Migration, Marriage and Social Mobility: Women in Sweden 1880-1900

Martin Dribe*, Björn Eriksson**, and Francesco Scalone***

* Centre for Economic Demography and Department of Economic History, Lund University, Sweden. Martin.Dribe@ekh.lu.se

** Centre for Economic Demography and Department of Economic History, Lund University, Sweden. Bjorn.Eriksson@ekh.lu.se

*** Department of Statistical Sciences, University of Bologna, Italy. francesco.scalone@unibo.it

Abstract

The aim of this paper is to study the connection between internal migration and marriage outcomes for women using complete count census data for Sweden. The censuses 1880, 1890 and 1900 have been linked at the individual level, which enables us to follow about 70,000 women from their parental home to their new marital household. . Marriage market imbalances do not seem to have been an important push factor in internal migration for women at the turn of the last century in Sweden. On the other hand, we find a strong association between migration distance and marriage outcomes, both in terms of overall marriage probabilities and in terms of partner selection. These results highlight the importance of migration for women's social mobility during industrialization.

Paper for the annual meeting of the Economic History Association, Nashville, TN, September 2015

This paper is part of the project "Towards the modern family. Socioeconomic stratification, family formation and fertility in a historical perspective", funded by the Swedish Research Council. We are grateful to the SweCens project at the Swedish National Archives, and especially Mats Berggren for providing the data and valuable assistance in developing the census linkage. A previous version of this paper was presented at the annual meeting of the Social Science History Association, Toronto, Canada, November 2014, and at the conference Women in Changing Labor Markets, Lund University, March 2015. We are grateful to participants at these occasions, and especially Ineke Maas, for valuable comments and suggestions.

Introduction

In historical contexts marriage and partner selection was an important factor for socioeconomic attainment. A micro-level study of a community in southern Sweden, for instance, showed that heterogamy was closely linked to social mobility (Dribe and Lundh 2009, 2010), which makes the workings of the marriage market a crucial factor for understanding reproduction of socioeconomic status (SES) in the past, and thus of inequality. Compared to being homogamously married, social hypergamy (marrying someone with higher socioeconomic status) had a positive effect on attaining higher social status for both men and women. Being hypogamously married, on the other hand, had a negative effect on status attainment for both sexes. Hypergamous marriages also increased the probability of upward social mobility for both men and women. Hypogamous marriages on the other hand lowered the likelihood of upward social mobility and increased the risk of downward mobility. One important factor in finding the right partner is migration. Leaving ones place of origin could be a way for people to enlarge the marriage market and thus to facilitate the search for a spouse with desirable characteristics (Choi and Mare 2012). In this way migration could be an important vehicle for social mobility, not only by promoting earnings or occupational mobility but also through the marriage market.

The aim of this paper is to study the connection between internal migration and marriage outcomes for women using complete count census data for Sweden. The censuses 1880 and 1900 have been linked at the individual level, which enables us to follow individuals from their parental home to their new marital household. We have information on socioeconomic status from occupation as well as place of birth and place of residence. In addition there is information about household context and a range of community-level socioeconomic and demographic indicators. We focus special attention on the association between migration and different marriage outcomes: overall marriage propensities, homogamy, hypergamy and hypogamy. In the following we will first analyze the connection between the marriage market and the likelihood of migration and then study the association between migration and marriage outcomes.

Background

Migration has been frequently associated with social mobility, and much of the literature on migration is concerned with the possible returns to migration and the way such potential returns affect the decision to move (e.g. Sjaastad 1962). At the same time migrants are often positively selected in terms of ability and react to labor market incentives such as wage

differentials or opportunities to climb the occupational ladder (e.g. Chiswick, 1978; Long, 2005; Eriksson and Stanfors 2015; see, however, also Borjas, 1987 and Stewart, 2006 for somewhat different views). By moving migrants can also sort themselves into labor markets where the returns to their specific skills are maximized (e.g. Ferrie 1999; Salisbury 2014). Thus, according to most accounts migration is assumed to be an important instrument for social advancement, and increasing migration fields during industrialization and urbanization are usually connected to changes in occupational structure and opportunity for social mobility (Dribe, Helgertz, and Van de Putte 2015).

However, migration could also be a way for people to enlarge the marriage market and thus to facilitate the search for a spouse with desirable characteristics (Choi and Mare 2012). In turn, partner selection was an important vehicle for social mobility in preindustrial and early industrial society (Dribe and Lundh 2009; 2010), and this was especially true for women who to a large extent depended on their husbands for their socioeconomic status. By searching for a spouse with higher socioeconomic status women could improve their own status without being occupationally mobile or finding a better paying job. Partner selection in historical times was, however, deeply embedded in local culture and traditions, which put limits to this route to social mobility.

Partner selection is often viewed as a function of three different factors: the structure of the marriage market, own preferences and third party influence (from parents, peers, church or other important local institutions, see, e.g., Kalmijn 1998). Hence, migration not only potentially increases chances of social mobility through occupational attainment or improved earnings, but could also have an indirect effect on social mobility through social heterogamy because migration can be expected to reduce parental and other third party influence over marriage decisions (Pélissier et al. 2005; Van de Putte et al. 2005). In particular for long-distance migration, we would expect migrants to lose much of the regular contact with their place of origin, which gradually should reduce the impact family and the local community of origin on marriage decisions. This diminishing third party influence could be expected to increase chances of heterogamy in general, as we know that much local custom in preindustrial, rural societies favored homogamy (see, e.g., Dribe and Lundh 2005).

Migration could also affect partner selection through the structure of the marriage market facing the migrants. To the extent that migration was not only circular within rural areas, but where sending and destination countries differed considerably in terms of age structure, gender composition and socioeconomic diversity, the marriage market facing the migrants in their new destinations might have been quite different from the ones in the

sending areas. This would in turn be an important factor in determining the search for a spouse (Blau, Blum, and Schwartz 1982; Abramitzky, Delavande, and Vasconceles 2011). More specifically, a more diverse marriage market would imply higher rates of heterogamy.

Finally, migration could also affect partner choice through simple selection. To the extent that migrants are positively selected in terms of ambition, ability, etc. they should probably also be expected to be more likely to find partners of a higher social origin (Pélissier et al. 2005).

