ORIGINAL PAPER

The Feminisation U, cultural norms, and the plough

Luca J. Uberti[1] · Elodie Douarin[2]

Abstract

The Feminisation U describes the tendency of female labour force participation (FLFP) to first decline and then rise in the process of economic development. While the Feminisation U is often presented as a 'stylised fact' of development, empirical support for it is mixed. Here, we show that cultural norms inherited from ancestral plough use exert a moderating influence on the shape of the Feminisation U. Specifically, we find a significantly U-shaped path of FLFP only in countries whose ancestors employed a plough-based agricultural technology. The shape of the U-curve becomes progressively more muted as the share of a country's ancestors that practiced plough agriculture decreases. In countries with little or no legacy of historical plough use, the time path of FLFP is effectively flat. This pattern of results is robust to correcting for dynamic panel bias, instrumenting for per-capita income, and controlling for other potential effect modifiers. Our findings are compatible with a nuanced reading of the main theoretical models proposed in the literature to explain the Feminisation U.

Keywords Female labour force participation · Economic development · Agricultural history · Cultural norms

JEL Classification J16 · J22 · O11

✉ Elodie Douarin
e.douarin@ucl.ac.uk

Luca J. Uberti
luca.uberti@uni.lu

[1] Department of Economics and Management, University of Luxembourg, Luxembourg, Luxembourg

[2] School of Slavonic and East European Studies, University College London, London, UK

🖄 Springer

1 Introduction

A long line of research suggests a U-shaped relationship between female labour force participation (FLFP) and economic development (Sinha 1967; Goldin 1995; Tam 2011): while the early stages of economic growth are accompanied by a de-feminisation of the labour force, women tend to become economically active again as income rises further. Figure 1 documents this pattern in a pooled cross-section of 172 countries during 1990–2013. Some of the classic contributions on the subject link this pattern to structural change: while industrialisation leads women to exit the labour force, the transition from an industry- to a service-based economy induces a re-entry (Boserup 1970; Goldin 1995). In other models, the Feminisation U is produced by changes in fertility (Galor and Weil 1996; Lagerlof 2003) and the gender education gap (Hiller 2014), both of which tend to first increase and then decline with rising income per capita.

Despite the status of the Feminisation U as a 'stylised fact' of development economics, recent empirical contributions have cast doubt on its veracity. Gaddis and Klasen (2014) argue that empirical support for a U-shaped pattern is feeble, showing that the estimated relationship vanishes under dynamic panel-data estimations. Sub-national studies have generally produced mixed results (Roncolato 2016; Lahoti and Swaminathan 2016), while some cross-country regressions provide evidence of an inverted U-shaped relationship (Çağatay and Özler 1995).

In this paper, we re-engage with the empirical controversy surrounding the Feminisation U by exploring variation in the dynamic path of FLFP across countries. We argue that the mechanisms that generate the theoretical U-curve—structural change, fertility dynamics, and gender differences in education—depend critically on initial conditions. In particular, these mechanisms play out more strongly in societies that espouse less gender-equal cultural norms when they embark on the process of modern economic development.

Income per capita begins to rise above subsistence levels when an economy emerges from the Malthusian trap (Galor and Weil 2000). If cultural norms at this point emphasise the role of women as homemakers, industrialisation is more likely to be accompanied by a withdrawal of women from the labour market; women will bear more of the brunt of childcare as fertility rises; and they will likely be prevented from accessing new educational opportunities on an equal footing with men. In these societies, the path of FLFP is strongly U-shaped. When initial cultural norms assign a more equal role to men and women, by contrast, the mechanisms that produce the U-curve are more muted: industrialisation does not completely deter women from seeking and obtaining manufacturing jobs; a rise in fertility is accompanied by a distribution of care labour that is less biased towards women; and new educational opportunities are less likely to be monopolised by men. As a result, fewer women quit the labour force as the economy travels through the middle-income band, leading to a less distinctly U-shaped path of FLFP.

Thus, the dynamic path of FLFP depends critically on the cultural norms that prevail when the economy enters a post-Malthusian growth regime. Here, we test

this proposition empirically. To measure pre-historical gender norms, we rely on the seminal work of Boserup (1970) and Alesina et al. (2013). These authors link the emergence of cultural norms about the appropriate role of women in the economy to the form of traditional agriculture practiced in pre-industrial society, when the economy is in the Malthusian regime and income per capita is constant. The key contrast is between plough-based agriculture and forms of shifting cultivation that rely on lighter, hand-held tools such as the hoe or the digging stick. Unlike the latter, the plough requires significant upper body strength, putting a productivity premium on male labour. Thus, ancestral societies with a plough-based agricultural technology developed an early specialisation of labour along gender lines. Over time, this division of labour facilitated the emergence of cultural norms that cast women as 'natural' home-makers.

By linking the plough argument to the mechanisms that generate the Feminisation U, we suggest that the legacy of ancestral plough use may be exerting a moderating influence on the dynamics of FLFP in the course of development. In societies that traditionally practiced plough agriculture, the gender norms that prevail at the transition to a post-Malthusian regime tend to disadvantage women. As a result, the path of FLFP is likely to be more strongly U-shaped as income per capita rises. The less a society's ancestors relied on a plough-based technology, the more equal are the norms that prevail when per-capita income begins to rise above subsistence, leading to a less markedly U-shaped path of FLFP.

To model the plough as an effect modifier, we allow the parameters of the U-curve to depend linearly on Alesina et al.'s (2013) measure of ancestral plough use, leading to a specification with interaction terms. Based on a global panel dataset covering 169 countries during 1990–2013, we present stylised facts that are consistent with initial cultural norms exerting a modifying influence on the dynamics of FLFP.[1] The pattern of results found in the data is extremely robust. A significant contrast between the dynamic paths of FLFP in plough vs. non-plough societies is observed in specifications that correct for dynamic panel bias (e.g. GMM) and treat economic development and historical plough use as endogenous to FLFP. We also consider other historical events (e.g. the Neolithic revolution) as potential correlates of initial gender-role cultural norms and hence as potential effect modifiers. In all cases, we find no evidence that these factors exert a significant effect-modifying influence once the time path of FLFP is allowed to depend on historical plough use. On the contrary, the observed heterogeneity in the shape of the Feminisation U appears to be robustly linked to the legacy of historical plough adoption only.

Our findings contribute to the literature in at least three ways. Recent studies have examined the historical emergence (Alesina et al. 2013; Hansen et al. 2015), transmission (Fernandez and Fogli 2009), persistence (Grosjean and Khattar 2019), and transformation (Fernandez 2013) of the cultural norms that assign

---

[1] We acknowledge that while it would be fascinating to test these patterns, stemming from differences when economies exited the Malthusian regime, on data on earlier periods, this is prevented by lack of data. We are indeed here making use of all the data currently available and based on consistent definitions throughout.
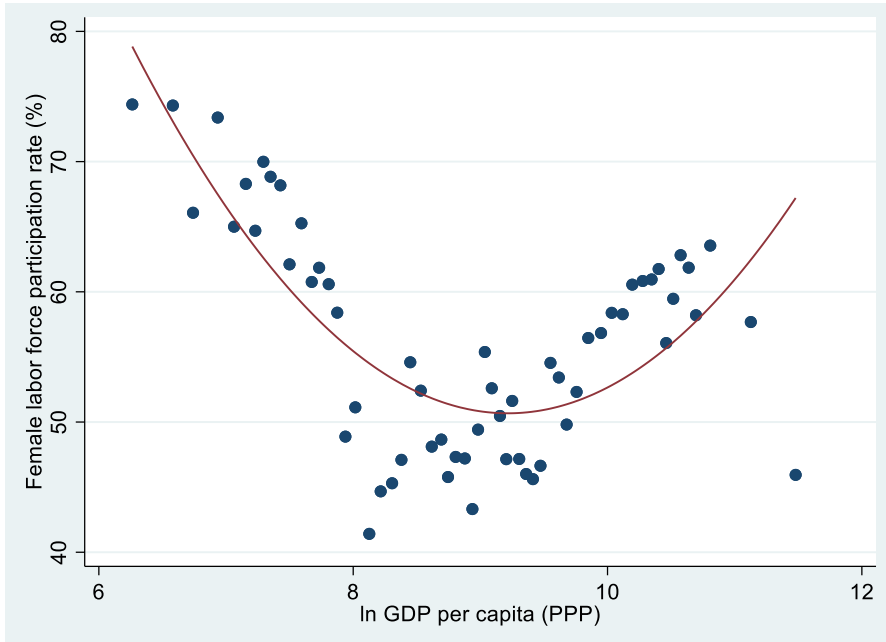
Fig. 1 FLFP and income per capita: binned scatterplot (1990–2013). Note: the diagram groups the observations into 60 equal-size bins (based on income per capita) and plots the means of income per capita and FLFP for the observations within each bin. The data refer to 169 countries during 1990–2013, for a total of 3,891 observations

different socio-economic roles to men and women. Yet, for the most part, this work has developed in isolation from the Feminisation U literature. Thus, our first contribution is to integrate two strands of research that have developed on parallel tracks.

Second, we suggest a possible way of solving the ongoing empirical controversy surrounding the Feminisation U: empirical studies may fail to observe a U-shaped relationship if they focus on 'non-plough' countries or if 'non-plough' settings dominate the sample. To our best knowledge, this is the first paper that investigates a potential source of heterogeneity in the relationship between economic development and FLFP.

Third, our results point to an important nuancing of Alesina et al.'s (2013) argument. Their seminal analysis provides estimates of the *average* effects of historical plough use on labour market outcomes today, providing broad evidence of the role of the plough in shaping gender roles. We nuance their results by revealing substantial heterogeneities along the income path. In particular, the detrimental effect of plough legacies is only observed in middle-income economies. Our findings imply that the labour-supply effects of the plough shock are neither immediate nor permanent: a substantial reduction in FLFP does not appear until a plough-based economy has reached a middling level of income, and this reduction is later reversed completely as more advanced levels of economic development are achieved. We think

that this qualification reflects more accurately Boserup's (1970) original formulation of the 'plough thesis'.

