UCL CENTRE FOR COMPARATIVE STUDIES OF EMERGING ECONOMIES

Working Paper Series

2020/2

The Feminisation U, Cultural Norms, and the Plough

Luca Uberti & Elodie Douarin

UCL School of Slavonic and East European Studies
16 Taviton Street, London WC1H 0BW
www.ucl.ac.uk/ssees/comparative-studies-emerging-economies

The Feminisation U, Cultural Norms, and the Plough

This version: October 2020

Luca Uberti[1] and Elodie Douarin[2]

Affiliation:
[1] University of Luxembourg
[2] School of Slavonic and East European Studies, University College London

Corresponding author:
luca.uberti@uni.lu

Abstract

The feminisation U describes the tendency of female labour force participation (FLFP) to first decline and then rise in the process of development. Long considered to be a 'stylised fact', the feminisation U is actually supported by mixed evidence. This research identifies an important source of heterogeneity in the shape of the feminisation U across countries – the cultural norms and values engendered by the adoption of the plough in pre-industrial times. In line with existing theoretical accounts of the U-curve, which suggest that initial conditions are critical, we find evidence that a tradition of plough use intensifies the U-shaped path of FLFP. Based on a dynamic panel-data estimator, we find evidence of a significantly U-shaped path of FLFP in countries with a history of ancestral plough use, but no such relationship in 'non-plough countries'. We also explore, and rule out, other potential drivers of heterogeneity (e.g. the timing of the Neolithic revolution), and investigate empirically the causal mechanisms that generate the feminisation U in plough countries. Our results address the empirical controversy surrounding the feminisation U, while shedding new light on the long-run effects of plough adoption.

Keywords

Female labour force participation, economic development, agricultural history, cultural norms

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 growth are accompanied by a de-feminisation of the labour force, women tend to become economically active in the market again as incomes rise further. Figure 1 documents this pattern in a pooled cross-section of 172 countries during 1990-2013, while Figure 2 provides illustrative evidence of a country, Egypt, where the time path of female labour supply is U-shaped in per-capita income. Some of the classic contributions on the subject have explained the feminisation of the labour force with reference to structural change: while industrialisation leads women to exit the labour force, the transition from industry to a service economy induces a re-entry (Boserup, 1970; Goldin, 1995). In other theoretical 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 as economies develop.

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

[Figure 1 and 2]

In this paper, we re-engage with this empirical controversy by exploring an important source of variation in the shape of the U-curve across different countries – the beliefs and values regarding the appropriate role of women in society. Recent contributions have examined the historical emergence (Alesina et al., 2013; Hansen et al., 2015), transmission (Fernandez and Fogli, 2009), persistence (Grosjean and Khattar, 2019), and change (Fernandez, 2013) of the cultural norms that assign different roles to men and women.[1] This research has largely developed in isolation from the Feminisation U literature. As such, it has not yet been used to shed light on the dynamics of female labour supply in the course of development. Here, we fill this gap by integrating these two strands of literature.

Our empirical analysis is motivated by recent theoretical work by Hiller (2014), who considers the link between economic development, cultural norms and FLFP in the context of a two-sex

[1] In the context of this paper, we understand 'cultural norms' to be synonymous with 'social norms', but we use the former throughout.

overlapping-generations model. Men have a comparative advantage in brawn and are initially more productive than women. Consequently, they supply more labour, and cultural norms emerge that favour boys in the household allocation of educational spending. As men accumulate human capital, the household's potential income increases, the gender productivity gap widens, and the cultural bias against girls' education intensifies, pushing down FLFP. When potential income crosses a given threshold, however, educational spending for girls generates a return, which induces an increase in female education and labour supply.

In Hiller's (2014) set-up, initial conditions are a critical parameter. If the productivity differential between uneducated men and women is very small, human-capital accumulation is (almost) gender-neutral from the outset, and FLFP does not exhibit a distinctly U-shaped dynamic path. If the initial male-female productivity differential is significant, by contrast, spending on girls' education only begins to rise above zero at a much higher level of income, and the dynamic path of FLFP is strongly U-shaped.

This argument points to an important source of heterogeneity in the shape of the U – the initial gender productivity gap. Hiller (2013), however, does not explicitly model the characteristics of the pre-industrial production technology that give rise of this gap. Instead, we refer to the seminal work of Boserup (1970) and Alesina et al. (2013). These authors distinguish between plough-based agriculture and other forms of shifting cultivation that rely on smaller hand-held tools such as the hoe. Since the plough requires a lot of physical strength to be operated, plough-based agricultural systems put a productivity and wage premium on male brawn. Thus, the degree to which ancestral societies used the plough may be used as a proxy for initial differences in productivity between uneducated men and women.

Combining Hiller's (2014) model with the plough argument, this paper formulates, and tests, an observable prediction that has so far remained unexplored – namely, that the legacy of ancestral plough use exerts an important moderating influence on the shape of the Feminisation U. In societies that traditionally practiced plough agriculture, the dynamic path of FLFP should be strongly U-shaped. The less a society's ancestors used the plough, the more attenuated the U-curve should be, potentially vanishing completely in societies whose ancestors practiced shifting cultivation.

Based on panel data on the rate of FLFP from 169 countries during 1990-2013, we present extensive empirical evidence in support of this prediction. 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 dynamic GMM estimator, we find strong evidence in support of our prediction: the legacy of the plough intensifies the U-shaped path of FLFP. While a statistically significant U-shaped relationship is observed in 'plough countries', our estimations in countries with no history of ancestral plough use reveal no such relationship.

This pattern is extremely robust. A significant contrast between the paths of FLFP in plough vs. non-plough societies is observed across a wide range of specifications, including semi-parametric models, and is robust to tests that treat economic development as endogenous to FLFP. We also test our hypothesis against rival explanations. Yet, we find no evidence that other competing historical factors – e.g. the timing of the Neolithic revolution – may exert a moderating influence on the path of FLFP in the course of development. Lastly, we investigate empirically the three mechanisms that have been proposed in the literature to explain the feminisation U – namely, structural change, fertility and education. Our evidence is consistent with all three mechanisms playing a role in generating a U-shaped path of FLFP. Yet, crucially, our findings suggest that these mechanisms are only operative in plough countries.

Our empirical findings make an important contribution in two ways. First, they suggest a possible way of solving the ongoing empirical controversy surrounding the Feminisation U. Individual country studies may fail to observe a U-shaped relationship if they focus on 'non-plough' countries. Relatedly, we also contribute to this literature by testing empirically the relative importance of the causal mechanisms suggested as explanations of the feminisation U. Second, our results contribute to the literature on the historical origins of gender roles, suggesting an important qualification to Alesina et al.'s (2013) argument. Their empirical analysis focused on the *average* effect of historical plough use, hiding important heterogeneities along the income distribution. In our (more flexible) specification, the detrimental effects of historical plough use are only observed in middle-income economies. These economies have travelled down a substantial portion of the downward-sloping branch of the U-curve, but have not yet begun to climb the upward-sloping branch. Thus, in the process of economic development, the unequalising effects of the plough shock are neither immediate nor permanent: they do not show up until a country has crossed a middling level of income; they also vanish completely as a country attains an advanced level of economic development.

The Feminisation U-Curve: A Review of the Literature

We begin by reviewing the main theoretical mechanisms that have been proposed to explain the Feminisation U, highlighting how their predictions depend on assumptions about gender-role norms. We then reconstruct the empirical controversy regarding its veracity.

Theory

Several theories predict a U-shaped relationship between FLFP and economic development. In particular, three possible mechanisms have been proposed: structural change, education and fertility dynamics. Here, we review each of them in turn.

Early arguments placed an emphasis on the sectoral shifts in production and employment that accompany the process of economic development (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 has the effect of marginalising women for two reasons. First, the physical separation between home and workplace makes it more difficult to reconcile productive and reproductive tasks[2] (Beneria, 1979). Second, since work in factories and industrial farms is generally considered 'dirty', a married woman engaged in paid manual labour outside the home brings 'stigma' on the family (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).[3] At more advanced stages of development, the availability of paid jobs that do not mark women with a stigma (notably, service-sector jobs) tends to increase, inducing an increase in FLFP. Ngai and Olivetti (2015) formalise this argument, showing that a model of structural transformation with home/market production and gender-specific comparative advantages leads to a U-shaped female labour supply.

In these models, the initial fall in FLFP comes from the emergence of *stigma* jobs, while the subsequent rise is due to an increased availability of non-stigma jobs. Thus, the U-pattern is implicitly sustained by the cultural norms that prescribe the appropriate role of women in society, and hence assign stigma to occupations, such as factory work, that are considered inappropriate for women. These norms are not a constant feature of human society but may be held to varying degrees. Accordingly, Goldin's (1995: 70) toy model of structural change comes in two variants – with and without social stigma. In the 'non-stigma case', Goldin shows that female labour supply is not necessarily be U-shaped.[4]

[2] We define reproductive tasks as including childrearing, care of the elderly and housework (amongst others).
[3] Here, social stigma has a utility value and appears in the household's utility function.
[4] To the extent that factory work is brawn-intensive, however, female labour supply may be U-shaped in the non-stigma case, too.

Other contributions have replicated the U-shaped pattern by considering the role played by fertility dynamics in the demographic transition. In a seminal paper building on Becker's (1960) early insights, 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') – a complement to physical capital. Men, however, have more 'brawn', and hence a comparative advantage in labour-intensive tasks. In poor economies, which use a labour-intensive technology, economic growth raises the male relative wage. The resulting income effect increases the demand for children, reducing FLFP. In richer, more capital-intensive economies, economic growth has a positive effect on the female wage, closing the gender wage gap and leading women to substitute out of childrearing and into market work.[5] As Galor and Weil acknowledge (1996: 384-5), simple extensions of the model can generate a non-monotonic U-shaped relationship.[6] These extensions were formalised in later contributions (Lagerlof, 2003; Kimura and Yasui, 2010).[7]

Implicitly, these models rely a set of gender-role cultural 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 depends (positively) on the female but not on the male wage, implying a pure income effect from a rise in the male wage. The intra-household allocation of care labour, however, is not the product of biological differences between men and women, and it may change across societies and cultures. Indeed, assuming a different allocation can modify the model's predictions. In societies where men and women contribute more equally to childrearing (so that the opportunity cost of having children depends also on the male wage), an increase in men's relative wages does not have a pure income effect on the demand for children. This is because a rise in male wages also has the effect of increasing the opportunity cost of childrearing, potentially offsetting any rise in fertility. Thus, in societies that start out with more equal gender norms, the dynamic path of female labour supply may be less strongly U-shaped.

Lastly, another argument, which goes back to Boserup (1970), links the Feminisation U to men's privileged access to education and technological knowledge.[8] There is no biological constraint

---

[5] The strong link between fertility behaviour and labour supply decisions is also emphasised by Bloom et al. (2009).

[6] Galor and Weil's (1996) choice of utility function leads to a model in which the income effect is never dominant. Thus, the relationship between income per capita and FLFP is positive and monotonic.

[7] While Galor and Weil 'assume that the only input required to raise children is time' (1996: 378), Lagerlof's (2003) version of the model allows for both time and goods costs of childrearing, which yields a strong positive effect on the demand for children (and hence a negative effect on FLFP) from rising male income. Kimura and Yasui (2010) explicitly incorporate a non-modern, home-production sector that competes for women's childrearing and market-labour time. This addition leads to an inverted N-shaped dynamic of fertility.

