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

Jesús M. Carro[†] and Matilde P. Machado[‡] and Ricardo Mora[‡]

[†] *Universidad Carlos III Madrid, C. Madrid 126, Getafe, Spain*
E-mail: jcarro@eco.uc3m.es, matilde.machado@uc3m.es, ricmora@eco.uc3m.es

Abstract Using Portuguese parish data from 1675-1925, we estimate the effect of a mother's labour force participation (FLFP) on that of her daughter. We adapt a methodology to avoid bias coming from potential non-random missing data which affects our historical data. We find that ignoring the missingness process results in substantial downward biased estimates of the effect, even for a proportion of missing values as low as 20 percent. In contrast, our methodology gives unbiased estimates regardless of the proportion of missing values. We document the existence of a large effect of a mother's FLFP on that of her daughter long before technological and cultural changes took place.

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

1. INTRODUCTION

A lot has been written about the rise of women's labour force participation during the second half of the 20th century. In particular, recent research highlights the potential multiplier effect of intergeneration transmission mechanisms on the dynamics of female labour force participation (FLFP) after 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 mother-daughter transmission mechanism of FLFP. Our finding of this early and large intergeneration link, which acts as a potent catalyst of shocks, contributes to the understanding of the long-term dynamics of FLFP.

A difficulty in studying the mother-daughter transmission of FLFP is to disentangle it from the transmission originated from other factors. We model a woman's decision to participate in the labor market as a function of her mother's participation and include controls for social status of the father, for the ownership of property, and a proxy for household income to avoid confoundedness with other sources of inertia within the family. Other factors, external to the family, may generate mother-daughter FLFP correlation.

[1] Key references are: 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).

Leading examples are the introduction or enforcement of mandatory education and the expansion of the service sector during the 20th century. We avoid the influence of these factors by investigating a place and time with close to no technological and cultural changes (Lains, 2006).[2] Moreover, we include time and location dummy variables aimed at capturing time and location specific factors such as labor market conditions and social multiplier effects (Maurin and Moschion, 2009, Olivetti et al., 2018).

Like most historical data, ours are characterized by abundant non-ignorable missing information, which affects both mothers' and daughters' participation status. In spite of this, the number of mother-daughter pairs without missing values ($n = 696$) is not small. An enticing approach is to estimate the model only with the observations for which we have complete information (hereafter we label this approach as Ignorability). But when the missing process is not random, Ignorability suffers from sample selection bias. We consider two approaches to address this non-ignorability problem in our estimation. First, informed by historical records and narratives, we conservatively impute missing values with the perceived predominant female labor market status at the time, i.e., non-participation, which sets the participation rate at an unrealistically low value relative to census data. We label this approach as 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 using all observations under the assumption that missing values are generated by the same missing process and allowing the missing values to be endogenous. We improve the identification of the parameters of the model by incorporating external information regarding aggregate female labor market participation from historical census data. We label this approach as the Likelihood Approach. We show that failure to recognize the endogeneity of the missing process reduces the estimates of the impact of mothers under Ignorability and Imputation by more than a half.

Using the Likelihood Approach we estimate a positive and statistically significant marginal effect of 26.8 pp (s.e. 0.033) of the mother's participation status on the daughter's probability of participation. A way to assess the importance of our estimated marginal effect for the evolution of FLFP is to simulate its long-run impact in the face of an exogenous shock. The average probability of participation in our sample is 28.8 percent but for women whose mother did not work is only 19.0 percent. Now imagine, there is a permanent shock that increases the latter by 50 percent to 28.5 percent. This level would be the expected long-run participation rate in the absence of a mother-to-daughter link. On the contrary, when mother-to-daughter transmission occurs, the long-run participation rate increases to 52.0 percent as an increasing proportion of descendants influence the decisions of their own daughters.

Our estimate of the transmission of labor market behavior from mothers to daughters that we find comprises a number of mother-to-daughter intergeneration 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 labor market from a mother to her daughter. There is evidence of such transmission within

[2]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 10-14 knew how to read. It was not until the 1960s that literacy of all children was achieved Gomes and Machado (2020).

the family during the 20th century (e.g. Fernández et al., 2004, Farré and Vella, 2013), and hence it is possible that it was also present previously. Our Likelihood Approach estimate of the impact of mothers is substantially higher than that found by Morrill and Morrill (2013) for the late 20th century—around 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, e.g access to formal education, decreases with economic and cultural development. Indeed, when we split our sample into different periods, we have evidence that the impact of mothers declines with time.

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

Finally, we evaluate using simulation methods the sensitivity of Ignorability and the Likelihood Approach to varying degrees in the incidence of missing information. We use the estimated parameters from the participation model and the missing process to simulate the data and apply different percentages of missing observations. We then re-estimate the model in these different samples. We find that ignoring the missing observations (Ignorability), results in a large downward bias of the mother's effect, 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 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 data set and the estimation methods, 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 concludes. 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, northeast of Portugal, and the city of Horta at Faial Island located in the Azores. A research team from University of Minho collected the main data sets using 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*

[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 http://www.ghp.ics.uminho.pt/genealogias.html

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

de confessados (literally, "the list of the confessed"). The latter originate from a parochial census of residents older than seven years of age. They were produced by the vicar during Lent to administer the sacrament of penance to the parishioners and contain occupation and/or social status information.

The original baptism, marriage, and death certificate records allowed family linkages to be reconstructed within each location beginning in the 1550s through the 20th century (Amorim, 1991). Altogether, after some basic cleaning, the data set includes entries for 34,897 women (50.24 percent of all records). It has information on birth, marriage, and death dates 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) as well as to compute the number of siblings for each woman.