2 The Feminisation U and initial conditions: bringing two strands of literature together

We begin by reviewing the main theoretical mechanisms that are known to lead to a Feminisation U—structural change, fertility dynamics, and gendered educational choices. In each case, we highlight how the prediction of a U-shaped path of FLFP depends on assumptions about initial conditions—in particular, the gender norms that prevail when the growth rate of per-capita income begins to rise above zero. This framework motivates the subsequent empirical analysis.

2.1 Structural change

Early explanations of the Feminisation U placed an emphasis on sectoral shifts in production and employment (Boserup 1970; Goldin 1995). Initially, economic growth shifts the locus of production from family farms and home workshops to factories, firms, and other places of wage work. This transition induces women to exit the labour force. For one thing, the physical separation between home and workplace makes it more difficult to reconcile productive and reproductive tasks (Benerìa 1979).[2] For another, work in factories and industrial farms is generally considered 'dirty'. A married woman engaged in paid manual labour outside the home brings 'stigma' to the family, lowering the household's utility (Boserup 1970). 'This stigma is a simple message: only a husband who is lazy, indolent, and entirely negligent of his family would allow his wife to do such labour' (Goldin 1995: 71). As per-capita income rises further, however, paid jobs that do not mark women with a stigma, such as service-sector jobs, become more widely available, and women re-enter the labour force.

Ngai and Olivetti (2015) formalise this argument, showing that a model of structural transformation with home/market production and gender-specific comparative advantages implies a U-shaped female labour supply. Ngai and Petrongolo (2017) and Rendall (2017) examine alternative models, putting greater emphasis on the rise of service-sector jobs in generating an increase in female participation.

In these models, the initial fall in FLFP comes from the emergence of occupations to which *stigma* is attached, while the subsequent rise is due to an increased

---

[2] Reproductive tasks include childrearing, care of the elderly, and housework.

availability of *non-stigma* jobs. Thus, the U-pattern is implicitly sustained by the cultural norms that prescribe the appropriate role of women in the economy.[3] Indeed, Goldin's (1995) original toy model came in two variants—with and without social stigma. In the 'non-stigma case', Goldin shows that female labour supply is not necessarily U-shaped.[4] Here, we link the cultural norms that prevail in the early stages of growth of per-capita income to the legacy of ancestral plough use.

2.2 Fertility dynamics

Other models reproduce a U-shaped path of FLFP by focusing on the role played by fertility dynamics in the demographic transition. In a seminal paper building on Becker (1960), Galor and Weil (1996) consider a growth model with gender heterogeneity and endogenous fertility. The relative wages of men and women have income and substitution effects on fertility and labour supply decisions. Both genders are equally endowed with mental human capital ('brains'). Men, however, have more 'brawn' and hence a comparative advantage in labour-intensive tasks. In low-income economies with a labour-intensive technology, economic growth raises the male relative wage. The resulting income effect increases the demand for children, reducing FLFP. At a higher level of income, the technology is less brawn-intensive, and economic growth has a positive effect on the female wage, closing the gender wage gap. At this point, women substitute out of childrearing and into market work. As Galor and Weil acknowledge (1996), simple extensions of their model can generate a non-monotonic U-shaped relationship between per-capita income and FLFP (see also Lagerlof 2003; Bloom et al., 2009; Kimura and Yasui 2010).

Implicitly, these models are premised on a specific set of gender norms. Although Galor and Weil 'do not assume that women are better at raising children than are men' (1996: 378), they nonetheless assume that as a matter of fact 'all childrearing is done by women' (1996: 375). Accordingly, the time (opportunity) cost of children is an increasing function of the female wage only, implying a pure income effect from a rise in the male wage. In some societies, however, cultural norms are more gender-equal, and men and women contribute more equally to childrearing. Here, the opportunity cost of having children is also a function of the male wage. A rising demand for children is partly offset by the opportunity cost of children, and the dynamic path of FLFP is less strongly U-shaped. Again, here we link the cultural norms affecting the gender allocation of care labour to the legacy of the plough.

---

[3] We note that Ngai and Petrongolo (2017) rely on sector-specific 'female-intensity' weights to capture gender comparative advantages, linking the U-curve to the process of structural change in economic development. Yet, these weights are derived from (or calibrated using) survey data reflecting revealed preferences, which are likely to be contaminated by social norms regarding gender roles.

[4] Of course, to the extent that factory work is brawn-intensive, however, female labour supply may be U-shaped in the non-stigma case, too.

2.3 The gender education gap

A final argument, first proposed by Boserup (1970), links the Feminisation U to men's privileged access to education and technological knowledge. Although there is no biological constraint on women and men attaining equal quantities of 'brains', gender-biased cultural norms favour the education of boys. This mechanism is examined formally by Hiller (2014), who considers a two-sex, overlapping generations model of the gender education gap in the course of development.[5] The household's utility is derived from both wages (returns to education) and status. The education of daughters negatively impacts status if cultural norms favour women's specialisation in housework. Parents maximise the household's utility by choosing how much to invest in their children's education. As in Fernandez (2013), cultural norms (and hence FLFP) are determined endogenously.

In this model, FLFP follows a U-shaped path. Yet, the model's dynamics depend critically on the initial difference in productivity between uneducated men and women. This difference leads to a cultural bias in the allocation of the household's education budget towards boys. When male and female labour are perfect substitutes, the allocation of the household's education budget is gender-neutral. As soon as the gender gap in labour productivity is even slightly larger than zero, however, 'there is room for an inegalitarian equilibrium [...]. Other things being equal, a higher [initial gender productivity differential] indicates a greater likelihood that an economy will fall into a gender-inequality trap', producing a more markedly U-shaped path of FLFP (Hiller 2014: 474).

In the model, the two genders are characterised by biological differences in physical strength. Whether these differences are economically salient, giving rise to a *productivity* differential depends on the extent to which the pre-industrial production technology relies on physically demanding labour (Hiller 2014). The more this technology is brawn-intensive—as is the case in plough-based agricultural systems—the greater the initial male advantage in productivity. This sets the stage for an education bias and a more distinctly U-shaped path of FLFP.[6]

---

[5] Other authors have modelled the dynamic relationship between cultural change and FLFP. Hazan and Maoz (2002) propose a simple model of labour supply with an intergenerational transmission of preferences. Without assuming any productivity or wage differentials between men and women, the model generates an S-shaped path of FLFP. Similarly, Fernandez examines cultural change as a 'rational, intergenerational learning process in which individuals are endogenously learning about married women's long-run payoff from working' (2013: 473). In a calibration exercise, she shows that this cultural learning mechanism is an important driver of the large increase in the proportion of working married women recorded in the USA after 1970. Fernandez (2013), however, focuses entirely on the upward-sloping branch of the curve, which is found to have the S-shape of a cumulative normal distribution function. Interestingly, if a woman's disutility of working is low (because no cultural norm is violated), Fernandez's (2013, 492) model predicts a much flatter (less S-shaped) path of FLFP.

[6] In an application to US data after WW2, Rendall (2017) uses a taxonomy of jobs to identify the extent to which they are brawn- or brain-intensive. She is then able to reproduce part of the closure of the gender gap in education through the rise of brain (i.e. more female friendly) jobs since the 1950s. Then she introduces a linearly diminishing parameter for female discrimination in the labour market, which combines with technological change and contributes to explaining even more of the education gap. She does not however explicitly model a situation where the discrimination against women in the labour market would in fact inflate the belief that women can only work brain jobs.

3 The empirical literature on the Feminisation U

Despite its strong theoretical rationale, the Feminisation U is supported by a mixed body of evidence. Early findings based on cross-country regressions were suggestive of a U-shaped relationship (Sinha 1967; Pampel and Tanaka 1986). Goldin's (1995) cross-sectional evidence from 100 countries in 1985 is also consistent with a U pattern in which female labour supply reaches its lowest point in countries with a per-capita GDP of around 3,000 US$ (1985). Based on a pooled cross-section of 193 countries in 1980 and 1990, however, Çağatay and Özler (1995) find statistically significant evidence of an *inverted* U-shaped relationship.[7]

Early studies also reported time-series evidence based on individual countries. Focusing on the USA, Goldin (1995) argues that FLFP probably traced out a U-shaped pattern over time, reaching a bottom in the 1920s, when per-capita GDP in the USA was around 5,000 US$ (PPP). Similar findings for England and France are documented by Tilly and Scott (1987).

As already noted by Durand (1975), the cross-sectional relationship may be biased, as the omission of unobserved country-level heterogeneities (e.g. cultural differences) may give rise to a Kuznets-type fallacy (Tam 2011). Meanwhile, results based on time series from individual countries may have limited external validity. For these reasons, the most recent work on the Feminisation U has turned to panel-data methods. Based on data from 90 countries during 1970–1985, Mammen and Paxson (2000) are the first to exploit variation within countries over time to identify the shape of the Feminisation U. Their fixed-effects panel estimates reveal a more muted U-shape than the corresponding pooled model, with a much lower (but statistically significant) turning point at 1,600 US$ per capita, as compared to 2,550 US$ in their pooled cross-section.

Luci (2009) and Tam (2011) estimate both static (OLS) and dynamic (GMM) panel-data specifications, confirming more formally that the Feminisation U shows up as an intertemporal relationship. Their models, however, do not control for potential time-varying confounders, nor do they address the potential endogeneity of income levels to female labour supply (although the GMM estimators they employ would allow them to instrument for per-capita GDP).

In a noteworthy contribution, Gaddis and Klasen (2014) employ more recent and comprehensive labour market data than either Luci (2009) or Tam (2011). They also use dynamic panel-data estimation techniques (GMM) and internal instruments to correct for the endogeneity of GDP per capita in 'GMM style'. In a break from most of the previous literature, they conclude that 'there is no clear evidence for the feminization U hypothesis from […] dynamic estimations' (Gaddis and Klasen, 2014: 660).

Gaddis and Klasen's (2014) findings are in line with recent studies that exploit variation in the level of development across sub-national units within countries. Using Indian state-level data spanning the period 1983–2012, Lahoti and

---

[7] The authors mistakenly claim that their findings are consistent with the Feminisation U hypothesis.

