changing norms of gender roles, contraception and improved schooling of women have helped to partly reverse the impact of history (Goldin, 1990; Goldin and Katz, 2002). But the journey is incomplete: FLFP – about 55 percent for the median OECD country on the eve of the pandemic – has remained well below MLFP (70 percent), and the gap is even larger for middle-income countries (48 versus 77 percent in median country terms).

This paper studies the theoretical and empirical relevance of extending one of the most prominent macroeconomic equations, the neo-classical production function, to account for gender. The production function is a building block of many macroeconomic models. It is used both in analyses of long-term growth, such as growth accounting, and to derive labour demand functions in short-term models. Growth accounting often disaggregates inputs to improve the analysis, for instance differentiating capital by type, such as public versus private, or machinery versus structures. However, for the most part, it has not been extended to address the gender dimension. Existing growth accounting methods thus assess the impact of an increase in female employment as not being fundamentally different from that of an increase in male employment – by construction, if the model assumes that production depends only on the sum of female and male workers, what matters is the total supply of workers. But it is increasingly recognised that social norms affect the way that women and men interact in the workplace, with women bringing a range of different skills and ideas to their work environments: in such a world, the mechanical exercise embedded in a model that assumes perfect gender substitutability may be wide of the mark.

Pew Research Center (2020) documented a range of social, management and negotiating settings in which women and men differ. Laboratory experiments have also highlighted gender differences in risk-aversion and competitive behaviour, and their relation to sociological and cultural contexts. Firm-level analyses have found that gender diversity contributes favourably to risk management, productivity and profits. Against this background, this paper makes four main contributions.

First, we clarify analytically the importance of gender diversity in cases in which the elasticity of substitution is finite. Through the prism of a simple aggregate model in which women and men are imperfect substitutes in production, it will be the case that, from an initially inefficient mix of women and men (that favours men over women), raising female employment boosts output by more than an equivalent increase in MLFP, as long as the productivity of the female entrants is not substantially lower than that of males. We show that the positive effect will be stronger the larger is the initial gender imbalance and the smaller is the elasticity of substitution.

Second, we apply linear and non-linear methods to estimate the elasticity of substitution between female and male labour, using firm-level data from China's manufacturing sector, cross-country data at the sectoral level and cross-country data at the aggregate level. To our knowledge, ours represents the first attempt at estimating this elasticity of substitution using production data, an exercise that permits us to compare the elasticity of substitution for different levels of aggregation. The production function model relates output to the stock of capital, male employment, and female employment, in a constant elasticity of substitution (CES) specification. We find that in most specifications, the elasticity of substitution between male and female labour is around 2-3 at the firm-level, between 1-2 for the sectoral sample, and below 1 for the aggregate sample, suggesting that men and women are indeed imperfect substitutes.

Third, we use our estimates to compute the potential economic gains from future increases in FLFP. For the median OECD country, simulations suggest that raising female employment to match male employment would increase GDP by 8-14 percent. A significant proportion of the benefit reflects the impact of imperfect substitutability on labour productivity. Our estimates suggest that the marginal product of male labour (and thus men's real wage) should be increasing in FLFP when the elasticity of substitution is below 2.5. Under this calibration, the complementarity effect of female participation outweighs the negative effect of a higher labour supply on capital intensity.

Fourth, a growth accounting exercise (calibrated with an elasticity of substitution consistent with our empirical results for macro data) suggests that some 0.2-0.3 percentage points of annual TFP growth over 1990-2019 is due to the gender diversity effect. This magnitude is economically large, as it corresponds to about one quarter of TFP growth for middle-income countries. The usual interpretation of TFP growth as driven by technology and innovation is a distortion that ignores other factors, notably, as we show, that of rising FLFP over time. Although economic and welfare gains resulting from women moving from non-market to market activities may differ (Ostry *et al*, 2018), the impact of rising FLFP on GDP is nevertheless salient for the overall macroeconomic picture[1] (Lagarde and Ostry, 2018).

Our findings underscore the importance of incorporating gender diversity into macroeconomic models in general and growth accounting exercises in particular. Macroeconomic models have also tended to be calibrated using either arbitrary values of the elasticity of substitution between female and male workers, estimates based on an early literature on labour demand (Freeman, 1979), or the estimates of Acemoglu *et al* (2004) based on US wage data from the 1950s. Such calibrations would benefit from new cross-country empirical evidence, which our paper provides.

The rest of the paper is organised as follows. Section 2 presents a framework to clarify how gender diversity relates to the empirical analysis of productivity growth. Section 3 discusses the data used in the empirical analysis. Section 4 reports our empirical results. Section 5 examines the implications for TFP and growth accounting. Section 6 concludes.