A connection between migration and social heterogamy has also been found in empirical studies of historical contexts, for example in the late nineteenth-century Netherlands where urban migrants were more likely to experience heterogamy than non-migrants (Bras and Kok 2005). For Stockholm, Sweden, Matovic (1990) noted a shortage of potential spouses in the upper class, which forced the local elite to marry in-migrants in order to avoid hypogamy. At the same time it has been argued that urban migrants sometimes faced difficulties integrating into city society, which may have made hypergamy, and even homogamy, more difficult (see, e.g., King 1997; Van de Putte 2003).

It is also possible that the local marriage market could be a factor in determining migration. If the structure of the marriage market is highly unfavorable in terms of availability of potential spouses of the right age and status, this might induce individuals to move to locations where it is easier to find a spouse. The marriage market can be viewed as similar to the labor market, where searching for a spouse is similar to searching for a job (Oppenheimer 1988). According to standard search theory, labor market imbalances affect the reservation wage, i.e. the wage at which a job searcher is willing to take a given job, and thus the matching of jobs and workers. A shortage of labor will lower the reservation wage while abundance of jobs will increase it (see Rogerson, Shimer, and Wright 2005 for a review). In a similar way a shortage of potential spouses could be expected to prolong search times and lower the requirements on spousal characteristics (see Lichter, Anderson, and Hayward 1995). One way of improving the search would be to move to destinations with a more favorable marriage market. We would expect women growing up in areas with a relative shortage of men in their own SES-group to have been more likely to move, and that migration was associated with higher chances of getting married (but perhaps at a later age) as well as higher chances of homogamy or even hypergamy, while risks of hypogamy would be largest for the stayers.

To summarize, we could hypothesize that migration had a positive effect on social heterogamy for several different reasons. Less parental influence over partner choice can be

expected to have been especially important for increasing hypogamy (marrying someone from a lower social status origin), as it could be expected that families were always more positive to children marrying a spouse from a higher socioeconomic origin while they tried to stop children from marrying down (see the discussion in Dribe and Lundh 2005). A more diverse marriage market should increase heterogamy overall, but it should not affect the direction to any larger degree. Migration as a deliberate way to alter the marriage market or positive selection of migrants in terms of ambition, ability etc. would both imply a positive effect of migration on hypergamy.

Data

In order to study migration and marriage, and test these hypotheses, we link micro-level data from two different Swedish censuses (1880 and 1900) using probabilistic matching techniques. The census data were digitized by the Swedish National Archives and are published by the North Atlantic Population Project (NAPP, www.nappdata.org) which adopts the same format as the Integrated Public Use Microdata Series (IPUMS). All registered individuals are grouped by household. In this way, each individual record contains a household index number and a person index within the household. Age, marital status and sex of each person are also registered. There are family pointer variables indicating the personal number within the household of the mother, father, or spouse, making it possible to link each woman to her own children and husband. In total, the 1880 census counts about 4.6 million persons in 1.2 million households from about 2,530 parishes, while the corresponding figures for 1900 are 5.2 and 1.4 million, respectively.

Any data linking project starts with identifying variables suitable for matching individuals. In order to not introduce bias, only variables that are time invariant over the life course should be considered (see Ruggles 2006). Disqualified variables therefore include information such as current location of residence, and civil or occupational status. Available variables that fulfil the criteria of being fixed over time include birth year, birth place, sex, and names. Birth year, sex, and birth place does not suffer from the problems of variation in spelling associated with names and are therefore used to index the data. In practice this means that individuals are only compared to potential matches between censuses if birth year, birth parish and sex matches exactly in two censuses.

Names (first names and surnames) thus remain as the only variables on which probabilistic linking is done. To make sure that the number of names held by a person does not influence the probability of being linked, only the first recorded first name and surname

are used (whatever name appears as the first entry in the census). Prior to linking, names were subjected to some very limited and basic standardization. The letters å, ä, ö, and w, were standardized to a, a, o, and v respectively while any non-alpha characters were removed. For patronymic surnames, the suffixes –sson and -sdotter was parsed out in order to decrease the higher homogeneity of patronymic surnames relative to family names. As a final standardization measure nobility particles (e.g. von and af) were eliminated from the surname string.

A peculiarity of the Swedish censuses is that a large proportion (about 46% in the 1900 census) of the population is missing surnames. This makes linking using surnames less straightforward than linking on first names. We remedy the problem of surnames in two ways: Firstly all children with missing surnames living in the same household as their fathers have had their fathers surname appended. Secondly, since the patronymic tradition was still followed by some families, patronymic surnames have been constructed using fathers first names. An individual can thus be linked by surnames in four possible ways:

1. own surname_t -> own surname
2. own surname_t -> fathers surname
3. fathers surname_t -> fathers surname
4. own surname_t -> patronym

The likeness of names is evaluated using the Jaro-Winkler algorithm (see Christen 2006 and Christen 2007:41-52 for a more detailed discussion of matching algorithms). The algorithm produces a similarity score by considering common characters, transpositions, common character pairs and increases the score if a string has the same initial characters, and checks for more agreement between long strings than between short and adjusts the score accordingly. For each potential match a similarity score is calculated that ranges from 0 (for completely dissimilar records) to 1.0 (for identical records). Because the true or false status will be unknown a classifier is required. A simple way of classifying a link is to set a threshold value that a potential link has to exceed in order to be classified as true. It is important to note that setting a higher threshold does not necessarily lead to an improvement in link quality. This is because when the threshold is increased the span within which links are compared for duplicates simultaneously narrows. Less restrictive criteria will thus initially yield more potential matches, but also result in an increased proportion of links being lost in the de-duplication stage due an increase in ambiguous links.

Given that some individuals will end up with multiple links, of which at least one will be false, a rule is required for how to deal with ambiguous links. A large share (about two-thirds) of all individuals in the censuses have a second name recorded. By evaluating the likeness of second names, different thresholds for the Jaro-Winkler score were assessed in terms of the share of links that could be confirmed and the number of links made. Based on these evaluations a Jaro-Winkler threshold of 0.85 for classifying a link as true was chosen.