Swaminathan ([2016]) estimate dynamic GMM models and find no systematic evidence of a U-shaped relationship between state-level income and FLFP. Similarly, Roncolato's ([2016]) study of South Africa employs micro data to investigate the effect of municipality-level income on a woman's probability of being in the labour force, concluding that in South Africa, the U-shaped relationship is more ambiguous than implied by theory.

Overall, these recent findings are mixed, and they strengthen the conclusion reached by Humphries and Sarasua ([2012]), based on their review of historical evidence from a number of now developed economies, that the Feminisation U cannot be assumed to hold universally.

A key limitation of all existing empirical studies is that they do not systematically investigate the potential heterogeneity of female labour supply dynamics across different contexts. Such an investigation could help reconcile some of the evidence summarised here, but also shed new light into the mechanisms driving the Feminisation U. In particular, we are not aware of any study examining how the cultural norms shaped by historical events from the distant past (e.g. the adoption of the plough) exert a modifying influence on the dynamic path of FLFP. The analysis that follows addresses this missing link in the empirical literature.

4 Empirical strategy

4.1 Data

To measure labour force participation (FLFP), we use the (most recent) sixth revision of the International Labour Organisation (ILO)'s *Estimates and Projections of the Economically Active Population* (EAPEP) database. This data, which was downloaded from the World Bank's website, is based on ILO staff estimates. FLFP is defined as the number of women in the labour force as a share of the total working-age (15–64) female population. In the System of National Accounts (SNA), the labour force is composed of all individuals who are working or seeking work. Any kind of employment for pay is included, as is self-employment and unpaid work that is performed to produce a *good* (as opposed to a service) auto-consumption (Klasen [2018]).

On this definition, subsistence farming counts as labour force participation. By contrast, housework and care work (e.g. childrearing and the care of the elderly) do not count as market work. This definition is consistent with our theoretical framework: women employed on the family farm or in household industries are classified as supplying their labour to market, while women engaged in childrearing or other forms of reproductive labour are considered to be economically inactive.

The EAPEP's sixth edition covers the period 1990–2019.[8] In 2013, however, the 19th Conference of Labour Statisticians, convened by the ILO, modified the

---

[8] Previous empirical studies on the Feminisation U (Luci [2009]; Tam [2011]) used data from earlier editions of the EAPEP, which provide greater coverage along the time dimension. Yet, the data based on these earlier editions are now considered to be unreliable (Gaddis and Klasen [2014]).

definitions of work and labour force, leading to a break in the ILO time series on labour force participation (Klasen 2018). Thus, to obtain estimates based on consistent definitions, we restrict our analysis to the period 1990–2013. In the sample available for estimation, which covers 169 countries, *FLFP* ranges between 10.1 (Jordan in 1991) and 91.5 percent (Burundi in 1991), with a mean (median) value of 56.3 (58.7) percent and a standard deviation of 17.5.

To measure the historical adoption of the plough (and the cultural norms it brought about), we use Alesina et al.'s (2013) original variable—*Plough*. The authors use comparable ethnographic information to construct 'estimates of the *fraction* of the population currently living in a […] country with ancestors that traditionally engaged in plough agriculture' (2013: 486).[9] *Plough* is a continuous variable ranging between 0 and 1, with a mean (median) value of 0.52 (0.73) and a standard deviation of 0.47. In 55.5 percent of the countries entering the estimation sample, either all (26.5 percent) or none (29.5 percent) of their present-day inhabitants had plough-using ancestors. Only 5.4 percent of countries in the sample have a value of *Plough* between 0.2 and 0.8.

An important limitation of Alesina et al. (2013)'s plough variable, which the authors acknowledge, is that it does not provide information on the exact timing of plough adoption. Yet, the introduction and diffusion of the plough certainly occurred well before the transition to a post-Malthusian growth regime—that is, when per-capita income starts to rise above subsistence levels and the economy begins to travel down the U-curve.[10] For this reason, Alesina et al.'s (2013) variable can be taken as an appropriate measure of initial conditions.

The data on income per capita, expressed in constant 2011 international dollars, are taken from the World Bank's *World Development Indicators* and are PPP-adjusted. The mean (median) value of $\ln GDPpc$ is 8.9 (9); and the standard deviation is 1.3, which provides us with sufficient variation to identify changes in FLFP across levels of income.

4.2 Model specification

We explore the heterogeneity in the dynamic path of FLFP in a panel-data framework. We model the relationship between FLFP and economic development as a quadratic function of (the log of) per-capita GDP—the standard approach in the literature. In contrast to previous studies, however, we allow the coefficients of the quadratic function to depend linearly on Alesina et al.'s (2013) *Plough* variable, leading to a specification with interaction terms:

$$FLFP_{it} = \rho FLPF_{it-1} + \beta_1 \ln GDPpc_{it} + \beta_2 \ln GDPpc_{it}^2 + \gamma_1 Plough_i \times \ln GDPpc_{it} + \gamma_2 Plough_i \times \ln GDPpc_{it}^2 + \varphi X_{it} + \sigma_i + \tau_t + \varepsilon_{it}$$

(1)

where *i* indexes countries and *t* time.[11]

---

[9] Emphasis added.

[10] On the history of the plough, see Fussell (1966) and Andersen et al. (2016).

[11] If the quadratic function is $\alpha_1 \ln GDPpc_{it} + \alpha_2 \ln GDPpc_{it}^2$, Eq. (1) is obtained by assuming that $\alpha_1 = \beta_1 + \gamma_1 Plough_i$ and $\alpha_2 = \beta_2 + \gamma_2 Plough_i$.

In line with previous contributions (Tam 2011; Gaddis and Klasen 2014), Eq. (1) is a dynamic specification in which the outcome variable is allowed to depend on its own previous period realisation. Substantively, the first lag of *FLFP* models the persistence of the cultural norms that shape and constrain women's labour supply decisions. A dynamic specification is motivated by the fact that cultural attitudes regarding married women's work are passed down from one generation to the next (Fernandez et al. 2004; Fernandez and Fogli 2009). As such, they have a strong tendency to persist unchanged over time (Grosjean and Khattar 2019). Statistically, the inclusion of a lagged dependent variable removes any serial correlation from the error term, leading to consistent panel-data estimates.

Our specification includes country-fixed effects (σ_i) which flexibly absorb all time-invariant country-level factors, including the main effect of historical plough use. Thus, Eq. (1) allows us to estimate the (heterogeneous) relationship between FLFP and per-capita income based on within-country variation only, comparing the average time path of FLFP in 'plough' vs. 'non-plough settings'.

τ_t denotes a full set of time-period dummies that capture labour market shocks affecting all countries simultaneously—e.g. global economic crises and business cycle fluctuations—as well as any global trend in gender equality.[12] In addition, the inclusion of τ_t prevents the idiosyncratic disturbances (ε_{it}) from being contemporaneously correlated across individuals, which would bias the variance estimator (Roodman 2009). In additional models, we also consider richer specifications of the time-trend component.

Furthermore, we also present specifications that condition the estimated relationship between FLFP and per-capita GDP on a set of *time-varying* observables. We choose these control variables based on a review of the literature on the determinants of female labour supply. In particular, we include a measure of armed conflict, which may create labour shortages and new labour-market opportunities for women (Goldin and Olivetti 2013): a country's dependence on oil exports, which have been shown to crowd out female-intensive tradable sectors (Ross 2008); a measure of aid dependence, on the assumption that aid agencies may favour male-biased technology transfer, feeding negative gender stereotypes (Boserup 1970; Jaquette and Summerfield 2006); an index capturing the extent to which a country's society and culture are globalised and hence exposed to the diffusion of new values and attitudes, some of which may concern gender roles (Potrafke and Ursprung 2012); and a measure of the quality of democracy (Beer 2009). A detailed description of these variables and of the associated data sources is provided in Appendix Table 7.

The parameters of interest in Eq. (1) are β_1, β_2, γ_1, and γ_2. The β's define the average curvature of the Feminisation U in countries in which *Plough* = 0, while the γ's measure the *difference* between the curvatures observed in countries with and without historical plough use. In post-estimation, we also compute the parameters of the Feminisation U in countries in which *Plough* = 1—that is, $\beta_1 + \gamma_1$ and $\beta_2 + \gamma_2$.

---

[12] See, for instance, Albanesi and Sahin (2018).

5 Main results

To explore the patterns in the data, we first present OLS estimates of Eq. (1). The results are summarised in Table 1. Column 1 reports the estimated coefficients of a pooled model. The specification in column 2 adds country (σ_i) and year (τ_t) fixed effects. This and all the following specifications use only time variation within countries to estimate the relationship between FLFP and economic development. In column 3, we enrich the specification by replacing the year fixed effects (which enter as jointly insignificant) with a full set of year×continent fixed effects, allowing for fully flexible trends in FLFP at the continent level. Column 4 presents an alternative specification that conditions the estimates on country-specific linear trends ($\sigma_i t$). Here, the functional form of the time trend is more restrictive, but each country is allowed to have a different trend in FLFP. In both models 3 and 4, the time trends are highly statistically significant.[13] Next, the specification in column 5 includes the set of time-varying observables discussed in Sect. 4.2. Lastly, column 6 reports a benchmark specification without $Plough_i \times \ln GDPpc_{it}$ and $Plough_i \times \ln GDPpc_{it}^2$, as in previous empirical studies of the Feminisation U. Due to data limitations, the sample available to estimate models 5 and 6 is substantially smaller.

Throughout models 1–5, the coefficient on the lagged dependent variable is close to (but always statistically distinguishable from) one, suggesting a high degree of persistence in the gender norms that shape and constrain women's decision to supply their labour to market. The estimates of β_1 and β_2 are always statistically insignificant, both individually *and* jointly, implying that when $Plough = 0$, the relationship between per-capita GDP and FLFP is statistically indistinguishable from a flat line.

The estimates of γ_1 and γ_2, by contrast, are always statistically significant (individually and jointly) at conventional levels, including in the more 'demanding' specifications that net out the influence of time trends in FLFP (columns 3–5). This finding indicates that in 'plough' countries, the expected relationship between GDP per capita and FLFP is significantly *different* from the expected relationship in 'non-plough' settings. In some specifications (for instance, model 3), the difference is only significant at the 10% level. This is arguably a consequence of multicollinearity. Indeed, in more parsimonious specifications that omit either $\ln GDPpc_{it}$ or $\ln GDPpc_{it}^2$, $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ are similar in magnitude to the estimates presented in Table 1 (column 3) but much more precisely estimated (see Appendix Table 8 columns 1 and 2, for the full results).