1 Jonathan Ostry, 'Economists' models miss the gains from more women in the workforce', *Financial Times*, 4 August 2022, https://www.ft.com/content/e545589d-907d-4961-ac13-a4aab45edb06

2 Framework

Our framework is centred on a CES production technology. Output, Y, is produced with a constant return to scale technology in a composite labour input L and the capital stock K:

$$Y = A\left(\delta_l L^{\rho_1} + \delta_k K^{\rho_1}\right)^{\frac{1}{\rho_1}}, \tag{1}$$

where the elasticity of substitution between labour and capital is $\sigma_1 = 1/(1-\rho_1)$ and A is a technology parameter. δ_l and δ_k are the share parameters. The labour variable L is itself a composite of female (F) and male (M) labour, nested in a CES function:

$$L = \left(\delta F^{\rho_2} + M^{\rho_2}\right)^{\frac{1}{\rho_2}}, \tag{2}$$

where the elasticity of substitution between male and female workers is $\sigma_2 = 1/(1-\rho_2)$. δ is a weight parameter, which may be necessary to re-scale female labour in a unit comparable to male labour. Statistics often record labour in terms of headcount whereas work hours would be more appropriate. δ could also be used to adjust for differences in skills or other aspects of human capital, though this is not investigated in this paper.

Expressing in lower case all variables in growth rates, and combining the log-linearisation of equations (1) and (2), output growth is:

$$y = \lambda\mu f + \lambda(1-\mu)m + (1-\lambda)k + a. \tag{3}$$

This equation accounts for output growth as the sum of four contributions: (i) growth in female labour (f) multiplied by the female labour share of income ($\lambda\mu$); (ii) the corresponding quantities for men (m multiplied by $\lambda(1-\mu)$); (iii) growth in the stock of capital (k) multiplied by the capital share of income ($1-\lambda$); (iv) and TFP growth (a). In general, μ is a non-linear function of σ_2, but it is possible to assess the influence of female-male labour substitutability by means of a Taylor approximation of equation (3), evaluated around the point of perfect substitutability between F and M. From an initial equilibrium with F < M and f >m, there are three implications:

- First, the Solow residual is increasing in M/F, ie the greater the initial gender imbalance, the larger is the Solow residual.
- Second, the Solow residual is increasing in f – m, ie the speed of reduction of the gender imbalance influences positively the Solow residual.
- And third, the Solow residual is decreasing in the EOS between F and M, ie the less substitutable are women and men in production, the larger is the Solow residual.

The logic of the Taylor approximation is clear. When women and men are imperfect substitutes, total factor productivity depends on the growth of female labour, an effect which standard growth accounting ignores. If women are added to the labour force at a rate faster than men (the gender gap is narrowing over time as is, and has been, the case in many countries), growth will depend on a gender diversity effect, which in turn is determined by the EOS between F and M (being larger the lower is this elasticity of substitution). Conversely, the gender diversity effect is zero if the initial gender imbalance is zero, or if the EOS is infinite: only in such cases are the assumptions and predictions of standard growth accounting exercises valid. In all but these highly unusual circumstances, TFP should not be understood as reflecting solely the influence of technological improvements, but rather as including the impact of a narrowing over time of the gender participation gap. And, in a forward-looking sense, the gains for conventionally-measured TFP from reducing gender inequalities will be larger than those assumed in models with perfect substitutability, because of the gender diversity effect that our richer model identifies.

3 A brief look at the data

Our data is in three buckets: firm level; sectoral level; and aggregate level. The firm-level data for China is a random subsample of 2528 firms taken from the Annual Surveys of Industrial Production conducted by the Chinese government's National Bureau of Statistics. The original data, which has been used in other studies of firm-level productivity (in particular Hsieh and Klenow, 2009, and Feenstra *et al*, 2014) covers the period 1998-2005, but data on the gender composition of the labour force of each firm is only available for the year 2004.

We next use a dataset of economic sectors, which includes sectoral value added, sectoral employment by gender, and sectoral capital stocks, taken from the OECD Structural Analysis Database (STAN). STAN covers OECD countries and includes sectors where male employment dominates (eg mining) as well as sectors where female employment dominates (eg education, health and social work). The dataset coverage is heterogeneous across countries and sectors, but sufficient to be representative of all sectors. The sectoral dataset includes 2831 annual observations, which are used to compute 513 non-overlapping five-year growth rates.

Finally, we use a macroeconomic dataset, which is smaller, although it allows to cover more countries. Labour force participation by gender is taken from the World Bank's World Development Indicators (WDI), which starts in the 1990s, whereas data on output and capital stocks (PPP) are taken from the Penn World Tables (version 9.0). We also check robustness to: using female and male employment from the OECD (Annual Labor Force Statistics, ALFS); using OECD data for GDP (PPP); and using IMF data on capital stocks in PPP (IMF, 2015). Depending on the exact series used, the macroeconomic annual dataset comprises around 1000 annual observations, which yields around 150 observations as 5-year non-overlapping growth rates.