[8] The importance of gender differences in education is also implicit in Goldin's (1995) toy model.

on women and men attaining equal quantities of 'brains', except that gender-biased cultural norms favour the education of boys. Thus, even if men's comparative advantage in 'brawns' is rewarded less and less as an economy develops, men can still command a higher relative wage by virtue of their privileged access to human capital. This argument, which reinforces a U-shaped dynamic path of FLFP, is explicitly developed by Hiller (2014) in a model we discuss in detail in the following section. For now, we only note that the model's prediction of a U-shape pattern in FLFP once again originates in gender-biased cultural norms.

Evidence

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, female labour supply reaching 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, Cagatay and Ozler (1995) find statistically significant evidence of an *inverted* U-shaped relationship.[9]

Early empirical studies also reported time-series results based on individual countries. Focusing on the United States, Goldin (1995) argues that FLFP probably traced out a U-shaped pattern over time, reaching a bottom in the 1920s, when per-capita GDP 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 not be representative of a more general, cross-country pattern. Thus, 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 U. Their fixed-effects panel estimates reveal a more muted U-shape than the corresponding pooled model, with a much earlier (but statistically significant) turning point at 1600 US$ per capita (rather than 2550 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.

---

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

The best paper employing panel-data methods to date is by Gaddis and Klasen (2014). The authors employ more recent and comprehensive labour market data than either Luci (2009) or Tam (2011), dynamic panel-data estimation techniques (GMM), and internal instruments to correct for the endogeneity of GDP per capita. In a break from most of the previous literature, their results lead them to conclude that 'there is no clear evidence for the feminization U hypothesis from […] dynamic estimations' (2014: 660).

Gaddis and Klasen's (2014) findings are also 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 Swaminathan (2016) estimate dynamic GMM models and find no systematic evidence of a U-patterned relationship between state-level income and FLFP. Similarly, Roncolato's (2016) study of South Africa employs micro data to investigate the effect of economic development at the municipality level on women'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 have not properly investigated 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 lights 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 exert a modifying influence on the path of FLFP in the course of development. This paper addresses this missing link in the literature.

Heterogeneous U-Curves

To guide our empirical analysis, we primarily rely on Hiller (2014), who provides a formal explanation of the Feminisation U that explicitly models the evolution of cultural norms in the course of development.[10] To our best knowledge, this is the only paper that explicitly includes a potential source of heterogeneity in the relationship between economic development, gender norms, and FLFP.

---

[10] 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 and wage differentials between men and women, the model generates an S-shaped path of FLFP. Similarly, Raquel 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 US after 1970.

Hiller's model

Hiller (2014) considers a two-sex, overlapping generations model of the gender education gap in the course of development. The household's utility is derived from both wages (returns to education) and status. The level of education of daughters negatively impacts status if there is a strong cultural norm against female participation in the labour force. Parents maximise the household's utility by choosing how much to invest in their children's education, besides choosing a consumption level and a time allocation between labour and care work. Two factors determine the parental decision to educate daughters. The first one relates to the wage they will be able to command given the level of development and their level of education, while the second relates to status and works in the opposite direction.

The cultural norms regarding female education (as measured by the utility cost of non-compliance) are endogenous. As in Fernandez (2013), the strength of the norm depends on its strength in the previous time period (reflecting the intergenerational transmission of cultural attitudes within the family) as well as on the observed levels of female education in the same time period, allowing for the possibility that norms may be updated over time.

The model exhibits three steady-state equilibria. Income levels can increase with human capital accumulation, but this dynamic kicks in only when the economy crosses a given poverty threshold. Below this threshold, households are trapped in a 'poverty regime', in which both boys and girls receive no education and income is stagnant. Above this threshold is a 'gender inequality regime'. Parents will first invest in their son's education, as doing so does not negatively impact on their status. A male bias in the allocation of the household's education budget widens the gender productivity gap, lowering the relative female labour supply. As human capital increases, raising potential income for women, female education catches up through a classic income effect. In this 'interior regime'[11], the female labour supply grows with rising female productivity, while gender-biased cultural norms lose strength. Hiller (2014: 473) proves formally that along the convergence path from the poverty regime to the interior regime, the female labour supply is U-shaped.

In the model, initial productivity differences between uneducated men and women play a critical role in determining subsequent dynamics as they are the crucial factor biasing the allocation

---

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. For this reason, it cannot provide the basis for an empirical investigation of the U-curve.
Interestingly, if a woman's disutility of working is low to start with (because no cultural norm is violated), Fernandez's (2013: 492) model predicts a much flatter (less S-shaped) path of FLFP.

[11] . This is termed the 'interior regime', as both son's and daughter's optimal levels of education are interior rather than corner solutions (i.e. none of them is 0).

of the household's education budget towards boys. As noted by Hiller (2014: 474), 'it is indeed this initial inequality that is reinforced by educational choices and that constitutes the starting point of the *vicious circle* that leads an economy into a [gender inequality] trap'.[12]

The original gender productivity gap (denoted by *s*) relates to biological differences in physical strength. Yet, whether these differences are economically salient (giving rise to a *productivity* differential) depends on the extent to which the pre-industrial technology of production relies on physically demanding labour. The more a technology relies on 'brawn', the greater the initial male advantage in the labour market, setting the stage for an education bias. Indeed, 'as soon as *s* is positive (even arbitrarily small), there is room for an inegalitarian equilibrium. Obviously, other things being equal, a higher value of *s* indicates a greater likelihood that an economy will fall into the gender-inequality trap' (Hiller, 2014: 474).

Thus, the exact shape of the U is a function of *s*, leading to a family of U-curves that are heterogeneous with respect to gender productivity differences at $t = 0$. This feature of the model allows us to formulate a testable prediction, which we explore in our empirical analysis. When *s* is large, the path of female labour supply in the course of development is predicted to be strongly U-shaped; the smaller the *s*, the 'shallower' the U-curve. In the limiting case of $s = 0$ (that is, a technology where female and male labour are perfectly substitutable inputs), the path of female labour supply should be flat (i.e. uncorrelated with per-capita income).

What drives variation in the gender productivity gap across pre-industrial societies? Why is *s* high in some societies and zero in others? In his conclusion, Hiller (2014) speculates that the initial conditions of the model may depend on the extent to which early agricultural forms relied on physical strength, as discussed by Alesina et al. (2013). This is the argument that we explore further here.

Sources of heterogeneity: the plough

Referring to the seminal work of Boserup (1970), Alesina et al. (2013) report evidence that traditional agricultural forms influenced the historical gender division of labour, shaping the subsequent evolution of gender norms. The key difference is between plough-based agriculture and shifting hoe cultivation. Unlike the latter, the former is capital-intensive and requires significant upper body strength, putting a productivity premium on male labour. Thus, societies that adopted the plough developed an early specialisation of labour along gender lines. As a result, women were less involved in agricultural production (in soil preparation but also planting, crop tending and harvesting), and retreated into the home.

Alesina et al. (2013) use comparable ethnographic information to construct estimates of the fraction of the present-day countries' population with ancestors that traditionally practiced plough agriculture.

[12] Emphasis added.

This variable displays considerable cross-country variation. For instance, while 99 percent of Egypt's present-day population has ancestors that practiced plough agriculture, the share is 87 percent for India, 54 percent for South Africa and 0 for Sierra Leone. In our analysis, we use this variable as a proxy for parameter *s*. Thus, in countries with high historical plough use, we expect a highly U-shaped path of female labour supply. In countries that traditionally practiced hoe cultivation, by contrast, we expect to observe a (near-) flat path.

An important limitation of Alesina et al.'s plough variable, which the authors acknowledge (2013: 479) is that it does not provide information on the exact date of plough adoption. Yet, it is clear the introduction and diffusion of the plough began well before the onset of modern economic development and industrialisation – that is, well before countries began transitioning from the 'poverty' to the 'gender inequality' regime (see, for instance, Fussell, 1966 and Andersen et al., 2016). Thus, Alesina et al.'s (2013) variable can be taken as an appropriate measure of initial conditions.

Alesina et al. (2013) argue that, over time, the historical gender division of labour influenced the emergence of corresponding gender norms. In particular, 'plough societies' developed 'the cultural belief that the natural place for women is within the home' (2013: 475). Cultural norms and beliefs are inherently sticky and slow-moving across generations (Fernandez and Fogli, 2009), perhaps because of family socialisation dynamics (Fernandez et al., 2004).[13] Consistent with this view, Alesina et al. (2013) report evidence that countries populated by descendants of plough societies have less equal gender norms today, as implied by lower aggregate levels of FLFP and less favourable attitudes towards women's participation in the economy.

In Alesina et al.'s (2013) analysis, the shock triggered by the introduction of the plough, and the consequent re-allocation of labour along gender lines, prompts a permanent change in norms, and hence a downward shift in the female labour supply curve. Combining Alesina et al. (2013) with Hiller (2014), we can explore how the adoption of the plough influenced the *dynamics* of FLFP. Rather than simply shifting the supply curve, the differential adoption and diffusion of the plough may have *moderated* the speed and intensity of female labour supply dynamics in the course of development.

An implication of combining Hiller (2014) with Alesina et al. (2013) is that, in the early phases of development, the effect of the introduction of the plough on FLFP should be small or null. Intuitively, this is because household income needs to have exited the 'poverty regime' for the family's educational expenditure to rise above zero, leading to a widening of the gender productivity gap and a progressive decrease in FLFP (Hiller, 2014). This prediction is plausible and is consistent with

[13] In addition, Hansen et al. (2015) argue that plough adoption made possible the rise of a non-food-producing class whose members contributed to the advancement of language and writing. This class contributed to the formal codification and entrenchment of cultural norms, including those about gender roles (Hansen et al., 2015). Alesina et al. (2013), lastly, suggest that cultural persistence over time may also arise from a complementarity between norms and beliefs and industrial structure.

anecdotal evidence suggesting that in pre-industrial plough societies women did not abandon productive labour to specialise in childrearing and other reproductive tasks (Boserup, 1970: 91). Rather, the plough initially gave rise to 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, albeit primarily in home industries.[14]

Similarly to poor economies, in rich economies that are in the 'interior regime' equilibrium, the long-run effect of ancestral plough adoption are expected to have vanished. At that level of income, women have fully caught up to men's level of human capital, and the FLFP has risen back to the level observed in (otherwise equal) 'non-plough' countries, where the 'vicious circle' described by Hiller was much more muted all along (if operative at all). It is at middling levels of income (in countries in the 'gender inequality regime'), that a negative long-run effect of the plough should be observed. In middle-income countries with a 'plough legacy', the decline in FLFP implied by the Feminisation U (and powered by plough adoption) is likely to be near the U-curve's bottom. In sum, the effect of plough adoption (as a proxy for parameter *s*) on FLFP should be null or very small in very poor and very rich countries, and significantly negative in middle-income economies. In the rest of this paper, we test these implications empirically.

Empirical Strategy

We test our hypotheses empirically in a panel-data framework. The U-shaped relationship between FLFP and economic development is modelled using 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[15]:

$$FLFP_{it} = \rho FLPF_{it-1} + \alpha Plough_i + \beta_1 \ln GDPpc_{it} + \beta_2 \ln GDPpc_{it}^2 + \gamma_1 Plough_i \times \ln GDPpc_{it} + \gamma_2 Plough_i \times \ln GDPpc_{it}^2 + \varphi \boldsymbol{X}_{it} + \sigma_i + \tau_t + \varepsilon_{it} \qquad (1)$$

[14] Boserup (1970: 91) writes that 'in many developing countries women form a large part of the home industries' labour force, and not only in those [non-plough] countries of Africa, South East Asia and Latin America where women are active in market trading, but also in some [plough] countries where women take little or no part in trade, such as Morocco and Iran. In these countries, with their tradition of female seclusion, women comprise one-fourth and one-third respectively of own-account work and family aids in industrial occupations […]. This is because women who live in seclusion have no way of earning money except by working in home industries […]. The social ban on other activities compel secluded women to take to the only activity by which they can earn money without the loss of social esteem within their community.'