In panel B, we also present estimates of the U-curve's parameters in 'plough' countries ($Plough = 1$). Throughout models 1–5, the relationship between GDP per capita and FLFP is consistent with a U-shaped pattern, with a positive and significant coefficient on the quadratic term of log GDP per capita and a negative and significant coefficient on the linear term. As noted by Lind and Mehlum (2010, 111), 'the requirement for a U shape is that the slope of the curve is [significantly] negative at the start and [significantly] positive at the end of a reasonably chosen interval of x-[$\ln GDPpc$] values' – say, [6, 11]. In models 1–5, this is always the case if $Plough_i = 1$, unequivocally indicating a non-monotonic relationship but never the case if $Plough_i = 0$.[14] In addition, we find that, in 'plough' countries, the 98 percent

---

[13] Coefficients not reported.

[14] Results not reported in full.

Table 1 Determinants of female labour force participation (1991–2013), OLS

	Pooled	FE	Trends	Trends	Controls	Controls
Dependent variable: FLFP$_t$	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Regression results						
FLFP$_{(t-1)}$	0.992***	0.935***	0.932***	0.802***	0.787***	0.790***
	(0.001)	(0.009)	(0.010)	(0.017)	(0.022)	(0.023)
Plough	4.156					
	(2.636)					
ln GDP pc [β_1]	0.006	−0.951	−0.932	1.465	1.100	−5.127***
	(0.433)	(0.744)	(0.778)	(1.438)	(2.394)	(1.866)
ln GDP pc^2 [β_2]	0.002	0.047	0.046	−0.117	−0.104	0.292**
	(0.026)	(0.042)	(0.044)	(0.099)	(0.158)	(0.114)
Plough × ln GDP pc [γ_1]	−1.102*	−3.312*	−3.200*	−8.039**	−13.007***	
	(0.606)	(1.861)	(1.992)	(4.000)	(4.252)	
Plough × ln GDP pc^2 [γ_2]	0.067*	0.204**	0.188*	0.502**	0.789***	
	(0.034)	(0.103)	(0.113)	(0.233)	(0.249)	
Country FE	No	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	No	No	No
Year x continent FE	No	No	Yes	No	No	No
Country FE x linear trend	No	No	No	Yes	Yes	Yes
Time-varying controls	No	No	No	No	Yes	Yes
N. of observations	3,738	3,738	3,738	3,738	2,943	2,943
N. of countries	169	169	169	169	147	147
Panel B: Parameters of the Feminisation U-curve—Plough = 1						
ln GDP pc [$\beta_1 + \gamma_1$]	−1.097***	−4.263**	−4.133**	−6.575*	−11.906***	
	(0.417)	(1.695)	(1.817)	(3.508)	(3.535)	
ln GDP pc^2 [$\beta_2 + \gamma_2$]	0.070***	0.251***	0.234**	0.384*	0.686***	
	(0.023)	(0.094)	(0.103)	(0.197)	(0.192)	
Turning point, PPP $	2,642**	4,882**	6,764**	5,199*	5,878***	
	(1,313)	(2,034)	(3,452)	(2,664)	(2,151)	
Panel C: Confidence intervals for the turning points—data range: [5.5, 11.8]						
Plough = 1, ln [98% CI]	[6.7, 9.0]	[7.5, 9.5]	[7.6, 10]	[7.4, 9.7]	[7.8, 9.5]	
Plough = 0, ln [98% CI]	[−247, 244]	[6.6, 13.7]	[6.5, 13.8]	[3.7, 8.8]	[−3.1, 13.7]	

Notes: OLS estimates with robust standard errors clustered at the country level in Panel A; delta-method standard errors in Panel B. The time-varying controls in specifications 5 and 6 are armed conflict, dependence on oil export, aid dependence, an indicator of socio-cultural globalisation, and a measure of the quality of democracy. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

confidence interval for the estimated turning point of the U (in logs) is always comfortably inside the data range (see panel C). In 'non-plough' countries, by contrast, the same confidence interval always exceeds the data range on either or both sides, leading to a rejection of a U-shaped relationship at the 1% level (Lind and Mehlum 2010, 113).

We also note some relevant differences across models with and without country-fixed effects. In 'plough' countries, the estimated parameters of the U-curve are larger in specifications that allow for country-specific effects (columns 2–5), implying a more pronounced U-shaped pattern (see panel B). Furthermore, the within-country estimates indicate that in 'plough' countries, women's labour-force participation reaches a minimum at a per-capita income of around 5,000–6,000 PPP dollars (around 8.6 on a log scale).[15] This estimated turning point is around twice as high as the turning point implied by a pooled model without country-fixed effects (column 1).

Lastly, in column 6, we present the estimates of a 'traditional' specification without interaction terms. The results imply that, *on average*, the time path of FLFP is U-shaped in GDP per capita, in line with previous findings in the literature on the Feminisation U. Yet, the shape of the 'average' U-curve (as captured by the coefficient on the squared income term, i.e. 0.292, *p*-value = 0.011) is *more* pronounced than in non- 'plough' countries (−0.104, *p*-value = 0.513) but *less* pronounced than in 'plough' countries (0.686, *p*-value = 0.000).

To further examine the heterogeneity in the relationship between FLFP and economic development, we also run OLS models on split samples, as an alternative to our preferred specification with interaction terms. In particular, we estimate a version of specification 5 in Table 1 on subsamples of countries where *Plough* is greater than 0.8 ('plough' countries) and lower than 0.2 ('non-plough' countries).[16] To facilitate a visual interpretation of the results, we plot the expected relationship between FLFP and per-capita GDP based on these two models (reported more extensively in Appendix Table 8, columns 6–7). The results, shown in Fig. 2, confirm our previous findings, although in this specification the estimated relationship in 'non-plough' countries (right-hand side panel) is negatively sloped and marginally distinguishable from a flat line at the 10% level (see Appendix Table 8, column 8).

Since almost all countries in Europe (Africa) have a history (no history) of ancestral plough use, we also run additional regressions (based on model 5 in Table 1) that exclude European (African) countries from the sample. In both cases, the results (reported in Appendix Table 8, columns 3–4) are qualitatively similar, although the coefficients of interest are much less precisely estimated. In additional tests, we also find that excluding the so-called neo-Europes (USA, Canada, Australia, New Zealand) from the sample leads to near-identical results (see Appendix Table 8 column 5). Lastly, we inspect a diagram plotting the leverage score for each observation against its normalised residual squared, finding no evidence that the results are driven by influential outliers.

Taken together, these findings suggest a stylised fact that is consistent with the theoretical mechanisms discussed in Sect. 2. The time path of FLFP is only significantly U-shaped in countries whose ancestors employed a plough-based agricultural technology. The shape of this Feminisation U becomes progressively more muted the lower the share of a country's ancestors that practiced plough agriculture. In

---

[15] $= \exp[-(\beta_1 + \gamma_1)/2(\beta_2 + \gamma_2)]$

[16] The results are robust to using different cut-offs.

countries with little or no legacy of historical plough use, the time path of FLFP is effectively flat in most specifications.

6 Robustness and extensions

In this section, we examine the robustness of the conditional relationships presented in Sect. 5 in a variety of ways.

6.1 Dynamic panel bias

The first issue concerns the dynamic nature of Eq. (1). In dynamic panel-data models, the lagged dependent variable is typically correlated with the disturbance, and the OLS (or the least-squares dummy variable or LSDV) estimator is subject to a finite-sample bias of order $1/T$ (Nickell 1981). To address this problem, the most recent studies of the Feminisation U (Tam 2011; Gaddis and Klasen 2014) employ Arellano and Bond's (1991) 'difference' GMM estimator. The 'difference' GMM estimator first-differences the estimating equation to expunge σ_i, while using lagged levels of $FLPF_{it-1}$ as internal ('GMM-style') instruments for $FLPF_{it-1}$. The problem is that when the outcome variable is highly persistent, as is the case with $FLFP$, 'difference' GMM suffers from a version of the weak-instrument problem, leading to severe small-sample bias (Blundell and Bond 1998; Soto 2009).

To address this problem, we take a two-fold approach. First, we turn to Kiviet's (1999) Nickell bias-corrected least-squares dummy variable estimator (or LSDVC), as extended and implemented by Bruno (2005). Kiviet (1999) derives expressions for the small-sample bias of the LSDV estimator, thus offering a method to correct the LSDV estimates.[17] Second, we divide the panel into 5-year intervals (1995, 2000, 2005, 2010). Using a panel dataset with gaps reduces the autocorrelation of FLFP, allowing us to apply 'difference' GMM consistently. Furthermore, the specification with 5-year intervals has the advantage of smoothing out the influence of short-term cyclical fluctuations in FLFP, e.g. those potentially induced by economic recessions (Albanesi and Sahin 2018).