After creating an initial sample of primary links, an additional sample of secondary links is created from the remaining unlinked pool of individuals by exploiting the indirect linking of households created by primary links in the first stage. The creation of secondary links exactly mirrors the method followed for primary links with two exceptions. A new identifier is created for every pair of households in two censuses connected through a primary link. This identifier is then added to the index variables (age, birth parish and sex), thereby narrowing the initial criteria for being considered a match to individuals of the same sex, born in the same parish in the same year and residing with a particular linked individual in both censuses. Because the new indexing severely reduces the size of the group that individuals are compared within between the censuses, only first names are used for probabilistic matching. Again, a threshold of 0.85 was set, and all ambiguous links were discarded. In total we end up with about 70,000 women in the analytical sample (see table 1).

Table 1 here

Table 1 shows the descriptive statistics for the whole population (I) and different stages of exclusions to arrive at the final sample of women in the population, married with information on both their fathers' and husbands' SES (V). Overall, there are no big differences across samples, indicating that the final sample is fairly representative of the whole population in terms of the variables included. Thus there does not appear to be any significant selection bias in the construction of the sample we use in the analysis.

We include a number of variables at individual/family level and community level (parish). Disability indicates whether there was any notation in the census about medical conditions (deafness, blindness, etc.). Only 0.4 percent of women suffered such conditions (0.1 percent in the final sample) but they could be expected to be relevant for both decisions to migrate and for marital outcomes. Migration distance indicates the distance in kilometers between the parish of residence in 1900 and 1880 based on parish centroids. On average women in the sample moved about 34 kilometers. For socioeconomic status we rely on

information about occupation of the household head, usually the father in case of daughters living at home and the husband for married women. Occupational notations in the censuses were coded according to the Historical International Standard Classification of Occupations (HISCO, see Van Leeuwen, Maas, and Miles 2002). Based on HISCO we have classified occupations into different classes following HISCLASS (Van Leeuwen and Maas 2011), which is a 12- category classification scheme based on skill level, degree of supervision, whether manual or non- manual, and whether urban or rural. It contains the following classes: 1) higher managers, 2) higher professionals, 3) lower managers, 4) lower professionals, and clerical and sales personnel, 5) lower clerical and sales personnel, 6) foremen, 7) medium skilled workers, 8) farmers and fishermen, 9) lower skilled workers, 10) lower skilled farm workers, 11) unskilled workers, 12) unskilled farm workers. In this paper we use a five-category version including (1) the elite and upper middle class (HC 1-5), (2) skilled workers (HC 6-7), (3) farmers (HC 8), (4) lower skilled workers (HC 9-10), and (5) unskilled workers (HC 11-12). In 1880 there was a predominance of women from farmer background (58 percent in the final sample), while 19 percent came from the unskilled working class and about 9 percent from the elite and upper middle class.

We look at heterogamy by comparing the SES of the woman's father in 1880 to the SES of her husband in 1900. Table 2 shows the transition matrices linking father SES and spouse SES by migration distance. Overall 63 percent of the women in the sample are married in 1900, and of these 41 percent are homogamously married, 27 percent hypogamously married (husband with lower SES), and 32 percent hypergamously married (husband with higher SES). Longer distance migration is associated with more homogamy in the elite group, and more heterogamy among farmers. For other social groups there is no consistent association between homogamy and migration distance.

Table 2 here

A crucial variable in the migration analysis is the sex ratio (M/F) in the class of origin. It is an indication of the structure of the marriage market and we would expect a higher sex ratio to be associated with less migration, as more potential spouses would be available locally, thus facilitating finding a spouse with preferred characteristics. Mean sex ratio in all parishes is 107 (standard deviation: 32.4; median: 105, inter-quartile range: 93-118). There is thus considerable variation in the sex ratios across different communities, allowing an analysis of its association with migration outcomes.

Methods

In the quantitative analysis we use OLS and linear probability models (LPM) to study the association between parish-level class-specific sex ratios and migration distance, and LPM, parish-level and sister fixed effects models, and instrumental variables (IV) models to study the association between migration and marital outcomes. We look both at migration as a dichotomous variable, representing different distances, and as a continuous variable indicating migration distance. For marriage outcomes we study both overall marriage probabilities and partner selection. As for the latter different models focus on homogamy, hypergamy, and hypogamy, respectively. In these analyses we only include the classes which have a possibility to enter the stage under consideration, which implies that the elite are excluded from the hypergamy estimation and the unskilled are excluded from the hypogamy estimations. We first estimate a basic model including only basic controls and then full models with all covariates and parish fixed effects. In the marriage analysis we also include a specification with sister fixed effects. These models control for all unobserved heterogeneity at the family level (i.e. all conditions shared between sisters in the family).

As a supplement to the fixed effects models, we also consider an alternative identification strategy based on instrumental variables. More specifically, variations in access to railway transportation is exploited as an instrument for migration. Our first measure of railway access is constructed by identifying all station masters in the 1880 census. Thereafter the distance from the location in 1880 of every women in our analytical samples to the nearest station master is calculated. The validity of the instrument rests on the assumption that access to railroads does not affect women's marital outcomes through any other channel than migration. One concern is that railway stations (and station masters) are not randomly dispersed, but rather the result of a carefully planned process. Many station masters were located in urban areas, a factor which in turn may have affected the marriage market. In order to address this concern, a control for distance to nearest urban area is included. Moreover, we construct a second instrument in which railway access is approximated by straight lines connecting the capital (Stockholm) and the second and third largest cities of Sweden (Gothenburg and Stockhom). These straight lines denote the lowest cost path between the major endpoints around which railway construction was undertaken. Again we use the minimum distance to one of these straight lines as the first stage instrument for migration.

Results

We begin by looking at the association between the marriage market, as measured by the class-specific sex ratio, and migration. Table 3 shows the estimates from all the different models. Overall the estimated coefficients are so small that they are insignificant in practice. For example, in the full model with fixed effects (panel C) a 25 percentage points higher sex ratio (corresponds to the entire inter-quartile range) lowers migration probabilities by between 0.15 and 0.52 percent. The corresponding effect on the migration distance is 950 meters (0.95 kilometers). Taken together, these results do not support the expectation that migration is sensitive to marriage market imbalances. In other words, there seems to be little connection between the local marriage market and migration propensities, implying that migrants are not selected from areas with a more unfavorable marriage market.