We have the year of birth for 59.3 percent of all records. By contrast, we have death information for only 33.3 percent of all records. Our observations include individuals who do not survive childhood and individuals who migrate to other locations and for whom no further information is available. In theory, the former group of individuals would be identified using the death date information, whereas the latter group could be indirectly inferred by the absence of information on their deaths. However, the profusion of missing death dates hinders the precise identification of both early death and migration.

We focus on exploiting birth-date information and attempt to complete records for which the date of birth is missing. To complete such records, we group all individuals for which the information is available into 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. When the cohort of the siblings or the spouse is identified, 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, 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. After all basic cleaning we are left with 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, marriage, and death information. Columns 1, 3, and 5 of Table 1 report the number of observations by location and quarter-century. Columns 2, 4, and 6 show the proportion (in percentage) of observations with occupation/social status information by period and location.

Occupations/social status coverage varies by period and across locations. For Ruivães from 1700 until 1800 and Horta in 1900, there is no occupations/social status information in the matched records. Hence, we do not use these location/period combinations in our analysis. We also discard the observations for Horta and Ruivães in 1675 and Horta in 1875 because 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 coverage. For example, whereas coverage in Horta in 1725 is 5.64 percent, it increases monotonically to 25.84 percent one century later. Coverage in Ronfe follows a U-shaped trend decreasing steadily from

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

	Horta		Ronfe		Ruivães	
	N.obs.	% non-missing	N.obs.	% non-missing	N.obs.	% non-missing
Year	(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: Sample includes women born between 1675-1925. 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.

15.72 percent in 1675 to 3.41 percent in 1775 to rise 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 is never larger than 36 percent. We can discard three reasons for these low figures. First, accounting for early deaths can at most only explain a small proportion of non-coverage. Second, gender bias in vicars recording practices does not seem to be important because the gender differential in coverage is only 2.2 percentage points and follows a similar location/period pattern. Third, the vicar could report occupation or social status to differentiate women with a common name. If there were an excessive number of, say, "Marias", 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 they are as likely among the reported as among the 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.

So, why would the vicar report some women's occupation/social status and not that of others? We must consider at least four potential sources of selection bias. The first one arises from the Lent census data collection practices. The censuses were organized by street and gathered by vicars door-to-door. Whenever vicars gave priority to nearby households, remote rural locations, where farmers and poor people would tend to live, would be less likely covered. The second one stems from the activity itself. Vicars 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 one originates from social status: parochial vicars are arguably more likely to register those parishioners who give large donations to the church. The fourth 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 identification code is unknown. By construction, in all these cases the participation status of the mother is missing. There are several reasons for this identification failure. When the mother's record precedes the first currently available registries, it is illegible, or it contains coding errors, 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, 35 percent of the observations from 1675 have an unidentified mother compared with 33 percent one hundred years later and only 30 percent two hundred years later. These registry issues are not systematically related to participation decisions or to the vicar'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 one 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 18 percent of the children baptised in 1700 were illegitimate. Illegitimate and orphans are more likely to be poor, landless, and to work for pay. The second one is the absence of the mother in the records: mothers who are 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 the mother remains in her original place. 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, -0.202 with p-value < 0.01). Hence, this source is likely non-ignorable.

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

2.3. Labour force participation

We construct women's occupation/social status using information from two variables from the original files: "profissão" and "título". Whereas the former reports professions, the latter reports social status—for example, a nobility title such as *countess* or an ecclesiastic position such as *abbess*. The available information on occupation/social status is not systematically classified across parochial vicars 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*.[5] 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

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

as housewives—and women to whom the vicar listed as *dona*, a term originally used to signal upper class and gradually adopted to also signal the bourgeoisie during the 18th and 19th centuries. The unproductive category includes the indigent or those classified as "very poor", individuals registered as nobility by the vicar, 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 vicar's registry of owners—from the Portuguese *proprietário*; 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 do not have any 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.)

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". Whereas the former reports professions, the latter reports social status—for example, a nobility title such as *countess* or an ecclesiastic position such as *abbess*. 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 compound of location-specific economic factors and differentials in the incidence of missing information. In some of the locations, certain categories appear under-represented (e.g., employee/farmer in Horta), whereas others seem over-represented (e.g., 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 by 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 effect of mother's labour force participation on the daughter's probability of participation. Table 3 shows participation rates by mother participation statuses in the subsample of mother-daughter pairs for whom the participation information is available. The raw estimate for the marginal effect ($69.72 - 7.33 = 62.39$ percentage points) is likely biased because of non-random missingness.

Table 3: Women Participation Frequencies by Mother's Participation

Subsample of mother-daughter pairs with observable participation status

	Mother		
Daughter	Does not participate	Participates	Total
Does not participate	544 (92.67)	33 (30.28)	577 (82.90)
Participates	43 (7.33)	76 (69.72)	119 (17.10)
Total	587 (100.00)	109 (100.00)	696 (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 female force participation rates as external sources of information to help identify the model. These census data, however, have at least three limitations. First, the first census in Portugal with information on labour market participation was taken in 1890. Second, for Ronfe and Ruivães we can only use information on the administrative regions they belong to, Guimarães and Vila Nova de Famalicão. Third, the definition of labour participation used in historical census data in Western countries has changed. The main difference is found in the concepts of "occupation" and "profession". The former concept, adopted early, classifies most women as "active" in the labour market, while the latter concept, adopted later, does not. The difference applies most notably to the case of women whose (at times irregular) work was developed in the household or the family farm/business, who would only be included in the labour force in early censuses (Goldin 1995). Thus, we should regard the rates such as the one obtained from the 1890 Census as approximations—relative to modern definitions. Reis (2005) provides a national-level estimate of the female labour market participation in 1864 that is substantially lower (19.12 percent) than the rate obtained for 1890 using census data (38.47 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 female labour force participation rates in our three locations, we construct predictions for labour force participation rates using a log functional specification with the 1864 data from Reis (2005) and data from all censuses up to 1991 (see Appendix B). For all years since 1890, we use these predictions as our external data. For the period before 1890, we adopt a conservative approach and consider for each location three scenarios—shown in Figure 1. In the "baseline scenario", we take the values of the predictions for our three locations in 1890 as the external female participation rates. In our "low scenario", we take the smallest local participation rate for which we have information in all censuses from 1890 to 1950. Similarly, in the "high scenario", we take the largest local participation rate for which we have information in all censuses from 1890 to 1950. More specifically, for the case of Horta we use the smallest and largest prediction for the borough of Horta, while for Ronfe and Ruivães, we take the largest prediction in the region (which occurs in the borough of Vila Nova de Famalicão).