The results are presented in Table 2. Models 1 and 2, based on the full dataset with panels pooled over consecutive years, are estimated using Kiviet's (1999) LSDVC, with and without time-varying controls. Model 1 (Table 2) can be compared with the corresponding pooled (model 1, Table 1) and LSDV specifications (model 2, Table 1). While the coefficient on the lagged dependent variable is known to be biased upwards in the former, it is biased downwards in the latter (Roodman 2009). Reassuringly, the LSDVC estimate of ρ (0.965) falls well within the pooled-LSDVC bracketing range (0.992–0.935), confirming the plausibility of our LSDVC results. In both models 1 and 2 (Table 2), the pattern of results on the coefficients of interest is qualitatively similar to our previous findings, with no U-shape

---

[17] In Monte Carlo simulations, Bruno (2005) shows that the LSDVC estimator outperforms other consistent estimators such as GMM in both bias and efficiency when N is small relative to T.
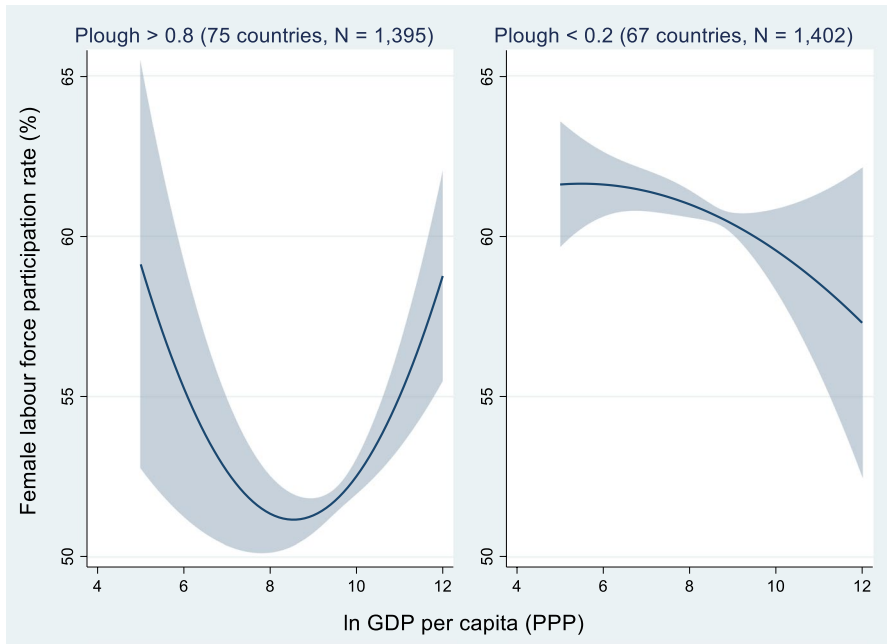
Fig. 2 FLFP and income per capita (1991–2013). Note: the diagrams plot the conditional mean (predictive margins) of FLFP in 'plough' (Plough > 0.8) and 'non-plough' (Plough < 0.2) countries, at different levels of per-capita GDP (in logs), averaging over the remaining co-variates. The 10% confidence interval is also shown. The plots are based on two regressions estimated on split samples. The model specification is similar to that of model 6, Table 1. The regressions control for country FE, country-specific linear trends, and five time-varying controls, including armed conflict, oil and aid dependence, globalization, and democracy

relationship between FLFP and per-capita GDP when *Plough* = 0 and a (statistically significantly different) U-shaped pattern when *Plough* = 1.

In models 3–5, we consider panels with 5-year intervals. Column 3 reports a benchmark LSDVC model, while columns 4 and 5 report 'difference' GMM estimates, with and without time-varying controls. The LSDVC results in column 3 are directly comparable (and, indeed, fairly similar) to the GMM estimates reported in column 4. Both GMM models pass the standard diagnostic tests.[18] Again, across models 3–5, the main pattern of results is qualitatively unaltered. We conclude that the stylised facts presented in Table 1 are not an artefact of dynamic panel bias.

---

[18] Using all available lags of $FLPF_{it-1}$ to form moment conditions, the total number of instruments (13 in model 4) is always lower than the number of countries (168 in model 4). Hansen's C test never rejects the null that the instruments are jointly valid. Also, based on Arellano and Bond's (1991) test, we never reject the null of no AR(2) autocorrelation in the residuals.

Table 2 Correcting for dynamic panel bias

Dependent variable: FLFP$_t$	Yearly panels		Panels with 5-year gaps		
	LSDVC	LSDVC	LSDVC	Δ-GMM	Δ-GMM
	(1)	(2)	(3)	(4)	(5)
FLFP$_{(t-1)}$	0.965***	0.957***	0.825***	0.799***	0.753***
	(0.006)	(0.007)	(0.048)	(0.097)	(0.112)
ln GDP pc [β$_1$]	−0.691	−2.031*	−2.618	0.003	−4.747
	(0.694)	(1.151)	(5.617)	(3.257)	(6.397)
ln GDP pc^2 [β$_2$]	0.036	0.107	0.098	−0.057	0.195
	(0.041)	(0.067)	(0.316)	(0.194)	(0.386)
Plough×ln GDP pc [γ$_1$]	−2.580**	−2.513*	−13.443*	−22.316*	−34.282***
	(1.099)	(1.504)	(8.108)	(13.739)	(11.890)
Plough×ln GDP pc^2 [γ$_2$]	0.166***	0.154*	0.815*	1.286*	1.951***
	(0.062)	(0.086)	(563)	(0.780)	(0.646)
Country FE	Yes	Yes	Yes	Yes	Yes
Time-period FE	Yes	Yes	Yes	Yes	Yes
Time-varying controls	No	Yes	No	No	Yes
N. of observations	3,738	2,943	640	640	522
N. of countries	169	147	168	168	139
N. of instruments (lag range)				13 (1 3)	18 (1 3)
AR(2) [p-value]				[0.528]	[0.495]
Hansen test [p-value]				[0.344]	[0.463]

Notes: Models 1–3 report estimates obtained using Kiviet's (1999) Nickell bias-corrected LSDV estimator (LSDVC). The bias correction, of order $N^{-1} T^{-2}$ (where N is the number of panel units and T is the number of time periods), was initialized using Arellano and Bond's (1991) one-step 'difference' GMM estimator (see Bruno (2005) for details). The standard errors are obtained from a parametric bootstrap procedure with 50 iterations, as recommended by Bruno (2005). Columns 4 and 5 report estimates based on a two-step, 'difference' GMM estimator. FLFP$_{(t-1)}$ is treated as pre-determined and instrumented for in GMM style. The standard errors are cluster-robust and Windmeijer-corrected. * $p<0.1$; ** $p<0.05$; *** $p<0.01$

6.2 Main effect of historical plough legacies

In all the models presented so far (except for model 1 in Table 1), the main effect of historical plough use was either absorbed by the country-fixed effects (LSDV and LSDVC) or removed by the first-difference transformation (Δ-GMM). In other words, while these models allow us to estimate the different *curvatures* of the time path of FLFP in 'plough' vs. 'non-plough' settings, they cannot identify the vertical distance between these time paths (e.g. the intercepts of the two U-curves). Thus, we also report estimates obtained using Blundell and Bond's (1998) 'system' GMM estimator, which permits the identification of coefficients on time-invariant regressors. While 'difference' GMM transforms the equation to eliminate σ_i, 'system'

GMM estimates the equation in levels, first-differencing the instruments to make them exogenous to σ_i (Roodman 2009).[19]

The regression results (reported in full in Appendix Table 9) are qualitatively consistent with our previous findings. In Table 3, we report the marginal effects of historical plough legacies—that is, the vertical distance between the time path of FLFP in 'plough' vs. 'non-plough' countries. We report these effects both on average and at varying levels of development, based on models with (column 1) and without (column 2) continent dummies. The point estimates are similar across models 1 and 2, although they are less precisely estimated in model 2. The *average* marginal effect of *Plough* is negative, with a coefficient implying a long-run effect of -10.40 ($= -1.716/(1 - \hat{\rho})$ where $\hat{\rho} = 0.835$). This is close to Alesina et al.'s (2013) estimate of -12.40 based on cross-section data.[20] This mean, however, conceals a substantially heterogeneous effect. In particular, the effect of historical plough use (the distance between the two curves) is only significantly negative and almost *twice* as large in magnitude as the average, at middling levels of income (8,100 PPP$).[21] At very low (245 PPP$) and very high levels of per-capita GDP (60,000 PPP$), by contrast, the effect is statistically insignificant (and indeed positively signed).

Consistent with our previous findings, these results suggest that the adverse legacy effects of the plough only appear when a country has travelled down a sufficient portion of the U-curve's downward-sloping branch. In addition, these negative effects disappear as a plough country develops further and women re-enter the labour force.

These findings are also consistent with anecdotal evidence suggesting that in pre-industrial plough societies, women did not immediately abandon productive labour to specialise in childrearing and other reproductive tasks (Boserup 1970). Rather, the initial consequence of plough adoption was a gender-based segmentation of the labour market: men were responsible for operating heavy agricultural machinery, while women specialised in food processing, drying, and storing tasks, as well as weaving and the production of other essentials (ibid.).

6.3 Omitted time-varying confounders

Our estimates of the relationship between FLFP and per-capita GDP may be confounded by the omission of unobserved time-varying factors. In the models reported in Table 1, this concern was addressed by presenting specifications that control for unobserved continent- or country-level trends in FLFP that may be correlated with GDP per capita and/or for a set of five time-varying observables that may exert a joint influence on FLFP and economic development.

---

[19] The results (see Appendix 3, column 3) are very similar if we estimate a 'system' GMM model that transforms the instruments by forward orthogonal deviations (FOD), rather than by first-differencing, as suggested by Arellano and Bover (1995). The FOD transformation has been shown to outperform the first-difference transformation in terms of both bias and efficiency (Hayakawa 2009).

[20] See their Table IV.

[21] This is the minimum point of the U-curve in 'plough' countries.

Table 3 Marginal effects of historical plough use (Plough), 'system' GMM

Dependent variable: $FLFP_t$	(1)	(2)
Average	− 1.716**	− 1.866
	(0.786)	(1.422)
GDP pc = 245	7.933	7.453
	(5.351)	(6.133)
GDP pc = 8,100	− 3.235***	− 3.300**
	(0.939)	(1.611)
GDP pc = 60,000	0.892	0.334
	(2.055)	(2.670)
Country FE	No	No
Continent FE	No	Yes
Time-period FE	Yes	Yes
Time-varying controls	Yes	Yes
N. of observations	529	529
N. of countries	146	146
N. of instruments (lag range)	23 (1 3)	28 (1 3)

Notes: The table reports marginal effects obtained in post-estimation (with delta-method standard errors). The estimates are based on a two-step, feasible 'system' GMM estimator applying the first-difference transformation on the instruments. We use the panel dataset with 5-year gaps. $FLFP_{(t-1)}$ is treated as pre-determined and instrumented for in GMM style. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Here, to further allay concerns of omitted variable bias, we present estimates that instrument for per-capita GDP and its square in GMM style (as in Gaddis and Klasen 2014), after splitting the sample in two groups as in the regressions underlying Fig. 2. Table 4 compares 'difference' GMM models that treat income and its square as strictly exogenous (columns 1 and 2) with corresponding specifications that treat them as endogenous (columns 3 and 4). In 'plough' settings, the U-shaped relationship between FLFP and per-capita income is still statistically significant, although slightly shallower and less precisely estimated, after instrumenting for per-capita income (compare models 1 and 3). In 'non-plough' settings, by contrast, the time path of FLFP is statistically indistinguishable from a flat line, again, whether or not the income terms are treated as endogenous (compare models 2 and 4). These findings reassure us that the conditional relationships we present are unlikely to result from omitted confounders.