Table 3 here

Table 4 displays the associations between migration distance and different marital outcomes: ever married (vs. never married), homogamy (vs. heterogamy), hypergamy (vs. homo/hypogamy), and hypogamy (vs. homo/hypergamy). We estimate six different models: model I controls for age and father SES; model II is the full model controlling for number of brothers and sisters, disabilities and the number of servants in the household in 1880; model III is the full models with parish-level fixed effects; model IV is the full model with sister fixed effects; model V is a two-stage model (2SLS) with distance to nearest station master as an instrument for migration distance; and model VI instead uses minimum distance to straight lines between Stockholm and Malmö or Gothenburg to instrument for migration.

Looking first at overall marriage probabilities in panel A, estimates are highly similar in the first three models, which indicates that neither the inclusion of control variables nor parish-level heterogeneity alter the association between migration distance and the probability of being married in 1900. Longer migration distance is associated with a higher propensity of being married. The estimate in model V, including sister fixed effects, is somewhat larger testifying that the association remains when controlling for all shared characteristics at the level of family of origin. When estimating a causal effect of migration, however, the effect is negative implying a lower likelihood of marriage as a result of longer migration.

For hypergamy (“marrying up”) the estimates of the first four models are remarkably similar and the causal estimates in model V and VI shows a somewhat stronger positive effect. Longer migration clearly seems to increase the chances of marrying a spouse from a

higher social origin, just as expected. The estimates for hypogamy are also largely consistent in showing a positive effect of migration distance. The estimates are also fairly similar across the different specifications, with the exception of the IV-estimate in model VI. In line with these results, estimates for homogamy in panel D show a clear and consistent negative effect of migration distance on the probability of homogamy. Longer migration implies a lower likelihood of marrying someone from the same background.

Table 4 here

Table 5 displays results for the first four models using a categorical measure of migration distance, in order to study possible non-linearities in the associations between migration distance and marital outcomes. Estimates are similar across different specifications, indicating that the associations found in a basic model are very robust to adding control variables and controlling for parish-level and sister fixed effects.

Overall there is an inverse U-shaped association between migration distance and overall marriage probabilities (panel A). Migration distance of 10-50 kilometers is associated with the highest marriage probabilities, and short-range migration is associated with the lowest, roughly 20 percentage points lower. Long-range migration, more than 100 kilometers, is associated with about 15 percentage points lower marriage probabilities (somewhat less than that in the sister-FE specification in model IV). This last finding is consistent with long-distance migrants (e.g. rural-urban migrants) marrying at a later age, and this may not be an indication of lower nuptiality overall (see, e.g., Oris 2000).

Table 5 here

Short-distance migration, or no migration, is also related to higher probabilities of homogamy in all specifications, while longer-distance migration is associated with lower chances of homogamy. Instead, long-range migration is positively related especially to hypergamy, but also to hypogamy, while stayers are less likely to experience both hypergamy and hypogamy. In other words, stayers are less likely to marry, but when they do they are more likely to match with someone from the same origin than with someone form either a higher or lower class. Long-distance migrants, on the other hand, are also less likely to marry overall, but if they do they are more inclined to marry upwards, than to marry downwards. Overall, these results are consistent with the hypothesis that migration could be an important

way to find a marriage partner of higher-class origin, which in turn could be an important way to advance socially. Of course, we cannot prove that this effect is causal in the sense that migration is actually promoting the marriage outcome. It may well be that some individuals who are selected to migration, are also selected in other ways that are conducive to finding a favorable marriage partner.

Sensitivity analysis

We have done several additional analyses to check the robustness of our results. One concern is that marriage takes place before migration, which could affect the interpretation of the patterns observed. To check this we linked the 1890 census to our sample and looked if migration between 1880 and 1890 predicted marriage outcomes in 1900, conditioning upon women being single in 1890. The results are displayed in tables 6 and 7, and overall the patterns are highly similar to the baseline results, which show that our results are not sensitive to post-marriage migration.

Tables 6 and 7 here

Another issue is with the farmers; a large and rather special group in this regard because of their strong connection to the land. To ascertain the extent to which our results are driven by the behavior of farmers, we re-estimated the models in Table 4 excluding farmers (see Table 8). For overall marriage probabilities (panel A) the exclusion of farmers does not change the results in any noticeable way. Also for hypergamy (panel B) the patterns are quite similar to the baseline results, while differences are larger for hypogamy (panel C) and homogamy (panel D). When excluding farmers there is a negative association between migration distance and hypogamy (marrying down), which is opposite to what was found in the baseline results. Even though the estimates are not statistically significant, a similar pattern is visible in the IV models, indicating that the negative association is explained by a causal effect. The results for homogamy are of the same sign as in the baseline estimations, but with a lower magnitude. Thus, the positive effect of migration on the risk of hypogamy was completely driven by the farmer group. When we exclude this group longer distance migration is clearly associated with less hypogamy as well as with more hypergamy. In both sets of estimates migration is negatively related to homogamy.

Table 8 here

Conclusion

The aim of this paper is to study the link between the structure of the marriage market, migration and heterogamy. This is important to increase our knowledge about both assortative mating and conditions for social mobility for women in industrializing societies, which is of particular relevance given the overwhelming focus on men in previous research.

Our results show no association between the structure of the marriage market, measured by the SES-specific sex ratios, and the likelihood of migration or the distance moved. Marriage market imbalances do not seem to have been an important push factor in internal migration for women at the turn of the last century in Sweden.

On the other hand, we find a strong association between migration distance and marital outcomes, both in terms of overall marriage probabilities (or at least timing of marriage) and in terms of partner selection. Longer-range migration is associated with lower nuptiality, but also with more hypergamy. For non-farmers longer distance migration is also connected to lower risks of hypogamy, while the opposite effect is found for farmers. Non-farmer migrants, in other words, may be less likely to marry, or at least to marry later, but are more likely to find a favorable match in terms of social origin. Stayers are also less likely to marry than medium-range migrants but are also more likely to be homogamous, and instead less likely to be both hypergamous and hypogamous. These patterns are consistent with several explanations for partner selection commonly mentioned in the literature. Most likely, migrants make more independent decisions on marriage than stayers because of a lower influence from parents or local authorities, which should contribute to heterogamy more generally. This we also see for the medium-distance migrants who have a higher propensity of hypergamy than stayers. However, the fact that hypergamy increases more or less linearly with migration distance, while hypogamy declines at longer distances indicates that longer-range migrants probably are selected in some way related to ambition and ability, or that they move to places with a higher supply of higher-status individuals. In any case we find little support for the idea that migrant women face difficulties integrating into the new communities, at least judged by their marriage patterns.