Reassuringly, our locations' rank in terms of reported participation rates (see Table 3) follows those obtained from the census i.e. Ronfe, Ruivães and Horta. Both the high participation rates and the fact that women in rural settings participated more in the labour market than urban women are consistent with many historical accounts for the 17th, 18th and 19th centuries (Humphries and Sarasúa, 2012).

3. THE ECONOMETRIC MODEL

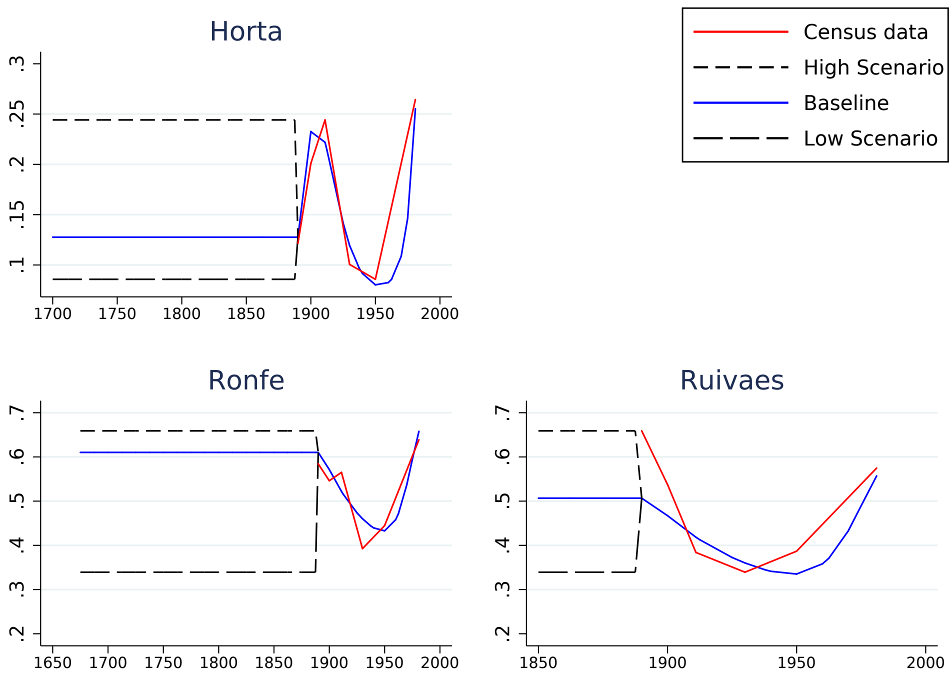
3.1. The participation model

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, $y_i = 0$. We assume that 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. Permanent and quasi-permanent factors are sources of intergenerational persistence in participation decisions. Our aim is to assess persistence arising exclusively from the mother-daughter link and other factors that induce persistence need to be controlled for. These include not only labour market conditions, sex ratios differentials, and social multiplier effects captured by location and time dummies, but also family characteristics. Within family characteristics, we consider family size–that we proxy by an indicator of more than four siblings, *siblings*—, and the father socio-economic status. Presumably, a woman who belongs to a large family is, through a negative income effect, more likely to participate in the labour market. We control for the father socio economic status with the use of a dummy of whether the vicar records that the father is a rentier, an owner, a merchant, a high-ranking civil servant, or an officer, *father SES*, and a dummy for whether the vicar reports the father to be an owner, *father owner*. The father's socio-economic status, as reported by the vicar, is important because it captures at least two sources of persistence within the family: the transmission of wealth (that likely leads to lower participation rates) and that of physical capital (that should lead to higher

Figure 1: Three Scenarios of Female Labour Force Participation



Note: Census data refers to female participations rates obtained using data from Recenseamentos Gerais da População 1864-1991 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). After 1890, "Baseline" refers to predictions of labour force participation rates for the closest locations using a log functional specification and all data available (see Appendix B). Before 1890, "Baseline" is the value of the prediction in 1890. "High Scenario" and "Low Scenario" coincide with "Baseline" after 1890. Before 1890, "Low Scenario" refers to the lowest and "High Scenario" to the largest local participation rate for which we have information in all censuses from 1890 to 1950. More specifically, for the case of Horta we use the lowest (largest) prediction for the borough of Horta, while for Ronfe and Ruivães, we take the lowest (largest) prediction in the region.

rates according to our participation measure). We include *father owner* with the aim to differentiate the latter effect of property transmission.

The transmission of wealth could also be reflected in migration flows. On the one hand, migrations of young women related to labour force participation decisions were likely associated to large families and low socio-economic status. On the other hand, marriages of individuals born in nearby parishes were common and marriages among wealthy families may have involved migration flows and could be related to 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. Hence, in Section 4 we discuss results controlling for migration status for the sample of Horta.