4 Results

4.1 Firm-level results

We start with the firm-level estimates. Because data is only available for 2004, we can only use cross-sectional information and we cannot difference the data to control for firm level total factor productivity. This also means that only the non-linear least squares estimation, based on the cross-section of firms, can be run. This regression is appropriate under the assumption that the firm's TFP is uncorrelated with the other explanatory variables.

The results are shown in Table 1, with the different columns presenting estimations for different sub-samples: the whole sample (column 1); the sample removing outliers (column 2); the sample of firms with larger (column 3) or smaller (column 4) capital stocks than the median; and the sample of firms with a level of employment larger (column 5) or smaller (column 6) than the median.

Bootstrapping simulations are used to estimate the confidence interval for the elasticity of substitution, and the median estimate, mean estimate, and one-standard-deviation confidence interval are shown in Table 1. The bootstrapping simulations indicate that the mode and the median of the distribution for the elasticity of substitution are below 2, with a 16th-84th percentile confidence interval at (1.2-2.9). The estimates appear quite stable to choosing sub-samples. Most estimates for the elasticity of substitution are between 1 and 3.5, broadly consistent with the evidence from the micro literature using labor supply and wage data which suggests relatively large effects of shocks to the gender composition of the labor force on relative wages (which effect is proportional to the degree of complementarity or the reciprocal of the elasticity of substitution).The mean estimate for firms with a large labour force (5.3, see column 5) is high, but this estimate is also the least precise. Finally, the labour share coefficient is estimated to be between 0.6 and 0.7.

Table 1: NLLS estimation, Chinese firm-level data, log-level estimation

	(1)	(2)	(3)	(4)	(5)	(6)
	All	**All: P1-99**	**K>Med**	**K<Med**	**L>Med**	**L<Med**
α (labour share)	0.75***	0.74***	0.58***	0.67***	0.72***	0.79***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
β (constant; ln(TFP))	2.82***	2.68***	1.97***	10.45	2.58***	2.87***
	(0.12)	(0.17)	(0.18)	(39.78)	(0.13)	(0.32)
δ (CES weight parameter)	0.52***	0.49***	0.51***	0.38***	0.53***	0.58***
	(0.06)	(0.05)	(0.11)	(0.07)	(0.09)	(0.09)
$\rho 2$ (=1-1/$\sigma 2$)	0.67***	0.48***	0.71**	-0.03	0.81***	0.43**
	(0.16)	(0.14)	(0.32)	(0.16)	(0.24)	(0.19)
$\sigma 2$ (elasticity of substitution)	**3.03**	**1.92**	**3.45**	**0.97**	**5.26**	**1.75**
Median bootstrapping	**2.89**	**1.84**	**2.04**	**1.08**	**2.58**	**1.72**
1 st.dev. confidence int.	**[1.91 - 6.06]**	**[1.25 - 2.89]**	**[1.12 - 5.71]**	**[0.97 - 2.08]**	**[1.52 - 6.22]**	**[1.28 - 2.57]**
Obs.	2528	2406	1203	1203	1200	1206
R-Sq.	0.47	0.38	0.36	0.08	0.36	0.05

Notes: standard errors in brackets; confidence interval for $\sigma 2$ obtained from 500 bootstrapping iterations

Column (1) All: All observations

Column (2) P1-99: Observations with growth rate below its 1st percentile or above its 99th percentile are excluded. Column (3) and (4): within the P1-99 group, subsample with K above and below its median value, respectively.

Column (5) and (6): within the P1-99 group, subsample with L above and below its median value, respectively.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

4.2 Sectoral-level results

We now turn to the sectoral data estimates. Table 2 reports the NLLS estimation. Because NLLS results appear less stable when small sector groupings are used, possibly because sample size falls, the NLLS estimations use broader sectors. In column (1), where the NLLS is estimated on the whole sample, the elasticity of substitution is found to be very high, but this regression suffers from too much heterogeneity in the sectors used (despite the use of sectoral fixed effects), and the homogeneity assumption for the CES parameters is almost certainly violated (see also the very low labour share coefficient α). In particular, it is known that the labour share varies across sectors, bottoming at 0.3-0.4 for the most capital-intensive sectors (mining, utilities) but exceeding 70 percent for several other sectors (hotels and restaurants; textiles) (Estrada and Valdeolivas, 2014). But the NLLS model assumes that this share, estimated by $\hat{\alpha}$, is constant across sectors. It is also possible that the elasticity of substitution between male and female employment σ_2 varies between agriculture, manufacturing, and services. Hence, we present in column (2) NLLS estimates for a panel comprising the services sector, in which employment is more balanced between genders.