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

where i indexes countries and t time. $FLFP$ denotes the female labour force participation rate – the number of women in the labour force as a share of the total working-age (15-64) female population. According to the standard definition, which is closely aligned with 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 (i.e. wage work) is included, as is self-employment (if it produces a marketed product or service) and unpaid work that is performed to produce a good for auto-consumption (Klasen, 2018: 162). It is important to note that, on this definition, subsistence farming counts as labour force participation. By contrast, housework and care work (including childrearing and care of the elderly) do not count as market work. This definition is consistent with our theoretical framework, in which that are 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 FLFP data, downloaded from the World Bank's website, is originally obtained from the (most recent) sixth revision of the International Labour Organisation (ILO)'s *Estimates and Projections of the Economically Active Population* (EAPEP) database, and is based on ILO staff estimates. The EAPEP's sixth edition covers the period 1990-2019.[16] In 2013, however, the 19th Conference of Labour Statisticians, convened by the ILO, modified the definitions of work and labour force, leading to a break in the ILO time series on labour force participation (Klasen, 2018: 164). 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 percent (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.

In line with previous contributions (Tam, 2011; Gaddis and Klasen, 2014), equation (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. Thus, our specification is in line with previous studies that document the intergenerational transmission of cultural attitudes regarding married women's work (Fernandez et al., 2004; Fernandez and Fogli, 2009), and their 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.

[16] The data from EAPEP's sixth edition is considered to be more reliable than that coming from earlier editions (Gaddis and Klasen, 2014: 648), which have been used in previous empirical studies on the feminisation U (Luci, 2009; Tam, 2011).

$Plough$ refers to Alesina et al.'s (2013) original measure of historical plough use. It provides 'an estimate of the *fraction* of the population currently living in a [...] country with ancestors that traditionally engaged in plough agriculture' (2013: 486).[17] It 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 (25.5 percent) or none (30 percent) of their present-day inhabitants had plough-using ancestors, while only 12.5 percent of the sampled countries have a value of $Plough$ between 0.1 and 0.9. The data on income per capita 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.

The parameters of interest in equation (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 curvature observed in countries with ($Plough = 1$) and without historical plough use ($Plough = 0$). The estimated coefficient on $Plough$ (that is, α) does not admit a meaningful interpretation as it measures the effects of plough legacies in countries with a per-capita GDP of 1 US$ ($\ln GDPpc = 0$).

In a set of alternative specifications, we condition our estimates of the U's curvature on a full vector $\boldsymbol{X}_{it}$ of potential confounders, which are described in more detail in Appendix A. The inclusion of these variables (which may be time-varying or country-specific) is justified by a number of alternative (or complementary) explanations of women's economic empowerment that are discussed in the recent literature. Conceptually, we classify these variables in three broad groups.

First, our 'geo-anthropological controls' are a sub-set of Alesina et al.'s (2013) control variables. They are intended to capture historical differences between the ancestral societies that had adopted plough agriculture and those that had not. Some of these characteristics (e.g. the soil's suitability for agriculture, the historical use of large domesticated animals) may have led to 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. Thus, their omission could lead to inconsistent estimates of α and, consequently, of γ_1 and γ_2.[18]

Second, a set of 'historical control variables' accounts for a number of more recent historical determinants of female participation in the economy. Some of these factors may be correlated with either ancestral plough adoption or contemporary economic development (or both), so that their omission would introduce bias in the estimates of the γ's and β's. In this group, we control for (in reverse chronological order): the experience of state socialism in Eastern Europe and Central Asia, which was accompanied by sustained efforts to promote gender equality across all spheres of social

[17] Emphasis added.

[18] By construction, $Plough_i \times \ln GDPpc_{it}$ and $Plough_i \times \ln GDPpc_{it}^2$ are highly correlated with $Plough_i$.

life (Lippman and Senik, 2018); the percentage of a present-day country's population that is of European descent, to account for the effects of the introduction of cash-crop agriculture by colonial settlers (Boserup, 1970; Beneria and Sen, 1981); the identity of a country's former colonial power, which might have transplanted its culture and/or labour market institutions to its former colonies; a country's exposure to the transatlantic slave trade, which gave rise to female-biased sex ratios, allowing women to take up occupations that were traditionally the preserve of men (Teso, 201); and, finally, an index of state antiquity, since early statehood has been associated with the emergence and consolidation of patriarchal norms (Lerner, 1986).

Lastly, our 'contemporary controls' account for a range of present-day factors which may be correlated with both economic development (or, less likely, historical plough use) and labour market outcomes for women. These include: armed conflict, which typically creates labour shortages and new opportunities for women (Goldin and Olivetti, 2013); a country's dependence on oil exports, which tend to crowd out female-intensive tradable sectors (Ross, 2008); aid dependence, since aid agencies may favour male-biased technology transfer, feeding negative gender stereotypes (Boserup, 1970; Jaquette and Summerfield, 2006); religious denomination, which has been shown to shape attitudes towards working women (Guiso et al., 2003); the extent to which a country's society and culture is globalised, and hence exposed to the diffusion of new values (Potrafke and Ursprung, 2012); and the quality of democracy (Beer, 2009).

σ_i captures any other unobserved, time-invariant individual effects that may be correlated with either income levels or plough use (e.g. other cultural, or geographical characteristics that are not included in vector $\boldsymbol{X}_{it}$). In our baseline analysis, we assume that after controlling for $\boldsymbol{X}_{it}$, income levels are exogenous to FLFP (a standard assumption in the feminisation U literature), so that we can ignore σ_i. In the robustness analysis, however, we relax this assumption, allowing for the possibility that per-capita income may be correlated with σ_i or ε_{it}. τ_t denotes a full set of time-period dummies that capture labour market shocks affecting all countries simultaneously – the likes of economic crisis and business cycle fluctuations[19]. 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.[20]

It is well known that OLS estimators of dynamic panel-data models are subject to finite-sample bias (Nickell, 1981). To address this problem, previous studies of the feminisation U (e.g. Gaddis and Klasen, 2014) employed Arellano and Bond's (1991) 'difference' GMM estimator. Yet, when the outcome variable is highly persistent (as is the case with $FLFP$), 'difference' GMM has been shown to

[19] See, for instance, Albanesi and Sahin (2018).
[20] See Roodman (2009: 121)

perform poorly (Blundell and Bond, 1998; Soto, 2009).[21] Thus, we estimate equation (1) using the 'system' GMM estimator proposed by Blundell and Bond (1998). Instead of transforming (e.g. differencing) the regressors to purge the individual effects (σ_i), and then using lagged levels as internal instruments (as in 'difference' GMM), 'system' GMM estimates the equation in levels, while transforming the instruments to make them exogenous to σ_i (Roodman, 2009: 114). In simulations, Soto (2009) shows that 'system' GMM has a lower bias and higher efficiency than both OLS and 'difference' GMM. An additional advantage of the 'system' estimator is that it can identify time-invariant regressors (e.g. $Plough_i$), which would otherwise disappear after first-differencing.[22]

Furthermore, instead of transforming the model's instruments by first-differencing (Δ), we choose to employ the 'forward orthogonal deviations' (FOD) transform proposed by Arellano and Bover (1995), which was shown to outperform the first-difference transformation in terms of both bias and efficiency (Hayakawa, 2009). We employ the standard two-step estimation process with Windmeijer-corrected cluster-robust errors. In Appendix B, however, we show that our main results are robust to using different versions of the 'system' GMM estimator. In all the models reported below, we restricted the instruments' lag ranges (or alternatively 'collapsed' the instrument matrix) with a view to: 1. Ensuring that the number of instruments is lower than the number of countries[23]; 2. Ensuring that the instruments are jointly valid (as shown by Hansen's test of the over-identifying restrictions) and the disturbance term ε_{it} is not subject to AR(2) autocorrelation, which would render some lags invalid as instruments (Roodman, 2009).

Main Results

GMM estimates of (alternative versions of) equation (1) are reported in Table 1. Model 1 is a simple specification without interaction terms that is intended to replicate previous findings in the literature. The coefficients on both $lnGDPpc$ and $lnGDPpc^2$ are 'correctly' signed and highly significant, indicating that for the 'average' country the path of FLFP is U-shaped, even after controlling for the legacy of the plough. This is line with previous results from the feminisation U literature (Luci, 2009; Tam, 2011). The coefficient on $Plough$ enters as negative and highly significant, even conditional on the income terms, with a coefficient implying a long-run effect of -14.546[24], which is close to Alesina et al.'s (2013: 494) estimate of -12.40 based on cross-section data.[25]

---

[21] This is because if the outcome variable changes slowly, 'past levels convey little information about future changes', leading to a weak instrument problem (Roodman, 2009: 114).
[22] Asymptotically, however, their inclusion does not affect the parameter estimates for the other regressors (Roodman, 2009: 115).
[23] This is a commonly accepted rule of thumb.
[24] $= \alpha/(1-\rho) = -0.349/(1-0.976)$
[25] See their table IV.

[Table 1]

Model 2-8 add two multiplicative interaction terms between $Plough$ and the income terms, allowing countries to have different paths of FLFP depending on their own history of ancestral plough use. Model 2 omits X_{it}, while models 2-5 enter our battery of control variables group by group. Model 6 is a kitchen-sink specification that includes all the potential confounders described in section 4 simultaneously. To avoid overfitting and to preserve sample size, model 7 is parsimonious specification that conditions on a more limited set of (six) control variables (those that enter significant in at least one of specifications 3-6) from across our three groups.[26] This is our preferred specification. To address the concern that variation in $Plough$ may be driven by broad differences between regions (e.g. Africa vs. Europe), model 8 also adds a full set of continent dummies (Alesina et al., 2013: 486).[27] As such, this last specification only uses within-continent variation for identification.

Models 1-8 are diagnostically sound.[28] Hansen's C test never rejects the null that the internal instruments used by GMM are jointly valid. Also, based on Arellano-Bond tests (Roodman, 2009: 119-121), we find no evidence of AR(2) autocorrelation in the residuals. The estimated coefficients on the lagged dependent variable (ρ) are always statistically significant and very large in magnitude, indicating that the gender norms that influence women's labour supply decisions are subject to strong inter-temporal persistence.[29] For comparison, Appendix C estimates models 2 and 7 (Table 1) using OLS rather than GMM. Models are presented both with and without country fixed-effects (FE). Reassuringly, the values of ρ estimated by GMM (e.g. 0.974 based on model 7) fall within the pooled-FE bracketing range defined by the corresponding OLS specifications (0.992-0.935 for model 7), confirming the plausibility of our GMM results.[30]

Coming now to our coefficients of interest, we note that once we allow for heterogeneous paths of FLFP (models 2-8), coefficients β_1 and β_2 decline in magnitude and lose statistical significance.[31] In countries with no tradition of plough use ($Plough_i = 0$), the average path of FLFP is statistically indistinguishable from a flat line. The estimated coefficients (γ_1 and γ_2) on the multiplicative terms, however, are 'correctly' signed and statistically significant across models 2-8.[32]

[26] These variables are: soil suitability, tropical climate, ancestral dependence on hunting for subsistence, number of years of socialist rule, state antiquity and oil dependence.
[27] Yet, a t-test cannot reject the null that the continent dummies are jointly equal to zero (*p*-value = 0.219).
[28] The only exception is (the extensively specified) model 6, where the number of instruments slightly exceeds the number of panels. This was the only specification that 'passed' the Hansen and AR(2) tests.
[29] The estimates, however, are always statistically distinguishable from one. Based on model 7, for instance, we can reject the null that $\rho = 1$ at the 1 percent level (*p*-value = 0.006).
[30] See Roodman (2009: 103) for a detailed discussion of these diagnostic checks.
[31] Across models 2-8, the coefficients on the individual income terms are also always *jointly* insignificant.
[32] In model 2, the coefficients on the interaction terms are only significant at the 10 percent level, which could be due to multicollinearity. In alternative specifications that omit either $\ln GDPpc_{it}$ or $\ln GDPpc_{it}^2$ (not reported, but available upon request), γ_1 and γ_2 are similar in magnitude but much more precisely estimated.