Throughout the sample period, any proxy for wealth is also a proxy for educational achievement because the Portuguese educational system remained extraordinarily elitist, with female illiteracy rates over 68 percent still 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.[6]

To summarize, in our main specification vector x_i includes location and time dummies as well as $siblings$, $father\, SES$, and $father\, owner$.

3.2. 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 vicar 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.

Missing completely at random arises when the probability of a missing observation is independent of y_i, in which case the missing observations are ignorable in the sense that their omission from estimation does not bias the results. As argued in Section 2, priests more likely recorded participation of those whose labour or social status was uncommon in the region and period. Thus, missing participation is likely related to individual characteristics and, in particular, to individual participation status. In this situation, the missing mechanism is non-ignorable.

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

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

for $v, w \in \{0, 1\}$. Assuming normality we have the conditional probit model:

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

The missingness mechanism and the participation process jointly define the event 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, i.e., $I_i = I_i^m = 1$, 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 equation (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:

[6] 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).

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

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

3.2.1. 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 (3.3) to

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

The probability of a non-missing observation is

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

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

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

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.

3.3. The Likelihood Approach to address 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] However, 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, as a special case, the situation in which the missingness mechanism is conditionally dependent on the outcome variable and a discrete partition of the covariates. Our empirical application calls for further adjustments to this strategy.

Vicars 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 was a differentiating characteristic. Hence, the chance of coverage is likely conditioned by the type of occupation of the woman. However, due to the high non-coverage rate, we don't have enough information to control for detailed occupations. We solve this problem by exploiting the information contained in the joint missing process for

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

mothers and daughters. As our data shows, daughters and mothers coverages correlate. Daughters of women whose occupation/social status is not reported by the vicar have a 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.

From equation (3.3) and without loss of generality,

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

The previous considerations warrant the following:

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

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

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. Because the mother's information is likely collected early on and through a similar process, an assumption closely related to DRCI 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 by Assumption 3.1, equation (3.6) warrants a simpler assumption for the mother:

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

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

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 vicars 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 coverages may be due to common unobservable factors affecting the vicars' decisions.[8] An important implication of Assumptions 3.1 and 3.2 is that covariates such as *father owner* affect the probability that the vicar 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 vicars' reporting decisions that the parameter of interest α is identified.

[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 we assume that the missing status of the daughter is independent from that of the mother, then we are not able to replicate the patterns in the data.

From now on, we are going to present the model assuming that control vector x_i is a vector of discrete variables, which is the case in our data.[9]

Assumptions 3.1-3.2 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 the discrete variables y_i, y_i^m are denoted $\Pr\{y_i = v\} = \Pi_v$ 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 that are observed in the data. Assumptions 3.1-3.2 imply that:

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

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 equation (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:

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

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

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

Thus, equation (3.4) simply becomes:

[9] 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.

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

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 (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:

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

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

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

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

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

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

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

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

Maximum Likelihood estimation will yield 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 in the number of different value combinations of variables y^m_i and x_i that are observed in the data. In our application, we may reduce the number of parameters by introducing additional restrictions. In particular, if we decompose vector x into location and time dummies, x_1, and *siblings*, *father SES*, and *father owner*, x_2, we have without loss of generality:

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

For our specific case, 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 $\Pi_v = \Pi_v^m$ for $v = 0, 1$. Although the model is identified, a very large proportion of missing observations likely affects the concavity of the likelihood function, impairing 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). In addition, we use this external information and estimates for $\Pi_{x_2|x_1}$ and Π_{x_1} to construct $\Pi_1^m = \Pi_1$ (i.e., the unconditional probability of participation) which, in conjunction with equation (3.18), simplifies the likelihood function and allows the identification of the constant β_0 in Equation (3.2).

As discussed in Section 2.4, we consider three alternative scenarios for female participation rates: baseline, low and high (see also Figure 1). We refer to this approach in addressing non-ignorable missing observations as the Likelihood Approach.

4. MAIN RESULTS

In Table 4 we present results under three major strategies to estimate equation (3.1). First, we assume Ignorability and present, in the first column, estimates using the sample obtained after dropping 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 not be identified when the mother's record precedes the first currently available registries, when there are coding errors, when the daughter is an illegitimate child, or when she or her mother change residency. Because we cannot rule out some sample selection 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 as non-participation all missing values. 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 three alternative scenarios 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 to the mother's participation status—parameter α in equation (3.1)—and of the corresponding average marginal effect (AME), i.e. the average change in the probability of participation when the mother participation status changes from no-participation to participation. 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 11.5 percentage points. This drop occurs because in the raw estimate we do not control for additional factors and in the probit estimate approximately 18 percent of the observed 696 cases are discarded because they predict the outcomes with probability one. Similarly, for Imputation A and

