The role of mothers on female labour force participation: an approach using historical parish records*

Jesús M. Carro†, Matilde P. Machado‡, Ricardo Mora§

Abstract

Using Portuguese parish data from 1675-1925, we estimate the relationship between a mother's participation in the labour force and that of her daughter. We adapt a methodology to prevent bias that originates from potentially non-random missing data. Ignoring the missingness process results in substantial downward-biased estimates of the relationship, even for a proportion of missing values as low as 20 percent. In contrast, our methodology yields unbiased estimates regardless of the proportion of missing values. We document the existence of a strong, positive association between the mother's participation and that of her daughter long before the 20th century's substantial changes in education and the labour market

Keywords: Female labour force participation, intergenerational transmission, historical family data, church registry data, non-ignorable missingness, econometric methods for missing data.

JEL Classification: J22, J24, J16, J12

*The authors acknowledge the financial support of the Ramón Areces Foundation, and the Spanish Ministry of Science and Innovation through Grants ECO2016-78652 (P. Machado), ECO2015-65204-P (Carro and Mora), RTI2018-095231-B-I00 (Carro), and PID2019-108576RB-I00 (P. Machado and Mora). They are grateful to M. Norberta Amorim and her research team at U. of Minho for the original data and assistance with historical information. They also thank Richard Blundell, Maristella Boticcini, Christian Dustmann, Ana Fernandes, Raquel Fernández, Eric French, Stefan Houpt, Steffi Huber, Nuno Palma, Imran Rasul, Jaime Reis, Jan Stuhler, Michele Tertilt, Anna Tur-Prats, Marcos Vera-Hernández and the audiences at the UCL Economics regular seminar, the Seminar at ISEG, the Economic History Seminar at UC3M, the 28th Annual Congress of the EEA, the CEPR conference on the Economics of Interactions and Culture, the Economic History Society conference, the APHES conference, and the Nuffield Historical Mobility Seminar, for helpful comments.

† jcarro@eco.uc3m.es. Universidad Carlos III de Madrid

‡ matilde.machado@uc3m.es. Universidad Carlos III de Madrid

§ ricmora@eco.uc3m.es. Universidad Carlos III de Madrid

1 Introduction

Much has been written regarding the rise of women's participation in the labour force during the second half of the 20th century. In particular, recent research highlights the potential multiplier effect of intergenerational transmission mechanisms on the dynamics of female labour-force participation (FLFP) subsequent to the second world war.[1] Yet much less is known about women's participation in earlier periods, arguably due to the difficulty in finding full participation information of mother-daughter pairs in historical registries. In this paper, we propose a methodology that addresses this problem and renders consistent estimates even in the presence of a high proportion of non-ignorable missing values. Using unique individual-level Portuguese data from 1675 to 1925 with a high proportion of missingness, we show the existence of an earlier and stronger association between mothers' and daughters' labour market participation decisions. Our finding of this early and substantial intergenerational link, which acts as a potent catalyst of shocks, contributes to the understanding of the long-term dynamics of FLFP.

We model a woman's decision to participate in the labour market as a function of her mother's participation. One difficulty in studying the mother-daughter transmission of FLFP involves disentangling it from the transmission originated from other factors. The influence of factors such as the introduction or enforcement of mandatory education and the expansion of the service sector is minimized by choosing a sample period before these factors became important.[2] Time and location dummy variables are included to capture factors such as labour market conditions and social multiplier effects (Maurin and Moschion, 2009; Olivetti et al., 2018). In an extended variable specification, controls are added for the social status of the father, for the ownership of property, and a for a

[1]Key references include: Fernández et al. (2004), Morrill and Morrill (2013), Alesina et al. (2013), Olivetti et al. (2018), Moen et al. (1997), Farré and Vella (2013), Kawaguchi and Miyazaki (2009), and theoretical approaches by Bisin and Verdier (2001) and Fernández (2013).

[2]There was no sustained industrialisation nor modern economic growth in Portugal before 1950 (Lains, 2006; Palma and Reis, 2019; Pedreira, 1990). At the turn of the 20th century, the agricultural sector accounted for 41.5 percent of GDP versus an estimated 33.6 percent for the service sector. It was not until the second decade of the century that the service sector outpaced the agricultural sector (Lains, 2006). Regarding education and according to the 1900 Portuguese census, only 19.5 percent of girls and 29 percent of boys aged 10-14 knew how to read. It was not until the 1960s that literacy of all children was achieved (Gomes and Machado, 2020).

proxy for household income to prevent confoundedness with other sources of inertia within the family.

Like most historical data, ours are characterised by abundant non-ignorable missing information, which affects both mothers' and daughters' participation statuses. Nevertheless, the number of mother-daughter pairs without missing values is not small ($n = 696$). An enticing approach is to estimate the model only with observations for which we have complete information (hereinafter, this approach is labelled "Ignorability"). However, when the missing process is not random, then Ignorability suffers from sample selection bias. We consider two approaches to address this non-ignorability problem in our estimation. First, informed by historical records, we conservatively impute missing values with the perceived predominant female labour market status at the time, that is, non-participation, which sets the participation rate at an unrealistically low value relative to census data. This approach is labelled "Imputation". Second, we adapt a methodology based on Ramalho and Smith (2013), which allows models to be estimated in contexts in which missing data are abundant and non-random. This methodology involves Maximum Likelihood (ML) estimation by using all observations and allowing the missing values to be endogenous. The identification of the parameters of the model is improved by incorporating external information regarding aggregate female labour market participation from historical census data. This approach is labelled "the Likelihood Approach". We show that failure to recognize the endogeneity of the missing process reduces the estimates of the mother-daughter link by more than a half.

Using the Likelihood Approach with only location and time dummies, we estimate a positive and statistically significant average marginal effect of the mother's participation status on the daughter's probability of participation (hereinafter, AME) of 39.7 percentage points (pp) with a standard error (s.e.) of 3.0 pp. This estimate comprises a number of mother-to-daughter intergenerational links: (a) the transmission of human capital such as that involved in certain crafts; (b) the transmission of physical capital, such as a small shop, or valuable craft-specific tools; and (c) the transmission of preferences, beliefs, traits, and values regarding the participation in the labour market from a mother to her daughter. There is evidence of such transmission within the family during the 20th century (Fernández et al., 2004; Farré and Vella, 2013), and hence it is possible that it was also present previously. We explore the importance of the transmission of human and physical capital in a variable specification where controls are added for several socio-economic factors. In this extended specification the AME estimates remain large and statistically significant at 26.7 pp (s.e. 2.8 pp).

Our Likelihood Approach estimates of the mother-daughter link are always substantially higher than those by Morrill and Morrill (2013) for the late 20th century: in the range of 2-7 pp. Although differences with Morrill and Morrill (2013) partly stem from using different periods, a reduction in the effect should be expected as the importance of the mother vis-a-vis external factors, such us access to formal education, decreases with economic and cultural development. Indeed, when we split our sample into different periods, the AME declines with time.

Estimating the model with the Likelihood Approach allows us to test whether the missing process is ignorable. We strongly reject Ignorability, which implies that observations with missing values have valuable information and should not be discarded. Moreover, the implicit assumption under Imputation, that is, that the missing occupations correspond to non-participation in the labour market, is inconsistent with our Likelihood Approach estimates.

Finally, by means of simulation methods we evaluate the sensitivity of Ignorability and the Likelihood Approach to varying degrees of the incidence of missing information. This is carried out by taking the estimated parameters to simulate the data and applying different percentages of missing observations. We then re-estimate the AME in these different samples. We find that ignoring the missing observations (Ignorability), results in estimates with large downward bias, which increases with the incidence of the missing information. For example, when the level of missing information matches that of the actual data, estimates are on average less than half of the true value in the data generating process. In contrast, the Likelihood Approach always delivers unbiased estimates, although, as expected, their standard errors increase with the incidence of missingness. By attaining unbiased estimation in the presence of non-random missing data, our approach confers considerable potential to incomplete historical data.

The remainder of this paper proceeds as follows. Sections 2 and 3 describe the dataset and the econometric model, respectively. Section 4 presents our main estimation results, and Section 5 discusses the robustness of the Likelihood Approach using simulated samples with varying prevalence of missing observations. Section 6 yields the conclusions, while Appendix A provides institutional background, Appendix B describes additional features of the data, and Appendices C, D, and E contain technical details.

2 Data

2.1 Parish data

The main data source is parish information that dates back to the end of the 16th century and was extracted from parish records in the villages of Ronfe and Ruivães in the Minho region, in the north-western part of Portugal, and the city of Horta on Faial Island located in the Azores. A research team from the University of Minho collected the main datasets based on all baptism, marriage, and death certificates found in the local churches.[3,4] The resulting dataset was matched with other church individual-specific records known as "rol de confessados" (literally, the list of the confessed). The latter originates from parochial censuses of residents over seven years of age. They were initially produced by the priest during Lent to administer the sacrament of penance to the parishioners and contain information regarding occupation and/or social status.

The original baptism, marriage, and death certificate records allowed family linkages to be reconstructed within each location beginning in the 1550s up until the mid 20th century (Amorim, 1991). Altogether, following basic cleaning, the dataset includes entries for 34,897 women (50.2 percent of all records). All observations of slaves and their children

[3] The research team was led by M. Norberta Amorim. Currently, the Grupo História das Populações (Universidade do Minho) in the Centro de Investigação Transdisciplinar «Cultura, Espaço e Memória», administers the genealogical database. See the genealogy web page http://www.ghp.ics.uminho.pt/genealogias.html.

[4] For a brief summary of the historical background see Appendix A.

(almost exclusively found in Horta) have also been excluded, and the number of female observations drops to 34,075.

The dataset holds information on the dates of birth, marriage, and death together with a family identification code. Gender is inferred from the individual's first name in the baptism Parish registry. We use the family identification code to link women with their mothers (and fathers) and to compute the number of siblings for each woman. The year of birth is available for 59.3 percent of all records. In contrast, death information is available for only 33.3 percent of all records. We complete records for which the date of birth is missing using birth-date information from other members of the family. To do so,all individuals for which the information is available are grouped into generational cohorts spanning 25 years. Observations for which the year of birth is missing are completed by sequentially examining the 25-year birth period of siblings, spouse, and children, in that order. Once the cohort of the siblings or the spouse is identified, then the record is completed with the 25-year birth period of the spouse or sibling. In the event that only cohorts of the children are identified, then the 25-year period previous to the cohort of the eldest child is assigned to the missing record. This procedure is repeated until no changes are produced. As a result, 82.4 percent of the original data can be associated with a given 25-year period.[5] Finally, records of individuals born before 1675 and after 1950 become sporadic in the original dataset and hence a restriction is imposed of only dates of birth between 1675 and 1950, resulting in a total of 21,645 female observations.

2.2 Non-ignorable missing occupations

The occupation/social status information contained in the matched records is not as complete as the baptism and marriage information. Columns 1, 3, and 5 of Table 1 report the number of observations per location and quarter-century (throughout the paper, each quarter-century is identified by its initial year). Columns 2, 4, and 6 show the proportion (as a percentage) of observations with occupation/social status information per quarter-

[5]Given the relatively high maternal mortality rates prevalent in Western countries before the 20th century, the date of death itself is unlikely to be very informative regarding the date of birth. Therefore, we consider our algorithm to be superior to that which uses dates of death even for those individuals for whom we have their date of death but not their date of birth.

century and location.

Table 1: Original Sample Sizes & Reported Occupations/Social Status

	Horta		Ronfe		Ruivães	
	N.obs.	% non-missing	N.obs.	% non-missing	N.obs.	% non-missing
Quarter-century	(1)	(2)	(3)	(4)	(5)	(6)
1675	926	2.70	318	15.72	196	0.51
1700	1557	3.66	324	5.25	195	n.a.
1725	1437	5.64	336	7.44	209	n.a.
1750	1390	6.47	379	7.65	199	n.a.
1775	1508	11.01	411	3.41	217	n.a.
1800	1733	14.25	438	4.34	293	n.a.
1825	1761	25.84	428	5.84	306	2.29
1850	1325	24.30	464	10.56	335	10.45
1875	860	2.79	581	14.97	396	23.99
1900	27	n.a.	730	17.95	489	34.36
1925	.	.	1082	12.38	795	35.35

Note: Quarter-century "1675" represents the 1675-1699 25-year period. Sample includes women born between quater-centuries 1675 and 1925. Women have their occupation/social status reported whenever the local vicar enters a description of her activity or social status in the Lent census (from the Portuguese "Rol de confessados") and that registry is matched by Amorim's research team (U. Minho) with the baptism, death or marriage certificates. "No. obs." is the number of observations. "% non-missing" is the percentage of observations with occupational/social status information reported. "n.a" indicates that there is no occupations/social status information in the matched records.

Occupations/social status coverage varies by quarter-century and across locations. For Ruivães from 1700 until 1800 and for Horta in 1900, there is no occupations/social status information in the matched records. These quarter-century/location combinations are therefore omitted from our analysis. We also discard the observations for Horta and Ruivães in 1675 and Horta in 1875 since both the number of observations and coverage are unusually low. Hence, our working sample includes only observations from Horta in the period 1700-1850, Ronfe in the period 1675-1925, and Ruivães in the period 1825-1925.

Our working sample shows location-specific trends in the coverage. For example, whereas coverage in Horta in 1725 is 5.6 percent, it increases monotonically to reach 25.8 percent one century later. Coverage in Ronfe follows a U-shaped trend decreasing steadily from 15.7 percent in 1675 to 3.4 percent in 1775 to rise again to around 18 percent levels in the beginning of the 20th century. Location-specific trends suggest that location factors are at work. This does not preclude the influence of individual-specific factors, which are a potential source of selection bias.