This finding indicates that the female labour supply in plough countries follows a statistically significantly *different* path from that of non-plough societies. For countries in which $Plough_i = 1$, the estimated relationship between per-capita income and FLFP is captured by the parameters reported in Panel A.[33] These estimates unequivocally describe a U-shaped pattern, with a positive a significant coefficient on the squared term, and a negative and significant coefficient on the linear term.

[Figure 3]

To aid interpretation, Figure 3 plots the estimated path of FLFP for a plough (red) and a non-plough (blue) country with 'average' observable characteristics.[34] The broad pattern – a flat path for $Plough_i = 0$ and a U-shaped path for $Plough_i = 1$ – is consistent throughout. The U-shaped pattern experienced by plough societies, however, is estimated to be particularly pronounced by specifications that include (some or all of) the variables from our group of 'historical controls' (models 4 and 6-8). In model 7, the turning point of the curve implies that, in expectation, women's participation in the labour force reaches a minimum as a country achieves a per-capita income of 5,195 US$ PPP (s.e. = 946 US$).[35]

In Appendix B, we show that these results are not sensitive to any of our GMM estimation choices. They hold if we use the one-step (rather than the two-step) estimation technique, or a first-difference (Δ), rather than the FOD, transformation. In addition, the results are robust to reducing substantially the instrument count by 'collapsing' the instrument matrix – a procedure which is equivalent to using standard (IV-style) instruments.

Based on the estimates in Table 1, we can also test our (related) hypothesis that the effect of plough adoption should be zero (or very small) in very poor and very rich economies, and significantly negative in middle-income economies. The marginal effects of historical plough use are reported in Table 2 for three possible per-capita GDP levels: US$ 245, US$ 8100, US$ 60,000, which are close to the sample minimum, mean and maximum, respectively.[36] The findings are in line with our theoretical priors – namely, that in very poor countries the detrimental effect of plough adoption on gender norms has yet to show through, while in very rich countries this negative effect has been completely offset by the forces of development. Accordingly, for countries at the level of development of Liberia (US$ 245) or Luxembourg (US$ 60,000) in 1995, we can always reject the null that $\partial FLFP_{it}/\partial Plough_i < 0$. In some specifications, we can even reject the null that

---

[33]For $Plough_i = 1$, equation (1) becomes $FLFP_{it} = (\beta_1 + \gamma_1)\ln GDPpc_{it} + (\beta_2 + \gamma_2)\ln GDPpc_{it}^2 + \alpha + \mu + \varepsilon_{it}$, where μ stands for a linear combination of the control variables evaluated at their sample means.

[34] In other words, we plot partial predictions with all covariates except for the income terms being held constant at their sample means.

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

[36] $\partial FLFP_{it}/\partial Plough_i = \alpha + \gamma_1 \ln GDPpc_{it} + \gamma_2 \ln GDPpc_{it}^2$

$\partial FLFP_{it}/\partial Plough_i = 0$, concluding that at these extreme levels of per-capita income the long-run effect of the plough is to *increase* female labour supply. These positive and significant coefficients run counter to our theoretical expectations, but may result from the imposition of an excessively rigid functional form on a region with sparse data.[37] Consistently with this interpretation, the estimates are very unstable across different specifications. We return to this issue in the robustness analysis.

[Table 2]

In contrast to this null (or even positive) effect of the plough at very low and very high levels of per-capita GDP, the impact of historical plough use is always *negative* and very precisely estimated at middling levels of per-capita GDP. According to model 7, the long-run effect of the plough for a country with a per-capita GDP of US$ 8100 is equal to -22.731. Thus, if a 'plough country' such as Ukraine had not had any history of traditional plough use, its FLFP rate in 2010 would be almost 23 percentage points higher – that is, 85 rather than 62 percent, implying that the Ukrainian labour force would count 3.85 million additional women.[38] At this level of development, the magnitude of the plough effect is almost twice as large as the magnitude reported by Alesina et al. (2013: 494). Their estimates, which are not allowed to vary by per-capita GDP and are best interpreted as 'averages', are thus missing substantial heterogeneities in the data.

In sum, the GMM estimates broadly confirm our theoretical expectations. The path of FLFP in the course of development is U-shaped, as suggested by the feminisation U hypothesis, *but* only in countries that traditionally practiced plough agriculture. Relatedly, the legacy effects of the plough on female labour supply (i.e. the 'distance' between the two curves reported in Figure 3) are only significantly negative in middle income economies.

Robustness Analysis

A possible concern with our estimates is that they may be confounded by short-term cyclical fluctuations, e.g. those induced by economic recessions (Albanesi and Sahin, 2018). In an alternative specification, we smooth out the influence of cyclical effects by dividing the panel into 5-year intervals (1995, 2000, 2005, 2010).[39] An additional advantage of this specification is that it reduces the number

[37] Below the lower intersection point of the two U-curves (which is at an income level of US$ 1,080 per head, or $e^{6.98}$), there are as few as four data points referring to *plough* countries (Cambodia and Tajikistan in the mid-1990s). Above the higher intersection point (at an income level of US$ 24,151, or $e^{10.1}$), there are only 9 *non-plough* countries (e.g. Australia, Saudi Arabia, Equatorial Guinea, etc.) providing data points.

[38] In 2010, the female population of Ukraine aged 15-64 was 16.76 million (World Bank, *World Development Indicators,* 2020).

[39] Thus, for instance, for an observation dated 2005, the lagged dependent variable refers to the year 2000. Observations dated 1990 are dropped.

of time periods while holding the number of countries constant, making the dataset closer to the panel form ('small T, large N') for which GMM estimators were originally designed. As a result, the instrument count is lower and the model diagnostics somewhat improved (see Table 3). The estimates, reported in column 1 of Table 3, are qualitatively consistent with our previous findings. Indeed, the U-shaped path is more pronounced than implied by our previous results, and the turning point somewhat higher (6,926 US$, s.e. = 1,025 US$). As expected, the estimate of the autoregressive coefficient ($\hat{\rho} = 0.692$) is considerably lower than the one obtained from a panel pooled over consecutive years (0.974 in model 7, Table 1).

[Table 3]

As an alternative way to examine the modifying influence of the plough on the evolution of FLFP, we drop the interaction terms and estimate our regression equation separately for a sample of plough ($Plough_i > 0.8$) and non-plough ($Plough_i < 0.2$) economies (columns 2 and 3, Table 3). Although the influence of the plough is now identified from dichotomous (plough vs. non-plough) rather than continuous variation, this approach is more flexible in that it allows the estimated coefficients of *all* variables (not just per-capita income) to change across the two groups.[40] The results are qualitatively consistent with the estimates reported in column 1 (Table 3). In fact, the contrast between the paths of FLFP in plough vs. non-plough countries is even more pronounced (2.915-0.468 = 2.447, instead of 1.848 in model 1).

[Figure 4]

Some (though not all) of the models summarised in Table 2 produce significantly *positive* coefficients for the effect of traditional plough use in very low- and very high-income countries – a theoretically puzzling result. We suspect that this effect may result spuriously from the use of an excessively rigid functional form for the relationship between FLFP and economic development. In particular, a polynomial approximation may not fit the data well at the tails of the income distribution, where the data points are sparse.

To check this possibility, we re-estimate models 2 and 3 in Table 3 using a more flexible, non-parametric function of per-capita GDP. Specifically, we replace the quadratic polynomial with a basis of Schoenberg B-splines, and estimate the resulting semi-parametric model by 'system' GMM

[40] This approach is equivalent to interacting all the model covariates with a dummy variable that takes the value one if $Plough_i > 0.8$ and the value 0 if $Plough_i < 0.2$, dropping the observations for which $0.2 < Plough_i < 0.8$.

(Newson, 2012). The estimated coefficients, which are not easy to interpret on their own, are not reported in full, but the predicted paths of FLFP are shown in Figure 4. Panel A, which serves as a benchmark, plots the U-curves implied by models 2 and 3 in Table 3; Panel B plots the corresponding semi-parametric estimates. The results are reassuring. Between an income level of US$ 3,000 ($= e^8$) and US$ 22,000 ($= e^{10}$) per capita, the paths are very similar across the two specifications. Yet, at very low and very high income levels, the positive effect of traditional plough use – the distance between the two U-curves (see Panel A) – is significantly reduced when the functional form is allowed to be flexible (see Panel B). An additional attractive feature of the semi-parametric specification is that conditional on plough use, the upward-sloping branch of the curve is now S-shaped, as predicted by Fernandez (2013).

[Table 4]

The theoretical U-curve results from the *joint* dynamics of per-capita income and gender inequality (Galor and Weil, 1996; Hiller, 2014). Thus, in the empirical analysis, we are not concerned about issues of reverse causality. That said, the observed relationship does not identify the theoretical U-curve if (conditional on $\boldsymbol{X}_{it}$) per-capita income is still correlated with unobserved individual effects (σ_i), or with other time-varying determinants of FLFP that are omitted from the regression, and hence subsumed in ε_{it}.

To assess the extent to which our results are affected by this potential source of endogeneity, we perform two robustness checks. First, we employ Arellano and Bond's (1991) 'difference' GMM estimator, which applies the first-difference transformation on eq. (1), purging the influence of *all* time-invariant individual effects. Although this estimator has been shown to perform poorly when the outcome variable is close to a random walk (Blundell and Bond, 1998; Soto, 2009), we are reassured by the fact that FLFP is less persistent in a 5-year than in a yearly panel. The results are shown in Table 4, which compares the 'system' (column 1) with the 'difference' estimates (column 2) for the same estimation sample.[41] Differencing out the individual effects does not alter our main conclusions. In fact, conditional on historical plough use, the estimated path of FLFP is now even more markedly U-shaped.

[Table 5]

[41] Since in 'difference' GMM, the regression equation is first-differenced, the first observation of each individual is lost.

Second, we split the sample in two groups (plough and non-plough countries) and instrument for per-capita GDP and its square in GMM style (as in Gaddis and Klasen, 2014). This approach has the advantage of correcting for any endogeneity bias resulting from the omission of time-varying determinants of FLFP, but the disadvantage of removing from the estimates any reverse causal effect from FLFP to income, which is part of the theoretical U-curve. Table 5 compares a set of models that treat income as exogenous (columns 1 and 2) with corresponding specifications that instrument for income (columns 3 and 4). Conditional on traditional plough use, the estimated relationship between FLFP and per-capita income remains significantly U-shaped (at the 10 percent level) even after instrumenting for income. Yet, the curve is now much more shallow (the estimated coefficient on $\ln GDPpc_{it}^2$ in model 3 drops by about a third compared to model 1), suggesting that the relationship observed in model 1 may indeed reflect a two-way causal effect.