Table 2: Non-linear model, sectoral level data, five-year growth rates

	(1) All sectors	(2) All sectors, excl. Agr., fishing, mining, manuf.	(3) All sectors, excl. Agr., fishing, mining, manuf., elect. and wholesale
No. of observations	513	395	324
α (labour share)	0.38***	0.46***	0.46***
	(0.04)	(0.05)	(0.05)
β (TFP growth)	0.02***	0.00	0.01
	(0.01)	(0.01)	(0.01)
δ (CES weight coef.)	0.30	0.68*	0.45
	(0.24)	(0.38)	(0.36)
ρ2 (1-1/σ2)	0.90	0.14	0.7
	(0.81)	(0.39)	(1.05)
σ2 (elasticity of substitution)	**10.00**	**1.16**	**3.33**
Median bootstrapping	***3.20***	***4.33***	***1.83***
1 st.dev confidence int.	***[0.7-9.3]***	***[2.5-12.0]***	***[0.9-3.8]***
No. of countries	32	32	32
No. of observations	513	395	324

Note: standard errors in brackets; confidence interval for $\sigma 2$ obtained from 500 bootstrapping iterations
P-values (not shown for $\sigma 2$): *** p<0.01, ** p<0.05, * p<0.1

The results with this sample suggest a labour share at 0.5, which is not unreasonable. The elasticity of substitution between male and female workers, σ_2 is then estimated at 1.16. However, the 16th-84th percentile (1 standard deviation) confidence interval for σ_2 remains large. The estimation in column (3) excludes two more sectors that differ from traditional services (electricity and wholesale trade). The results appear to be more robust. The median bootstrapping estimate of the elasticity of substitution between female and male employment is 1.8, and the 16th-84th percentile confidence interval is narrower, at [0.9-3.8].

As far as the bootstrapping simulations of the model presented in column (3) of Table 2, the mode of the distribution for σ_2 is 1, and σ_2 was only found to be higher than 4 in 15 percent of the simulations. Nevertheless, these simulations confirm that the elasticity of substitution is likely to be below 4, thus in the range where gender diversity effects are quantitatively relevant.

Heterogeneity in the ratio *F/M* is an obstacle to linear estimation, the more so for sectoral data since the ratio is highly variable across sectors. Hence, we group sectors into broader categories according to the ratio $F/(F+M)$. The results, available on request, are broadly consistent with the NLLS results above, suggesting relatively low estimates of the elasticity of substitution.

4.3 Aggregate-level results

Table 3 reports results using the macroeconomic panel and the linear fixed-effects model. Column (1) shows that the coefficient on growth for female labour supply is higher than that for male labour supply (this latter coefficient is also not statistically different from zero). Independently of whether δ is calibrated so as to match the lower working hours of women (ie $d = \delta^{1/\rho} = 0.87$) or not, the estimation implies a low elasticity of substitution between female and male employment (around 0.6), since the effect of adding female workers is stronger than the effect of adding male workers – if male and female were perfect substitutes, increasing FLPF by 1 percent would increase growth by less than increasing MLPF by 1 percent since FLFP is smaller than MLPF. The regressions results are robust to imposing constant returns to scale, ie $\beta_f + \beta_m + \beta_k = 1$ (column (2)). The dataset can be expanded by replacing the World Bank labour force data with OECD employment data; see column (3). This estimation leads to a higher elasticity of substitution, although when the IMF data is used for the capital stock and OECD data is used for GDP, the elasticity of substitution is found again to be below 1 (see column (4)).

Table 3: Linear model, aggregate level data, five-year growth rates

Variables	(1) Fixed effects (FE) WB data	(2) FE with CRS WB data	(3) FE OECD data	(4) FE OECD and alternate data
Female labour supply, WB (pc change)	0.649*** (2.795)	0.691*** (3.022)		
Male labour supply, WB (pc change)	0.531	0.225		
	(1.403)	(0.987)		
Capital stock, PWT (pc change)	0.0712	0.0838*	0.111**	
			(2.498)	
Female employment, OECD (pc change)	(1.486)	(1.812)	0.427**	0.328**
			(2.160)	(2.044)
Male employment, OECD (pc change)			0.497*	0.313
			(1.740)	
Capital stock, PPP, IMF (pc change)				(1.278) 0.308** (2.544)
				0.0708***
Constant	0.0863***	0.233***	0.101***	
	(6.713)	(5.793)	(8.208)	(4.827)
No. of observations	140	140	158	172
Number of countries	35		35	32
R-squared	0.331		0.313	0.218
Avg. female to male empl	0.77	0.77	0.77	0.77
σ (when $\delta^{\rho} = 0.87$)	0.67	0.26	1.62	0.90
σ (when $\delta^{\rho} = 1$)	0.57	0.19	2.40	0.85
σ (when $\delta^{\rho} = 1$), bootstrap 5th percentile	0.08	0.07	0.08	0.07
σ (when $\delta^{\rho} = 1$), bootstrap 95th percentile	6.06	4.08	5.57	4.61

Note: t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

Since these estimations are based on a small number of observations, caution is warranted in interpreting them. Although bootstrapping simulations confirm that the point estimate of the elasticity of substitution is low, estimation is imprecise. The confidence interval includes the existing estimates based on micro data, eg Acemoglu *et al* (2004), whose estimate is around 3.