Taken together, our analysis shows the importance of migration for social mobility not only through earnings mobility and occupational career as has been the focus of much previous research, but also through assortative mating in the marriage market. Migration was clearly a way for many women in early industrial Sweden to improve their socieoeconomic attainment by searching for, and finding, the right partner.

References

Abramitzky, R., Delavande, A., &Vasconcelos, L. (2011). Marrying up: the role of sex ratio in assortative matching. *American Economic Journal: Applied Economics* 3: 124–157.

Blau, P. M., Blum, T. C., & Schwartz, J. E. (1982). Heterogeneity and intermarriage. *American Sociological Review* 47:45–62.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *American Economic Review* 77: 531–553.

Bras, H., & Kok, J. (2005). "They live in indifference together": marriage mobility on Zeeland, The Netherlands, 1796–1922. *International Review of Social History*: 50: S247–S274.

Chiswick, B. R. (1978). The effect of Americanization on the earnings of foreign-born men. *Journal of Political Economy* 86: 897–921.

Christen, P. (2006). *A Comparison of Personal Name Matching: Techniques and Practical Issues. Joint Computer Science Technical Report Series: TR-CS-06-02,* The Australian National University.

Christen, P. (2007). *Febrl - Freely Extensible Biomedical Record Linkage*, Release 0.4.01, The Australian National University.

Choi, K. H., & Mare, R. D. (2012). International migration and educational assortative mating in Mexico and the United States. *Demography* 49: 449–476.

Dribe, M., & Lundh, C. (2005). Finding the right partner: rural homogamy in nineteenth-century Sweden. *International Review of Social History*: 50: S149–S178.

Dribe, M., & Lundh, C. (2009). Partner choice and intergenerational occupational mobility. The case of nineteenth century rural Sweden. *Continuity and Change* 24: 487–512.

Dribe, M., & Lundh, C. (2010). Marriage choices and social reproduction. The interrelationship between partner selection and intergenerational socioeconomic mobility in 19th century Sweden. *Demographic Research* 22: 347–382.

Dribe, M., Helgertz, J., & van de Putte, B. (2015). Did social mobility increase during the industrialization process? A micro-level study of a transforming community in southern Sweden 1828–1968. *Research in Social Stratification and Mobility* 41: 25–39.

Eriksson, B., & Stanfors, M. (2015). Industrious migrants: The earnings of male and female migrants in Swedish manufacturing around 1900. In: Eriksson, B. Dynamic Decades (PhD diss). Lund: Lund Studies in Economic History 72.

Ferrie, J. P. (1999). *Yankees Now: Immigrants in the Antebellum United States, 1840-1860*. New York: Oxford University Press.

Kalmijn, M. (1998). Intermarriage and homogamy: causes, patterns, trends. *Annual Review of Sociology* 24: 395–421.

King, S. (1997). Migrants on the margin? Mobility, integration and occupations in West Riding, 1650-1820. *Journal of Historical Geography* 23: 284–303.

Lichter, D. T., Anderson, R. N., & Hayward, M. D. (1995). Marriage markets and marital choice. *Journal of Family Issues* 16: 412–431.

Long, J. (2005). Rural-urban migration and socioeconomic mobility in Victorian Britain. *Journal of Economic History* 65: 1–35.

Matovic, M. (1990). Migration, family formation, and choice of marriage partners in Stockholm, 1860–1890. In: A. van der Woude, J. de Vries, and A. Hayami (eds.), *Urbanization in History. A Process of Dynamic Interactions*. Oxford: Clarendon.

Oppenheimer, V. K. (1988). A theory of marriage timing. *American Journal of Sociology* 94: 563–591.

Oris, M. (2000). The age at marriage of migrants during the industrial revolution in the region of Liège. *History of the Family* 5: 391–413.

Pélissier, J-P., Rébaudo, D., Van Leeuwen, M. H. D., & Maas, I. (2005). Migration and endogamy according to social class: France, 1803-1986. *International Review of Social History*: 50: S219–S246.

Rogerson, R., Shimer, R., & Wright, R. (2005). Search-theoretic models of the labor market: A survey. *Journal of Economic Literature* 43: 959–988.

Salisbury, L. (2014). Selective migration, wages, and occupational mobility in nineteenth century America. *Explorations in Economic History* 53: 40–63.

Sjaastad, L. (1962). The costs and returns to human migration. *Journal of Political Economy* 70: 80–93.

Stewart, J. I. (2006). Migration to the agricultural frontier and wealth accumulation, 1860–1870. *Explorations in Economic History* 43: 547–577.

Van de Putte, B. (2003). Homogamy by geographical origin: segregation in nineteenth-century Flemish cities (Gent, Leuven, and Aalst). *Journal of Family History* 28: 364–390.

Van de Putte, B., Oris, M., Neven, M., & Matthijs, K. (2005). Migration, occupational identity, and societal openness in nineteenth-century Belgium. *International Review of Social History*: 50: S179–S218.

Van Leeuwen, M. H. D. and Maas, I. (2011). *HISCLASS. A historical international social class scheme.* Leuven: Leuven University Press.

Van Leeuwen, M. H. D., Maas, I., and Miles, A. (2002). *HISCO. Historical international standard classification of occupations*. Leuven: Leuven University Press.

Table 1. Descriptive statistics.

	Census		Linked sample		
	I	II	III	IV	V
Age	14.972	14.328	14.984	14.338	14.532
	(3.105)	(2.970)	(3.150)	(3.011)	(2.986)
Sisters	1.363	1.557	1.396	1.541	1.502
	(1.290)	(1.309)	(1.301)	(1.314)	(1.306)
Brothers	1.490	1.690	1.537	1.681	1.643
	(1.348)	(1.354)	(1.357)	(1.359)	(1.351)
Male servants	0.208	0.148	0.256	0.186	0.185
	(0.602)	(0.483)	(0.661)	(0.543)	(0.531)
Female servants	0.365	0.182	0.408	0.223	0.193
	(0.833)	(0.586)	(0.860)	(0.649)	(0.600)
Disability	0.004	0.004	0.004	0.004	0.001
Father social status					
1		0.072		0.081	0.066
2		0.104		0.098	0.091
3		0.540		0.572	0.583
4		0.075		0.070	0.074
5		0.209		0.179	0.186
Spouse social status					
1					0.107
2					0.149
3					0.406
4					0.136
5					0.202
Migration distance (km)					33.771
					(86.325)
N	483539	320088	167527	123244	69815

Note: I: Entire population; II: population with information on father SES; III: Linked sample 1880-1900; IV: Linked sample with information on father SES; V: Linked sample with information on father and spouse SES (analytical sample).