Coverage never exceeds 36 percent. Three reasons for these low figures can be discarded. First, accounting for early deaths can at most only explain a small proportion of non-coverage (Brettell, 1998). Second, gender bias in priests' recording practices appears to be minor because the gender differential in coverage is only 2.2 pp and follows a similar location/period pattern. Third, the priest could report occupation or social status to differentiate between women with a common name. If there were an excessive number of, say, "Marias", then their occupation or social status would help distinguish between them. However, common names, such as Maria and Ana, are typically followed by a second given name, and are as likely among those reported as among those unreported. For example, among the reported, 18.3 percent are named "Maria" and 4.7 percent "Ana" compared with 16.5 percent and 5.2 percent, respectively, among the unreported.

Why would the priest report some women's occupation/social status and not that of others? We believe that there are at least four potential sources of selection bias. The first arises from the Lent census data-collection practices. The censuses were organised per street and gathered by priests door-to-door. Whenever priests gave priority to nearby households, then remote rural locations (where farmers and poorer people would tend to live) would be less likely to be covered.

The second source of selection bias stems from the activity itself. Priests might tend to only record the activity of those whose labour or social status was uncommon in the region and period, such as that of civil servants or of the miller in a village of farmers. Indeed, in the two rural locations, the share of farmers among reported occupations is unrealistically low: 10.3 percent in Ronfe and 8.2 percent in Ruivães.

The third source originates from social status: parochial priests are arguably more likely to register those parishioners who give large donations to the church.

The fourth and last source of non-coverage is due to the absence of identification of the mother-daughter pair. For 31.5 percent of the daughters ($5,839$ observations) the mother's identification code is unknown. By construction, in all these cases, the participation status of the mother is missing. There are several reasons for this failure in identification. When the mother's record precedes the first currently available registries, or it is illegible, or it contains coding errors, then the match with the daughter's record is impossible. These problems are slightly more likely to occur the older the records are: in our working sample, around 35 percent of the observations from 1675 have an unidentified mother compared with 33 percent one hundred years later and 30 percent two hundred years later. These registry issues affecting the identification of the mother-daughter pair are not systematically related to participation decisions or to the priest's recording practices and are thus not a likely source of selection bias in our estimates. However, there are at least two sources of non-identification of the mother that may lead to non-ignorability. The first source is illegitimacy. Annotations from the University of Minho's team suggest that at least 4 percent of all cases with an unidentified mother were orphans or illegitimate children. Historical accounts suggest that there were more cases. For example, Scott (1999) reports that in Ronfe, 20.7 percent of the heads of households were single females in 1750 and at least 18 percent of the children baptised in 1700 were illegitimate. Illegitimate children and orphans are more likely to be poor, landless, and to work for pay. The second source involves the absence of the mother in the records: mothers who were not born, married, or deceased in the parish leave no personal records there and, thus, cannot be found in our dataset. This situation arises, for example, when, for reasons of work or marriage, a woman migrates from another location to one of our three locations and her mother remains in her original place of residence. The raw data show that daughters with identified mothers are less likely to participate (identification of the mother correlates with the woman's participation conditional on reported, correlation value of -0.202 with *a* p-value smaller than 0.01). Hence, this source might be non-ignorable.

To sum up, lack of coverage, that is, missingness, might be associated with non-random individual factors such as social status and activity choices and, thus, is probably non-ignorable.

2.3 Labour force participation

We construct women's occupation/social status using information from two variables from the original files: "profissão" (profession) and "título" (title). In most cases, "profissão" reports professions and provides information on participation. Although "título" could be interpreted as social status, in many cases it also provides information on participation. Therefore, we infer participation from all the information available, including that in "título".

The information on occupation/social status is not systematically classified across parochial priests and across time. As a result, the original data include more than 500 occupations/social status categories, many of which are close substitutes for one another. To make this information tractable, we group all categories into four major classes: *employee/farmer*, *professional/capital owner*, *domestic production*, and *unproductive*.[6] The class *employee/farmer* includes all paid and unqualified jobs. *professional/capital owner* includes landlords, liberal professions, traders, businesswomen, the self-employed and qualified and managerial jobs. *Domestic production* includes observations classified as "doméstica" in Portuguese (a term that can be interpreted as a housewife) and women to whom the priest listed as "dona", a term originally employed to signal a woman of the upper class and this was gradually adopted to also indicate the bourgeoisie during the 18th and 19th centuries. The unproductive category includes the indigent or those described as very poor, and individuals registered as nobility by the priest, and others. Based on these four major categories, we define labour force participation as being an *employee/farmer* or a *professional/capital owner*. Our definition mostly differs from the usual labour market participation measure due to the lack of precision in the priest's registry of owners, given in Portuguese as "proprietário" (owner), whereas we correctly include as participants all small capital owners who are self-employed, such as shop owners, the term "proprietário" also pertains to landlords who have no labour market involvement. As a robustness check, we also conduct our main analysis under an alternative definition of participation that excludes property owners as participants. (See footnote 11.)

[6] Only five women from Horta have a second profession recorded in the registry. We conservatively adopted only their first profession as valid, which in all cases was that of housewife.

Table 2: Distribution (%) of Women Across Four Major Economic Categories

Observations with Recorded Economic Activity or Social Status

	Horta	Ronfe	Ruivães	Overall
Employee/farmer	6.28	43.62	41.81	22.72
Professional/capital owner	9.94	34.83	7.34	14.94
Domestic production	83.57	4.14	50.68	58.28
Unproductive	0.21	17.41	0.17	4.06
Total	100	100	100	100
Participating (as % reported)	16.22	78.45	49.15	37.65

Note: The table is based on the sample of 2,584 women from Horta (1700-1850), Ronfe (1675-1925), and Ruivães (1825-1925) who have their occupation/social status reported by the vicar. We classify women's occupation/social status using information from two variables from the original files: "profissão" and "título". Because the original data had so many different descriptions of professions and occupations, we aggregated them into the four categories described in this table. The last row shows the percentage of those with reported activity/social status who we consider participant in the labour force and corresponds to the sum of the proportions for *employees/farmers* and *professional/capital owner*.

In Table 2, we report the distribution of women across our four major categories by location. Distributions vary significantly across locations and may be the combination of location-specific economic factors and differentials in the incidence of missing information. Certain categories appear sometimes under-represented (i.e. *employee/farmer* in Horta), whereas others seem over-represented (i.e. *professional/capital owner* in Ronfe). As shown in the last row of Table 2, these disparities lead to large differences in the proportion of women participating per location.

Observed participation rates are presumably contaminated by non-random missingness. More importantly for our objective, disregarding observations with missing participation status may bias estimates of the association between the mother's labour force participation and the daughter's probability of participation. Table 3 shows participation rates in terms of mother participation statuses in the subsample of mother-daughter pairs for whom the participation information is available. The raw estimate for the marginal effect ($69.72 - 7.33 = 62.39$ pp) is probably biased due to non-random missingness.

Table 3: Women Participation Frequencies by Mother's Participation

Subsample of mother-daughter pairs with observable participation status

Daughter	Mother: Does not participate	Mother: Participates	Mother: Total
Does not participate	544	33	577
	(92.67)	(30.28)	(82.90)
Participates	43	76	119
	(7.33)	(69.72)	(17.10)
Total	587	109	696
	(100.00)	(100.00)	(100.00)

Note: The table is based on the subsample of 696 mother-daughter pairs from Horta (1700-1850), Ronfe (1675-1925), and Ruivães (1825-1925) who have their occupation/social status reported by the vicar. A woman participates if she is an employee/farmer or a professional/capital owner. Percentage values in parenthesis.

2.4 External data on female labour market participation rates

We use census data on local and national FLFP rates as external sources of information to enhance model identification. These census data, however, suffer from at least three limitations. First, the first census in Portugal with information on labour market participation was taken in the year 1890. Second, for Ronfe and Ruivães only information on the administrative regions that they belong to, Guimarães and Vila Nova de Famalicão, can be used. Third, the definition of labour participation used in census data in Western countries has changed over time. The main difference is found in the concepts of occupation and profession. The concept of occupation, adopted early on, classifies most women as active in the labour market, while the concept of profession, adopted later, does not. The difference applies most notably to the case of women whose (at times irregular) work was carried out in the home or the family farm/business, who would only be included as part of the labour force in early censuses (Goldin, 1995). The rates such as that obtained from the 1890 Census should therefore be regarded as approximations relative to modern definitions. Reis (2005) provides a national-level estimate of the female labour market participation in 1862 that is substantially lower (19.1 percent) than the rate obtained for 1890 using census data (38.5 percent); for a detailed account of the census data, see Appendix B. On the other hand, it has been observed that women's work has been vastly under-reported in historical official statistics during the same period in other countries (Humphries and Sarasúa, 2012; Goldin, 1995).

For the purpose of obtaining FLFP rates for our three locations, we construct predictions for participation rates using a log functional specification with the 1862 data from Reis (2005) and data from all censuses up to 1991 (see Appendix B). These predictions rank our three locations (Ronfe, Ruivães, and Horta) in the same way as the reported participation rates in our data (see Table 2). The high participation rates and the fact that women in rural settings participated more in the labour market than urban women are consistent with historical accounts for the 17th, 18th, and 19th centuries (Humphries and Sarasúa, 2012).

For all the years since 1890, we use our predictions as external data (see Fig. B1 in Appendix B). Before 1890, neither census data nor predictions are available. In their stead, we postulate alternative scenarios. We initially report results under what we refer to as the "Baseline scenario". In this scenario, we take the values of the predictions for our three locations in 1890 as the external female participation rates for all previous periods. In order to gauge the validity of these results, we estimate the model for a very large number (over 65,000) of alternative scenarios in Section 4.1.

3 The econometric model

We model a woman's participation decision as dependent on her mother's participation. Woman i chooses either to participate in the labour market, $y_i = 1$, or not, $y_i = 0$. The discrete choice is expressed in the following linear specification:

$$y_i = \mathbf{1}\left\{\alpha y_i^m + x_i\beta + \epsilon_i > 0\right\} \tag{3.1}$$

where dummy variable y_i^m indicates the labour force participation status of the mother. Parameter α captures the effect of a mother's participation status on that of her daughter. Vector x_i includes additional controls.

3.1 Likelihood with missing information

Most observations have missing entries in $\{y_i, y_i^m\}$ (86.1 percent for y_i and 87.9 percent for y_i^m). Let us define a binary indicator I_i, which takes value 1 if the daughter's occupation or social status is reported by the priest and available in the dataset and 0 otherwise. Similarly, let I_i^m take value 1 if the mother's occupation or social status is reported and 0 otherwise.

When the probability of a missing observation is independent of y_i, missing observations are ignorable in the sense that their omission from estimation does not bias the results. As argued in Section 2, missing participation is likely to be related to non-random individual factors such as social status and activity choices and is, therefore, non-ignorable.

The aim is to estimate parameter vector $\theta \equiv \{\alpha, \beta\}$ where:

$$\Pr\{y_i = v \,|y_i^m = w, x_i\} = F(v, w, x_i; \theta) \tag{3.2}$$

where v and w are values that y and y^m can take, that is $v, w \in \{0, 1\}$. By assuming normality, the conditional probit model is obtained:

$$F(v, w, x_i; \theta) \equiv \begin{cases} \Phi(\alpha w + x_i\beta) & \text{if } v = 1 \\ 1 - \Phi(\alpha w + x_i\beta) & \text{otherwise} \end{cases} .$$

The missingness mechanism and the participation process jointly define the event's probability:

$$\begin{aligned} \Pr\{I_i = r, I_i^m = s, y_i = v, y_i^m = w, x_i\} = \; & \Pr\{I_i = r, I_i^m = s \,|y_i = v, y_i^m = w, x_i\} \times \\ & F(v, w, x; \theta) \times \Pr\{y_i^m = w, x_i\}. \end{aligned} \tag{3.3}$$

for $r, s \in \{0, 1\}$. For an observation with non-missing information, the joint probability of non-missingness, that is, $I_i = I_i^m = 1$, and the vector variables $\{y_i, y_i^m, x_i\}$ is $\Pr\{I_i = I_i^m = 1, y_i = v, y_i^m = w, x_i\}$, which is a particular case of Eq. (3.3).

There are three situations in which a given observation may have missing information: when the daughter's information is missing but the mother's is not, when the mother's information is missing but the daughter's is not, and when information for both the daughter and the mother is missing. This implies the following probability of observation i:

$$\begin{aligned} p_i = \; & \left(\Pr\{I_i = I_i^m = 1, y_i, y_i^m, x_i\}\right)^{I_i I_i^m} \times \\ & \left(\Pr\{I_i = 0, I_i^m = 1, y_i^m, x_i\}\right)^{(1-I_i)I_i^m} \times \\ & \left(\Pr\{I_i = 1, I_i^m = 0, y_i, x_i\}\right)^{I_i(1-I_i^m)} \times \\ & \left(\Pr\{I_i = I_i^m = 0, x_i\}\right)^{(1-I_i)(1-I_i^m)}. \end{aligned} \tag{3.4}$$

Appendix C shows the expression of the different terms of p_i with probabilities given by the model in Eq. (3.3).

3.1.1 Ignorability of the Missing Process

If, conditional on vector x_i, the missing mechanisms of the mother and the daughter are independent of their participation decisions, then we can simplify Eq. (3.3) to

$$\Pr\{I_i = r, I_i^m = s, y_i = v, y_i^m = w, x_i\} = \Pr\{I_i = r, I_i^m = s\} F(v, w, x; \theta) \Pr\{y_i^m = w, x_i\}.$$

The probability of a non-missing observation is

$$\Pr\{I_i = 1, I_i^m = 1, y_i = v, y_i^m = w, x_i\} = \Pr\{I_i = I_i^m = 1\} F(v, w, x; \theta) \Pr\{y_i^m = w, x_i\}$$

and the probability conditional on the observation being non-missing is

$$F(v, w, x_i; \theta) \times \Pr\{y_i^m = w, x_i\}. \tag{3.5}$$

Thus, θ can be consistently estimated by Maximum Likelihood using only the observations for which no information is missing, and the missing process is ignorable.