The results for the NLLS approach are shown in Table 4. The estimate for σ_2, the elasticity of substitution between male and female labour force, is consistent with those of the linear estimation, varying between 0.2 and 0.6. Using World Bank data (column (1)), the NLLS model estimates for the other parameters of the production function are also reasonable. The estimated labour share of income is 0.82, on the high side.

Table 4: Non-linear model, aggregate level data, five-year growth rates

	(1) WB data	(2) OECD data (post 1995)	(3) OECD data (whole sample)
α (labour share)	0.83	0.84	0.79
	(0.05)	(0.05)	(0.04)
β (5-year growth in TFP)	0.10	0.11	0.11
	(0.01)	(0.01)	(0.01)
δ (CES weight coef.)	0.40	0.07	0.11
	(0.79)	(0.22)	(0.25)
ρ2 (1-1/σ2)	-0.70	-2.89	-2.70
	(3.33)	(4.60)	(3.41)
σ2 (elasticity of substitution)	**0.59**	**0.26**	**0.27**
Median bootstrapping	***0.51***	***0.26***	***0.29***
1 st.dev confidence int.	***[0.22 - 1.64]***	***[0.06-0.54]***	***[0.20-0.51]***
No. of observations	140	135	178
Proportion of runs converged	0.92	0.76	0.76

Notes: standard errors in brackets; confidence interval for σ2 obtained from 500 bootstrapping iterations

The 500 bootstrapping simulations corresponding to the results for the OECD data (column 3 of Table 4) show that the production function estimates are not very sensitive to the sample. The labour share and TFP growth coefficients are always close to the mean estimate, and the elasticity of substitution is most often found to be between 0.2 and 0.75. The elasticity of substitution is below 1 in 93.5 percent of the simulations.

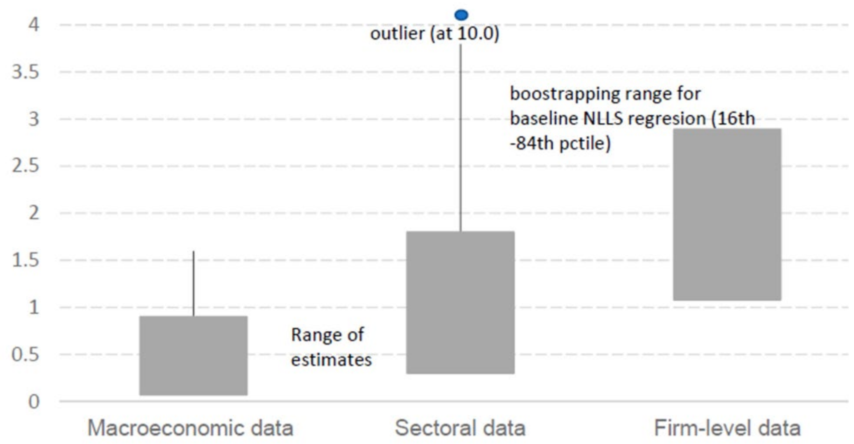
Figure 1 summarises the range of empirical results using the median estimates from the NLLS estimations.

5 Implications

We assess the implications of our finding of a relatively low elasticity of substitution, first with a calibration exercise, second with a growth accounting exercise applied to data over 1990–2019. We first calibrate a production function and simulate output for different values of σ_2, assuming that the gender gap in labour force participation is closed over time. Although it is difficult to tell whether closing completely the gender gap in employment is efficient, as this will depend in particular on gender differences in preferences and productivity, such an exercise corresponds often to stated policy objectives.

The production function is calibrated using a labour share of income of 60 percent, in line with the literature ($\alpha = 0.6$), and female employment is rescaled to reflect the lower working hours of women, which use of time statistics indicated to be around 17 percent lower than men, ie $\delta^{1/\rho} = 0.83$. Our calibration thus assumes no gender difference in human capital, though future research could investigate separately the role of gender-based differences in human capital and their consequences for economic growth. The exercise consists in quantifying the effect on GDP of increasing FLFP to the level of MLFP. The results are shown in blue in Figure 1, for different elasticities of substitution between female and male labour (along the x-axis), to assess the importance of the gender diversity effect, and comparing to a benchmark of a traditional headcount exercise with no gender differences accounted for in the measure of the labour force. This benchmark is shown in red, and it is obtained under the assumption of no gender diversity effect (perfect substitutability) and no 'scale' effect ($\delta = 1$), to also help us quantify the relative importance of differences in work hours and human capital.