Table 4: Main Results

Estimates, AMEs, and Ignorability Tests

	Ignorability	Imputation		Likelihood Approach		
		A	B	baseline	high	low
y^M	0.743***	0.303***	-0.082	1.224***	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)	(0.075)
father SES	-0.118	0.119	0.154	-0.325**	-0.167	-0.381***
	(0.220)	(0.125)	(0.122)	(0.147)	(0.147)	(0.145)
father owner	0.323	-0.054	-0.113	0.387***	-0.031	0.652***
	(0.341)	(0.173)	(0.165)	(0.109)	(0.093)	(0.115)
AME	0.115***	0.019***	-0.008	0.268***	0.256***	0.263***
	(0.035)	(0.005)	(0.006)	(0.033)	(0.035)	(0.029)
$\widehat{\Pi}_1 = \widehat{\Pr}\{y=1\}$	.	.	.	0.288	0.369	0.213
$\widehat{\Pr}\{I^m=1\,\lvert y^m=0\}$	.	.	.	0.098	0.110	0.088
$\widehat{\Pr}\{I^m=1\,\lvert y^m=1\}$	.	.	.	0.181	0.141	0.248
$\widehat{\Pr}\{I=1\,\lvert I^m=0,y=0\}$	.	.	.	0.103	0.105	0.101
$\widehat{\Pr}\{I=1\,\lvert I^m=0,y=1\}$	.	.	.	0.136	0.132	0.142
$\widehat{\Pr}\{I=1\,\lvert I^m=1,y=0\}$	.	.	.	0.703	0.689	0.713
$\widehat{\Pr}\{I=1\,\lvert I^m=1,y=1\}$	.	.	.	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. See Section 3.1 for definitions of remaining controls. All models include quarter-century dummies 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 3.3. The three scenarios only differ in participation rates for periods before 1890. In the baseline scenario external participation rates before 1890 take the value of the 1890 predictions using census data. In the low (high) scenario, we use as external information for the period before 1890 the smallest (largest) local participation rate for which we have information in all censuses from 1890 to 1950 (see Figure 1). "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 $\Pr\{I_i=1\,\lvert I_i^m, y_i\}$ and $\Pr\{I_i^m=1\,\lvert y_i^m\}$ do not vary with y, y^m, and I^m. $\bar{y}$ is the sample average of the dependent variable.

B estimates, the AME drops from 5.2 to 1.9 and from 2.9 to -0.8 percentage points, respectively. Following the patterns found in the raw data (Table 3 and Appendix D), AMEs under Ignorability are larger than under Imputation.

The size of the AME using the Likelihood Approach is almost invariant by scenario and is substantially larger than under Ignorability or Imputation.[10] According to the results in the baseline scenario, a woman whose mother participates in the market has a probability of participation which is 26.8 pp larger than a woman whose mother does not participate.[11] This value is relatively large as it represents 71.1 percent of the sample average participation and implies an increase of 127.5 percent in the probability of participation for those women whose mothers do not participate.[12]

Regarding the estimates of the other dummy variables coefficients, 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 $\widehat{\alpha}$ and their signs (in all but one non-significant case in the high scenario) are as expected: a large family, 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.

Naturally, sample averages of the dependent variable differ across Ignorability and Imputation. They also differ between Ignorability and the Likelihood Approach because the sample used to compute $\bar{y}$ 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). In the baseline scenario, $\widehat{\Pi}_1$ is around 9 percentage points smaller than $\bar{y}$. In the high and low scenarios the disparities are 1 and 16 percentage points, respectively. As anticipated, priests seem to selectively report participants.

4.1. Ignorability tests and Imputation Assumptions

To formally test Ignorability in model (3.13), we must test that $\Pr\left\{I_i = 1\,|I_i^m, y_i\right\}$ and $\Pr\left\{I_i^m = 1\,|y_i^m\right\}$ 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 each of the three scenarios.[13] In all scenarios the differences in conditional probabilities are sizeable, congruent with the hypothesis that priests selectively report participants. For example, in the baseline scenario the probability that the mother's participation status is observed almost doubles

[10] This result is observed even if Ignorability is only imposed partially and then the Likelihood Approach is conducted. For example, if the model is estimated for a sample with only women whose mothers are identified—ignoring over 30 percent of the sample—our AME estimate using the Likelihood Approach is reduced to 0.094 (with standard error 0.03). This highlights that ignoring missing values leads to endogenous sample selection.

[11] If we exclude property ownership from the definition of participation in the labour market, property becomes a source of intergenerational transmission of non-participation status. Hence, in the baseline scenario the role of the mother is actually reinforced as the AME increases to 45.9 percent with a 95 percent confidence interval of $(33.5, 58.2)$ while the sign of the coefficient associated with *father owner* becomes negative and strongly significant (from 0.3887, s.e. 0.108, to -1.213, , s.e. 0.312). Results are available upon request.

[12] 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 $\frac{\text{AME}}{(1-\text{AME})\times\Pi_1} = \frac{0.268}{(1-0.268)\times 0.288} \approx 1.271$.

[13] These probabilities are non-linear functions of the parameters of the model. For example, $\Pr\left\{I_i^m = 1\,|y_i^m = 0\right\} = \frac{H_{10}^m}{\Pi_0^m}$ 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.

when she participates. Similarly, the probability that the daughter's participation status is observed increases from 10.3 percent to 13.6 percent when she participates if the mother 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 8.3 percent whereas it reaches 70.3 percent when she is not. 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 a *Dona*, i.e. that she does not participate. These large differences in conditional probabilities lead to strong rejection of the Ignorability test for all scenarios.

In addition, the ML estimates discard the plausibility of the implicit assumption under Imputation, i.e. that missing participation implies non-participation in the labour market. For example, given that

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

the ML estimate for the probability that mothers participate if their information is missing is $(1 - 0.181) \times \frac{0.288}{0.8787} \times 100 \approx 26.84$ percent, a value which is well over zero.