Table 2. Transition matrices

	Father social status					
Spouse social status	1	2	3	4	5	Row obs.
			Non-migrants (< 10 km)			
1	39.6	12.9	4.1	7.2	5.3	2779
2	16.8	28.1	8.7	16.1	16.1	4961
3	22.7	19.2	66.7	22.8	25.0	20685
4	10.0	17.6	7.7	33.3	19.0	4929
5	11.0	22.2	12.8	20.6	34.6	7069
Column obs.	1956	3068	26048	2806	6545	40423
			Migrant (10-50 km)			
1	38.7	13.5	8.1	8.9	5.8	1839
2	16.5	25.9	14.1	18.7	16.5	3029
3	20.4	17.2	45.0	20.3	22.3	6460
4	11.4	17.5	12.3	23.8	17.7	2760
5	13.0	25.9	20.5	28.3	37.7	4654
Column obs.	997	1752	10642	1367	3984	18742
			Migrant (50-100 km)			
1	59.9	24.6	17.4	15.6	12.0	1004
2	14.2	30.3	19.1	26.4	22.1	989
3	8.1	6.9	22.7	7.7	10.8	680
4	9.3	18.7	14.0	22.7	23.7	793
5	8.5	19.6	26.8	27.5	31.5	1151
Column obs.	528	598	1960	454	1077	4617
			Migrant (> 100 km)			
1	65.4	30.4	24.3	24.7	15.1	1866
2	17.8	27.6	21.8	27.1	26.9	1423
3	6.0	4.5	12.2	5.3	7.4	489
4	6.4	19.4	17.0	23.1	21.3	1020
5	4.4	18.2	24.7	19.8	29.4	1235
Column obs.	1137	942	2034	546	1374	6033

Table 3. Associations between sex ratio in own class and migration in different models.

A. Basic model

	Migration distance (km)				**Binary migration outcome**					
	Raw		log		>10 km		>50 km		>100 km	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Sex ratio	-0.053	0.000	-0.0005	0.013	0.00002	0.696	-0.0002	0.000	-0.0002	0.000

B. Full model, no county-level FE

	Migration distance (km)				**Binary migration outcome**					
	Raw		log		>10 km		>50 km		>100 km	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Sex ratio	0.012	0.156	0.0001	0.601	0.0001	0.166	0.00002	0.651	-0.00001	0.858

C. Full model with county-level FE

	Migration distance (km)				**Binary migration outcome**					
	Raw		log		>10 km		>50 km		>100 km	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Sex ratio	-0.038	0.093	-0.00021	0.671	0.000060	0.633	-0.000088	0.347	-0.00021	0.006

Note: Basic model only control for age, full model control for age, disability, place of birth, labor force participation, household SES 1880, # unmarried brothers 1880, #unmarried sisters 1880, #male servants 1880; #female servants 1880, household head place of birth 1880; sex of household head 1880, urban residence, parish-level socioeconomic characteristics (share employed in transport, share employed in agriculture, proportion migrants, female labor force participation).

Table 4. Associations between migration distance and marital outcomes in 1900.

	I		II		III		IV		V		VI	
						A. Married						
Log distance	0.036	***	0.036	***	0.037	***	0.050	***	-0.201	***	-0.212	***
	(0.001)		(0.001)		(0.001)		(0.002)		(0.019)		(0.014)	
R^2	0.035		0.044		0.041		0.060		.		.	
N	118855		118855		118855		118855		118855		118855	
						B. Hypergamy (“marrying up”)						
Log distance	0.034	***	0.034	***	0.033	***	0.031	***	0.080	***	0.060	***
	(0.001)		(0.001)		(0.001)		(0.003)		(0.011)		(0.008)	
R^2	0.200		0.206		0.184		0.019		0.174		0.196	
N	65197		65197		65197		65197		65197		65197	
						C. Hypogamy (“marrying down”)						
Log distance	0.020	***	0.022	***	0.022	***	0.019	***	0.018		0.001	
	(0.001)		(0.001)		(0.001)		(0.004)		(0.013)		(0.010)	
R^2	0.073		0.096		0.098		0.012		0.096		0.089	
N	56835		56835		56835		56835		56835		56835	
						D. Homogamy						
Log distance	-0.047	***	-0.048	***	-0.047	***	-0.045	***	-0.090	***	-0.058	***
	(0.001)		(0.001)		(0.001)		(0.003)		(0.013)		(0.009)	
R^2	0.087		0.093		0.077		0.035		0.069		0.092	
N	69815		69815		69815		69815		69815		69815	
Fixed effect	-		-		Parish		Sisters		-		-	

Note: Estimates from OLS (I, II), FE (III, IV) and 2SLS (V, VI) regressions. I: Controls for age and father SES; II: Full model; III: Full models with parish FE; IV: Full model with sister FE; V: Distance to railway IV; VI: Straight line IV.
*p<0.1, **p<0.05, ***p<0.01

Table 5. Associations between categorical migration distance and marital outcomes in 1900.