Figure 1: Elasticity of substitution: range of estimates



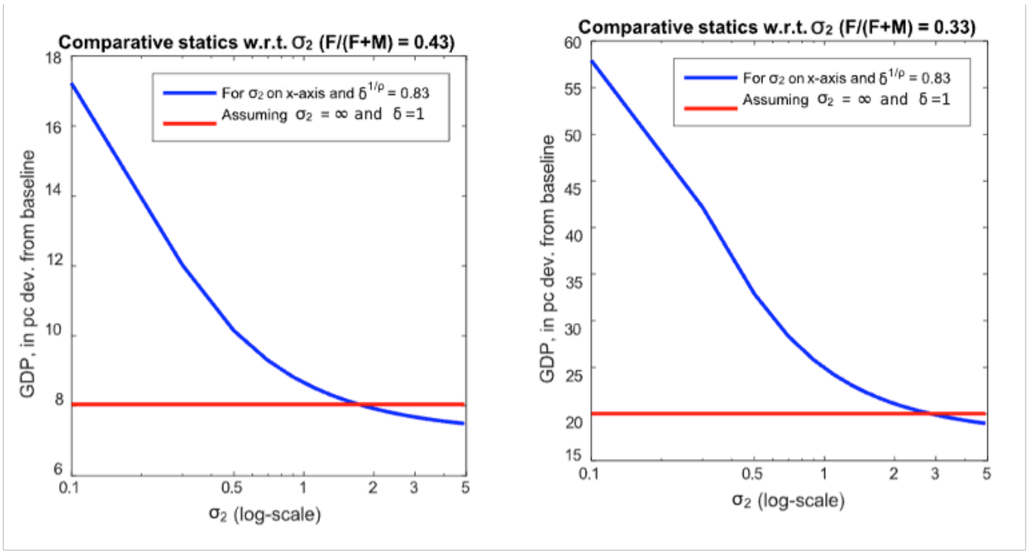
Source: author's calculations. Notes: range of estimates uses median estimate for NLLS

The LHS chart shows that, for countries with small gender gap in labour force participation ($F/(F+M) = 0.43$, the median value in the OECD), closing the gender gap would increase GDP by 8 to 17 percent, depending on the elasticity of substitution between female and male labour. Increasing FLFP affects GDP by increasing the overall labour force, although the 'scale' channel reduces GDP per worker since $\delta^{1/\rho} = 0.83$. When the gender diversity channel is sufficiently strong ($\sigma_2 < 2$, consistent with our estimations), the diversity effect more than outweighs the 'scale' channel. In that case, traditional growth accounting would underestimate the GDP gains due to increases in FLFP, ie the Solow residual estimated using traditional growth accounting misses the benefits of increasing FLFP. The RHS chart shows the results of the same numerical exercise for countries that start further away in terms of female labour force participation ($F/(F+M) = 0.33$), the median value for middle-income countries). The effect of aligning FLFP with MLFP would be much larger, around or above 20 percent of GDP. In addition, for these countries, the gender diversity effect dominates the scale effect even for an elasticity of substitution as high as 2.5.

Overall, the simulations confirm that the gender diversity effect is economically meaningful: for an elasticity of substitution between 0.2 and 2, complementarity effects would contribute to an overall increase in production of 1 to 6 percent of GDP, and this effect is even stronger for countries that start with a low level of FLFP. We also note that even the marginal product of male labour (and thus men's real wage, under the classical modelling of wages) should be increasing in FLFP when $\rho < \alpha$ (ie when $\sigma_2 < 2.5$, assuming $\alpha = 0.6$). Indeed, this is a case where the gender diversity effect of female participation outweighs the negative effect of a higher labour supply on capital intensity. In most of our estimations, the elasticity of substitution between female and male labour is below 2.5, which would imply that the positive effect of gender diversity on male wages could outweigh the negative effect of increased competition in the labour market.