4.2. The stability of the role of mothers

We divide our working sample into three sub-periods to assess the extent to which the mother's effect is stable across time. Each period corresponds roughly to one third of the working sample. The first period ranges from 1675 to 1750, the second period starts in 1775 and ends in 1825, and the remaining quarters constitute the last period (i.e. 1850 : 1925).[14]

We have sample identification problems in the first period because all reported mother-daughter pairs are non-participants. This makes it impossible to estimate the model under Ignorability and results in non-convergence and instability for the Likelihood Approach. Hence we disregard the first period and in the remaining part of this section we focus on the second and third periods. We present results under all scenarios for the second period and, for brevity, only the baseline scenario for the third period.[15]

Table 5 shows results using different sample periods. All AMEs are large and statistically significantly different from zero. Results using the sample without the first period, i.e. years 1775 : 1925, are very similar to those shown in Table 4 using all periods. In the second period 1775 : 1825—which ends before 1890—, the external information for each observation varies across scenarios. Unsurprisingly, it is for this period where we find the largest differences across scenarios in the AME point estimates. For the third period, i.e., 1850 : 1925, we find a significantly smaller AME, 0.223 vs. 0.314, which is consistent with a decreasing role of mothers in time as participation increases from a $\widehat{\Pi}_1$ estimate of 23.9 percent to an estimate of 38.5 percent.

[14] As the number of missing values decreases in 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.

[15] 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: AME estimates for the three scenarios in the third period range from 0.223 to 0.263 with standard errors around 0.065.

Table 5: ML Estimates for Different Periods
Estimates, AMEs, and Ignorability Tests

	1675 − 1925	1775 − 1925			1775 − 1825			1850 − 1925
	baseline	baseline	high	low	baseline	high	low	baseline
y^M	1.224***	1.094***	1.059***	1.138***	1.434***	1.544***	1.376***	0.642***
	(0.119)	(0.119)	(0.116)	(0.120)	(0.230)	(0.219)	(0.234)	(0.186)
siblings	0.697***	0.800***	0.786***	0.791***	0.408***	0.404***	0.452***	0.554***
	(0.074)	(0.071)	(0.068)	(0.074)	(0.105)	(0.108)	(0.105)	(0.121)
father SES	-0.325**	-1.062***	-1.011***	-1.174***	-1.278***	-1.310***	-1.289***	-2.319***
	(0.147)	(0.176)	(0.177)	(0.183)	(0.201)	(0.201)	(0.202)	(0.276)
father owner	0.387***	1.702***	1.433***	1.979***	1.465***	1.269***	1.726***	2.528***
	(0.108)	(0.145)	(0.157)	(0.157)	(0.175)	(0.171)	(0.191)	(0.266)
AME	0.268***	0.269***	0.282***	0.250***	0.314***	0.403***	0.210***	0.223***
	(0.033)	(0.035)	(0.035)	(0.033)	(0.039)	(0.041)	(0.031)	(0.060)
$\widehat{\Pi}_1 = \widehat{\Pr}\{y=1\}$	0.288	0.310	0.381	0.243	0.239	0.344	0.147	0.385
$\widehat{\Pr}\{I^m=1\,\lvert y^m=0\}$	0.098	0.117	0.130	0.106	0.089	0.103	0.079	0.157
$\widehat{\Pr}\{I^m=1\,\lvert y^m=1\}$	0.181	0.213	0.174	0.275	0.066	0.046	0.106	0.302
$\widehat{\Pr}\{I=1\,\lvert I^m=0,y=0\}$	0.103	0.172	0.188	0.160	0.163	0.183	0.152	0.242
$\widehat{\Pr}\{I=1\,\lvert I^m=0,y=1\}$	0.136	0.135	0.125	0.145	0.038	0.035	0.040	0.214
$\widehat{\Pr}\{I=1\,\lvert I^m=1,y=0\}$	0.703	0.785	0.807	0.757	0.997	0.992	0.995	0.234
$\widehat{\Pr}\{I=1\,\lvert I^m=1,y=1\}$	0.083	0.091	0.090	0.093	0.108	0.109	0.108	0.110
Ignorability test	108527.23	196687.77	258874.58	145187.35	6.54e+06	3.19e+06	3.14e+07	5525.98
	[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. See Section 3.1 for definitions of remaining controls. All models include quarter-century dummies and a location dummy for Horta. The 1675 − 1925 sample includes Horta (1700-1850), Ronfe (1700-1925) 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 3.3. The three scenarios only differ in participation rates for periods before 1890. In the baseline scenario external participation rates before 1890 take the value of the 1890 predictions using census data. In the low (high) scenario, we use as external information for the period before 1890 the smallest (largest) local participation rate for which we have information in all censuses from 1890 to 1950 (see Figure 1). "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 $\Pr\{I_i=1\,\lvert I_i^m,y_i\}$ and $\Pr\{I_i^m=1\,\lvert y_i^m\}$ do not vary with y, y^m, and I^m. $\bar{y}$ is the sample average of the dependent variable.

The bottom half of Table 5 deserves a number of comments. Ignorability tests are again strongly rejected in both sub-periods. Naturally, as the number of missing values decreases in time, all but one of the estimates of the conditional probabilities of observability increase. Participation of the daughter is always associated with a lower probability that the priest reports her activity regardless of the observability of the participation of the mother, I^m. More interestingly, by distinguishing between the two periods, we find evidence of a decreasing influence of social status on the priest reporting activity at the turn of the century: the role of I^m on $\Pr(I=1|y=0,I^m)$ decreases substantially from $0.997-0.163=0.834$ in the $1775:1825$ period to $0.234-0.242=-0.008$ in the $1850:1925$ period.

4.3. Migration

Family migration decisions may reflect wealth and, hence, migration status is a potential control in our model. Unfortunately, in our data only Horta has sufficient coverage of this information. Here we discuss results under the Likelihood Approach adding a dummy for migrant status for the sample of Horta. Reassuringly the AMEs for the three scenarios are similar to those obtained without the dummy and in line with the results with the full sample. In the baseline scenario, our AME estimate using only Horta observations and without adding migrant status as additional control is 0.229 (with a standard error of 0.120). Adding the migrant status dummy hardly changes the AME estimate to 0.225 (0.118). In the high and low scenarios the estimates are 0.298 (0.050) vs. 0.385 (0.052) and 0.193 (0.101) vs. 0.196 (0.103), respectively. The coefficient for migrant status is in all scenarios negative and significant. For example, in the baseline scenario it is -0.281 (0.105). This suggests that migrants into Horta were on average non-participants and, hence, migration was likely associated to marriage decisions.