In addition, we note that, in the models by Galor and Weill (1996) and Hiller (2014), female labour enters the aggregate production function as a separate factor input, so that per-capita income is endogenous to FLFP. Thus, a disadvantage of the models with 'endogenous income' (Table 4) is that they remove from the estimates any reverse causal effect running from FLFP to income, which is part of the theoretical U-curve. This might explain why the specifications treating income as endogenous (columns 3 and 4) reveal a more muted U-curve.

Table 4 Endogenous development: instrumenting for per-capita income, Δ-GMM

Dependent variable: FLFP$_t$	Income treated as exogenous		Income treated as endogenous	
	Plough > 0.8	Plough < 0.2	Plough > 0.8	Plough < 0.2
	(1)	(2)	(3)	(4)
FLFP$_{(t-1)}$	0.508***	0.929***	0.522***	0.939***
	(0.180)	(0.159)	(0.178)	(0.253)
ln GDP pc	−39.772***	2.211	−37.469**	9.006
	(12.156)	(8.548)	(18.083)	(37.488)
ln GDP pc^2	2.061***	−0.201	1.821*	−0.337
	(0.650)	(0.548)	(1.052)	(2.032)
Country FE	Yes	Yes	Yes	Yes
Time-period FE	Yes	Yes	Yes	Yes
Time-varying controls	Yes	Yes	Yes	Yes
N. of observations	238	255	238	255
N. of countries	66	65	66	65
N. of instruments	16	16	20	20

Notes: The estimates are based on a two-step, 'difference' GMM estimator applied to the dataset with 5-year gaps. Windmeijer-corrected cluster-robust errors in parentheses. FLFP$_{(t-1)}$ is always treated as predetermined and instrumented for in GMM style (the lag range is 1–3 throughout). In models (3) and (4), ln GDP pc and ln GDP pc^2 are also treated as endogenous and instrumented for in GMM style (the lag range is 2–3 throughout). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

6.4 Alternative measures of historical plough legacies

Next, we present reduced-form estimates that replace *Plough* with the instrumental variables used by Alesina et al. (2013). These variables capture the suitability of ancestral environments to the cultivation of crops that are technically compatible with the plough. More specifically, they measure 'the proportion of the [present-day country's] population with ancestors that lived in climates that could grow plough-positive cereals (wheat, barley and rye), and the proportion that lived in climates that could grow plough-negative cereals (sorghum, foxtail millet, and pearl millet)' (Alesina et al., 2013, 516). In Table 5, we present the estimated coefficients on the income terms (obtained in post-estimation) when the interactant is the measure of plough-positive crop suitability (column 1) or the measure of plough-negative crop suitability (column 2) instead of *Plough*.

Confirming the pattern of results presented so far, we find in column 1 that countries whose ancestors lived in climates that were suitable to the cultivation of crops benefitting from plough use (plough-positive crops = 1) experience a significantly U-shaped path of FLFP. By contrast, countries whose ancestors lived in climates that were *not* suitable to the cultivation of crops displaying plough complementarities (plough-positive crops = 0) experience a path of FLFP that is, statistically speaking, flat. Similarly, in column 2, we find that countries whose ancestors lived in climates that were suitable to the cultivation of crops that did *not* benefit from plough use (plough-negative crops = 1) experience a flat time path of FLFP. By contrast, in countries whose ancestors lived in climates that were *not* suitable to the cultivation

Table 5 Alternative measures of historical plough use, Δ-GMM

Dependent variable: $FLFP_t$	(1)	(2)
Interactant	Plough-positive crops	Plough-negative crops
Interactant $=0$, Parameters of the Feminisation U-curve:		
ln GDP pc	4.467	-26.266
	(9.095)	(18.543)
ln GDP pc^2	-0.409	1.601*
	(0.576)	(0.963)
Interactant $=1$, Parameters of the Feminisation U-curve:		
ln GDP pc	-50.906***	-3.641
	(14.982)	(14.112)
ln GDP pc^2	2.823***	-0.107
	(0.791)	(0.791)
N. of observations	510	510
N. of countries	136	136

Notes: The table reports marginal effects obtained in post-estimation (with delta-method standard errors). The estimates are based on a two-step, 'difference' GMM estimator applied to the panel dataset with 5-year gaps. The lagged dependent variable is always treated as pre-determined and instrumented for in GMM style. * $p<0.1$; ** $p<0.05$; *** $p<0.01$

of crops requiring other tools than the plough (plough-negative crops $=0$), the time path of FLFP is (marginally) U-shaped at the 10% level.

6.5 Other effect modifiers

Lastly, we investigate the potential effect-modifying role of other variables than *Plough*. Conceivably, the U-shaped path of FLFP may be modified also by other historical events, other features of pre-industrial societies, or other correlates of historical gender norms. If so, the heterogeneous shape of the Feminisation U may result partly from other factors than historical plough legacies. Furthermore, if these alternative factors are correlated with historical plough use, the coefficients on the interaction terms ($Plough_i \times \ln GDPpc_{it}$ and $Plough_i \times \ln GDPpc_{it}^2$) could be picking up their unobserved influence, spuriously attributing their effect-modifying role to the plough. To address these two related concerns, we present models that control for two additional interaction terms—$Z_i \times \ln GDPpc_{it}$ and $Z_i \times \ln GDPpc_{it}^2$, where Z_i is a time-invariant country characteristic that may exert a moderating influence on the dynamic path of FLFP.

As Z_i, we consider several potential factors. The first one is the pre-historical transition from a hunter-gatherer to an agricultural society. Based on cross-sectional regressions and archaeological evidence, Hansen et al. (2015) argue that historical gender inequalities pre-date the introduction of the plough. Indeed, these authors emphasise the transition to sedentary agriculture more generally as a critical historical juncture that (a) accelerated a rise in fertility and (b) assigned a premium on male brawn, promoting the emergence of a gendered division of labour and, over

time, a system of norms and beliefs that downgrade the status of women in the productive sphere.[22] If so, the timing of the Neolithic revolution may exert a moderating influence on the subsequent evolution of FLFP. Since the shift to sedentary agriculture and the introduction of the plough are correlated, part of this influence may be spuriously picked up by our interaction terms. To capture the timing of the Neolithic revolution, we use a measure of the number of years of settled agriculture in 1500 AD compiled by Putterman and Trainor (2006).

Second, we examine the potential modifying influence of the cultural norms that are ingrained in languages. Linguists have shown that the rules governing grammatical gender, which differ greatly across languages, 'arose from evolutionary pressures concerning specialisation, reproduction and the division of labor' (Shoham and Lee 2018, 1217).[23] Indeed, the degree to which a language grammatically emphasises gender 'acts as a [stable] cultural marker for *historical* gender norms' (ibid., see Shoham and Lee 2018 for a review of the literature).[24] Thus, we consider the possibility that the cultural norms crystallised in languages may exert an additional moderating influence on the path of FLFP and that part of this influence may be spuriously picked up by the estimated coefficients on $Plough_i \times \ln GDPpc_{it}$ and $Plough_i \times \ln GDPpc_{it}^2$. To measure the degree of grammatical gender marking, we use the Gender Intensity Index (GII) developed by Gay et al. (2013).

Third, we investigate the potential role of a country's exposure to the transatlantic slave trade. This variable is non-zero in a number of African countries, many of which have little or no tradition of historical plough use. Recent contributions suggest that the slave trade gave rise to female-biased sex ratios, allowing women to take up occupations that were traditionally the preserve of men (Teso 2019). If this occupational reallocation was responsible for the emergence of more equal gender-role attitudes, the legacy of the slave trade may exert an additional moderating influence on the path of FLFP, potentially confounding the effect-modifying role of the plough. We use a measure of a country's total slave exports normalised by land area (Nunn 2008).

Fourth, we consider the potential effect-modifying impact of state antiquity—that is, the number of years each country was ruled by state institutions above tribal level (Borcan et al. 2018). In the literature, early statehood has been associated with the emergence and consolidation of patriarchal norms (Lerner 1986), as well as with the introduction of settled agriculture.

---

[22] We note here that Boserup's (1970) evidence suggesting that hoe-agriculture is not associated with a withdrawal of women from the labour force seems incompatible with Hansen et al.'s (2015) conjecture.

[23] Recent contributions have documented a link between the strength of gender marking in a given language and various dimensions of gender inequality, including educational attainment (Davis and Reynolds 2018), popular attitudes and beliefs about gender (Liu et al. 2018), and the gender wage gap (Shoham and Lee 2018). More relevant for us, Hicks et al. (2015) and Gay et al. (2018) show that female immigrants in the USA who speak a more strongly gendered language (e.g. Arabic, Spanish) also participate less in the labour force, and work less, than migrants whose mother tongue is less strongly gendered (e.g. Mandarin, Armenian).

[24] Emphasis added.