	I		II		III		IV	
				A. Married				
Migration distance (reference: <10 km)								
10-50 km	0.197	***	0.196	***	0.194	***	0.218	***
	(0.003)		(0.003)		(0.004)		(0.009)	
50-100 km	0.112	***	0.114	***	0.107	***	0.156	***
	(0.006)		(0.006)		(0.008)		(0.015)	
> 100 km	0.033	***	0.038	***	0.046	***	0.107	***
	(0.005)		(0.005)		(0.007)		(0.014)	
R^2	0.043		0.052		0.047		0.059	
N	118855		118855		118855		118855	
				B. Hypergamy ("marrying up")				
Migration distance (reference: <10 km)								
10-50 km	0.057	***	0.058	***	0.063	***	0.064	***
	(0.004)		(0.004)		(0.004)		(0.012)	
50-100 km	0.156	***	0.154	***	0.155	***	0.147	***
	(0.008)		(0.008)		(0.008)		(0.026)	
> 100 km	0.214	***	0.208	***	0.197	***	0.195	***
	(0.007)		(0.007)		(0.008)		(0.024)	
R^2	0.201		0.207		0.185		0.021	
N	65197		65197		65197		65197	
				C. Hypogamy ("marrying down")				
Migration distance (reference: <10 km)								
10-50 km	0.099	***	0.098	***	0.090	***	0.079	***
	(0.005)		(0.004)		(0.005)		(0.014)	
50-100 km	0.084	***	0.091	***	0.088	***	0.074	***
	(0.009)		(0.009)		(0.010)		(0.027)	
> 100 km	0.030	***	0.051	***	0.067	***	0.055	**
	(0.008)		(0.008)		(0.009)		(0.026)	
R^2	0.076		0.098		0.098		0.013	
N	56835		56835		56835		56835	
				D. Homogamy				
Migration distance (reference: <10 km)								
10-50 km	-0.134	***	-0.133	***	-0.132	***	-0.129	***
	(0.004)		(0.004)		(0.005)		(0.014)	
50-100 km	-0.210	***	-0.213	***	-0.212	***	-0.195	***
	(0.007)		(0.007)		(0.009)		(0.025)	
> 100 km	-0.210	***	-0.219	***	-0.221	***	-0.216	***

	(0.007)	(0.007)	(0.008)	(0.023)
R^2	0.084	0.089	0.074	0.034
N	69815	69815	69815	69815
Fixed effect	-	-	Parish	Sisters

Note: See Table 4.

Table 6. Associations between migration distance in 1890 and marital outcomes in 1900 (1890 migration distance).

	I		II		III		IV		V		VI	
						A. Hypergamy ("marrying up")						
Log distance	0.021	***	0.022	***	0.021	***	0.023	***	0.125	***	0.066	***
	(0.002)		(0.002)		(0.002)		(0.009)		(0.033)		(0.017)	
R^2	0.161		0.169		0.149		0.008		0.05		0.148	
N	34407		34407		34407		34407		34407		34407	
						B. Hypogamy ("marrying down")						
Log distance	0.026	***	0.024	***	0.021	***	-0.002		0.026		-0.006	
	(0.002)		(0.002)		(0.002)		(0.010)		(0.040)		(0.021)	
R^2	0.059		0.082		0.083		0.009		0.082		0.072	
N	29952		29952		29952		29952		29952		29952	
						C. Homogamy						
Log distance	-0.039	***	-0.038	***	-0.035	***	-0.019	**	-0.143	***	-0.058	***
	(0.002)		(0.002)		(0.002)		(0.008)		(0.037)		(0.018)	
R^2	0.055		0.059		0.048		0.01		.		0.055	
N	36856		36856		36856		36856		36856		36856	
Fixed effect	-		-		Parish		Sisters		-		-	

Note: See Table 4.

Table 7. Associations between categorical migration distance in 1890 and marital outcomes in 1900. (1890 migration distance).

	I		II		III		IV	
				A. Hypergamy ("marrying up")				
Migration distance (reference: <10 km)								
10-50 km	0.035	***	0.041	***	0.045	***	0.048	
	(0.007)		(0.007)		(0.008)		(0.037)	
50-100 km	0.127	***	0.129	***	0.126	***	0.029	
	(0.014)		(0.014)		(0.015)		(0.068)	
> 100 km	0.144	***	0.145	***	0.124	***	0.173	***
	(0.013)		(0.013)		(0.014)		(0.067)	
R^2	0.162		0.170		0.150		0.010	
N	34407		34407		34407		34407	
				B. Hypogamy ("marrying down")				
Migration distance (reference: <10 km)								
10-50 km	0.152	***	0.138	***	0.118	***	0.007	
	(0.009)		(0.009)		(0.010)		(0.043)	
50-100 km	0.077	***	0.068	***	0.058	***	-0.074	
	(0.017)		(0.016)		(0.017)		(0.090)	
> 100 km	0.023		0.022		0.030	*	-0.005	
	(0.015)		(0.015)		(0.015)		(0.071)	
R^2	0.066		0.087		0.087		0.010	
N	29952		29952		29952		29952	
				C. Homogamy				
Migration distance (reference: <10 km)								
10-50 km	-0.146	***	-0.141	***	-0.131	***	-0.052	
	(0.007)		(0.007)		(0.008)		(0.037)	
50-100 km	-0.175	***	-0.171	***	-0.162	***	0.024	
	(0.013)		(0.013)		(0.014)		(0.068)	
> 100 km	-0.149	***	-0.148	***	-0.136	***	-0.152	***
	(0.012)		(0.012)		(0.014)		(0.057)	
R^2	0.057		0.061		0.049		0.012	
N	36856		36856		36856		36856	
Fixed effect	-		-		Parish		Sisters	

Table 8. Associations between migration distance and marital outcomes in 1900. Farmers excluded from the sample.

	I		II		III		IV		V		VI	
						A. Married						
Log distance	0.035	***	0.035	***	0.035	***	0.046	***	-0.081	***	-0.187	***
	(0.001)		(0.001)		(0.002)		(0.004)		(0.028)		(0.029)	
R^2	0.038		0.043		0.038		0.052		.		.	
N	44571		44571		44571		44571		44571		44571	
						B. Hypergamy (“marrying up”)						
Log distance	0.025	***	0.024	***	0.027	***	0.034	***	0.093	***	0.006	
	(0.002)		(0.002)		(0.002)		(0.007)		(0.026)		(0.018)	
R^2	0.111		0.128		0.140		0.027		0.060		0.131	
N	19820		19820		19820		19820		19820		19820	
						C. Hypogamy (“marrying down”)						
Log distance	-0.016	***	-0.011	***	-0.013	***	-0.016	**	-0.074	*	-0.018	
	(0.002)		(0.002)		(0.002)		(0.008)		(0.042)		(0.036)	
R^2	0.038		0.082		0.078		0.020		0.020		0.089	
N	13441		13441		13441		13441		13441		13441	
						D. Homogamy						
Log distance	-0.011	***	-0.013	***	-0.013	***	-0.016	**	-0.051	**	-0.004	
	(0.002)		(0.002)		(0.002)		(0.006)		(0.026)		(0.019)	
R^2	0.031		0.041		0.040		0.010		0.018		0.041	
N	23681		23681		23681		23681		23681		23681	
Fixed effect	-		-		Parish		Sisters		-		-	

Note: See Table 4.
*p<0.1, **p<0.05, ***p<0.01

Table A1. First stage estimations, IV regressions.