5. SENSITIVITY OF RESULTS TO PREVALENCE OF MISSINGNESS

Our working sample is very large ($n = 18,523$) and the absolute number of mother-daughter pairs with observed participation (696, see Table 3) is not unusual for historical data. Under Assumptions 3.1-3.2 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 about the robustness of the results. In order to evaluate the impact of the high proportion of missing values in our estimates, we perform a Monte Carlo experiment with $R = 250$ simulations based on our data and model. (See Appendix E for the algorithm details.)

In Figure 2 we present smoothed densities of all estimators of the AME obtained from the Monte Carlo experiment, both under the Likelihood Approach and ignoring missing observations. They 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 (see the upper panel in Figure 2). In contrast, under the Likelihood Approach the AME is effectively unbiased regardless of the proportion of missing values (see the lower panel in Figure 2). The larger the proportion of missing values, the larger the standard errors. However, the AME is quite precisely estimated even for the largest missing incidence (with a standard deviation, 0.0246, ten times smaller than the true AME, 0.267), probably because of the large sample size. In addition to the results shown in Figure 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 96 percent.

To sum up, our Monte Carlo analysis highlights the fundamental role played by the information contained in I and I^m and 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 86 percent.

6. CONCLUSIONS

In this paper, we use historical parish registry data from three Portuguese locations from the late 17th century until the beginning of the 20th centuries to estimate a female labour force participation model that identifies the effect of mothers' labour market participation on that of their daughters. Importantly, we address the problem of missing values, which frequently affects historical records, 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 effect of the mother's working status on the daughter's decision to participate in the labour market. In our preferred specification, the probability that a woman participates in the labour market increases by 26.8 percentage points if her mother also works, a very large effect given that the probability of female participation in the estimated model is 28.8 percent. The existence of such an early transmission mechanism acting as a catalyst of shocks contributes to the understanding of the long-term dynamics of the female labour force participation.

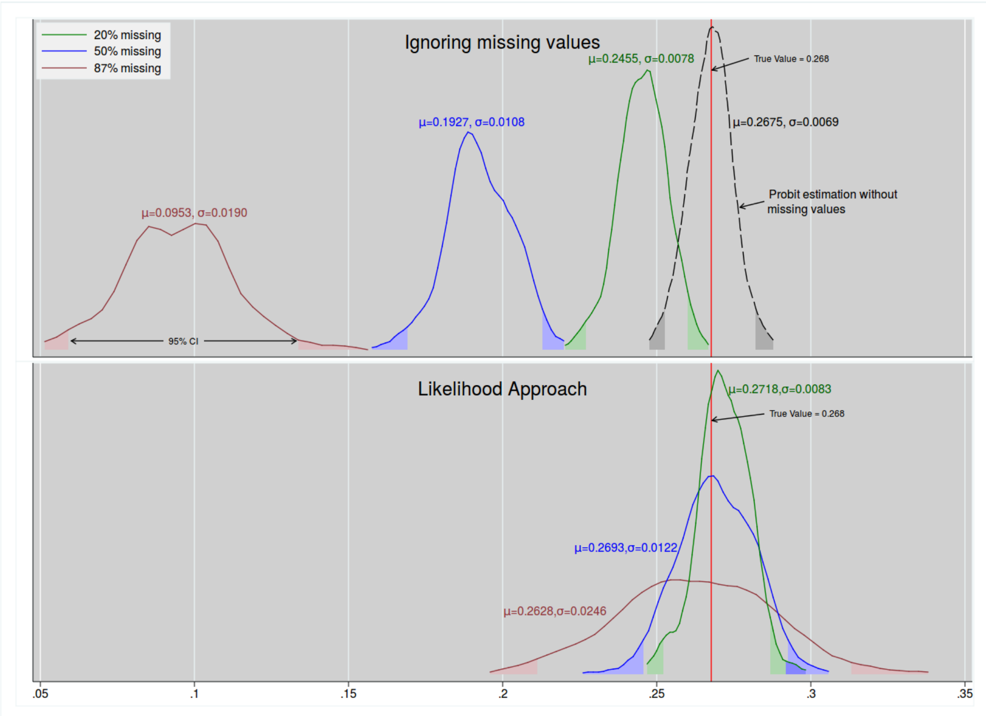


Figure 2: Smoothed densities for AME estimators.

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 4 (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 3.3 using the external information under the baseline scenario (see Figure 1).

ACKNOWLEDGEMENTS

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, Imran Rasul, Jaime Reis, Jan Stuhler, Michèle Tertilt, 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.

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 reconstituiçao 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.

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.

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 9 kilometres apart and strategically located between the two historical administrative centres, Guimarães and Braga, in the Minho region in the northeast of Portugal. The coastal city of Horta at Faial Island—a major stopping port on the journey to Brazil—is located in the Azores Archipelago.

Most of the period studied pre-dates any technological change that occurred in Portugal (e.g., Lains, 2006). Additionally, the legal and social background of Portuguese society during the sample period does not favour the economic independence of women. The most relevant changes in the Portuguese legal system regarding women's rights occurred only after the proclamation of the First Republic in 1910 (Solsten, 1993). Women were not only legally discriminated against but 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—is 13.4 percent and 18.0 percent, respectively. By contrast, the corresponding shares for girls are only 5.0 percent and 1.3 percent. Although some young women began to receive higher education late in the 19th century, a law was passed to allow for the creation of all-girls public schools for secondary education only in 1888. Women were, however, 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 name (and title) of the property owner. The *Morgadio* was practised only among 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 favour spouses over children, male over female children, and older over younger children. In contrast to the *Morgadio*, daughters could inherit life-long rentals, as frequently occurred in the Minho region (Durães, 2009).