Alternative Explanations

Here, we test the plough hypothesis against rival explanations. In our theoretical framework, which builds on Hiller (2014), the evolution of FLFP in the course of development depends critically on initial conditions – specifically, the productivity differential between men and women in pre-industrial societies. We argue that this differential arose from the introduction of the plough. Yet, conceivably, other historical events could have played a role in modifying the U-shaped path of FLFP. If these events 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 modifying effect on the shape of the U-curve to the plough.

With this in mind, we consider two alternative factors that could affect how FLFP evolves in the course of development. 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 gender inequalities could pre-date the introduction of the plough. Rather, these authors emphasise the transition to sedentary agriculture more generally as a crucial historical event that: a. translated into higher fertility; and b. created a premium on male brawn, promoting the emergence of a gendered division of labour and, over time, a system of norms and beliefs downgrading the status of women.[42] Latching this argument onto Hiller's (2013) framework, we conjecture that it may have been the Neolithic revolution, rather than the adoption of the plough, to open up a gender productivity differential, shaping the subsequent evolution of FLFP. In societies with earlier Neolithic revolutions, the exit of women from the labour force could have begun earlier, leading to a more markedly U-shaped path of FLFP.

---

[42] Hansen et al. (2015: 377) also report evidence that the Neolithic transition led to the emergence of a non-food-producing class of intellectuals and rulers that played a key role in codifying and enforcing cultural norms.

Second, we examine the potential modifying influence of the cultural norms that are ingrained in language. 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). Thus, the degree to which a language grammatically emphasises gender 'acts as a [stable] cultural marker for historical gender norms' (ibid., se Shoham and Lee, 2018 for a review of the literature). 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 relevantly for us, Hicks et al. (2015) and Gay et al. (2018) show that female immigrants in the US who speak a more strongly gendered language (e.g. Arabic, Spanish) also participate less in the labour force, and if they participate, work less, than migrants whose mother tongue is less strongly gendered (e.g. Mandarin, Armenian). Thus, we consider the possibility that the moderating influence that we attribute to the plough may actually be picking up the role played by the historical events that gave rise to the cultural norms crystallised in language.

[Table 4]

To test these two competing hypotheses, in Table 4 we add two additional country-level variables to our reference specification (i.e. model 7 in Table 1): the number of years of settled agriculture in 1500 AD (Putterman and Trainor,2006), and the Gender Intensity Index (GII) developed by Gay et al. (2013) to measure the degree of grammatical gender marking (see Appendix A). The variables are entered both individually (odd-numbered columns) and interacted with per-capita GDP and its square (even-numbered columns). In all cases, the additional variables enter as insignificant, while the estimated coefficients on $Plough$ and its interaction terms remain unaltered. These findings reassure us that the plough variable (and its interactions with the income terms) are unlikely to be capturing other omitted influences. They also provide suggestive evidence that it was specifically the adoption of the plough, rather than a combination of factors, that drove a wedge between the productivity of men and women, modifying the subsequent evolution of FLFP.

Causal Mechanisms

To complete our analysis, we provide some evidence of the causal mechanisms underlying our results. As mentioned earlier, the theoretical literature suggests three mechanisms as possible explanations of a U-shaped path of FLFP in the course of development – structural change from an agricultural to an industrial and service-based economy (Goldin, 1995; Ngai and Olivetti, 2015); the fertility transition

(Galor an Weil, 1996); and changes in the gender education gap (Hiller, 2014). In Hiller (2014), the extent to which the education mechanism is operative depends on initial conditions, which we link to a tradition of plough agriculture. In section 2, we also suggested that the structural change and fertility mechanisms may not work as expected in the absence of gender-biased cultural norms, which again may be the legacy of traditional plough use.

[Table 7]

Here, we test: 1. the extent to which each of these three mechanisms accounts for the feminisation U; and 2. whether each causal mechanism is activated, or at least intensified by the legacy of plough agriculture. The results are presented in Table 7. Our strategy consists of adding to a benchmark model (column 1) based on our preferred specification (model 7 in Table 1) a set of control variables that capture the mechanisms linking per-capita income to FLFP (see Appendix A for details on data sources). To measure the structural composition of the economy, we include the share of agricultural, manufacturing and services gross value added (GVA) in total GDP (column 2). To capture fertility dynamics, the model reported in column 3 includes a measure of the number of children born to a woman – that is, the total fertility rate. Lastly, to account for the education mechanism, model 4 adds UNESCO's Gender Parity Index, which measures the ratio of girls to boys enrolled at primary level. Model 5 includes all three sets of controls simultaneously.

These variables capture a set of factors that shape opportunities (e.g. availability of 'clean' service-sector jobs) and constraints (e.g. childcare obligations) faced by women deciding whether or not to join the labour force. These opportunities and constraints may arise to varying degrees at different levels of development (e.g. service-sector jobs are typically more widely available in advanced economies), explaining the path of FLFP. Yet, since our theoretical expectation is that opportunities and constraints only matter *as such* in the presence of gender-biased cultural norms, each variable is entered individually and also interacted with the $Plough$ variable. (For instance, having more children, or the absence of 'clean' service-sector jobs, only act as a constraint on women's labour-force participation in the presence of norms that allocate care work primarily to women, and that assign stigma to factory work). If mechanisms 'X' is operative, we expect the inclusion of variable 'X' in the regression to control away (or at least mute) the estimated relationship between per-capita GDP and FLFP.[43]

In line with our theoretical priors, the inclusion of each of these additional variables, controls away part of the mechanism linking per-capita income to FLFP in plough countries, making the

[43] This is because, if mechanism 'X' is operative, variable 'X' should be associated with both $\ln GDPpc$ and FLFP.

estimated relationship less U-shaped. Specifically, controlling for structural change, fertility and the education gap individually leads to the estimated coefficient on $\ln GDPpc^2$, conditional on $Plough = 1$ (Panel B), to drop by about 20, 28 and 5 percent, respectively. When all three mechanisms are controlled for simultaneously, the coefficient on $\ln GDPpc^2$ drops by more than 53 percent and loses statistical significance altogether (as does the coefficient on $\ln GDPpc$). By contrast, in non-plough countries (Panel A) the estimated coefficients on the income terms are practically unaffected by the inclusion of these additional variables, and the path of FLFP in the course of development is always statistically indistinguishable from a flat line.

Taken together, these findings provide suggestive evidence in support of two conclusions. First, all three mechanisms proposed in the literature as explanations of the feminisation U contribute to generating a U-shaped path of FLFP in the course of development. Collectively, they are likely to account for most of the observed U-shaped relationship between per-capita income and FLFP (in plough countries). Specifically, the evidence is consistent with fertility dynamics 'doing the bulk of the work' in tracing out the U-shaped path of FLFP in the course of development, consistent with the coefficient on $\ln GDPpc^2$ dropping by more than one quarter when the total fertility rate alone is included in the regression. By contrast, the role of the education mechanisms appears to be at best secondary or complementary. Second, the evidence is only consistent with these mechanisms being operative in societies that inherited highly gender-biased cultural norms – namely, plough societies. Where cultural norms are more gender-equal to start with – that is, in non-plough societies – these mechanisms do not appear to be driving the relationship between per-capita income and FLFP, which is null regardless.

Conclusion

Although it enjoys the status of a 'stylised fact' in development economics, the feminisation U hypothesis has received at best mixed confirmation in the empirical literature, especially in recent contributions that employ panel-data estimation techniques (e.g. Gaddis and Klasen, 2014) and in sub-national studies. This paper suggests a partial solution of this empirical controversy by identifying an important source of heterogeneity in the shape of the U-curve. It does so by integrating the feminisation U literature with recent work on the long-run historical determinants of gender-role cultural norms.

Our findings provide evidence of substantial heterogeneities in the shape of the feminisation U-curve. Based on recent theoretical work (Hiller, 2014), we conjecture that for a given country, the path of FLFP in the course of development depends critically on initial conditions – specifically, on the agricultural production technology the country employed in pre-historic times. To test this hypothesis empirically, we draw on the seminal work of Boserup (1970) and Alesina et al. (2013). Their argument focuses specifically on whether pre-industrial agriculture relied on the plough, which was physically demanding and hence bestowed a productivity and wage premium on men, or on smaller hand-held tools such as the hoe or the digging stick, which could be operated indifferently by either men or women.

We find that the more a present-day country's ancestors relied on traditional plough agriculture, the more the country's expected path of FLFP is U-shaped. In countries with no tradition of ancestral plough use, the path of FLFP is indistinguishable from a flat line. These results are extremely robust and unlikely to be driven by omitted influences. We also present evidence consistent with the mechanisms that explain the feminisation U (structural change, fertility dynamics, gender gaps in education) being operative in plough countries only.

Our findings have important implications for both the feminisation U literature in development and labour economics, and for the more recent literature on the origins, transmission and evolution of cultural norms. As for the former, our findings 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. As for the latter, our results point to an important qualification of the '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). This legacy, however, is likely to be considerably more nuanced than previously suggested.[44] Rather than leading to a permanent *shift* in gender norms and female labour supply, our evidence suggests that the plough shock had the effect of *modulating* the evolution of labour supply (and perhaps, attitudes and beliefs about working women) in the course of development. 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 an economy attains high-income status.

Our findings raise important questions about the interaction of culture and structure in shaping development outcomes. They also point to further avenues of research. While we show that (pre-)historical legacies can modify the effects of development on labour market outcomes for women, we have not presented direct evidence of a cultural transmission channel. Thus, it remains to be seen whether the evolution *of* cultural beliefs and values concerning the appropriate economic role of men and women is also U-shaped in the course of development, as Hiller's (2014) model implies. Further research will also have to investigate the impact of the timing of plough adoption, as well as how subsequent technological innovations (e.g. the invention of the heavy plough in medieval Northern Europe) influenced labour market outcomes for women.

[44] Indeed, Alesina et al. (2013: 500) seem to be aware of this when they note that 'despite the evidence of persistence over time on average, […] there are well-documented exceptions to this rule'.

References:

Alesina, A., Giuliano, P., Nunn, N. (2013) “On the origins of gender roles: Women and the plough” Quarterly Journal of Economics, 128 (2), pp. 469-530.

Albanesi, S., Sahin, A. 2009. The gender unemployment gap. Rev. Econ. Dyn. 30, 47-67

Andersen, T.B., Jensen, P.S., Skovsgaard, C.V. 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, 1995. Another look at the instrumental variable estimation of error-components models, J. Econometrics 68, 29-51

Becker, Gary S. "An Economic Analysis of Fertility," in Ansley J. Coale, ed., Demographic and economic change in developed countries. Princeton, NJ: Princeton University Press, 1960, pp. 209-40.

Beer, C. 2009. Democracy and Gender Equality. St. Comp. Int. Dev. 44, 212-227

Beneria, L. 1979. Reproduction, production and the sexual division of labour. Camb. J. Econ. 3, 203-225

Beneria, L., Sen. G. Accumulation, Reproduction and “Women’s Role in Economic Development”: Boserup Revisited. Signs: J. Women in Culture and Soc. 7(2), 279-298

Bloom, D.E., 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. Econometrics 87, 115-143

Borcan, O., Olsson, O., Putterman, L. 2018. State History and Economic Development: Evidence from six Millennia, 23, 1-40

Boserup, E. (1970) Woman's Role in Economic Development. London: George Allen and Unwin Ltd

Çağatay, N. and Ş. Özler (1995) “Feminization of the labor force: The effects of long-term development and structural adjustment”. World Development, Volume 23, Issue 11, pages 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/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, J.D. 1975. The labor force in economic development: A comparison of international census data, 1946.1966. Princeton University Press, Princeton

Lagerlof, N. P. 2003. Gender equality and long-run growth. J. Econ. Growth. 8, 403-426

Fernandez, R., Fogli, A., and Olivetti, C. (2004), Mothers and Sons: Preference Formation and Female Labor Force Dynamics, Quarterly Journal of Economics 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. The American Economic Review. Vol. 103, No. 1, pp. 472-500.