	Married				Hypergamy				Hypogamy				Homogamy			
	V		VI		V		VI		V		VI		V		VI	
Father social status																
1	ref.		ref.						ref.		ref.		ref.		ref.	
	(.)		(.)						(.)		(.)		(.)		(.)	
2	-0.186	***	-0.195	***	ref.		ref.		-0.262	***	-0.275	***	-0.263	***	-0.276	***
	(0.027)		(0.027)		(.)		(.)		(0.036)		(0.036)		(0.036)		(0.036)	
3	-0.901	***	-0.908	***	-0.721	***	-0.718	***	-0.981	***	-0.991	***	-0.983	***	-0.993	***
	(0.022)		(0.022)		(0.024)		(0.024)		(0.030)		(0.030)		(0.030)		(0.030)	
4	-0.380	***	-0.386	***	-0.288	***	-0.287	***	-0.547	***	-0.558	***	-0.548	***	-0.560	***
	(0.029)		(0.029)		(0.034)		(0.033)		(0.038)		(0.038)		(0.038)		(0.038)	
5	-0.209	***	-0.212	***	-0.099	***	-0.090	***					-0.357	***	-0.361	***
	(0.025)		(0.025)		(0.028)		(0.027)						(0.033)		(0.033)	
Age (reference: 10)																
11	-0.007		-0.006		0.004		0.007		-0.015		-0.012		0.007		0.010	
	(0.023)		(0.023)		(0.031)		(0.031)		(0.034)		(0.034)		(0.031)		(0.031)	
12	0.023		0.023		-0.002		-0.001		-0.014		-0.014		-0.007		-0.006	
	(0.023)		(0.023)		(0.032)		(0.032)		(0.035)		(0.035)		(0.031)		(0.031)	
13	0.003		0.003		-0.031		-0.029		-0.052		-0.051		-0.038		-0.036	
	(0.023)		(0.023)		(0.031)		(0.031)		(0.034)		(0.034)		(0.030)		(0.030)	
14	0.033		0.034		-0.028		-0.025		-0.043		-0.041		-0.030		-0.027	
	(0.023)		(0.023)		(0.030)		(0.030)		(0.033)		(0.033)		(0.030)		(0.030)	
15	0.023		0.024		-0.053	*	-0.051	*	-0.076	**	-0.073	**	-0.048		-0.046	
	(0.023)		(0.023)		(0.031)		(0.031)		(0.034)		(0.033)		(0.030)		(0.030)	
16	-0.025		-0.023		-0.076	**	-0.069	**	-0.084	**	-0.079	**	-0.082	***	-0.076	**
	(0.024)		(0.024)		(0.031)		(0.031)		(0.034)		(0.034)		(0.031)		(0.031)	
17	-0.079	***	-0.077	***	-0.117	***	-0.113	***	-0.123	***	-0.121	***	-0.112	***	-0.108	***
	(0.025)		(0.025)		(0.033)		(0.033)		(0.035)		(0.035)		(0.032)		(0.032)	

18	-0.140	***	-0.140	***	-0.202	***	-0.199	***	-0.207	***	-0.205	***	-0.194	***	-0.191	***
	(0.026)		(0.026)		(0.034)		(0.034)		(0.036)		(0.036)		(0.033)		(0.033)	
19	-0.169	***	-0.167	***	-0.222	***	-0.218	***	-0.226	***	-0.222	***	-0.219	***	-0.215	***
	(0.027)		(0.027)		(0.036)		(0.036)		(0.038)		(0.038)		(0.035)		(0.035)	
20	-0.218	***	-0.219	***	-0.243	***	-0.244	***	-0.222	***	-0.224	***	-0.235	***	-0.235	***
	(0.028)		(0.028)		(0.036)		(0.036)		(0.038)		(0.038)		(0.036)		(0.036)	
Sisters	0.085	***	0.086	***	0.079	***	0.081	***	0.089	***	0.090	***	0.086	***	0.087	***
	(0.004)		(0.004)		(0.006)		(0.006)		(0.006)		(0.006)		(0.005)		(0.005)	
Brothers	0.033	***	0.034	***	0.058	***	0.059	***	0.058	***	0.058	***	0.054	***	0.055	***
	(0.004)		(0.004)		(0.005)		(0.005)		(0.006)		(0.006)		(0.005)		(0.005)	
Male servants	-0.038	***	-0.046	***	-0.045	***	-0.058	***	-0.087	***	-0.098	***	-0.089	***	-0.101	***
	(0.011)		(0.011)		(0.016)		(0.016)		(0.015)		(0.015)		(0.015)		(0.015)	
Female servants	0.117	***	0.116	***	0.115	***	0.115	***	0.192	***	0.191	***	0.194	***	0.193	***
	(0.010)		(0.010)		(0.017)		(0.017)		(0.014)		(0.013)		(0.014)		(0.014)	
Disability	-0.987	***	-0.987	***	0.431	*	0.444	**	0.411		0.432	*	0.360		0.371	*
	(0.081)		(0.081)		(0.222)		(0.222)		(0.255)		(0.254)		(0.222)		(0.222)	
Distance to urban area	-0.002	***	-0.000		-0.002	***	0.001	*	-0.002	***	-0.000		-0.002	***	0.000	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
Distance to station master	-0.001	***			-0.002	***			-0.002	***			-0.002	***		
	(0.000)				(0.000)				(0.000)				(0.000)			
Distance to straight line			-0.001	***			-0.001	***			-0.001	***			-0.001	***
			(0.000)				(0.000)				(0.000)				(0.000)	
Constant	2.119	***	2.161	***	2.209	***	2.253	***	2.468	***	2.525	***	2.463	***	2.521	***
	(0.028)		(0.028)		(0.033)		(0.033)		(0.040)		(0.040)		(0.038)		(0.038)	
R^2	0.052		0.055		0.045		0.051		0.063		0.068		0.059		0.064	
N	118855.000		118855.000		65197.000		65197.000		56835.000		56835.000		69815.000		69815.000	