The third norm, and 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 (or part of it) and the adjacent land, either became the property of the first marrying child (in which case daughters were more likely to receive it) or was given to a spouse or to unmarried children—who were frequently daughters (Brettell, 1991, Durães, 2009, de Pina Cabral, 1986, Matos, 2009). Daughters might have been favoured by the *terço* for several reasons. First, having land—or the promise of it—increased a woman's chances in the marriage market. The reason for the thin marriage market was the heavy male emigration to Brazil. Second, because 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.

A feature of the three locations is the predominantly male emigration to Brazil beginning in the 16th century. As a consequence, the Minho region—to which Ronfe and Ruivães belong—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 in the city of Horta, 75 men to 100 women, was the third lowest among the 32 largest Portuguese cities, and in Minho's Braga District, there were only 81 men for every 100 women. 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). Researchers report, however, that there are differences in the sex ratio over time (see, e.g. Scott, 1999 for the evidence on Ronfe) and differences in emigration across regions and time (e.g. Reis, 2005).

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 as the "baseline" scenario. The first Portuguese census was administered in 1864, and since then, 11 censuses have been conducted more or less periodically. 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 even the professions of men and women above a certain age at various levels of geographical aggregation, which unfortunately varies across censuses. The census collects data for all regions of Portugal.

For the purpose of obtaining female labour force participation rates, we assemble data from Portugal as a whole, the two largest cities (Lisbon and Oporto), 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 belong; 4) the cities of Lisbon and Oporto; 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 reports our best approximate calculation of the female labour market participation rates for eight regions in Portugal over the 1864-1991 period based on the census data. Although we use mostly a single data source (the exception being 1864, 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 suggesting that the data are still not fully comparable. To correct for this shortcoming, 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, the $D_r's$ represent regional dummies, specifically, a dummy for Braga district locations (encompassing

Table B1: Female Participation Rates by Place of Birth

	Country	City			District		Borough	
	Portugal	Lisbon	Oporto	Horta	Braga	Horta	Guimarães	V. N. 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 Oporto. In 1864, 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 "populaçao 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).

Braga District, Guimarães and V.N. Famalicão), a dummy for Horta district locations (encompassing, Horta District and the city of Horta), and a dummy for the borough of Guimarães.

The unknown parameters of function B.1 are obtained by OLS. To obtain smooth predicted participation rates for our three locations—Horta, Ronfe, and Ruivães—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.

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 equation (3.3) in Section 3.2:

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

The treatment of the second case, i.e., 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, i.e., 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, equation (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 equation (3.3), we can write p_i in equation (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., $\Pr\{y_i = 1 | I_i = 0, I_i^m, y_i^m, x_i\} = 0$, ($\Pr\{y_i^m = 1 | I_i^m = 0, I_i, y_i, x_i\} = 0$).[16] Then, $\overline{v} = \overline{w} = 0$ and the likelihood is:

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

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 equation (D.5). If the remaining terms in equation (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 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 7.33 percent to 69.72 percent —or 62.39 percentage points—when the mother also participates. That increase is substantially lower after the imputation, i.e., 2.86 percentage points. 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.

[16] 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, i.e., $\{I_i = 0, I_i^m = 1, y_i = 1, y_i^m, x_i\}$, $\{I_i = 1, I_i^m = 0, y_i, y_i^m = 1, x_i\}$, $\{I_i = 0, I_i^m = 0, y_i = 1, y_i^m = 1, x_i\}$, $\{I_i = 0, I_i^m = 0, y_i = 1, y_i^m = 0, x_i\}$, and $\{I_i = 0, I_i^m = 0, \;\; y_i = 0, y_i^m = 1, x_i\}$.

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 4. 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 $y_{i,r}^m$ 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 4.

3 The missing information I_i and I_r^m. 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 of 4 evaluated at the simulated values, i.e. we use $\widehat{\Pr}\left\{I^m = 1 \left| y^m = y_{i,r}^m \right.\right\}$ to generate $I_{i,r}^m$ and $\widehat{\Pr}\left\{I = 1 \left| y = y_{i,r}, I^m = I_{i,r}^m \right.\right\}$ to generate $I_{i,r}$.

(b) The proportion of missing values is $q = \{20, 50\}$ percent. We must re-scale the probabilities $\widehat{\Pr}\left\{I^m = 1 \left| y^m = y_{i,r}^m \right.\right\}$ and $\widehat{\Pr}\left\{I = 1 \left| y = y_{i,r}, I^m = I_{i,r}^m \right.\right\}$ to reach the target missing proportion. Specifically, let $\widetilde{\Pr}\{y^m = 1\}$ and $\widetilde{\Pr}\{y^m = 0\}$ 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 $\widetilde{\Pr}\{I^m = 1\} = \widehat{\Pr}\{I^m = 1 \left| y^m = 1 \right.\} \times \widetilde{\Pr}\{y^m = 1\} + \widehat{\Pr}\{I^m = 1 \left| y^m = 0 \right.\} \times \widetilde{\Pr}\{y^m = 0\}$. Hence, we re-scale $\widehat{\Pr}\left\{I^m = 1 \left| y^m = y_{i,r}^m \right.\right\}$ by multiplying it by the factor $\left(\frac{1-q}{\widetilde{\Pr}\{I^m=1\}}\right)$. Similarly, let

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

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

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 4 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.