Fussell, G.E. 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. and Weil, D. (1996), The Gender Gap, Fertility, and Growth, American Economic Review 86 (3), 374–387.

Gay, V., Hicks, D.L., Santacreu-Vasut, E. and 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
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
Guiso, L., Sapienza, P., Zingales, L. 2003. People's opium? Religion and economic attitudes. J. Mon. Econ. 50(1), 225-282
Hansen, C.W., Jensen, P.S., Skovsgaard, C.V. (2015) "Modern gender roles and agricultural history: the Neolithic inheritance". Journal of Economic Growth, 20 (4), pp. 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. and Maoz, Y. (2002), Women's Labor Force Participation and the Dynamics of Tradition, Economics Letters 75, 193–198.
Hicks, D. L., Santacreu-Vasut, E., & Shoham, A. (2015). Does mother tongue make for women's work? linguistics, household labor, and gender identity. Journal of Economic Behavior & Organization, 110, 19–44.
Hiller, V. (2014) "Gender Inequality, Endogenous Cultural Norms, and Economic Development" Scandinavian Journal of Economics 116(2), pages 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, J.S., Summerfield, G. (eds.) 2006. Women and gender equity in development theory and practice: Institutions, resources and mobilization. Durham and London: Duke University Press
Kimura, M. and Yasui, D. (2010), The Galor–Weil Gender Gap Model Revisited: From Home to Market, Journal of Economic 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. & H. Swaminathan (2016) Economic Development and Women's Labor Force Participation in India, Feminist Economics, 22:2, 168-195
Lerner, G. 1986. The creation of patriarchy. New York: Oxford University Press
Lippmann, Q., Senik, C. 2018. Math, girls and socialism. J. Comp. Econ. 46(3), 874-888
Liu, A., Shair-Rosenfield, S., Vance, L.R., 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" International Journal of Innovation and Sustainable Development (IJISD), Vol. 4, No. 2/3.
Mammen, K., Paxson, C. 2000. Women's Work and Economic Development. J. Econ. Perspect. 14(4), 141-164
Newson, R. 2012. Sensible parameters for univariate and multivariate splines. Stata J. 12(3), 479-504
Ngai, L.R., Olivetti, C. 2015. Structural Transformation and the U-Shaped Female Labor Supply, unpublished manuscript
Nickell, S. 1981. Biases in Dynamic Models with Fixed Effects. Econometrica, 49(6), 1417-1426
Nunn, N. 2008. The Long-Term Effects of Africa's Slave Trades, Q. J. Econ. 123(1), 139-176

Nunn, N. Puga D. 2012. Ruggedness: The Blessings of Bad Geography in Africa, Rev. Econ. Stat., 94(1), 20-36

Pampel, F.C., Tanaka, K. 1986. Economic development and female labor force participation: A reconsideration. Soc. Forces 64 (3), 599-619

Potrafke, N., Ursprung, H. W. 2012. Globalization and gender equality in the course of development. Eur. J. Pol. Econ. 28, 399-413

Putterman, L., Trainor, C.A. 2006. Agricultural transition year country data set, Providence: Brown University

Roncolato, L. (2016) "The Feminization U in South Africa: Economic Structure and Women's Labor Force Participation" Feminist Economics, 22 (4), pp. 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. 2009. Oil, Islam and Women. Am. Pol. Sci. Rev. 102(2), 1-18

Shoham, A., Lee, S.M. 2018. The Causal Impact of Grammatical Gender Marking on Gender Wage Inequality and Country Income Inequality. Business & Soc. 57(6), 1216-1251

Sinha, J.N., 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

Uberti, L.J. 2018. Corruption in Transition Economies? Ottoman, Socialist or Structural? Econ. Sys. 42, 533-555

Figure 1 – The Feminisation U with 95% C.I.: Cross-country evidence (1990-2013)

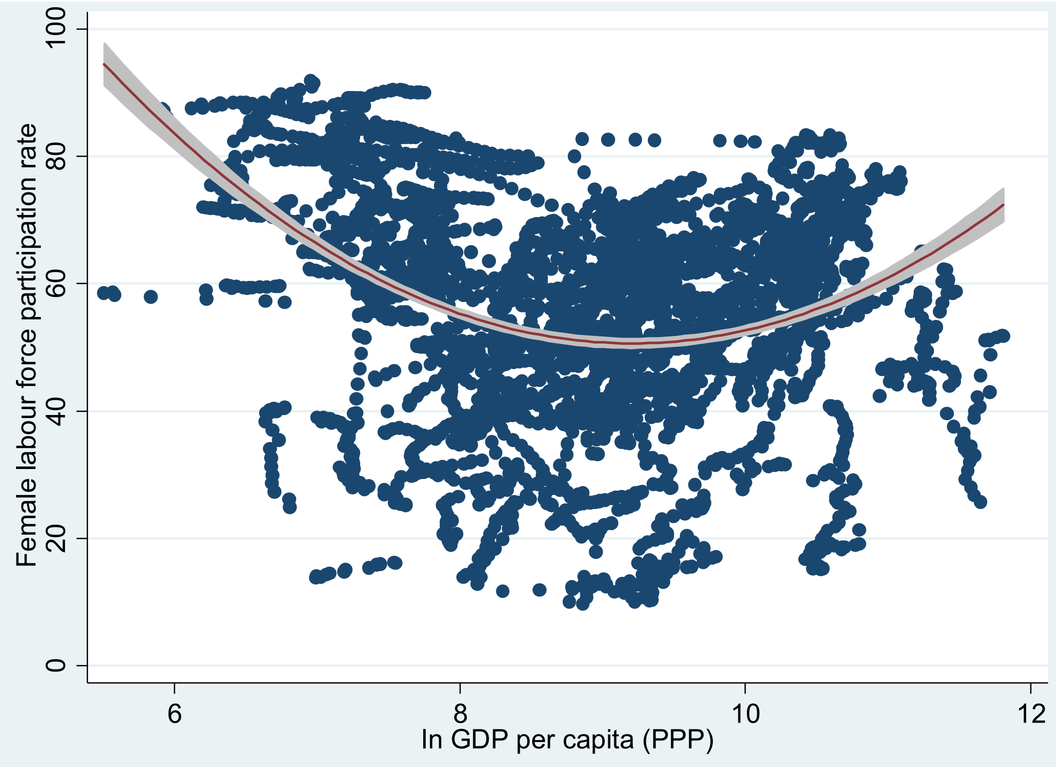


Figure 2 – The Feminisation U with 95% C.I.: Egypt (1990-2013)

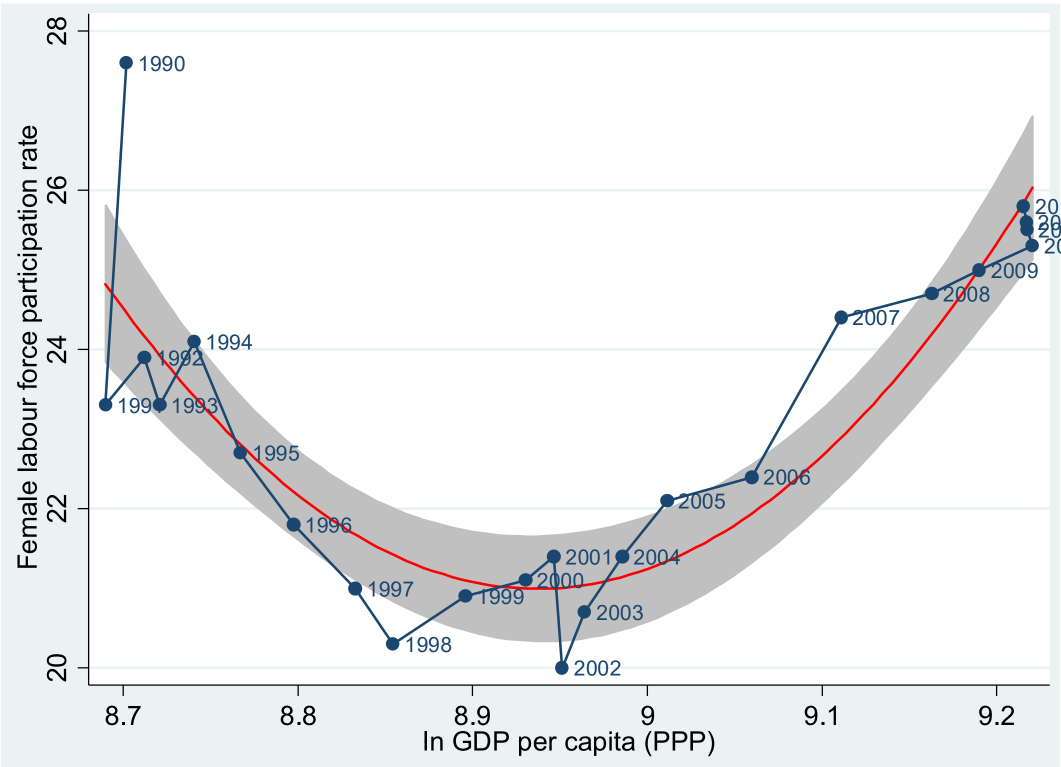


Table 1 – Main results: determinants of Female Labour Force Participation (1991-2013), System GMM models

Dependent variable: $FLFP_t$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$FLFP_{(t-1)}$	0.976***	0.976***	0.977***	0.975***	0.982***	0.976***	0.974***	0.973***
	(0.006)	(0.006)	(0.007)	(0.008)	(0.011)	(0.017)	(0.009)	(0.009)
Plough	-0.349***	6.169	8.538**	15.770***	8.462*	17.192***	18.925***	17.523***
	(0.086)	(4.218)	(4.233)	(5.046)	(4.679)	(6.480)	(6.296)	(5.806)
ln GDP pc [β_1]	-1.697***	-0.700	-0.320	-0.303	-0.998	-0.537	-0.332	-0.785
	(0.459)	(0.701)	(0.633)	(0.618)	(0.925)	(1.099)	(0.800)	(0.784)
ln GDP pc^2 [β_2]	0.101***	0.040	0.020	0.014	0.058	0.032	0.019	0.043
	(0.026)	(0.042)	(0.038)	(0.036)	(0.049)	(0.063)	(0.047)	(0.047)
Plough × ln GDP pc [γ_1]		-1.594*	-2.067**	-3.887***	-2.057*	-4.014***	-4.585***	-4.242***
		(0.952)	(0.949)	(1.177)	(1.058)	(1.489)	(1.453)	(1.351)
Plough × ln GDP pc^2 [γ_2]		0.095*	0.118**	0.231***	0.119**	0.228***	0.268***	0.248***
		(0.054)	(0.053)	(0.068)	(0.058)	(0.085)	(0.083)	(0.078)
Geo-anthropological controls			Yes			Yes		
Historical controls				Yes		Yes		
Contemporary controls					Yes	Yes		
Parsimonious set of controls							Yes	Yes
Continent dummies								Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of observations	3738	3738	3579	3151	3038	2699	3132	3132
N. of countries	169	169	162	142	153	133	146	146
N. of instruments (lag range)	153 (1 5)	136 (1 4)	143 (1 4)	142 (1 4)	142 (1 4)	136 (1 3)	138 (1 4)	143 (1 4)
AR(2) [*p*-value]	[0.988]	[0.988]	[0.993]	[0.741]	[0.530]	[0.754]	[0.674]	[0.674]
Hansen test [*p*-value]	[0.167]	[0.110]	[0.155]	[0.259]	[0.135]	[0.203]	[0.118]	[0.222]
Panel A: Parameters of the Feminisation U-curve - Plough = 1								
		-2.294***	-2.387***	-4.189***	-3.055**	-4.551**	-4.917***	-4.989***
ln GDP pc [$\beta_1 + \gamma_1$]		(0.790)	(0.852)	(1.318)	(1.525)	(1.756)	(1.498)	(1.329)
		0.135***	0.138***	0.245***	0.176**	0.259***	0.287***	0.291***
ln GDP pc2 [$\beta_2 + \gamma_2$]		(0.042)	(0.045)	(0.073)	(0.080)	(0.095)	(0.083)	(0.073)

Notes: The estimates are based on a two-step, feasible "system" GMM estimator employing the Forward Orthogonal Deviations (FOD) transform. Windmeijer-corrected cluster-robust errors in parentheses (delta-method standard errors in Panel A). *** p<0.01, ** p<0.05, * p<0.1. FLFP(t -1) is treated as pre-determined and instrumented for in GMM style.