Table 6 Ruling out alternative effect modifiers, Δ-GMM

	Z=							
	Timing of Neolithic revolution	Language gender marking	Slave trade	State antiquity	Extended family structure	Private property	Large domesticated animals	Hunting
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: $FLFP_t$								
$FLFP_{(t-1)}$	0.799***	0.659***	0.769***	0.785***	0.738***	0.804***	0.742***	0.734***
	(0.113)	(0.180)	(0.108)	(0.117)	(0.117)	(0.115)	(0.116)	(0.112)
ln GDP pc	−2.977	−8.713	−6.894	−2.778	−9.826	0.775	−14.514	−19.558
	(12.902)	(17.608)	(7.264)	(6.333)	(11.492)	(8.863)	(52.639)	(17.704)
ln GDP pc^2	0.106	0.386	0.309	0.092	0.461	−0.176	0.568	1.096
	(0.745)	(0.956)	(0.437)	(0.387)	(0.664)	(0.544)	(2.659)	(0.972)
Plough × ln GDP pc	−29.507*	−36.535**	−32.223***	−35.174**	−34.035***	−38.648***	−31.170**	−23.298
	(17.717)	(16.149)	(11.838)	(14.394)	(12.203)	(12.193)	(12.224)	(14.936)
Plough × ln GDP pc^2	1.695*	2.033**	1.847***	2.008***	1.927***	2.255***	1.752***	1.275
	(0.961)	(0.922)	(0.647)	(0.767)	(0.666)	(0.683)	(0.670)	(0.818)
Z × ln GDP pc	−0.001	−3.114	2.392	−0.014	7.512	−11.989	7.400	141.628
	(0.004)	(4.759)	(2.014)	(0.032)	(14.585)	(13.089)	(55.960)	(178.033)
Z × ln GDP pc^2	0.000	0.185	−0.161	0.001	−0.388	0.809	−0.219	−8.674
	(0.000)	(0.255)	(0.129)	(0.002)	(0.815)	(0.787)	(2.825)	(9.625)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-varying controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of observations	497	433	522	494	518	518	518	518
N. of countries	132	115	139	131	138	138	138	138

Notes: The estimates are based on a two-step, 'difference' GMM estimator applied to the panel dataset with 5-year gaps. Windmeijer-corrected, cluster-robust errors in parentheses. $FLFP(t-1)$ is treated as pre-determined and instrumented for in GMM style. Throughout columns 1–8, the number of internal instruments used is 20, and the lags used to form moment conditions are 1–3. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Lastly, we interact income with some of the anthropological variables used by Alesina et al. (2013) as controls. These are measures of the share of a present-day country's inhabitants whose ancestors live in societies with an extended family structure, private property (land inheritance) rules, an economy that relied on large domesticated animals, and an economy that relied on hunting. These characteristics of ancestral societies may have facilitated the emergence of slow-moving gender norms that in turn made it more likely for a pre-historical society to invent or adopt the plough. A detailed description of all the variables used as interactants, and the associated data sources, is provided in Appendix Table 7.

The results are presented in Table 6. All models control for country and time-period (5-year) fixed effects, as well as our set of time-varying controls. Across specifications 1–8, the estimated coefficients on $Z_i \times \ln GDPpc_{it}$ and $Z_i \times \ln GDPpc_{it}^2$ are always statistically insignificant, after controlling for $Plough_i \times \ln GDPpc_{it}$ and $Plough_i \times \ln GDPpc_{it}^2$. We find no evidence that any of the Z_i's we considered exerts an additional moderating influence on the relationship between FLFP and per-capita GDP over and above the effect-modifying role of the plough. At the same time, including $Z_i \times \ln GDPpc_{it}$ and $Z_i \times \ln GDPpc_{it}^2$ in the regression leaves our results concerning the moderating influence of historical plough use qualitatively unaltered. Throughout models 1–7, the estimated coefficients on $Plough_i \times \ln GDPpc_{it}$ and $Plough_i \times \ln GDPpc_{it}^2$ are always statistically significant, while in model 8, the estimates are marginally insignificant at conventional levels (p-values = 0.119). We conclude that the estimated effect-modifying influence of the plough is unlikely to be spuriously picking up the impact of other historical correlates of the plough. Meanwhile, the observed heterogeneity in the path of FLFP across countries, while consistently related to the legacy of the plough, appears to be unrelated to these other historical factors.

7 Conclusion

In the literature, the evidence in support of the Feminisation U is mixed. In this paper, we investigated empirically the observed heterogeneity in the dynamic path of FLFP in the course of development. Building on the theoretical literature, we argued that initial conditions are critical in shaping the subsequent path of FLFP in the course of development. In particular, the cultural norms that prevail when per-capita income begins to grow above subsistence levels determine whether the mechanisms that are known to generate the Feminisation U (structural change, fertility dynamics, and the gender education gap) are operative or not.

Building on the previous literature on the historical origins of gender roles, we presented stylised facts that are consistent with initial gender-role cultural norms exerting a moderating influence on the dynamics of FLFP in the course of development. The conditional relationship between FLFP and per-capita GDP is only significantly U-shaped in countries with a history of plough agriculture, where historical

gender norms were initially shaped by brawn-induced productivity differences. In countries without such history, by contrast, the time path of FLFP is effectively flat. These conditional relationships are robust to controlling for time-varying observables and unobservables, to using estimators that correct for dynamic panel bias, to instrumenting for per-capita GDP, to using alternative measures of historical plough use, and to controlling for other plausible historical effect modifiers.

Our findings have important implications. First, they integrate two literatures—on the Feminisation U and on the historical origins of gender roles—that have so far developed on parallel tracks. Second, they suggest that empirical studies may fail to estimate a significant U-shaped path of FLFP if they focus on countries or regions without a history of traditional plough use. For instance, a possible reason why Roncolato's (2016) study of South Africa does not reveal a U-shaped relationship between FLFP and per-capita income is that South Africa is not a plough country (*Plough* = 0.54) Third, our results point to an important nuancing of Alesina et al.'s (2013) version of Boserup's (1970) 'plough argument'. The introduction and diffusion of the plough leaves a legacy that persists up to the present day, as argued by Alesina et al. (2013). Yet, rather than leading to a permanent *shift* in gender norms and FLFP, the evidence presented here suggests that the plough shock had the effect of *modifying the evolution* of female labour supply (and arguably, the evolution of attitudes and beliefs about working women) in the post-Malthusian era. The labour market effects of plough adoption are neither immediate nor permanent, as they only appear as an economy reaches a middling level of income and tend to vanish completely as a high level of income is attained.

Overall, our research raises important questions about the interaction of culture and structure in shaping development outcomes. It also points to further avenues of research. While we show that historical legacies can modify the effects of post-Malthusian development on labour market outcomes for women, we have not presented direct evidence of a cultural transmission channel. While it seems difficult to conceive of a variable measuring gender-role norms and attitudes coherently over time and space to directly demonstrate cultural transmission, there is well-documented indirect evidence that supports our interpretation of the stylised facts presented here (e.g. Fernandez and Fogli 2009; Fernandez 2013).

Future research should also look for direct evidence of the mechanisms that, according to theory, are responsible for generating a U-shaped time path of FLFP. In particular, evidence should be sought that these mechanisms are triggered and sustained by the historical adoption of the plough. Answering these questions will also shed light on the time allocation of women as they leave and re-enter the labour force in the course of development. Is women's time used to raise more children? Is girls' education at its lowest level near the U-curve's turning point? Is there a stronger son preference in middle-income 'plough' countries? Future research will also have to pay more attention to the timing of plough adoption and to how subsequent technological innovations (e.g. the invention of the heavy plough in mediaeval Northern Europe) influenced labour market outcomes for women.

Appendix

Table 7 Control variables

Variable	Description	Source
1. Time-varying controls		
Armed conflict	Number of active interstate, internal and internationalised armed conflicts at the country-year level	Uppsala Conflict Data Program (UCDP)
Oil dependence	Net oil exports value per capita, constant 2000 dollars	Ross (2008)
Aid dependence	Net ODA received (% GNI)	World Bank, *World Development Indicators*
Index of globalisation	KOF Index of Globalisation (measuring economic, social and cultural dimensions of globalisation)	Dreher (2006)
Quality of democracy	V-Dem index of electoral democracy (polyarchy)	Varieties of democracy (V-Dem) project
2. Historical country characteristics—potential effect modifiers		
Timing of Neolithic revolution	Years of agriculture in 1500 CE	Putterman and Trainor, 2006
GII Index	Intensity of gender marking in the country's dominant language, focusing on sex-based gender, number of genders, gender assignment, and gendered pronouns	Gay et al. (2013)
Slave exports	Log of total slave exports normalized by land area (non-zero values for African countries only)	Nunn (2008)
State antiquity index	State History Index of accumulated state experience since 3500 BCE in the territories defined by present-day political borders	Borcan et al. (2018)
Extended family structure	Proportion of a country's ancestors with an extended family structure	Alesina et al. (2013)
Private property	Proportion of a country's ancestors without land-inheritance rules	Alesina et al. (2013)
Presence of large animals	Fraction of population with ancestors using large domesticated animals	Alesina et al. (2013)
Reliance on hunting	Ancestral dependence on hunting for subsistence, pop. Weighted	Alesina et al. (2013)

Table 8 Alternative specifications mentioned in Sect. 5, OLS

Dependent variable: $FLFP_t$	Parsimonious specifications		w/o Africa	w/o Europe	w/o Neo-Europes	Plough > 0.8	Plough <0.2	Plough <0.2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$FLFP_{(t-1)}$	0.933***	0.933***	0.759***	0.795***	0.787***	0.754***	0.811***	0.811***
	(0.010)	(0.010)	(0.026)	(0.025)	(0.022)	(0.031)	(0.031)	(0.031)
ln GDP pc [β_1]	−0.150*		2.439	0.425	1.750	−10.875***	1.129	−0.526*
	(0.083)		(6.208)	(2.484)	(2.482)	(3.694)	(2.435)	(0.288)
ln GDP pc^2 [β_2]		−0.008*	−0.219	−0.061	−0.146	0.637***	−0.103	
		(0.005)	(0.380)	(0.163)	(0.165)	(0.201)	(0.159)	
Plough×ln GDP pc [γ_1]	−4.008**	−4.164**	−15.208**	−6.294	−13.409***			
	(1.817)	(1.808)	(7.325)	(5.422)	(4.323)			
Plough×ln GDP pc^2 [γ_2]	0.236**	0.244**	0.967**	0.393	0.817***			
	(0.103)	(0.102)	(0.436)	(0.298)	(0.255)			
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year×continent FE	Yes	Yes	No	No	No	No	No	No
Country FE×linear trend	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Time-varying controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes
N. of observations	3,738	3,738	1,963	2,368	2,851	1,395	1,402	1,402
N. of countries	169	169	99	119	143	75	67	67

Notes: OLS estimates with robust standard errors clustered at the country level. The time-varying controls include measures of armed conflict, dependence on oil export, aid dependence, an indicator of socio-cultural globalisation, and a measure of the quality of democracy. * $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table 9 System' GMM estimates