3.2 The Likelihood Approach to addressing non-ignorable missingness

The traditional solution to non-ignorable missingness is to perform a procedure in which the missing values are imputed (see, among others, Little and Rubin, 2002). This presumes that certain events have zero probability. In Section 4, we present estimations under the assumption that daughters (mothers) with missing participation do not participate in the labour force. This is consistent with assuming that all missing observations engage in domestic production (either as housewives or as unpaid farmers).[7]

These imputation procedures are *ad hoc*. An alternative approach is to propose a model for the missing data mechanism and jointly estimate the participation model together with the missing data generation process. In this section, we follow Ramalho and Smith (2013) and state weaker assumptions regarding the missing data mechanism to identify participation while controlling for potentially non-ignorable missing information. Ramalho and Smith (2013) consider the situation in which the missingness mechanism is conditionally dependent on the outcome variable and on a discrete partition of the covariates. Our empirical application calls for further adjustments to this strategy. From Eq. (3.3) and without any loss of generality,

$$\Pr\left\{I_i, I_i^m \left| y_i, y_i^m, x_i \right.\right\} = \Pr\left\{I_i \left| I_i^m, y_i, y_i^m, x_i \right.\right\} \times \Pr\left\{I_i^m \left| y_i, y_i^m, x_i \right.\right\}. \tag{3.6}$$

Our strategy depends on additional assumptions regarding both $\Pr\left\{I_i \left| I_i^m, y_i, y_i^m, x_i \right.\right\}$ and $\Pr\left\{I_i^m \left| y_i, y_i^m, x_i \right.\right\}$.

Priests might have been more likely to under-report the incidence of occupations that were common, such as farmers, and more likely to record the occupations for those whose labour status presented a differentiating characteristic. Hence, the chance of coverage is

[7] See Appendix D for details regarding the Imputation approach.

probably conditioned by the type of occupation of the woman. However, due to the high non-coverage rate, insufficient information is available to control for detailed occupations. We solve this problem by exploiting the information contained in the joint missing process for mothers and daughters. As our data show, the coverage of mothers and daughters are correlated. Daughters of women whose occupation/social status is not reported by the priest have an 88.4 percent chance of not being reported. Likewise, daughters of women whose status is reported have a much larger chance of having theirs also reported compared to the average daughter (31.0 percent versus 13.9 percent). By assuming that observability of the daughter's participation depends both on her participation status and on the observability of her mother, which can be interpreted as a proxy of social status, the model can capture different effects on coverage by occupation. These considerations warrant the following:

Assumption 3.1 *(Daughter's Response Conditional Independence) Non-missingness in y_i is conditionally independent of y_i^m and x_i, that is,*

$$\Pr\{I_i\,|I_i^m, y_i, y_i^m, x_i\} = \Pr\{I_i\,|I_i^m, y_i\}\,. \tag{3.7}$$

Assumption 3.1 is not an imputation procedure because it does not replace the missing observations with any set of values. All information affecting the probability of daughter's occupation coverage is contained jointly in the daughter's occupation status and the mother's coverage. Since the mother's information is probably collected early on and through a similar process, an assumption closely related to Assumption 3.1 but referring to the availability of the mother's participation decision can also be made. Note, however, that although the mother's coverage status is allowed to correlate with the daughter's through Assumption 3.1, Eq. (3.6) warrants a simpler assumption for the mother:

Assumption 3.2 *(Mother's Response Conditional Independence) Non-missingness in y_i^m is conditionally independent of y_i, and x_i, that is,*

$$\Pr\{I_i^m\,|y_i, y_i^m, x_i\} = \Pr\{I_i^m\,|y_i^m\}\,. \tag{3.8}$$

Given the values of y_i and I_i^m (y_i^m), Assumptions 3.1 and 3.2 state that the missing process is independent of the covariates. This implies, for example, that priests of any place and location in our sample use the same criteria to decide whether to report or not the labour status of women. This restriction is mitigated because Assumption 3.1 allows for the missing processes for daughter and mother to be related. In this way, the correlation of daughters' and mothers' coverage may be due to common unobservable factors affecting the priests' decisions.[8] An important implication of Assumptions 3.1 and 3.2 is that covariates in x_i affect the probability that the priest reports the information only through their effects on y_i and y_i^m. Indeed, it is precisely through this implied correlation between controls and priests' reporting decisions that the parameter of interest α is identified.[9]

Henceforth, we assume that control vector x_i is a vector of discrete variables, which is the case in our data.[10] Assumptions 3.1 and 3.2 are sufficient conditions to identify the parameter vector θ. Let $H_{rsv} \equiv \Pr\{I_i = r, I_i^m = s, y_i = v\}$ and $H_{sw}^m \equiv \Pr\{I_i^m = s, y_i^m = w\}$ with $r, s, v,$ and $w \in \{0, 1\}$. Furthermore, the unconditional probabilities of discrete variables y_i and y_i^m are denoted by $\Pr\{y_i = v\} = \Pi_v$ and $\Pr\{y_i^m = w\} = \Pi_w^m$, respectively. Finally $\Pi_{w,x} \equiv \Pr\{y_i^m = w, x_i\}$, where the number of parameters in $\Pi_{w,x}$ is given by the number of different value combinations of variables y_i^m and x_i observed in the data. Assumptions 3.1 and 3.2 imply that:

$$\Pr\{I_i = I_i^m = 1, y_i = y_i^m = 1, x_i = x\} = \left(\frac{H_{111}}{\Pi_{11}}\right)\left(\frac{H_{11}^m}{\Pi_1^m}\right) F\{1, 1, x; \theta\}\, \Pi_{1,x}, \tag{3.9}$$

[8]In Section 5, we present Monte Carlo simulation results using ML estimates of the model. We successfully replicate the average missing and participation status patterns from the data. In contrast, if the missing status of the daughter is assumed to be independent from that of the mother, then we are unable to replicate the patterns in the data.

[9]The exclusion restrictions contained in Assumptions 3.1 and 3.2 can be relaxed in two ways. First, identification is still possible even if I_i and I_i^m are not conditionally independent of certain elements of vector x_i. Second, estimating the model for a sub-sample defined by a set of values in x_i automatically weakens the assumptions. We follow this strategy in Section 4, where it is found that AME estimates for different sub-samples are not statistically significantly different. Hence, either our assumptions regarding the missing process are correct, or the missing model is flexible enough to produce only small biases.

[10]It is not difficult to allow for continuous variables in x_i, although additional assumptions for $\Pr\{y_i^m = w, x_i\}$ would be required.

where $\Pi_{11} = \Pr\{I_i^m = 1, y_i = 1\}$ and $\Pi_{1,x}$ is the parameter of the matrix $\Pi_{w,x}$ that corresponds to the specific combination of values of variables $(y_i^m, x_i) = (1, x)$.

Consider the case in which only the daughter's information is missing. The joint probability for $\{I_i = 0, I_i^m = 1, y_i^m = 1, x_i = x\}$, which corresponds to the second term in Eq. (3.4), is:

$$
\begin{aligned}
\Pr\{I_i = 0, I_i^m = 1, y_i^m = 1, x_i = x\} &= \{\Pr\{I_i = 0, I_i^m = 1, y_i = 1, y_i^m = 1, x_i\} \\
&\quad + \Pr\{I_i = 0, I_i^m = 1, y_i = 0, y_i^m = 1, x_i\}\} \\
&= \left(\tfrac{H_{011}}{\Pi_{11}}\right)\left(\tfrac{H_{11}^m}{\Pi_1^m}\right) F\{1, 1, x_i; \theta\}\,\Pi_{1,x_i} + \left(\tfrac{H_{010}}{\Pi_{10}}\right)\left(\tfrac{H_{11}^m}{\Pi_1^m}\right) F\{0, 1, x_i; \theta\}\,\Pi_{1,x_i}
\end{aligned}
\tag{3.10}
$$

Following a similar argument, the joint probability of an observation in which the daughter participates and the mother's participation decision is missing is:

$$
\begin{aligned}
\Pr\{I_i = 1, I_i^m = 0, y_i = 1, x_i\} = &\left(\frac{H_{101}}{\Pi_{01}}\right)\left(\frac{H_{01}^m}{\Pi_1^m}\right) F\{1, 1, x_i; \theta\}\,\Pi_{1,x_i} \\
&+ \left(\frac{H_{101}}{\Pi_{01}}\right)\left(\frac{H_{00}^m}{\Pi_0^m}\right) F\{1, 0, x_i; \theta\}\,\Pi_{0,x_i}.
\end{aligned}
\tag{3.11}
$$

The joint probability of an observation in which both participation decisions are missing is:

$$
\begin{aligned}
\Pr\{I_i = 0, I_i^m = 0, x_i\} = \;& \left(\tfrac{H_{001}}{\Pi_{01}}\right)\left(\tfrac{H_{01}^m}{\Pi_1^m}\right) F\{1, 1, x_i; \theta\}\,\Pi_{1,x_i} + \\
& \left(\tfrac{H_{000}}{\Pi_{00}}\right)\left(\tfrac{H_{01}^m}{\Pi_1^m}\right) F\{0, 1, x_i; \theta\}\,\Pi_{1,x_i} + \\
& \left(\tfrac{H_{001}}{\Pi_{01}}\right)\left(\tfrac{H_{00}^m}{\Pi_0^m}\right) F\{1, 0, x_i; \theta\}\,\Pi_{0,x_i} + \\
& \left(\tfrac{H_{000}}{\Pi_{00}}\right)\left(\tfrac{H_{00}^m}{\Pi_0^m}\right) F\{0, 0, x_i; \theta\}\,\Pi_{0,x_i}.
\end{aligned}
\tag{3.12}
$$

Thus, Eq. (3.4) simply becomes:

$$
\begin{aligned}
p_i = \quad & \left(\left(\frac{H_{11y_i}}{\Pi_{1y_i}}\right)\left(\frac{H^m_{1y^m_i}}{\Pi^m_{y^m_i}}\right)F\left\{y_i, y^m_i, x_i;\theta\right\}\Pi_{y^m_i,x_i}\right)^{I_i I^m_i} \times \\
& \left(\sum\nolimits_{v\in\{0,1\}}\left(\frac{H_{01v}}{\Pi_{1v}}\right)\left(\frac{H^m_{1y^m_i}}{\Pi^m_{y^m_i}}\right)F\left\{v, y^m_i, x_i;\theta\right\}\Pi_{y^m_i,x_i}\right)^{(1-I_i)I^m_i} \times \\
& \left(\sum\nolimits_{w\in\{0,1\}}\left(\frac{H_{10y_i}}{\Pi_{0y_i}}\right)\left(\frac{H^m_{0w}}{\Pi^m_w}\right)F\left\{y_i, w, x_i;\theta\right\}\Pi_{w,x_i}\right)^{I_i(1-I^m_i)} \times \\
& \left(\sum\nolimits_{v,w\in\{0,1\}}\left(\frac{H_{00v}}{\Pi_{0v}}\right)\left(\frac{H^m_{0w}}{\Pi^m_w}\right)F\left\{v, w, x_i;\theta\right\}\Pi_{w,x_i}\right)^{(1-I_i)(1-I^m_i)} .
\end{aligned}
\tag{3.13}
$$

A clarification concerning our notation is perhaps in order. The meaning of subscript i in a given variable is that the function is to be evaluated at the value of the variable at observation i. For example, $F\{y_i, w, x_i;\theta\}$ in the third row of Eq. (3.13) should be evaluated at the value that variables y and x have at observation i and a running value w over the two potential values $\{0,1\}$ for y^m_i. In contrast, whenever the subscript i is used in parameters, it indicates that the relevant parameter is that which corresponds to the value at that observation. For example, if, for observation i, $y^m_i = a$ and $x_i = b$, then $\Pi_{y^m_i,x_i} \equiv \Pr\{y^m_i = a, x_i = b\}$.

The vector of parameters includes, in addition to θ, the probabilities $\{H_{rsv}\}$, for $r,s,v \in \{0,1\}$,$\{H^m_{sw}\}$, for $s,w \in \{0,1\}$, and $\{\Pi_{w,x}\}$, which has as many parameters as the combinations of the values of y^m_i and x_i in the data. Equation (3.13) represents the likelihood $\mathcal{L}_i$ for any given observation i as a function of the expanded parameter vector $\{\theta, \{H_{rsv}\}, \{H^m_{sw}\}, \{\Pi_{w,x}\}\}$. The log-likelihood function results from the sum of the log of $\mathcal{L}_i$, $\log(\mathcal{L}) = \sum_{i=1}^N \log(\mathcal{L}_i)$, subject to the following constraints:

$$\sum\nolimits_{r,s,v} H_{rsv} = 1 \tag{3.14}$$

$$\sum\nolimits_{s,w} H^m_{sw} = 1 \tag{3.15}$$

$$\sum\nolimits_{r,v} H_{r,s,v} = \sum\nolimits_{w} H^m_{sw} \text{ for } \quad s = 0,1 \tag{3.16}$$

$$\sum\nolimits_{w,x} \Pi_{w,x} = 1 \tag{3.17}$$

$$\Pi_v = \sum\nolimits_{w,x} F\{v, w, x;\theta\}\Pi_{w,x} \text{ for } v = 0,1 \tag{3.18}$$

$$\Pi_v = \sum_{r,s} H_{rsv} \text{ for } v = 0, 1 \tag{3.19}$$

$$\Pi^m_w = \sum_s H^m_{sw} \text{ for } w = 0, 1 \tag{3.20}$$

$$\Pi^m_w = \sum_x \Pi_{w,x} \text{ for } w = 0, 1 \ . \tag{3.21}$$

Maximum Likelihood estimation yields consistent and asymptotically efficient estimates of θ. One of the difficulties associated with this model is that the vector of parameters $\Pi_{w,x}$ grows with the number of different value combinations of variables y^m_i and x_i that are observed in the data. In our application, the number of parameters may be reduced by introducing additional restrictions. In particular, if we decompose vector x into location and time dummies, x_1, and additional controls, x_2, then we have without any loss of generality:

$$\Pi_{w,x} = \Pi_{x_2|w,x_1} \Pi_{w|x_1} \Pi_{x_1}$$

In order to reduce the number of parameters, we assume that the distribution of x_2 depends only on the location and time dummies and not on the mother's working status:

$$\Pi_{w,x} = \Pi_{x_2|x_1} \Pi_{w|x_1} \Pi_{x_1}.$$

Finally, we also assume that $\Pi_v = \Pi^m_v$ for $v \in \{0, 1\}$. Although the model is identified, a very large proportion of missing observations probably affects the concavity of the likelihood function, thereby impairing the identification of the parameters of the participation process. To improve sample identification, external information that provides direct values for $\Pi_{w|x_1}$ can be plugged into the likelihood function resulting in a reduction of the set of parameters (Imbens and Lancaster, 1994). Furthermore, we use this external information and estimates for $\Pi_{x_2|x_1}$ and Π_{x_1} to construct $\Pi^m_1 = \Pi_1$ (i.e., the unconditional probability of participation) which, in conjunction with Eq. (3.18), simplifies the likelihood function and allows the identification of the constant β_0 in Eq. (3.2).