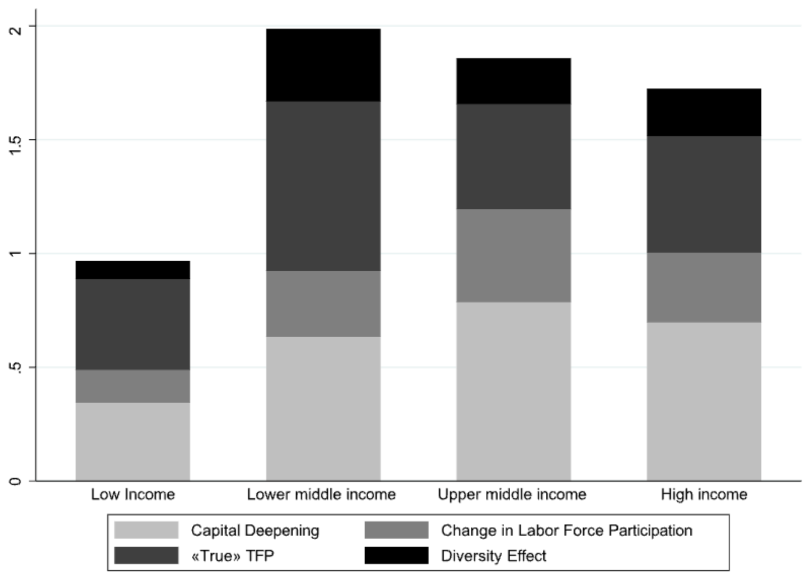
The second exercise is a standard growth accounting, using cross-country data from the Penn World Tables 10.01, over 1990-2019. The growth accounting uses the same parameters as the calibration exercise, as well as an elasticity of substitution of 0.5, to assess the effect of gender diversity on the mis-measurement of TFP growth. This calibration is in line with our macroeconomic estimates and it implies that the gender diversity effect outweighs the scale effect. Figure 2 shows the results of the growth accounting, averaged by income levels. Figure 3 focuses on the effect of gender diversity on TFP growth (equivalent to $\lambda(f - m)(\mu - F/N)$ as shown by the Taylor approximation. The exercise shows that for the average middle-income country, the gender diversity effect contributed some 0.2–0.3 percentage points to annual growth, a substantial part of the TFP growth obtained by standard growth accounting. The contribution is smaller in high-income countries, where gender gaps were smaller in 1990. In low-income countries, the exercise is likely to underestimate substantially the benefits of gender diversity because gaps in the labour force participation statistics are small, but other gender gaps (eg in legal rights, or in the quality of employment) are pervasive and they are also obstacles to growth.

Figure 2: CES production function – comparative statics



Source: author's calculations. Note: The blue locus shows the steady-state gains, in percent of GDP, obtained by increasing FLFP to the level of MLFP, for different values of the elasticity of substitution (x-axis) and different initial value of FLFP (LHS chart vs RHS chart). The red locus shows the same gains, under the assumption that both the gender diversity channel and the s cale channel (reflecting differences in work hours or human capital) are null, which would be the outcome of a traditional growth accounting exercise. When the blue locus is higher than the red locus, the gender diversity channel outweighs the scale channel and thus traditional growth accounting underestimates the benefits of increasing FLFP to close gender gaps in LFP.

Figure 3: Growth accounting over 1990-2019: contribution of diversity effect

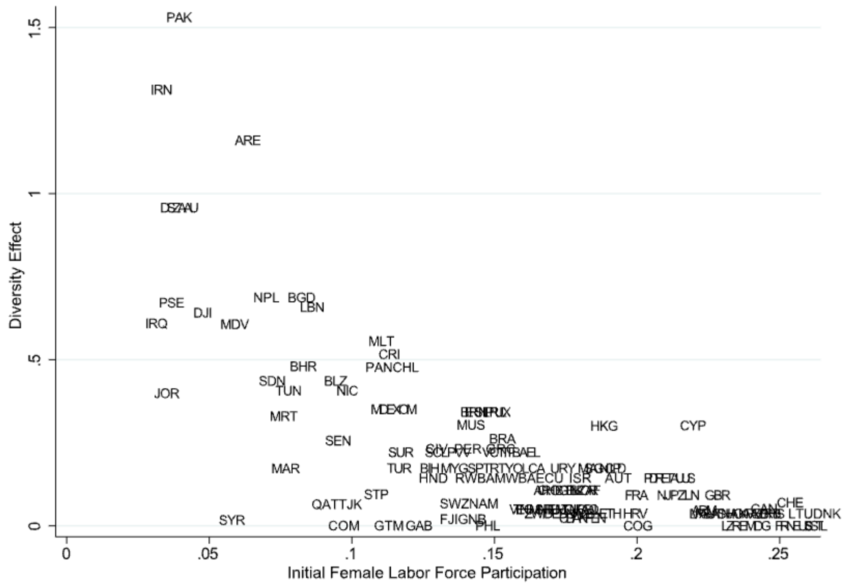


Source: PWT 10.01, World Development Indicators, and authors' calculations. Note: increases in FLFP and in MLFP also raise the total Labor Force Participation ratio.