	Δ transform	Continent FE	FOD transform
Dependent variable: FLFP$_t$	(1)	(2)	(3)
FLFP$_{(t-1)}$	0.835***	0.828***	0.836***
	(0.049)	(0.074)	(0.048)
Plough	72.769**	68.359*	69.707*
	(36.241)	(40.961)	(35.664)
ln GDP pc [β$_1$]	−9.441	−10.657	−9.198
	(5.810)	(7.105)	(5.774)
ln GDP pc^2 [β$_2$]	0.516	0.586	0.504
	(0.322)	(0.388)	(0.317)
Plough×ln GDP pc [γ$_1$]	−17.043**	−15.963*	−16.400**
	(8.158)	(9.181)	(8.006)
Plough×ln GDP pc^2 [γ$_2$]	0.955**	0.889*	0.923**
	(0.455)	(0.509)	(0.445)
Country FE	No	No	No
Continent FE	No	Yes	No
Time-period FE	Yes	Yes	Yes
Time-varying controls	Yes	Yes	Yes
N. of observations	529	529	529
N. of countries	146	146	146
N. of instruments (lag range)	23	28	23
AR(2) [p-value]	[0.438]	[0.440]	[0.464]
Hansen test [p-value]	[0.501]	[0.499]	[0.524]

The regressions are based on a two-step, feasible 'system' GMM estimator employing the first-difference (columns 1–2) or the forward-orthogonal deviation (FOD) transformation (column 3) on the instruments. We use the panel dataset with 5-year gaps. FLFP$_{(t-1)}$ is treated as pre-determined and instrumented for in GMM style. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Declarations

Conflict of interest The authors declare no competing interests.

References

Alesina A, Giuliano P, Nunn N (2013) On the origins of gender roles: women and the plough. Q J Econ 128(2):469–530

Albanesi S, Sahin A (2018) The gender unemployment gap. Rev Econ Dyn 30:47–67

Andersen TB, Jensen PS, Skovsgaard CV (2016) The heavy plough and the agricultural revolution in Medieval Europe. J Dev Econ 118:113–149

Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev Econ Stud 58:277–297

Arellano M, Bover O (1995) Another look at the instrumental variable estimation of error-components models. J Economet 68:29–51

Becker GS (1960) An Economic Analysis of Fertility. In: Coale AJ (ed) Demographic and economic change in developed countries. Princeton University Press, Princeton, NJ, pp 209–240

Beer C (2009) Democracy and gender equality. St Comp Int Dev 44:212–227

Benerìa L (1979) Reproduction, production and the sexual division of labour. Camb J Econ 3:203–225

Bloom DE, Canning D, Fink G, Finlay J (2009) Fertility, female labor force participation, and the demographic dividend. J Econ Growth 14:79–101

Blundell R, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. J Economet 87:115–143

Borcan O, Olsson O, Putterman L (2018) State history and economic development: evidence from six millennia. J Econ Growth 23:1–40

Boserup E (1970) Woman's role in economic development. George Allen and Unwin Ltd, London

Bruno G (2005) Estimation and inference in dynamic unbalanced panel-data models with a small number of individuals. Stata J 5(4):473–500

Çağatay N, Özler S (1995) Feminization of the labor force: the effects of long-term development and structural adjustment. World Dev 23(11):1883–1894

Davis, L S and Reynolds, M (2018) Gendered language and the educational gender gap (February 21, 2018). Available at SSRN: https://ssrn.com/abstract=2782540 or http://dx.doi.org/https://doi.org/10.2139/ssrn.2782540

Dreher A (2006) Does globalization affect growth? evidence from a new index of globalization. Applied Econ 38(10):1091–1110

Durand, JD (1975) The labor force in economic development: a comparison of international census data, 1946.1966. Princeton University Press, Princeton, NJ

Kiviet JF (1999) Expectation of expansions for estimators in a dynamic panel data model; some results for weakly exogenous regressors. In: Hsiao C, Lahiri K, Lee L-F, Pesaran MH (eds) Analysis of Panels and Limited Dependent Variable Models. Cambridge University Press, Cambridge, pp 199–225

Lagerlof NP (2003) Gender equality and long-run growth. J Econ Growth 8:403–426

Fernandez R, Fogli A, Olivetti C (2004) Mothers and sons: preference formation and female labor force dynamics. Q J Econ 119:1249–1299

Fernandez, R, Fogli, A (2009) Culture: an empirical investigation of beliefs, work and fertility. Am. Econ. J.: Macroecon. 1, 146–177

Fernandez R (2013) Cultural change as learning: the evolution of female labor force participation over a century. Am Econ Rev 103(1):472–500

Fussell GE (1966) Plough and ploughing before 1800. Agri Hist 40(3):177–186

Gaddis I, Klasen S (2014) Economic development, structural change, and women's labor force participation: a re-examination of the Feminisation U hypothesis. J Popul Econ 27:639–681

Galor O, Weil D (1996) The gender gap, fertility, and growth. Am Econ Rev 86(3):374–387

Galor O, Weil D (2000) Population, technology and growth: from malthusian stagnation to the demographic transition and beyond. Am Econ Rev 90(4):806–828

Gay V, Hicks DL, Santacreu-Vasut E, Shoham A (2018) Decomposing culture: an analysis of gender, language, and labor supply in the household. Rev Econ Household 16:879–909

Gay, V, Santacreu-Vasut, E, & Shoham, A (2013) The grammatical origins of gender roles. Berkeley Economic History Economic Laboratory working paper series WP2013–03.

Goldin, C (1995) The U-shaped female labor force function in economic development and economic history. In: Schultz, T.P. (ed) Investment in women's human capital. University of Chicago Press, Chicago, IL

Goldin C, Olivetti C (2013) Shocking labor supply: a reassessment of the role of World War II on women's labor supply. Am Econ Rev 103(3):257–262

Grosjean P, Khattar R (2019) It's raining men! Hallelujah? The long-term consequences of male-biased sex ratios. Rev Econ Studies 86:723–754

Hansen CW, Jensen PS, Skovsgaard CV (2015) Modern gender roles and agricultural history: the Neolithic inheritance. J Econ Growth 20(4):365–404

Hayakawa K (2009) First difference or forward orthogonal deviation: which transformation should be used in dynamic panel data models?: A simulation study. Econ Bul 29(3):2008–2017

Hazan M, Maoz Y (2002) Women's labor force participation and the dynamics of tradition. Econ Letters 75:193–198

Hicks DL, Santacreu-Vasut E, Shoham A (2015) Does mother tongue make for women's work? linguistics, household labor, and gender identity. J Econ Behav Org 110:19–44

Hiller V (2014) Gender inequality, endogenous cultural norms, and economic development. Scand J Econ 116(2):455–481

Humphries J, Sarasua C (2012) Off the record: reconstructing women's labor force participation in the European past. Fem Econ 18(4):39–67

Jaquette JS, Summerfield G (eds) (2006) Women and gender equity in development theory and practice: Institutions, resources and mobilization. Duke University Press, Durham and London

Kimura M, Yasui D (2010) The Galor-Weil gender gap model revisited: from home to market. J Econ Growth 15:323–351

Klasen S (2018) What explains uneven female labour force participation levels and trends in developing countries? World Bank Res Obs 34(2):161–197

Lahoti R, Swaminathan H (2016) Economic development and women's labor force participation in India. Fem Econ 22(2):168–195

Lerner G (1986) The creation of patriarchy. Oxford University Press, New York

Lind JT, Mehlum H (2010) With or without u? The appropriate test for a U-shaped relationship. Ox Bul Econ Stat 72(1):109–118

Liu A, Shair-Rosenfield S, Vance LR, Csata Z (2018) Linguistic origins of gender equality and women's rights. Gender & Soc 32(1):82–108

Luci, A (2009) Female labour market participation and economic growth Intl. J. Innov. Sust. Dev. 4(2/3)

Mammen K, Paxson C (2000) Women's work and economic development. J Econ Perspect 14(4):141–164

Ngai, LR, Olivetti, C (2015) Structural transformation and the U-shaped female labor supply, unpublished manuscript

Ngai LR, Petrongolo B (2017) Gender gaps and the rise of the service economy. Am Econ J Macro 9(4):1–44

Nickell S (1981) Biases in dynamic models with fixed effects. Economet 49(6):1417–1426

Nunn N (2008) The long-term effects of Africa's slave trades. Q J Econ 123(1):139–176

Pampel FC, Tanaka K (1986) Economic development and female labor force participation: a reconsideration. Soc Forces 64(3):599–619

Potrafke N, Ursprung HW (2012) Globalization and gender equality in the course of development. Eur J Pol Econ 28:399–413

Putterman L, Trainor CA (2006) Agricultural transition year country data set. Brown University, Providence

Rendall, M (2017) Brain versus brawn: the realization of women's comparative advantage. University of Zurich, Institute for Empirical Research in Economics, Working Paper No. 491, Revised version, Available at SSRN: https://ssrn.com/abstract=1635251 or https://doi.org/10.2139/ssrn.1635251

Roncolato L (2016) The Feminization U in South Africa: economic structure and women's labor force participation. Fem Econ 22(4):54–81

Roodman D (2009) How to do xtabond2: an introduction to difference and system GMM in Stata. Stata J 9(1):86–136

Ross M (2008) Oil, Islam and women. Am Pol Sci Rev 102(2):1–18

Shoham A, Lee SM (2018) The causal impact of grammatical gender marking on gender wage inequality and country income inequality. Business & Soc 57(6):1216–1251

Sinha, JN (1967) Dynamics of female participation in economic activity in a developing economy. In: United Nations Department of Economic and Social Affairs. (ed), Proceedings of the World Population Conference, vol.IV. United Nations, New York

Soto, M (2009) System GMM estimation with a small sample, unpublished manuscript

Tam H (2011) U-shaped female labor participation with economic development: some panel data evidence. Econ Lett 110:140–142

Teso E (2019) The long-term effect of demographic shocks on the evolution of gender roles: evidence from the transatlantic slave trade. J Europ Econ Asoc 17(2):497–534

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.