4 Main results

In our initial specification, vector x_i includes location and time dummies. We present estimates of α and AME under three major strategies in Table 4. First, in the column labelled "Ignorability", we report estimates using the sample obtained after having dropped missing observations. Second, we use the sample obtained after imputing non-activity to missing values on both mother's and daughter's participation status. The mother is not identified for 5,839 women, or 31.5 percent of our sample. Mothers may remain unidentified: (*i*) when the mother's record precedes the first currently available registries, (*ii*) when there are coding errors, (*iii*) when the daughter is an illegitimate child, or (*iv*) when she or her mother change residency. Since sample selection cannot be ruled out due to these reasons, we consider here two approaches under Imputation. In the column labelled "Imputation A" we show results using the sample with identified mothers and impute all missing values as non-participation. In column "Imputation B" we additionally impute non-participation for the missing values when the mother's record is not identified within the sample. Third, we present results using the Likelihood Approach developed in the previous section for the Baseline scenario described in Section 2.4.

In all but one of the estimations (Imputation B), we find a positive and statistically significant estimate of the parameter associated with the mother's participation status (parameter α in Eq. (3.1)) and of the corresponding average marginal effect (AME), that is, the average change in the probability of participation when the mother's participation status changes from non-participation to participation. By comparing the AME from column "Ignorability" in Table 4 with the raw marginal effect obtained using the ratios computed from the information in Table 3, we observe a drop from 62.4 to 12.9 pp (12.9 pp corresponds to an AME estimate in Table 4 of 0.129). This drop occurs because no factors are controlled for in the raw estimate. Similarly, for Imputation A and B estimates, the AME drops from 5.2 and 2.9 pp in the raw data to 2.0 and -1.9 pp, respectively. Following the patterns found in the raw data (Table 3 and Appendix D), the AME under Ignorability is larger than under Imputation.

The size of the AME using the Likelihood Approach is substantially larger than under

Table 4: Benchmark results

Estimates, AMEs, and Ignorability Tests

	Ignorability	Imputation A	Imputation B	Likelihood Approach
y^M	0.827***	0.308***	-0.199***	1.610***
	(0.214)	(0.073)	(0.065)	(0.126)
AME	0.129***	0.020***	-0.019***	0.397***
	(0.033)	(0.005)	(0.006)	(0.030)
$\widehat{\Pi}_1 = \widehat{\Pr}\{y = 1\}$	.	.	.	0.288
$\widehat{\Pr}\{I^m = 1 \,\vert y^m = 0\}$	0.121	0.075	0.074	0.097
$\widehat{\Pr}\{I^m = 1 \,\vert y^m = 1\}$	0.121	0.685	1.000	0.182
$\widehat{\Pr}\{I = 1 \,\vert I^m = 0, y = 0\}$	0.140	0.065	0.067	0.086
$\widehat{\Pr}\{I = 1 \,\vert I^m = 0, y = 1\}$	0.140	1.000	1.000	0.203
$\widehat{\Pr}\{I = 1 \,\vert I^m = 1, y = 0\}$	0.140	0.271	0.271	0.515
$\widehat{\Pr}\{I = 1 \,\vert I^m = 1, y = 1\}$	0.140	1.000	1.000	0.106
Ignorability test	.	.	.	704.04
				[0.000]
$\bar{y}$	0.209	0.032	0.053	0.377
No. obs.	570	12684	18523	18523

Note: Standard errors are in parenthesis and $p-$values in brackets. The dependent variable is the daughter's participation decision. Dummy y^m indicates the mother's participation status. All models include quarter-century dummies (reference quarter-century is 1675-1699) and a location dummy for Horta. The results are based on the sample from Horta (1700-1850), Ronfe (1675-1925), and Ruivães (1825-1925). "Ignorability" reports results dropping observations with missing participation status of either mothers or daughters. "Imputation" displays results after imputing non-participation to all missing values. In "Imputation A" our sample is restricted to those women whose records are successfully linked with those of their mothers. In "Imputation B" we include all women. "Likelihood Approach" reports ML estimates for the model in Section 3.2 under the Baseline scenario described in section 2.4. For the period where external information is available, we use as external information the predicted participation rates shown in Figure B1 which are predictions from the estimation of Eq. (B.1) in Appendix B. For the period where external information is not available, we take the value of the 1890 predictions. "AME" refers to the estimated average marginal effect, i.e. the sample average change in the estimated probability of participation when the mother participation status changes from no-participation to participation. $\widehat{\Pi}_1$ is the estimated unconditional probability of participation. $\widehat{\Pr}$'s are estimates of the conditional probabilities of missing values. "Ignorability test" is the Wald statistic for the null that probabilities $\Pr\{I_i = 1 \,\vert I_i^m, y_i\}$ and $\Pr\{I_i^m = 1 \,\vert y_i^m\}$ do not vary with y, y^m, and I^m. $\bar{y}$ is the sample average of the dependent variable.

Ignorability or Imputation. According to the results using the Likelihood Approach, a woman whose mother participates in the labour market has a probability of participation which is 39.7 pp larger than a woman whose mother does not participate.[11] Naturally,

[11] If we change the definition of participation in the labour market by excluding property owners from participation, then the results remain unchanged: the AME estimate is 39.7 pp (s.e. 2.9).

sample averages of the dependent variable differ across Ignorability and Imputation. They also differ between Ignorability and the Likelihood Approach because the sample employed to compute $\overline{y}$ expands from 570 under Ignorability to additionally include observations with missing participation status of the mother to a total of $2,584$ under the Likelihood Approach. For the latter, we also report the ML estimate for the unconditional probability of participation (parameter Π_1 in our model), which is approximately 9 pp smaller than $\overline{y}$. As anticipated, priests seem to selectively report participants.

In order to formally test Ignorability in the model (3.13), we test whether $\Pr\{I_i = 1 \,|I_i^m, y_i\}$ and $\Pr\{I_i^m = 1 \,|y_i^m\}$ do not vary with y, I^m, and y^m. In the bottom half of Table 4 we report these conditional probabilities and the Wald test (which we label "Ignorability test") for the equality of all conditional probabilities in the Likelihood Approach.[12] The differences in conditional probabilities are sizable and congruent with the hypothesis that priests selectively report participants. For example, the probability that the mother's participation status is observed almost doubles when she participates. Similarly, the probability that the daughter's participation status is observed increases from 8.6 percent to 20.3 percent when she participates if the mother's participation status is not observed. Notably, this statistical association between participation and its observability reverses and amplifies when the participation of the mother is observed. In this case, the probability that the participation of the daughter is observed when she is participating is 10.6 percent whereas it reaches 51.5 percent when she is not participating. When the priest reports the mother, signalling her as socially important, the fact that her daughter does not participate is statistically associated with a very high probability that the priest reports her as "dona", that is, she does not participate. These large differences in conditional probabilities lead to a strong rejection of the Ignorability test.

In addition, the ML estimates discard the plausibility of the implicit assumption under Imputation, that is, that missing participation implies non-participation in the labour

[12] These probabilities are non-linear functions of the parameters of the model. For example, $\Pr\{I_i^m = 1 \,|y_i^m = 0\} = \frac{H_{10}^m}{\Pi_0^m}$ where $\Pi_0^m = \Pr\{y_i^m = 0\} = H_{00}^m + H_{10}^m$ and H_{10}^m, and H_{00}^m are parameters in the model.

market. For example, given that

$$\Pr\left(y^m = 1 | I^m = 0\right) = \Pr\left(I^m = 0 | y^m = 1\right) \times \frac{\Pr\left(y^m = 1\right)}{\Pr\left(I^m = 0\right)},$$

the ML estimate for the probability that mothers participate if their information is missing is $(1 - 0.182) \times \frac{0.288}{0.8787} \times 100 \approx 26.8$ percent, a value which is well over zero.

4.1 Alternative scenarios

Hitherto, we have reported results under the Baseline scenario. In this scenario, the local predictions of female participation rates in 1890 from the estimation of Eq. (B.1) in Appendix B are used for all quarter-centuries before the year 1890. This is perhaps restrictive. In this section, we propose two alternative strategies to assess the validity of the results under the Baseline scenario.

The first strategy considers two scenarios with constant but extreme local participation rates before 1890. In what we refer to as the "Low Participation scenario", we take the smallest local participation rate for which we have information in all censuses from 1890 to 1950. Similarly, in the "High Participation scenario", the largest local participation rate is taken. Specifically, we take the smallest and largest rates in the borough of Horta for Horta and the smallest and largest values in the nearby borough of Vila Nova de Famalicão for Ronfe and Ruivães.

In the second strategy, a large number of paths for each location are simulated backwards under the weak assumption that 25-year changes before 1890 were no larger, in absolute value, than the average changes from 1890 to 1950. To simulate the change in the local participation rate from one quarter-century to the previous quarter-century, we consider two possibilities: increase or decrease. In the case of an increase, the change is drawn from the uniform distribution between zero and the local mean increase in the 1890-1950 period. In the case of a decrease, the value is drawn from the uniform distribution between the local mean decrease and zero. For Horta, there are a total of $2^7 = 128$ combinations of possible increases and decreases in the seven quarter-centuries from 1700-1724 to 1850-1874. The number of possible combinations for Ronfe is 256 because it has an additional quarter-century, while for Ruivães it is 2 because there is only one quarter-century before 1890. The interaction of all the combinations of the three locations generates a total of $128 \times 256 \times 2 = 65,536$ scenarios, each with a different participation path per location. Figure 1 shows several examples of simulated paths for Horta and Ronfe.

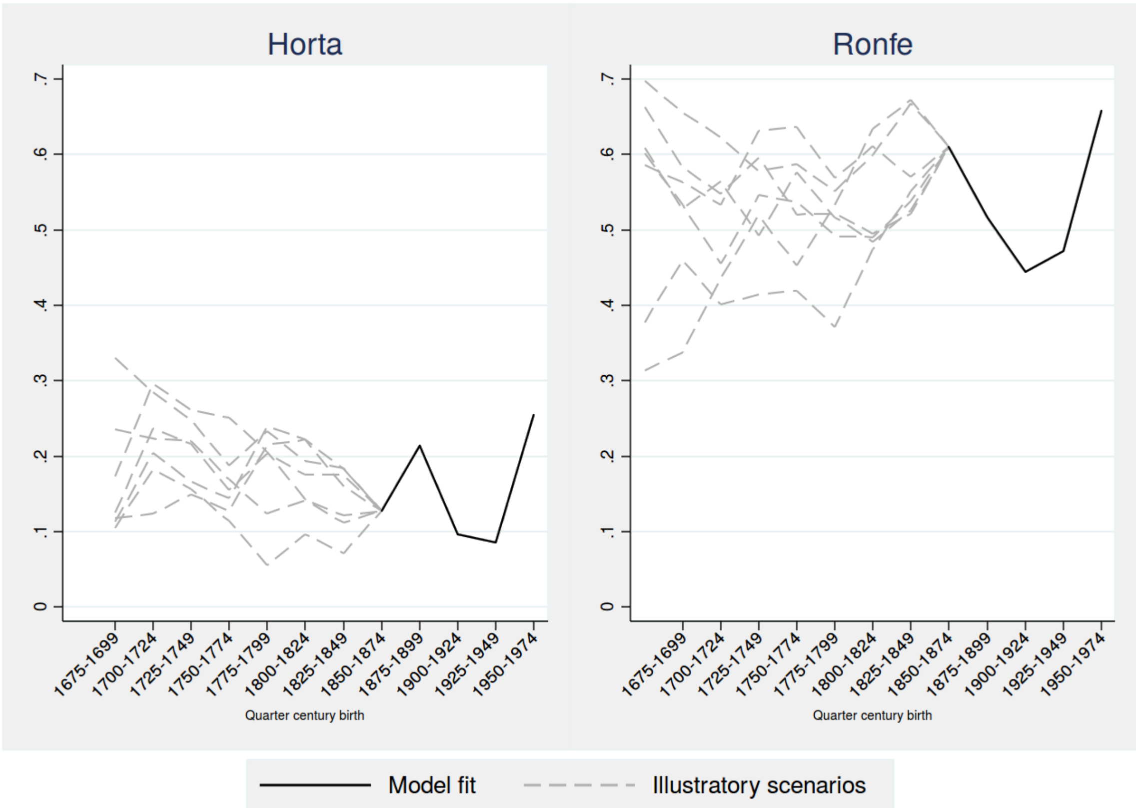

Figure 1: Local female participation rates.

Note: Female participation rates in the quarter century following the quarter century birth. "Model fit" refers to predictions in midpoint years of each quarter century from the estimation of Eq. (B.1) in Appendix B. "Illustratory scenarios" are seven examples of simulated paths obtained from the algorithm explained in Section 4.1 for the period before any external sources are available.

We estimate the model using the Likelihood Approach for all scenarios under the two strategies. Table 5 shows the Ignorability and Baseline results from Table 4, the estimates under the High and Low Participation scenarios, and the minimum and maximum AME estimates out of the 65,536 simulated paths (columns "Minimum AME" and "Maximum AME", respectively).