Figure 4 shows that there is substantial heterogeneity in the effect of gender diversity on TFP. For some countries, the gender diversity effect of reducing LFP gaps may have contributed more than 1 percentage point to annual growth, although for most countries, the contribution is around 0.1–0.2

percentage points. There is also heterogeneity within income levels. For instance, the gender diversity effect would have been strongly positive in Panama, where the FLFP rate rose by around 12 percentage points in 30 years, but not in Turkey, where FLFP stayed flat, even though the two countries have now similar GDP per capita.

Figure 4: Growth accounting over 1990-2019: contribution of diversity effect



Source: PWT 10.01, World Development Indicators, and authors' calculations. Note: increases in FLFP and in MLFP also raise the total Labor Force Participation ratio.

6 Concluding remarks

In this paper, we estimated production functions at different levels of aggregation, with linear and non-linear methods, to gauge the extent to which the assumption of perfect substitutability between female and male employment, incorporated in many mainstream models, is validated in the data. We found, using firm-level, sectoral, and aggregate data, that the elasticity of substitution between male and female workers is low, between 2-3 in the firm-level data, in the range of 1-2 in the sectoral data, and below 1 in the aggregate data. Variation according to the degree of aggregation may reflect greater substitution within firms and sectors than across sectors (Espinoza *et al*, 2019).

We would highlight a couple of important implications of our findings for macroeconomics. The first is that closing the gender gap in the labour force would lead to larger increases in GDP (by 8 percent to 17 percent for the median advanced economy) than standard models suggest, and even to an increase in men's real wages: the complementarity between women and men in production, ignored by the previous literature, accounts for about one fourth to one third of the GDP gain according to our estimates(where we do not assume any inherent differences in the human capital of women and men). An important assumption underlying our finding that gains from gender inclusion are increasing in the complementarity between women and men is that the quantity of female labour actually in production may matter for GDP gains over and above the quantity of total labour. A second implication is that growth accounting exercises are misleading because they ignore the role that past reductions in gender inequality have had on TFP growth (and thus overstate the contribution of technology when the participation gap narrows). Until now, TFP growth has been interpreted as originating in technological improvements, but it should also be understood that worker diversity contributes to macroeconomic efficiency gains over time.

Our framework and our estimations signal that the aggregate production function has been mis-specified because one of its inputs, female labour, may have distinct effects on GDP that need to be properly measured. Because such a misspecification has profound implications for our understanding of the determinants of growth and of the effect of gender policies, it will be important that macroeconomic models disaggregate labour inputs by gender and calibrate the impacts appropriately. We hope that the empirical contributions in this paper facilitate such progress in the future.

References

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

An, Z., A. Kangur and C. Papageorgiou (2019) 'On the substitution of private and public capital in production', *European Economic Review* 118: 296–311

Espinoza R., J.D. Ostry, and C. Papageorgiou (2019) 'The Armistice of the Sexes: Gender Complementarities in the Production Function', *CEPR Discussion Paper* 13792

Estrada, A. and E. Valdeolivas (2014) 'The fall of the labor income share in advanced economies', *Essayos Economicos* 1(70): 17-81

Feenstra, R.C., Z. Li and M. Yu (2014) 'Exports and credit constraints under incomplete information: Theory and evidence from China', *Review of Economics and Statistics* 96(4): 729-744

Freeman, R.B. (1979) 'The effect of demographic factors on age-earnings profiles', *Journal of Human Resources* 14(3): 289–318

Goldin, C. (1990) *Understanding the gender gap: An economic history of American women*, Oxford University Press

Goldin, C. and L.F. Katz (2002) 'The power of the pill: Oral contraceptives and women's career and marriage decisions', *Journal of political Economy* 110(4): 730–770

Hsieh, C.-T. and P.J. Klenow (2009) 'Misallocation and manufacturing TFP in China and India', *The Quarterly Journal of Economics* 124(4): 1403–1448

IMF (2015) 'Estimating public, private, and PPP capital stocks', appendix to IMF Board paper *Making public investment more efficient*, International Monetary Fund

Lagarde, C. and J.D. Ostry, (2018) 'The macroeconomic benefits of gender diversity', *VoxEU Column*, 5 December, Centre for Economic Policy Research

Ostry, J.D., J. Alvarez, R. Espinoza and C. Papageorgiou (2018) 'Economic gains from gender inclusion: New mechanisms, new evidence', IMF Staff Discussion Note 18/06, International Monetary Fund

Pew Research Center (2020) Women Make Gains in the Workplace Amid a Rising Demand for Skilled Workers, *Technical Report*, Pew Research Center

World Bank (2012) *2012 World Development Report on Gender equality and development*, World Bank

bruegel

Bruegel, Rue de la Charité 33, B-1210 Brussels
(+32) 2 227 4210
info@bruegel.org
www.bruegel.org