Figure 3 – Estimated time path of FLFP in plough (red) and non-plough(blue) countries

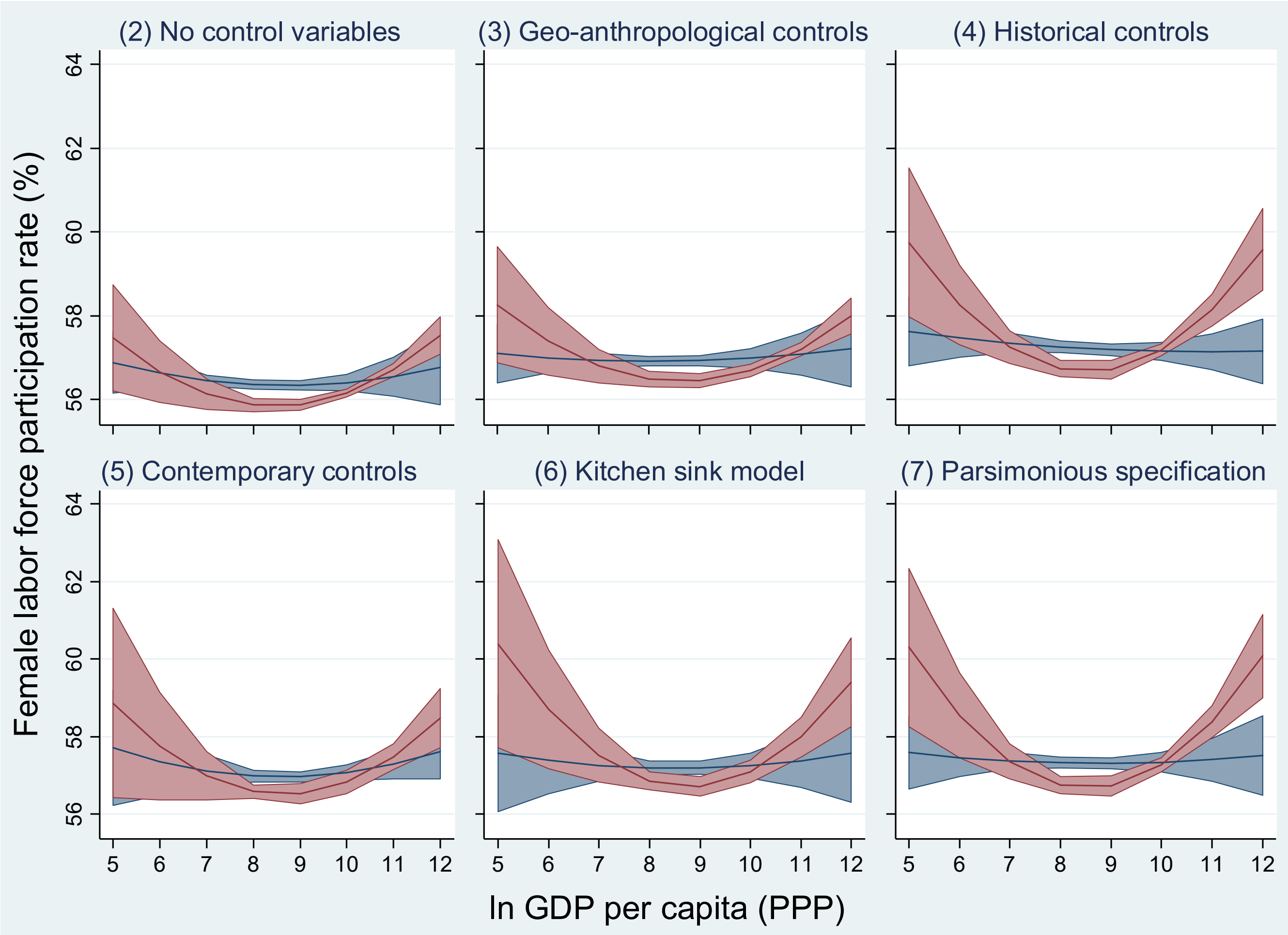


Table 2 - Marginal effects of traditional plough use at different levels of development

GDP p.c. (PPP) [in logs]	Country (e.g.)	(2)	(3)	(4)	(5)	(6)	(7)
Average marginal effect		-0.320***	-0.293**	-0.176	-0.264***	-0.140	-0.210
		(0.100)	(0.149)	(0.123)	(0.094)	(0.152)	(0.152)
		[0.001]	[0.048]	[0.154]	[0.005]	[0.357]	[0.170]
245 [5.5]	Liberia (1995), Mozambique (1992)	0.284	0.748	1.388**	0.739	2.008**	1.831**
		(0.630)	(0.648)	(0.663)	(0.650)	(0.882)	(0.846)
		[0.652]	[0.248]	[0.036]	[0.256]	[0.023]	[0.031]
8,100 [9]	Egypt (2005), Ukraine (2010)	-0.461***	-0.479***	-0.480***	-0.440***	-0.476**	-0.591***
		(0.097)	(0.144)	(0.148)	(0.151)	(0.199)	(0.179)
		[0.000]	[0.001]	[0.001]	[0.004]	[0.017]	[0.001]
60,000 [11]	Norway (2010), Luxembourg (1995)	0.161	0.121	0.996***	0.192	0.611	0.978**
		(0.289)	(0.339)	(0.357)	(0.227)	(0.509)	(0.179)
		[0.578]	[0.721]	[0.005]	[0.398]	[0.230]	[0.022]

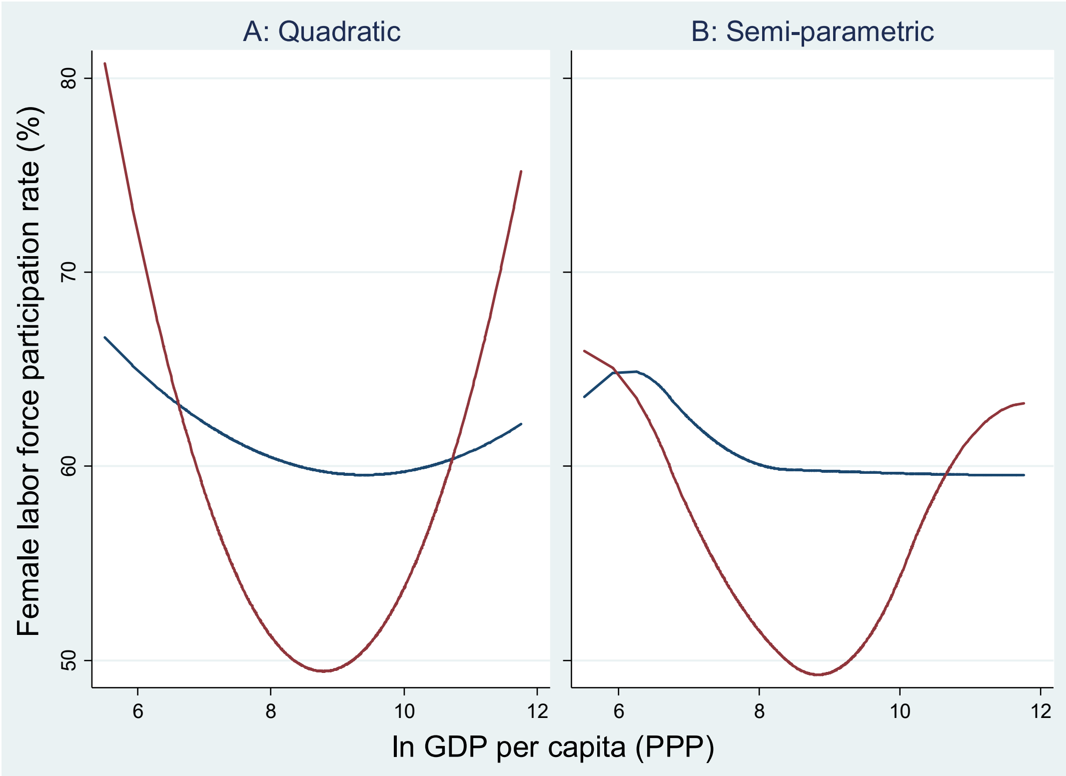
Notes: based on Table 1, models (2)-(7)

Table 3 – 5-year panel (1995-2010) and split samples

Dependent variable: $FLFP_t$	(1)	Plough(>.8) (2)	Nonplough (<.2) (3)
$FLFP_{(t-1)}$	0.692***	0.667***	0.725***
	(0.111)	(0.161)	(0.074)
Plough	138.7**		
	(58.722)		
ln GDP pc	-11.876*	-51.215**	-8.800
	(7.280)	(23.412)	(5.685)
ln GDP pc^2	0.659	2.915**	0.468
	(0.417)	(1.295)	(0.333)
Plough × ln GDP pc	-32.460**		
	(13.416)		
Plough × ln GDP pc^2	1.848**		
	(0.756)		
Parsimonious set of controls	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
N. of observations	552	278	245
N. of countries	145	75	62
N. of instruments (lag range)	24 (1 3)	21 (1 3)	21 (1 3)
AR(2) [*p*-value]	[0.949]	[0.149]	[0.256]
Hansen test [*p*-value]	[0.664]	[0.318]	[0.993]

Notes: The estimates are based on a two-step, feasible GMM estimator employing a Forward Orthogonal Deviations (FOD) transform. Windmeijer-corrected cluster-robust errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. FLFP(t -1) is treated as pre-determined and instrumented for in GMM style. "Plough" countries are defined as those for which *Plough* > 0.8; "Nonplough" countries are those for which *Plough* < 0.2. The models are variations of model (2) in Table 1.

Figure 4 – Semi-parametric analysis: path of FLFP in plough (red) and non-plough (blue) countries



Notes: the diagrams plot the predicted expected value of FLFP against the log of GDP per capita, holding the other covariates constant at their sample mean. Panel A is based on models (2) and (3) in Table 3.