In contrast to the Ignorability case, the Baseline results lie comfortably within the

Table 5: Alternative Scenarios for External Information

Estimates and AMEs

	Ignorability	Likelihood Approach				
		Baseline	High Participation	Low Participation	Minimum AME	Maximum AME
y^M	0.827***	1.610***	1.323***	1.927***	1.327***	1.920***
	(0.214)	(0.126)	(0.122)	(0.126)	(0.125)	(0.131)
AME	0.129***	0.397***	0.325***	0.466***	0.323***	0.486***
	(0.033)	(0.030)	(0.029)	(0.029)	(0.029)	(0.034)

Note: Standard errors are in parenthesis. The dependent variable is the daughter's participation decision. All models include quarter-century dummies (reference quarter-century is 1675-1699) and a location dummy for Horta. The results are based on the sample from Horta (1700-1850), Ronfe (1675-1925), and Ruivães (1825-1925). Dummy y^m indicates the mother's participation status. "AME" refers to the estimated average marginal effect, i.e. the sample average change in the estimated probability of participation when the mother participation status changes from no-participation to participation. "Ignorability" reports results dropping observations with missing participation status of either mothers or daughters. We report "Likelihood Approach" ML estimates for the model in Section 3.2 under five scenarios for the external information on participation rates. They only differ in participation rates for periods before 1890 for which we do not have external information. "Baseline" reports ML estimates taking the value of the 1890 predictions from the estimation of Eq. (B.1) in Appendix B. In the "High Participation" scenario we take the largest local participation rates available from 1890 to 1950. In the "Low Participation" scenario we take the smallest local participation rates available from 1890 to 1950. Columns "Minimum AME" and "Maximum AME" report estimates from the two scenarios with minimum and maximum AME results out of 65,536 random participation paths for the periods before 1890 as explained in Section 4.1.

range defined by the Minimum and Maximum AMEs. Furthermore, the latter two values are surprisingly similar to those of the High and Low Participation scenarios, which strongly suggests that the High and Low Participation scenarios provide reasonable extremes to capture the potential variability of the AME given the lack of external information before 1890. Henceforth, for computational reasons, we only report Baseline and High and Low Participation results.

4.2 The stability of the role of mothers

Results from the Likelihood Approach in Table 4 are obtained under the assumption that α in Eq. (3.1) is invariant across periods. To assess the validity of this assumption, our working sample is divided into three periods: The first period ranges from 1675 to 1750; the second period starts in 1775 and ends in 1825; , and the remaining four quarter-centuries constitute the last period (i.e. 1850-1925).[13]

Sample identification problems arise in the first period because all reported mother-daughter pairs are non-participants. This makes estimation under Ignorability impossible and leads to non-convergence and instability under the Likelihood Approach. The first period is therefore disregarded and in the remaining part of this section the focus is on the second and third periods. Table 6 presents the Baseline and High and Low Participation results for the second period and, for brevity, only the Baseline results for the third period.[14]

Table 6: ML Estimates for Different Periods

Estimates, AMEs, and Ignorability Tests

	1675-1925	1775-1925			1775-1825			1850-1925
	Baseline	Baseline	High Participation	Low Participation	Baseline	High Participation	Low Participation	Baseline
y^M	1.610***	1.645***	1.422***	1.862***	1.718***	1.489***	2.166***	1.021***
	(0.126)	(0.129)	(0.123)	(0.133)	(0.237)	(0.228)	(0.243)	(0.217)
AME	0.397***	0.459***	0.396***	0.518***	0.499***	0.402***	0.618***	0.379***
	(0.030)	(0.033)	(0.031)	(0.034)	(0.071)	(0.060)	(0.067)	(0.075)
$\widehat{\Pi}_1 = \widehat{\Pr}\{y=1\}$	0.288	0.310	0.381	0.243	0.239	0.344	0.147	0.385
$\widehat{\Pr}\{I^m=1\,\vert y^m=0\}$	0.097	0.116	0.130	0.106	0.089	0.103	0.079	0.156
$\widehat{\Pr}\{I^m=1\,\vert y^m=1\}$	0.182	0.215	0.174	0.278	0.065	0.046	0.105	0.304
$\widehat{\Pr}\{I=1\,\vert I^m=0, y=0\}$	0.086	0.114	0.127	0.103	0.097	0.114	0.086	0.148
$\widehat{\Pr}\{I=1\,\vert I^m=0, y=1\}$	0.203	0.254	0.198	0.340	0.084	0.059	0.139	0.358
$\widehat{\Pr}\{I=1\,\vert I^m=1, y=0\}$	0.515	0.501	0.543	0.458	0.833	0.868	0.801	0.151
$\widehat{\Pr}\{I=1\,\vert I^m=1, y=1\}$	0.106	0.121	0.114	0.133	0.163	0.145	0.186	0.156
Ignorability test	703.53	517.47	255.39	874.58	242.59	385.17	344.72	326.18
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
$\bar{y}$	0.377	0.399	0.399	0.399	0.160	0.160	0.160	0.570
No. obs.	18523	12782	12782	12782	6585	6585	6585	6197

Note: Standard errors are in parenthesis and $p-$values in brackets. The dependent variable is the daughter's participation decision. Dummy y^m indicates the mother's participation status. All models include quarter-century dummies and a location dummy for Horta. The "1675-1925" sample includes Horta (1700-1850), Ronfe (1675-1925) and Ruivães (1825-1925) and replicates the results shown in the fourth column of Table 4. The "1775-1925" sample includes Horta and Ronfe (1775-1850) and Ruivães (1825-1925). The "1775-1825" sample includes Horta and Ronfe (1775-1825), and Ruivães (1825). The "1850-1925" sample includes Horta (1850) and Ronfe and Ruivães (1850-1925). Results are ML estimates for the model in Section 3.2. We consider three scenarios for the external information for participation rates. They only differ in participation rates for periods before 1890 for which we do not have external information. For the period where external information is available, we use as external information the predicted participation rates shown in Figure B1 which are predictions from the estimation of Eq. (B.1) in Appendix B. For the period where external information is not available: (a) in the baseline scenario we take the value of the 1890 predictions; and (b) in the Low (High) Participation scenario, we take the smallest (largest) local participation rates available from 1890 to 1950. "AME" refers to the estimated average marginal effect, i.e. the sample average change in the estimated probability of participation when the mother participation status changes from no-participation to participation. $\widehat{\Pi}_1$ is the estimated unconditional probability of participation. $\widehat{\Pr}$'s are estimates of the conditional probabilities of missing values. "Ignorability test" is the Wald statistic for the null that probabilities $\Pr\{I_i=1\,\vert I_i^m, y_i\}$ and $\Pr\{I_i^m=1\,\vert y_i^m\}$ do not vary with y, y^m, and I^m. $\bar{y}$ is the sample average of the dependent variable.

All AME estimates are large and statistically significantly different from zero. For

[13] Since the number of missing values decreases over time, this selection of periods results in samples with different proportions of missing values. In the next section, we show via Monte Carlo simulation that under the Likelihood Approach variations in these proportions do not bias our results.

[14] Scenarios only apply to observations prior to 1890 and therefore they affect only a small number of observations in the third period sample with only negligible effects on the estimates (available upon request).

the sample without the first period, that is, for 1775-1925, the results increase but the confidence intervals overlap with those shown in Table 4 using all periods. In the second period 1775-1825, which ends before 1890, the external information varies across scenarios for all observations. Unsurprisingly, it is for this period where the largest differences are found across scenarios in the AME estimates. For the third period, that is, for 1850-1925, we find a smaller effect: 39.7 vs 49.9 pp. This is consistent with a decreasing role of mothers over time as participation increases from a $100 \times \widehat{\Pi}_1$ estimate of 23.9 percent to an estimate of 38.5 percent.

The bottom half of Table 6 deserves several comments. Ignorability tests are again strongly rejected in both periods. Naturally, as the number of missing values decreases over time, most of the estimates of the conditional probabilities of observability increase. More interestingly, by distinguishing between the two periods, we find evidence of decreasing influence of social status on the reporting activity of the priest at the turn of the century: for example, for $y = 0$, the role of I^m on $\Pr(I = 1|y, I^m)$ decreases substantially from $83.3 - 9.7 = 73.6$ pp in the 1775-1825 period to $15.1 - 14.8 = 0.3$ pp in the 1850-1925 period.

4.3 Additional controls

4.3.1 Family size and father's social status

Permanent and quasi-permanent factors induce intergenerational persistence in participation decisions. Inasmuch as the goal is to assess persistence arising from the mother-daughter link, other factors that induce persistence need to be controlled for. These include labour market conditions, sex-ratio differentials, and social multiplier effects captured by the location and time dummies used in our specification in Table 4. In this section, we add to Eq. (3.1) family characteristics that may induce persistence. Within family characteristics, we consider family size, which is proxied by an indicator of more than four siblings, *siblings*, and two dummy variables related to the father's socio-economic status: (i) whether the priest records that the father is a rentier, an owner, a merchant, a high-ranking civil servant, or an officer, *father SES*, and (ii) whether the priest reports the father to be an owner, *father owner*. These two dummy variables are important because they capture at least two sources of persistence within the family: the transmission of wealth (which probably leads to lower participation rates) and that of physical capital (which should lead to higher rates according to our participation measure). The dummy variable *father owner* identifies the latter effect of property transmission.

One could argue that a measure of human capital should also be added. Throughout the sample period, however, any proxy for wealth is also a proxy for educational achievement: female illiteracy rates remained over 68 percent in Portugal in the 1930s and formal education was only accessible to a small and privileged group (Candeias et al., 2004). Hence, the controls for wealth indirectly control for human capital.[15]

Table 7 replicates Table 4 with the additional controls. As in Table 4, the results under Ignorability and Imputation differ widely from those obtained with the Likelihood Approach. The AME estimates with the Likelihood Approach are smaller because the additional controls capture part of the inertia within the family. Contrary to the results in Table 5, the AME estimates under the Baseline, and the High and Low Participation scenarios are very similar. Under the Baseline scenario, a woman whose mother participates in the labour market has a probability of participation which is 26.7 pp greater than a woman whose mother does not participate. This value is relatively large as it represents over 70 percent of the sample average participation and implies an increase of 126.5 percent in the probability of participation for those women whose mothers do not participate.[16] If the actual effect were zero and our AME estimate were biased due to selection on unobservables, this bias would double that from selection on observables (that is, the difference between 39.7 and 26.7). The magnitude of the estimated effect suggests that this is not the situation.[17]

[15] Educational reforms initiated in 1822, 1835, and 1844 were primarily targeted at boys' education and were left incomplete and largely unimplemented (see Appendix A for literacy figures for boys and girls in 1864).

[16] Given that $\Pi_1 = (1-\Pi_1) \times \Pr(y=1|y^m=0) + \Pi_1 \times \Pr(y=1|y^m=1)$ and $\Pr(y=1|y^m=1) = \Pr(y=1|y^m=0) + \text{AME}$, then the estimated probability of participation for those women whose mothers do not participate is $\Pr(y=1|y^m=0) = \Pi_1 \times (1-\text{AME})$. Since $\widehat{\Pi}_1 = 0.288$, then $\frac{\text{AME}}{(1-\text{AME})\times\Pi_1} \times 100 = \frac{0.267}{(1-0.267)\times 0.288} \times 100 \approx 126.5$.

[17] These results are robust to different specifications of the controls. In particular, the results for AME lie between 24.4 and 29.6 pp when the following dummy variables are included: (i) dummy variables for

Table 7: Results controlling for SES

Estimates, AMEs, and Ignorability Tests

	Ignorability	Imputation A	Imputation B	Likelihood Approach Baseline	 High Participation	 Low Participation
y^M	0.743***	0.303***	-0.082	1.223***	1.128***	1.338***
	(0.227)	(0.074)	(0.066)	(0.119)	(0.120)	(0.115)
siblings	0.303*	0.040	-0.410***	0.697***	0.718***	0.663***
	(0.167)	(0.051)	(0.043)	(0.074)	(0.075)	(0.074)
father SES	-0.118	0.119	0.154	-0.328**	-0.167	-0.380***
	(0.220)	(0.125)	(0.122)	(0.148)	(0.147)	(0.146)
father owner	0.323	-0.054	-0.113	0.389***	-0.031	0.651***
	(0.341)	(0.173)	(0.165)	(0.109)	(0.093)	(0.115)
AME	0.115***	0.019***	-0.008	0.267***	0.256***	0.263***
	(0.035)	(0.005)	(0.006)	(0.028)	(0.028)	(0.025)
$\widehat{\Pi}_1 = \widehat{\Pr}\{y=1\}$	.	.	.	0.288	0.369	0.213
$\widehat{\Pr}\{I^m=1\,\vert y^m=0\}$	.	.	.	0.098	0.110	0.088
$\widehat{\Pr}\{I^m=1\,\vert y^m=1\}$	.	.	.	0.181	0.141	0.248
$\widehat{\Pr}\{I=1\,\vert I^m=0,y=0\}$	.	.	.	0.103	0.105	0.101
$\widehat{\Pr}\{I=1\,\vert I^m=0,y=1\}$	.	.	.	0.136	0.132	0.142
$\widehat{\Pr}\{I=1\,\vert I^m=1,y=0\}$	.	.	.	0.703	0.689	0.713
$\widehat{\Pr}\{I=1\,\vert I^m=1,y=1\}$	.	.	.	0.083	0.084	0.083
Ignorability test	.	.	.	108280.97	90353.74	121734.54
				[0.000]	[0.000]	[0.000]
$\bar{y}$	0.209	0.032	0.053	0.377	0.377	0.377
No. obs.	570	12684	18523	18523	18523	18523

Note: Standard errors are in parenthesis and $p-$values in brackets. The dependent variable is the daughter's participation decision. Dummy y^m indicates the mother's participation status. All models include quarter-century dummies (reference quarter-century is 1675-1699) and a location dummy for Horta. See the main text in Section 4.3.1 for definitions of other controls. The results are based on the sample from Horta (1700-1850), Ronfe (1675-1925), and Ruivães (1825-1925). "Ignorability" reports results dropping observations with missing participation status of either mothers or daughters. "Imputation" displays results after imputing non-participation to all missing values. In "Imputation A" our sample is restricted to those women whose records are successfully linked with those of their mothers. In "Imputation B" we include all women. We report "Likelihood Approach" ML estimates for the model in Section 3.2 under three scenarios for the external information for participation rates. They only differ in participation rates for periods before 1890 for which we do not have external information. "Baseline" reports ML estimates taking the value of the 1890 predictions from the estimation of Eq. (B.1) in Appendix B. In the "High Participation" scenario we take the largest local participation rates available from 1890 to 1950. In the "Low Participation" scenario we take the smallest local participation rates available from 1890 to 1950. "AME" refers to the estimated average marginal effect, i.e. the sample average change in the estimated probability of participation when the mother participation status changes from no-participation to participation. $\widehat{\Pi}_1$ is the estimated unconditional probability of participation. $\widehat{\Pr}$'s are estimates of the conditional probabilities of missing values. "Ignorability test" is the Wald statistic for the null that probabilities $\Pr\{I_i=1\,|I_i^m,y_i\}$ and $\Pr\{I_i^m=1\,|y_i^m\}$ do not vary with y, y^m, and I^m. $\bar{y}$ is the sample average of the dependent variable.

Regarding the estimates of the additional controls, their signs are not always consistent across the three approaches. Under the Likelihood Approach, they are always smaller in magnitude than the estimate for the mother participation coefficient $\widehat{\alpha}$. Conditional on the participation of the mother, a large number of siblings, a low SES status of the father, and an owner status of the father, increase the probability that the woman participates in the labour market.

family size, (ii) interactions of the size of the family with the periods defined in Section 4.2, and (iii) two dummy variables related to the father's profession (whether he is a merchant or a soldier).

The estimated probabilities of observability are similar to the ones presented in the Likelihood Approach in Table 4. Moreover, we also strongly reject that these probabilities do not vary with y, I^m, and y^m.

4.3.2 Migration

The transmission of wealth could also be reflected in migration flows. On the one hand, large families and low socioeconomic status may increase the probability of migration for economic reasons. On the other hand, marriages among wealthy families may involve migration flows and be associated with non-participation. Hence, migration status could be an additional control for transmission of wealth. Unfortunately, place of birth (which identifies migration status) is not reported in all records. In Ronfe and Ruivães the place of birth is missing for 77.7 percent and 70.9 percent of the observations, respectively. In contrast, in Horta only 6.1 percent of observations have this information missing.

Here we discuss results under the Likelihood Approach when (*i*) adding a dummy variable for migrant status for the sample of Horta and (*ii*) estimating the model for the subgroup of migrant women in Horta.

When we restrict the sample to immigrants in Horta, the number of observations drops to 3,571, and we lose the sample identification for the extended model with socioeconomic variables. We can nevertheless get estimates by ML in this small sample with a simpler variable specification. Specifically, we replace the quarter-century dummies with century dummies and, given that the average family size in Horta is smaller, redefine the dummy variable *siblings* to the existence of more than two siblings. With this new specification, the AME for the complete Horta sample becomes 36.8 pp (s.e. 6.5 pp). Addition of the migrant status dummy variable only slightly changes the AME: 32.4 pp (s.e. 6.5 pp). The estimate of the coefficient for migrant status is negative and significant: -0.349 (s.e. 0.090). This suggests that migrants into Horta were, on average, non-participants and, hence, migration was probably associated with marriage decisions. Finally, for the sample of migrants, the AME is 27.2 pp (s.e. 11.8 pp). To sum up, all these AME estimates overlap, which suggests that the results are robust to migration flows.

5 Sensitivity of results to prevalence of missingness

Our working sample is very large ($n = 18,523$) and the number of mother-daughter pairs with observed participation (696, see Table 3) is not unusual for historical data. Under Assumptions 3.1 and 3.2, the ML estimator is consistent. However, the proportion of missing values for y (86.1 percent) and for y^m (87.9 percent) might raise doubts regarding the robustness and reliability of the results.

In order to evaluate the impact of the high proportion of missing values on our estimates, we perform a Monte Carlo experiment with $R = 250$ simulations based on our data and the extended model with additional controls from Section 4.3.1. (See Appendix E for the algorithm details.)

Figure 2 presents the smoothed densities of all estimators of the AME obtained from the Monte Carlo experiment, both under Ignorability and the Likelihood Approach. These show that ignoring the missingness process results in downward biased estimates of the effect, even for a proportion of missing values as low as 20 percent (upper panel in Fig. 2). In contrast, under the Likelihood Approach, the AME is effectively unbiased regardless of the proportion of missing values (lower panel in Fig. 2). The larger the proportion of missing values, the larger the standard errors. However, the AME is precisely estimated even for the largest missing incidence (with a standard deviation, 2.39 pp, more than ten times smaller than the true AME, 26.74 pp), probably due to the large sample size. In addition to the results shown in Fig. 2, the proportion of simulations in which the true value falls within the 95 percent confidence interval in the Likelihood Approach with the largest missing incidence is 94 percent.

To sum up, our Monte Carlo analysis highlights the fundamental role played by the information contained in I and I^m, and by the modelling of the missing process on the robustness and reliability of the Likelihood Approach even when the incidence of missing values is as large as 87 percent.

6 Conclusions

In this paper, we use historical parish registry data from the late 17th century until the beginning of the 20th century from three Portuguese locations to estimate the relation between mothers' labour market participation and that of their daughters. Our main data source is drawn from baptism, marriage, and death certificates found in the local churches.

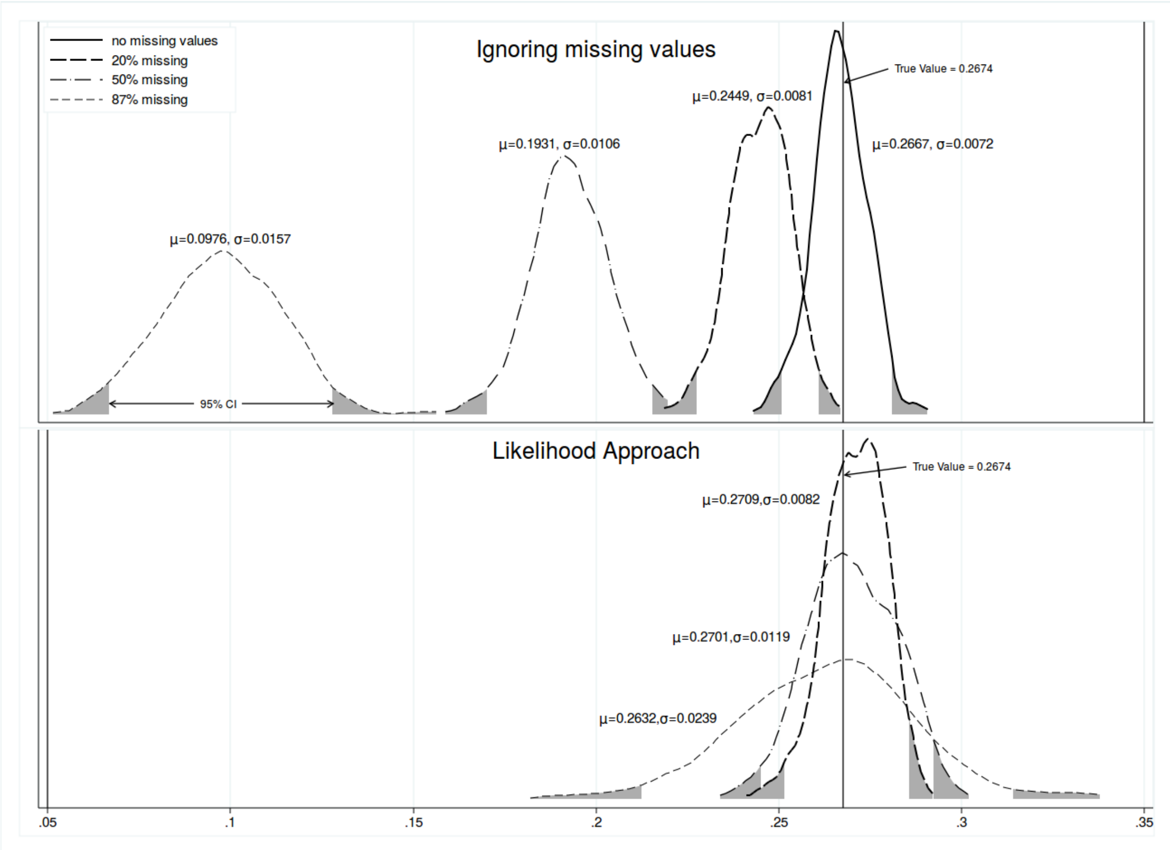

Figure 2: Smoothed densities for AME estimators.

Note: Smoothed densities for AME estimators using simulation results with 250 replications. “AME” refers to the estimated average marginal effect, i.e. the sample average change in the estimated probability of participation when the mother participation status changes from no-participation to participation. The smoothed densities are obtained using the Epanechnikov kernel function computed using Stata® `kdensity` command with default bandwith parameter. Parameters of the DGP are the estimates in Baseline column in Table 7 (AME value in the DGP referred to as “True Value”). Participation and missing statuses for daughters and mothers are simulated as described in the text. “20% missing” (“50%”) means that on average 20% (50%) of mothers and daughters have their participation status unreported. “87%” means that missing statuses for mothers and daughters are simulated to target the actual proportion of missing values in the original dataset. In Panel “Ignoring missing values” we show smoothed densities for probit estimators ignoring the simulated observations with missing participation status of mothers and daughters. We also show probit estimates obtained with the simulated data without missing values. In Panel “Likelihood Approach” we show smoothed densities from ML estimates of the model in Section 3.2 using the external information under the Baseline scenario.

In addition, our data was matched with information from parochial censuses carried out by the priests during Lent to administer the sacrament of penance to the parishioners. Although these censuses contain invaluable information on occupation and/or social status, the coverage never exceeds 36 percent and, in certain locations and periods, this falls to below 10 percent. We argue that coverage of occupation/social status is associated with non-random individual factors such as social status and activity choices. Therefore, any restriction of the estimating sample to only those observations with complete coverage would result in a selected sample and estimation biases. We address the problem of non-random missing values by adapting the methodology proposed by Ramalho and Smith (2013). By allowing the estimation of models in contexts in which missing data are abundant and non-random, this methodology confers considerable potential to the examination of historical data.

Our results show a large and positive statistically significant relation between the mother's working status and the daughter's decision to participate in the labour market. After having controlled for location and time dummies and socio-economic characteristics, the probability that a woman participates in the labour market increases by 26.7 pp if her mother also works. One way to assess the potential importance of this estimate is to simulate the long-term evolution of the FLFP process. The average probability of participation in our sample is 29 percent, although for women whose mothers did not work, this is only 19 percent. Now imagine a 50 percent increase of the latter to 29 percent. This level would be the expected long-term participation rate in the absence of a mother-to-daughter association. In contrast, with the estimated association between mothers and daughters' FLFP, the long-term participation rate would be expected to increase to 52 percent.

The existence of such an early transmission mechanism acting as a catalyst of change contributes towards the understanding of the long-term dynamics of female labour force participation.

Compliance with Ethical Standards

Funding: This study was funded by Ramón Areces Foundation, and the Spanish Ministry of Science and Innovation through Grants ECO2016-78652 (P. Machado), ECO2015-65204-P (Carro and Mora), RTI2018-095231-B-I00 (Carro), and PID2019-108576RB-I00 (P. Machado and Mora).

Conflict of Interest: All Authors declare that they have no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors

References

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

Amorim, M. N. (1991). Uma metodologia de reconstituição de paróquias. Technical report, Universidade do Minho.

Amorim, N. and C. Santos (2009). Marriage strategies in azorean communities of pico island (19th century). In A. Margarida Duraes, L. Fauve-chamoux, and J. K. Ferrer (Eds.), *The Transmission of Well-Being: Gendered Marriage Strategies and Inheritance Systems in Europe (17th-20th Centuries)*. Bern: Peter Lang Pub Incorporated.

Bisin, A. and T. Verdier (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory 97*(2), 298–319.

Brettell, C. B. (1991). Kinship and contract: Property transmission and family relations in northwestern portugal. *Comparative Studies in Society and History 33*(03), 443–465.

Brettell, C. B. (1998). Historical perspectives on infant and child mortality in northwestern portugal. In N. Scheper-Hughes and C. Sargent (Eds.), *Small Wars: The Cultural Politics of Childhood*. University of California Press Berkeley, CA.

Candeias, A., A. L. Paz, and M. Rocha (2004). *Alfabetização e escola em Portugal nos séculos XIX e XX: os censos e as estatísticas*. Fundação Calouste Gulbenkian.

Carrilho, M. J. (1996). População activa: Conceito e extensão através dos censos. Technical report, Lisbon: INE.

de Pina Cabral, J. (1986). *Sons of Adam, daughters of Eve. The peasant worldview of the Alto Minho*. Oxford (UK) Clarendon Press.

Durães, M. (2009). Providing well-being to women through inheritance and succession. In A. Margarida Durães, L. Fauve-chamoux, and J. K. Ferrer (Eds.), *The Transmission of Well-Being: Gendered Marriage Strategies and Inheritance Systems in Europe (17th-20th Centuries)*. Bern: Peter Lang Pub Incorporated.

Farré, L. and F. Vella (2013). The intergenerational transmission of gender role attitudes and its implications for female labour force participation. *Economica 80*(318), 219–247.

Fernández, R. (2013). Cultural change as learning: The evolution of female labor force participation over a century. *The American Economic Review 103*(1), 472–500.