Table 4 - Endogenous development: difference vs. system GMM (5-year panel)

Estimator:	System	Difference
Transformation:	Δ	Δ
Dependent variable: $FLFP_t$	(1)	(2)
$FLFP_{(t-1)}$	0.753***	0.399*
	(0.090)	(0.239)
Plough	134.25**	
	(55.53)	
ln GDP pc	-9.799	-3.563
	(6.108)	(5.909)
ln GDP pc^2	0.549	0.220
	(0.351)	(0.349)
ln GDP pc × Plough	-31.225**	-40.715***
	(12.600)	(13.047)
ln GDP pc^2 × Plough	1.771**	2.185***
	(0.706)	(0.700)
Parsimonious set of controls	Yes	Yes[a]
Year dummies	Yes	Yes
N. of observations	407	407
N. of countries	145	145
N. of instruments (lag range)	23 (1 3)	10 (1 2)
AR(2) [*p*-value]	[0.958]	[0.932]
Hansen test [*p*-value]	[0.303]	[0.613]

Notes: The estimates are based on two-step, feasible GMM estimators ('system' and 'difference') employing a first-difference (Δ) transformation (on the instruments and on the regression equation, respectively). Windmeijer-corrected cluster-robust errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. FLFP(t -1) is treated as pre-determined and instrumented for in GMM style; [a]All the time-invariant control variables are purged by the first-difference transformation.

Table 5- Endogenous development: instrumenting for per-capita income (yearly panel)

	Income treated as exogenous		Income treated as endogenous	
	Plough	Nonplough	Plough	Nonplough
Dependent variable: $FLFP_t$	(1)	(2)	(3)	(4)
$FLFP_{(t-1)}$	0.829***	0.943***	0.959***	0.943***
	(0.078)	(0.041)	(0.033)	(0.028)
ln GDP pc	-24.373**	-1.437	-7.191*	-1.080
	(11.300)	(3.336)	(4.338)	(2.282)
ln GDP pc^2	1.374**	0.078	0.440*	0.069
	(0.626)	(0.198)	(0.242)	(0.130)
Parsimonious set of controls	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
N. of observations	1604	1359	1604	1359
N. of countries	75	63	75	63
N. of instruments	37	37	45	45
$FLFP_{(t-1)}$ (lag range)	(1 5) collapse	(1 5) collapse	(1 5) collapse	(1 5) collapse
ln GDP pc & ln GDP pc^2 (lag range)			(2 5) collapse	(2 5) collapse
AR(2) [*p*-value]	[0.469]	[0.717]	[0.437]	[0.729]
Hansen test [*p*-value]	[0.951]	[0.368]	[0.786]	[0.332]

Notes: The estimates are based on a two-step, feasible GMM estimator employing a Forward Orthogonal Deviations (FOD) transformation. Windmeijer-corrected cluster-robust errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. FLFP(t -1) is treated as pre-determined and instrumented for in GMM style. In models (3) and (4), ln GDP pc and ln GDP pc^2 are treated as endogenous and instrumented for in GMM style. The instrument matrix is collapsed throughout to contain instrument proliferation. "Plough" countries are defined as those for which *Plough* > 0.8; "Nonplough" countries are those for which *Plough* < 0.2.

Table 6 - Timing of the Neolithic revolution and language gender marking (yearly panel)

	Neolithic revolution		Language gender marking	
Dependent variable: $FLFP_t$	(1)	(2)	(3)	(4)
$FLFP_{(t-1)}$	0.974***	0.974***	0.978***	0.978***
	(0.010)	(0.010)	(0.010)	(0.010)
Plough	18.499***	21.408***	16.260***	16.216***
	(5.792)	(7.804)	(5.987)	(6.040)
ln GDP pc	-0.261	-0.508	-0.654	-0.633
	(0.738)	(1.107)	(0.836)	(0.874)
ln GDP pc^2	0.014	0.023	0.037	0.036
	(0.044)	(0.063)	(0.050)	(0.052)
Plough × ln GDP pc	-4.538***	-5.124***	-3.945***	-3.952***
	(1.344)	(1.792)	(0.075)	(1.371)
Plough × ln GDP pc^2	0.269***	0.298***	0.232***	0.234***
	(0.077)	(0.101)	(0.075)	(0.077)
Years of agriculture (x1000)	-0.019	-0.683		
	(0.025)	(1.244)		
Years of agriculture × ln GDP pc		0.121		
		(0.280)		
Years of agriculture × ln GDP pc^2		-0.005		
		(0.016)		
GII			0.001	-0.048
			(0.001)	(0.078)
GII × ln GDP pc				0.009
				(0.017)
GII × ln GDP pc^2				-0.000
				(0.001)
Parsimonious set of controls	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
N. of observations	3003	3003	2874	2874
N. of countries	140	140	133	133
N. of instruments (lag range)	120 (1 3)	120 (1 3)	120 (1 3)	120 (1 3)
AR(2) [*p*-value]	[0.690]	[0.690]	[0.660]	[0.661]
Hansen test [*p*-value]	[0.151]	[0.173]	[0.079]	[0.071]

Notes: The estimates are based on a two-step, feasible "system" GMM estimator employing the Forward Orthogonal Deviations (FOD) transform. Windmeijer-corrected cluster-robust errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; FLFP(t -1) is treated as pre-determined and instrumented for in GMM style.

Table 7 - Exploring the Mechanisms: Structural change, Fertility and Education

	Benchmark (1)	Structural change (2)	Fertility (3)	Education (4)	All (5)
Panel A: Parameters of the Feminisation U-curve - Plough = 0					
ln GDP pc	-0.761	-1.009	-0.606	-0.744	-0.635
	(0.654)	(0.693)	(0.679)	(0.911)	(0.800)
ln GDP pc^2	0.042	0.052	0.031	0.041	0.031
	(0.038)	(0.041)	(0.038)	(0.052)	(0.046)
Panel B: Parameters of the Feminisation U-curve - Plough = 1					
ln GDP pc	-4.084***	-3.007**	-2.804	-3.876***	-1.446
	(1.496)	(1.505)	(1.811)	(1.464)	(1.591)
ln GDP pc^2	0.246***	0.198**	0.176*	0.234***	0.115
	(0.084)	(0.084)	(0.098)	(0.083)	(0.087)
Sectors GVA, % GDP (interacted with plough)		Yes			Yes
Fertility rate (interacted with plough)			Yes		Yes
Education, Gender Parity Index (interacted with plough)				Yes	Yes
Parsimonious set of controls	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes
N. of observations	1932	1932	1932	1932	1932
N. of countries	115	115	115	115	115
N. of instruments (lag range)	99 (1 2)	105 (1 2)	101 (1 2)	101 (1 2)	109 (1 2)
AR(2) [*p*-value]	[0.937]	[0.877]	[0.934]	[0.889]	[0.937]
Hansen test [*p*-value]	[0.253]	[0.287]	[0.293]	[0.307]	[0.287]

Notes: The estimates are based on a two-step, feasible "system" GMM estimator employing the Forward Orthogonal Deviations (FOD) transformation. The table reports coefficients obtained in post-estimation, with delta-method standard errors in parentheses. The sectors are agriculture (including forestry and fishing), manufacturing and services; *** p<0.01, ** p<0.05, * p<0.1.

Appendix A - Control variables

Variable	Description	Source
1. Geo-anthropological Controls		
Tropical climate	Fraction of land with (sub)tropical climate, pop. weighted	Alesina et al., 2013
Distance from the coast	Average distance to nearest ice-free coast (1000 km.)	Nunn and Puga, 2012
Agricultural suitability	Share of land defined as suitable for cultivation of wheat, barley, rye, millets, sorghum (as defined by FAO)	Alesina et al., 2013
Presence of large animals	Fract. of pop. 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
Absence of private property	Proportion of a country's ancestors without land-inheritance rules	Alesina et al., 2013
Extended family structure	Proportion of a country's ancestors with an extended family structure	Alesina et al., 2013
2. Historical Controls		
Duration of socialist rule	Number of years of socialist rule before 1991	Uberti, 2018
Fraction of European descent	% of pop. with European descent	Alesina et al., 2013
Identity of former colonizer	Two dummies identifying countries ruled by a Northern and a Southern European colonizer	Nunn, 2008
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
3. Contemporary 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, constant 2000 dollar	Ross, 2009
Aid dependence	Net ODA received (% GNI)	World Bank, *World Development Indicators*
Religious denomination	Set of variables capturing the share of population that is Catholic, Protestant, Orthodox Christian, Muslim and Hindu.	Alesina et al., 2013
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
4. Additional variables used in the analysis		
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

Sectors shares of GVA	Three variables capturing the share of gross value added generated by agriculture, manufacturing and the service sector, % GDP	World Bank, *World Development Indicators*
Fertility rate	Number of children that would be born to a woman if she were to live to the end of her childbearing years and bear children in accordance with age- and year-specific fertility rates.	World Bank, *World Development Indicators*
Gender parity index, gross school enrolment	Ratio of girls to boys enrolled at primary level in public and private schools (based on UNESCO estimates)	World Bank, *World Development Indicators*

Appendix B - OLS Results

	Cf. table 1, model 2		Cf. table 1, model 7	
	Pooled	FE	Pooled	FE
Dependent variable: $FLFP_t$	(1)	(2)	(3)	(4)
$FLFP_{(t-1)}$	0.992***	0.935***	0.990***	0.925***
	(0.001)	(0.009)	(0.002)	(0.011)
Plough	4.095		10.502***	
	(2.647)		(2.883)	
ln GDP pc	0.006	-0.951	-0.006	-1.329
	(0.433)	(0.744)	(0.442)	(0.940)
ln GDP pc^2	0.002	0.047	0.003	0.081
	(0.026)	(0.042)	(0.026)	(0.056)
Plough × ln GDP pc	-1.090*	-3.312*	-2.623***	-5.922***
	(0.608)	(1.861)	(0.669)	(1.881)
Plough × ln GDP pc^2	0.067*	0.204**	0.157***	0.337***
	(0.034)	(0.102)	(0.038)	(0.104)
Parsimonious set of controls	No	No	Yes	Yes
Country FE	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes
R-squared	0.99	0.94	0.99	0.94
N. of observations	3738	3738	3130	3130
N. of countries	169	169	146	146

Notes: OLS estimates with robust standard errors clustered at the country level reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1; FLFP(t -1) is treated as strictly exogenous. The R-squared refers to the "within" R-squared in models (2) and (4).

Appendix C – Alternative System GMM estimators

Estimator:	One-step	Two-step	Two-step
Transformation:	FOD	Δ	FOD
Instrument matrix:			Collapsed
Dependent variable: $FLFP_t$	(1)	(2)	(3)
$FLFP_{(t-1)}$	0.974***	0.945***	0.842***
	(0.009)	(0.014)	(0.057)
Plough	14.714***	27.655***	58.959**
	(4.517)	(10.143)	(29.536)
ln GDP pc	-0.760	-1.619	-7.350
	(0.730)	(1.499)	(4.932)
ln GDP pc^2	0.044	0.092	0.408
	(0.043)	(0.090)	(0.287)
Plough × ln GDP pc	-3.614***	-6.515***	-13.763**
	(1.044)	(2.339)	(6.805)
Plough × ln GDP pc^2	0.214***	0.374***	0.779**
	(0.060)	(0.134)	(0.389)
Parsimonious set of controls	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
N. of observations	3132	3132	3132
N. of countries	146	146	146
N. of instruments (lag range)	138 (1 4)	138 (1 4)	45 (1 10)
AR(2) [*p*-value]	[0.669]	[0.683]	[0.751]
Hansen test [*p*-value]	[0.118]	[0.347]	[0.820]

Notes: Cluster-robust errors in parentheses (Windmeijer-corrected in models 2-3). FOD stands for "Forward Orthogonal Deviations"; Δ stands for "first-difference". *** p<0.01, ** p<0.05, * p<0.1. FLFP(t -1) is treated as pre-determined and instrumented for in GMM style. The models are variations of model (7) in Table 1.