Fernández, R., A. Fogli, and C. Olivetti (2004). Mothers and sons: Preference formation and female labor force dynamics. *The Quarterly Journal of Economics 119*(4), 1249–1299.

Goldin, C. (1995). *The U-Shaped Female Labor Force Function in Economic Development and Economic History*, pp. 61–90. University of Chicago Press.

Gomes, P. and M. P. Machado (2020). Literacy and primary school expansion in portugal: 1940-62. *Revista de Historia Económica-Journal of Iberian and Latin American Economic History 38*(1), 111–145.

Humphries, J. and C. Sarasúa (2012). Off the record: Reconstructing women's labor force participation in the european past. *Feminist Economics 18*(4), 39–67.

Imbens, G. W. and T. Lancaster (1994). Combining micro and macro data in microeconometric models. *The Review of Economic Studies 61*(4), 655–680.

Kawaguchi, D. and J. Miyazaki (2009, Jan). Working mothers and sons' preferences regarding female labor supply: direct evidence from stated preferences. *Journal of Population Economics 22*(1), 115–130.

Lains, P. (2006). Growth in a protected environment: Portugal, 1850–1950. *Research in Economic History 24*, 119–160.

Little, R. J. and D. B. Rubin (2002). *Statistical analysis with missing data*, Volume 333. Wiley-Interscience.

Matos, P. L. (2009). Female life courses and property transmission in the azorean periphery (portugal). the case of the island of são jorge in the 19th century. In A. Margarida Duraes, L. Fauve-chamoux, and J. K. Ferrer (Eds.), *The Transmission of Well-Being: Gendered Marriage Strategies and Inheritance Systems in Europe (17th-20th Centuries)*. Bern: Peter Lang Pub Incorporated.

Maurin, E. and J. Moschion (2009). The social multiplier and labor market participation of mothers. *American Economic Journal: Applied Economics*, 251–272.

Mesquita, M. and J. Leite (2013). Agregados domésticos na paróquia da sé de angra no século xviii. uma abordagem a partir dos registos paroquiais e dos róis de confessados. In *Actas do I Congresso Histórico Internacional*.

Moen, P., M. A. Erickson, and D. Dempster-McClain (1997). Their mother's daughters? the intergenerational transmission of gender attitudes in a world of changing roles. *Journal of Marriage and the Family*, 281–293.

Moreira da Silva, R. F. (1983). Contraste e mutações na paisagem agrária das planícies e colinas minhotas. *Studium Generale. Estudos contemporâneos 5*, 9–115.

Morrill, M. S. and T. Morrill (2013). Intergenerational links in female labor force participation. *Labour Economics 20*(1), 38–47.

Olivetti, C., E. Patacchini, and Y. Zenou (2018, 12). Mothers, Peers, and Gender-Role Identity. *Journal of the European Economic Association 18*(1), 266–301.

Palma, N. and J. Reis (2019). From convergence to divergence: Portuguese economic growth, 1527-1850. *The Journal of Economic History 79*(2), 477–506.

Pedreira, J. (1990). Social structure and the persistence of rural domestic industry in nineteenth century portugal. *Journal of European Economic History 19*(3), 521–547.

Ramalho, E. A. and R. J. Smith (2013). Discrete choice non-response. *The Review of Economic Studies 80*(1), 343–364.

Reis, J. (2005). O trabalho. In P. Lains and A. F. da Silva (Eds.), *História Económica de Portugal: 1700-2000*. Imprensa de Ciências Sociais, Lisboa.

Scott, A. S. V. (1999). *Famílias, formas de união e reprodução social no noroeste português (séculos XVIII e XIX)*. NEPS-Universidade do Minho.

Solsten, E. (1993). Portugal: A country study. Federal Research Division Library of Congress.

Appendix A Historical Background

The three Portuguese locations from which our data are obtained are São Tiago de Ronfe (hereafter Ronfe), Ruivães, and Horta. The villages of Ronfe and Ruivães are only nine kilometers apart and strategically located between the two historical administrative centers, Guimarães and Braga, in the Minho region in the northwest of Portugal. The coastal city of Horta is located in the Azores Archipelago and was a major stopping port on the journey to Brazil.

The legal and social background of Portuguese society during the sample period did not favor the economic independence of women. The most relevant changes in the legal system regarding women's rights occurred only after the proclamation of the First Republic in 1910 (Solsten, 1993). Women were also excluded from the main educational system. According to the population census of 1864, the share of boys aged between 6 and 15 attending primary educational institutions in Horta and Braga—the regions ("distritos") of our locations—were 13.4 percent and 18.0 percent, respectively. In contrast, the corresponding shares for girls were only 5.0 percent and 1.3 percent. It was not until 1888 that a law was passed to allow for the creation of all-girls public schools for secondary education. Women were, nevertheless, allowed to own and inherit property.

During our sample period, there were three succession systems. The first was a male primogeniture system referred to as "morgadio" through which the oldest son inherited the land and the title of the property owner. The "morgadio" applied only to the wealthiest families of landlords and aristocrats from the 13th century until it was abolished in 1863 (Moreira da Silva, 1983). The second norm, far more common than the "morgadio", applied to life-long rentals of aristocratic or ecclesiastic land. Life-long rentals had to be transmitted to a single heir and tended to favor spouses over children, male over female children, and older over younger children. In contrast to the "morgadio", daughters could inherit life-long rentals, as was frequently the case in the Minho region (Durães, 2009).

The third norm, that constituted a general rule for divisible property transmission, was to divide two-thirds of the property (the "legitima") equally among the legitimate heirs and to dispose of one-third (the "terço") to benefit one of the children or the surviving spouse. Scholars describing the local customs report that the "terço", which typically included the main house and the adjacent land, either became the property of the first marrying child or was given to a spouse or to unmarried children. Daughters might have been favored by the "terço" for several reasons. First, having land (or the promise of it) increased a woman's chances of marriage since they faced a thin marriage market due to the heavy male emigration to Brazil. Second, since married daughters tended to live with their parents for a period of time (at least until the couple had their own house and land and/or until the next daughter married), they were more welcome in the house than daughters-in-law. Third, it was also common for single daughters to inherit the "terço", which, on the one hand, would guarantee them the means of survival and, on the other hand, would also guarantee that the parents would be cared for in their old age (Brettell, 1991, Durães, 2009, de Pina Cabral, 1986, Matos, 2009).

A feature of the three locations is the predominantly male emigration to Brazil beginning in the 16th century. As a consequence, the Minho region and the Azores were atypical in Portugal in terms of the population's gender composition, with women substantially outnumbering men. According to the 1864 Census, the male-to-female ratio was 0.75 in the city of Horta (the third lowest among the 32 largest Portuguese cities), and 0.81 in Braga's district. Mesquita and Leite (2013) reports a male to female ratio for a parish in Angra do Heroismo (one of the major cities in the Azores) of 0.83 and 0.76 for 1725 and 1750, respectively. Scott (1999) reports a ratio of 0.64 for Ronfe in 1740. A similar pattern has also been documented for other locations in the Azores (Amorim and Santos, 2009).

Appendix B External Information on Female Participation Rates

In this appendix, we describe the external data used to obtain aggregate female participation rates that are included in the Likelihood Approach. The first Portuguese census was administered in 1864, and since then, censuses have been conducted more or less periodically every ten years. In most censuses, the smallest geographical area for which demographic data are collected is the borough ("concelho"), followed by the district ("distrito") and the province. Most censuses also publish information regarding economic activity and professions of men and women above a certain age at various levels of geographical aggregation (which unfortunately varies across censuses). Censuses collect data for all regions of Portugal.

For the purpose of obtaining FLFP rates, we assemble data from Portugal as a whole, the two largest cities (Lisbon and Porto), and the boroughs and districts to which the locations in our sample belong. Specifically, the regions are: 1) Portugal (including the Azores and Madeira and excluding colonial territories); 2) the district of Braga to which Ronfe and Ruivães belong; 3) the district of Horta to which the city of Horta belongs; 4) the cities of Lisbon and Porto; 5) the borough of Guimarães to which Ronfe belongs; 6) the borough of Vila Nova de Famalicão to which Ruivães belongs; and 7) the borough (city) of Horta. Carrilho (1996) provides a detailed guide for the different definitions of active female population across censuses. One of the main differences regards the reference population. For example, whereas until 1930 the reference population was considered the "present" population ("população *de facto*"), after 1940 the censuses considered the "resident" population. Moreover, although until 1930 all individuals were included in the reference population, after 1940 only those above a certain age (either age 10 or 12, depending on the census) were considered. The other main difference refers to the definition of active population. The earlier censuses (until circa 1950) defined as active population everyone with an occupation regardless of whether that occupation was a profession (for example, domestic or agricultural unpaid work was considered an occupation although not a profession). This definition implied very high and unrealistic participation rates for women, particularly in the districts and boroughs that were less urban. Since 1960, unemployed individuals seeking a job are included in the active population.

Table B1: Female Participation Rates by Place of Birth

	Country	City			District		Borough	
	Portugal	Lisbon	Porto	Horta	Braga	Horta	Guimarães	Vila Nova de Famalicão
1862	.19745							
1890	.40595	.2425	.3533	.12159	.54256	.11328	.58439	.65898
1900	.32397	.28975	.39791	.20099	.56568	.16286	.54608	.53761
1911	.30842	.2954	.42098	.24418	.43609	.21686	.56513	.38395
1925		.29703	.27013					
1930	.19702	.27457	.33186	.10055	.3546	.08802	.39256	.33923
1940	.21688	.27423	.35084		.34882	.07255		
1950	.22935	.30734	.36484	.08575	.31977	.06705	.44455	.38677
1960	.18386	.34668	.3972		.25592	.06145		
1970	.26467	.36846	.43142		.36383	.085		
1981	.43168	.51124	.52948	.26445	.50987	.19658	.63888	.57485
1991	.45258							

Note: Authors computations from Recenseamentos Gerais da Populaçao (INE, Lisbon) and Reis (2005). The 1925 census was a special census restricted to the cities of Lisbon and Porto. In 1862, the active female population is taken from the total number of working women (mão-de-obra feminina) in Reis (2005). Although the population target changed across censuses, we compute participation rates with the most similar population definitions. For all years, the data refers to women younger than 70. Until 1950, the rates include women older than 10. For 1950, the minimum age becomes 12. For 1960 and 1970, it raises to 15, while for 1981 and 1991 it decreases back to 14 and 12, respectively. In 1890, 1900, 1911 the active female population is defined as females (including the servants) minus the housewives (unpaid domestic work) and the unproductive (data taken from Table V in 1890 and 1900 and from Table 3 in 1911). In 1925, the actives are the females minus the female children between 0-9 years old, the housewives (unpaid domestic work), those without profession, and the female beggars. In 1930, the data comes from Table 1 and the actives are the sum of the three first columns under "população activa" (i.e. females who work for the administration, in the private sector, and the self-employed) minus the housewives (unpaid domestic work) who are considered in the column of active and self-employed. In 1940, the actives are the number of active females ("activas") minus the number of females in non-professional activities (information from Table 24). In 1950, the actives are the active females ("activas") with profession (information from Table 1). In 1960, the actives are the active females ("activas") with profession minus active females with profession less than 15 years of age (Table 1, Tomo 5, vol III). In 1970, the actives are the active females ("activas") with profession (Table 8). In 1981 and 1991, the actives are the active females ("activas") (Tables 6.13 and 6.16.1, respectively).

Table B1 reports our calculation of the FLFP rates for eight regions in Portugal over the 1862-1991 period based on the census data. Although we use mostly a single data source (the exception being 1862, for which we also use data published in Reis, 2005), and we compute participation rates with the most similar population definitions, the figures reported in Table B1 still exhibit jumps across censuses and substantial differences across regions for a given census, which suggests that the data are still not fully comparable. To correct for these shortcomings, we first fit the following function:

$$y_{it} = c + \delta_1 f(t) + \delta_2 f(t) \times D_{Horta} + \sum_{r=1}^{R} \gamma_r D_r \tag{B.1}$$

where $y_{it} = ln\left(\frac{p_{it}}{1-p_{it}}\right)$ and p_{it} correspond to the values of the female participation rate presented in Table B1 for region i and census year t. The function $f(t)$ is a trend polynomial of order 10, D_{Horta} is a dummy for Horta district or borough, D_r are regional dummies (specifically, a dummy variable for Braga district locations, encompassing Braga District, Guimarães and V.N. Famalicão, a dummy variable for Horta district locations, encompassing Horta District and the city of Horta, and a dummy variable for the borough of Guimarães).

We compute OLS estimates for Eq. B.1. To obtain smooth predicted participation rates for our three locations, we take the predicted values for y_{it} of the closest borough, transform them into participation rates p_{it}, and correct for the non-linearity of the transformation. For all years since 1890, we use these predictions (shown in Fig. B1) as our external data.

Figure B1: Predicted participation rates from the estimation of Eq. (B.1).
Notes: Census data refers to female participation rates obtained using data from Recenseamentos Gerais da População 1890-1991 for the closest locations (i.e. the borough of Horta for Horta; the borough of Guimarães for Ronfe; the borough of Vila Nova de Famalicão for Ruivães) and the national female participation rate in 1862 published in Reis (2005). Predictions are of labour force participation rates for the closest locations using a log functional specification and all data available.

Appendix C Probabilities of non-missing and participation status in the sample

In this appendix, we derive the probabilities of the different observable situations of the missing and participation processes. There are three situations in which a given observation may have missing information: (a) when the daughter's information is missing but the mother's is not, (b) when the mother's information is missing but the daughter's is not, and (c) when information for both the daughter and the mother is missing. Consider the first case. The joint probability for observation $\{I_i = 0, I_i^m = 1, y_i^m = w, x_i\}$ decomposes into two event probabilities that are, again, particular cases of Eq. (3.3) in Section

3.1:

$$\begin{aligned}\Pr\{I_i=0, I_i^m=1, y_i^m, x_i\} = \Pr\{I_i=0, I_i^m=1, y_i=1, y_i^m, x_i\} \\ + \Pr\{I_i=0, I_i^m=1, y_i=0, y_i^m, x_i\}.\end{aligned} \tag{C.1}$$

The treatment of the second case, that is, when the mother's information is missing but the daughter's is not, is similar to that of the first case:

$$\begin{aligned}\Pr\{I_i=1, I_i^m=0, y_i, x_i\} = \Pr\{I_i=1, I_i^m=0, y_i, y_i^m=0, x_i\} \\ + \Pr\{I_i=0, I_i^m=1, y_i, y_i^m=1, x_i\}.\end{aligned} \tag{C.2}$$

In the third case, that is, when information for both the mother and the daughter is missing, the joint probability for observation $\{I_i=0, I_i^m=0, x_i\}$ decomposes into four event probabilities:

$$\begin{aligned}\Pr\{I_i=0, I_i^m=0, x_i\} = \Pr\{I_i=0, I_i^m=0, y_i=1, y_i^m=1, x_i\} \\ + \Pr\{I_i=0, I_i^m=0, y_i=0, y_i^m=1, x_i\} + \Pr\{I_i=0, I_i^m=0, y_i=1, y_i^m=0, x_i\} \\ + \Pr\{I_i=0, I_i^m=0, y_i=0, y_i^m=0, x_i\}.\end{aligned} \tag{C.3}$$

Appendix D Imputation Approach

In this appendix, we derive the likelihood function for the traditional imputation procedure. From equations (C.1), (C.2), and (C.3) in Appendix C, we can rewrite p_i as

$$\begin{aligned}
p_i = \; & \left(\Pr\left\{I_i = I_i^m = 1, y_i, y_i^m, x_i\right\}\right)^{I_i I_i^m} \times \\
& \left(\sum_{v\in\{0,1\}} \Pr\left\{I_i = 0, I_i^m = 1, y_i = v, y_i^m, x_i\right\}\right)^{(1-I_i)I_i^m} \times \\
& \left(\sum_{w\in\{0,1\}} \Pr\left\{I_i = 1, I_i^m = 0, y_i, y_i^m = w, x_i\right\}\right)^{I_i(1-I_i^m)} \times \\
& \left(\sum_{v\in\{0,1\}}\sum_{w\in\{0,1\}} \Pr\left\{I_i = I_i^m = 0, y_i = v, y_i^m = w, x_i\right\}\right)^{(1-I_i)(1-I_i^m)} .
\end{aligned} \tag{D.1}$$

The traditional imputation procedure in which missing values are filled in implies that certain events are known to have zero probability. After imputing the missing observations, Eq. (D.1) simplifies to

$$\begin{aligned}
p_i = \; & \left(\Pr\left\{I_i = I_i^m = 1, y_i, y_i^m, x_i\right\}\right)^{I_i I_i^m} \times \\
& \left(\Pr\left\{I_i = 0, I_i^m = 1, y_i = \overline{v}, y_i^m, x_i\right\}\right)^{(1-I_i)I_i^m} \times \\
& \left(\Pr\left\{I_i = 1, I_i^m = 0, y_i, y_i^m = \overline{w}, x_i\right\}\right)^{I_i(1-I_i^m)} \times \\
& \left(\Pr\left\{I_i = I_i^m = 0, y_i = \overline{v}, y_i^m = \overline{w}, x_i\right\}\right)^{(1-I_i)(1-I_i^m)}
\end{aligned} \tag{D.2}$$

where $y_i = \overline{v}$ ($y_i^m = \overline{w}$) denotes that the only admissible value $\overline{v}$ ($\overline{w}$) is imputed in the observation, and y_i and y_i^m denote observed values. By Eq. (3.3), we can write p_i in Eq. (D.2) as:

$$\begin{aligned}
p_i = \; & \left(\Pr\left\{I_i = 1, I_i^m = 1 \left| y_i, y_i^m, x_i\right.\right\} \times F\left(y_i, y_i^m, x_i; \theta\right) \times \Pr\left\{y_i^m, x_i\right\}\right)^{I_i I_i^m} \times \\
& \left(\Pr\left\{I_i = 0, I_i^m = 1 \left| y_i = \overline{v}, y_i^m, x_i\right.\right\} \times F\left(\overline{v}, y_i^m, x_i; \theta\right) \times \Pr\left\{y_i^m, x_i\right\}\right)^{(1-I_i)I_i^m} \times \\
& \left(\Pr\left\{I_i = 1, I_i^m = 0 \left| y_i, y_i^m = \overline{w}, x_i\right.\right\} \times F\left(y_i, \overline{w}, x_i; \theta\right) \times \Pr\left\{y_i^m = \overline{w}, x_i\right\}\right)^{I_i(1-I_i^m)} \times \\
& \left(\Pr\left\{I_i = 0, I_i^m = 0 \left| y_i = \overline{v}, y_i^m = \overline{w}, x_i\right.\right\} \times F\left(\overline{v}, \overline{w}, x_i; \theta\right) \times \Pr\left\{y_i^m = \overline{w}, x_i\right\}\right)^{(1-I_i)(1-I_i^m)}
\end{aligned} \tag{D.3}$$

such that, rearranging those factors, the likelihood is equal to

$$\begin{aligned}
\prod_i p_i = & \prod_i \Big\{ F\left(y_i, y_i^m, x_i; \theta\right)^{I_i I_i^m} F\left(\overline{v}, y_i^m, x_i; \theta\right)^{(1-I_i) I_i^m} \times \\
& \quad F\left(y_i, \overline{w}, x_i; \theta\right)^{I_i (1-I_i^m)} F\left(\overline{v}, \overline{w}, x_i; \theta\right)^{(1-I_i)(1-I_i^m)} \Big\} \times \\
& \prod_i \Big\{ \left(\Pr\left\{I_i = 1, I_i^m = 1 \left| y_i, y_i^m, x_i \right.\right\} \Pr\left\{y_i^m, x_i\right\}\right)^{I_i I_i^m} \times \\
& \quad \left(\Pr\left\{I_i = 0, I_i^m = 1 \left| y_i = \overline{v}, y_i^m, x_i \right.\right\} \Pr\left\{y_i^m, x_i\right\}\right)^{(1-I_i) I_i^m} \times \\
& \quad \left(\Pr\left\{I_i = 1, I_i^m = 0 \left| y_i, y_i^m = \overline{w}, x_i \right.\right\} \Pr\left\{y_i^m = \overline{w}, x_i\right\}\right)^{I_i (1-I_i^m)} \times \\
& \quad \left(\Pr\left\{I_i = 0, I_i^m = 0 \left| y_i = \overline{v}, y_i^m = \overline{w}, x_i \right.\right\} \Pr\left\{y_i^m = \overline{w}, x_i\right\}\right)^{(1-I_i)(1-I_i^m)} \Big\} .
\end{aligned} \tag{D.4}$$

In Section 4, we present estimations under the assumption that, independent of everything else, daughters (mothers) with missing participation do not participate in the labour force, i.e. $\Pr\left\{y_i = 1 | I_i = 0, I_i^m, y_i^m, x_i\right\} = 0$, $\left(\Pr\left\{y_i^m = 1 | I_i^m = 0, I_i, y_i, x_i\right\} = 0\right)$.[18] Then, $\overline{v} = \overline{w} = 0$ and the likelihood is:

$$\begin{aligned}
\prod_i p_i = & \prod_i \Big\{ F\left(y_i, y_i^m, x_i; \theta\right)^{I_i I_i^m} F\left(0, y_i^m, x_i; \theta\right)^{(1-I_i) I_i^m} \times \\
& \quad F\left(y_i, 0, x_i; \theta\right)^{I_i (1-I_i^m)} F\left(0, 0, x_i; \theta\right)^{(1-I_i)(1-I_i^m)} \Big\} \times \\
& \prod_i \Big\{ \left(\Pr\left\{I_i = 1, I_i^m = 1 \left| y_i, y_i^m, x_i \right.\right\} \Pr\left\{y_i^m, x_i\right\}\right)^{I_i I_i^m} \times \\
& \quad \left(\Pr\left\{I_i = 0, I_i^m = 1 \left| y_i = 0, y_i^m, x_i \right.\right\} \Pr\left\{y_i^m, x_i\right\}\right)^{(1-I_i) I_i^m} \times \\
& \quad \left(\Pr\left\{I_i = 1, I_i^m = 0 \left| y_i, y_i^m = 0, x_i \right.\right\} \Pr\left\{y_i^m = 0, x_i\right\}\right)^{I_i (1-I_i^m)} \times \\
& \quad \left(\Pr\left\{I_i = 0, I_i^m = 0 \left| y_i = 0, y_i^m = 0, x_i \right.\right\} \Pr\left\{y_i^m = 0, x_i\right\}\right)^{(1-I_i)(1-I_i^m)} \Big\} .
\end{aligned} \tag{D.5}$$

Under the assumption that the imputation is correct, the ML estimator of the participation model (3.2) is consistently estimated by maximizing only the first two lines of Eq. (D.5). If the remaining terms in Eq. (D.5) depend on the parameters of the participation model (3.2), this Maximum Likelihood estimator will still be consistent but may not be efficient. Otherwise, it will be identical to the full Maximum Likelihood estimator.

In the original data, the participation rate of those women whose mothers also work is

[18]These assumptions set as impossible those events in which either the mother or the daughter (or both) participate in the labour market and for which information is missing, that is, $\{I_i = 0, I_i^m = 1, y_i = 1, y_i^m, x_i\}$, $\{I_i = 1, I_i^m = 0, y_i, y_i^m = 1, x_i\}$, $\{I_i = 0, I_i^m = 0, y_i = 1, y_i^m = 1, x_i\}$, $\{I_i = 0, I_i^m = 0, y_i = 1, y_i^m = 0, x_i\}$, and $\{I_i = 0, I_i^m = 0, \ \ y_i = 0, y_i^m = 1, x_i\}$.

very high at 69.72 percent (see Table 3). The figure declines to 7.97 percent for the sample with imputed values for all missing records. Moreover, in the original data, the proportion of participating women increases from 7.3 percent to 69.7 percent (or 62.4 pp) when the mother also participates. That increase is substantially lower after the imputation: 2.86 pp. By reducing the differential in the transition of participation rates from mother to daughters between mothers who participated and mothers who did not, the imputation might bias downwards the effect of mothers' labour market participation on that of their daughters.

Appendix E The Monte Carlo Algorithm

In order to evaluate the impact of the high proportion of missing values in our estimates, we perform the following Monte Carlo experiment based on our data and model.

STEP 1. Initialization. Take our working sample used under the Likelihood Approach in Table 7. We use the information on location and time dummies together with *siblings*, *father SES*, and *father owner* for all observations.

STEP 2. Simulation (r). We sequentially simulate, for every observation i in our working sample:

1. The mothers' participation status $y^m_{i,r}$ using the conditional participation probabilities $\left\{\widehat{\Pi}_{w,x}\right\}$, $w = \{0, 1\}$.
2. The daughter's participation status $y_{i,r}$ using the participation model ML estimates $\widehat{\theta} \equiv \left\{\widehat{\alpha}, \widehat{\beta}\right\}$ reported in the baseline specification of Table 7.
3. The missing information I_i and I^m_r. We consider two cases:
   (a) The proportion of missing values is equal to the actual proportion of missing values in the working sample. We use the missing probabilities estimated in our baseline model reported in the bottom half of Table 7 evaluated at the simulated values, i.e. we use $\widehat{\Pr}\left\{I^m = 1 \left| y^m = y^m_{i,r}\right.\right\}$ to generate $I^m_{i,r}$ and $\widehat{\Pr}\left\{I = 1 \left| y = y_{i,r}, I^m = I^m_{i,r}\right.\right\}$ to generate $I_{i,r}$.
   (b) The proportion of missing values is $q = \{20, 50\}$ percent. We must re-scale the probabilities $\widehat{\Pr}\left\{I^m = 1 \left| y^m = y^m_{i,r}\right.\right\}$ and $\widehat{\Pr}\left\{I = 1 \left| y = y_{i,r}, I^m = I^m_{i,r}\right.\right\}$ to reach the target missing proportion. Specifically, let $\widetilde{\Pr}\left\{y^m = 1\right\}$ and $\widetilde{\Pr}\left\{y^m = 0\right\}$ be the unconditional proportions of y^m being equal to one and equal to zero in the simulated data, respectively. The simulated unconditional probability of $I^m = 1$ is $\widetilde{\Pr}\left\{I^m = 1\right\} = \widehat{\Pr}\left\{I^m = 1 \left| y^m = 1\right.\right\} \times \widetilde{\Pr}\left\{y^m = 1\right\} + \widehat{\Pr}\left\{I^m = 1 \left| y^m = 0\right.\right\} \times \widetilde{\Pr}\left\{y^m = 0\right\}$. Hence, we re-scale $\widehat{\Pr}\left\{I^m = 1 \left| y^m = y^m_{i,r}\right.\right\}$ by multiplying it by the factor $\left(\frac{1-\frac{q}{100}}{\widetilde{\Pr}\{I^m=1\}}\right)$. Simi-

larly, let

$$\widetilde{\Pr}\{I=1\} = \sum_{v,w\in\{0,1\}} \widehat{\Pr}\{I=1\,|y=v, I^m=w\} \times \widetilde{\Pr}\{y=v, I^m=w\},$$

and $\widetilde{\Pr}\{y=v, I^m=w\}$ be the simulated proportions of $\{y=v, I^m=w\}$. We re-scale $\widehat{\Pr}\left\{I=1\,\middle|y=y_{i,r}, I^m=I^m_{i,r}\right\}$ by multiplying it by the factor $\left(\frac{1-\frac{q}{100}}{\widetilde{\Pr}\{I=1\}}\right)$.

STEP 3. Estimation (r). With the simulated data we estimate the conditional participation model both ignoring the missing observations as in column Ignorability in Table 7 and under the Likelihood Approach in the baseline scenario as in the corresponding column in the said table.

STEP 4. Repeat STEP 2 and STEP 3 R times.