Dynamic Decades: A micro perspective on late nineteenth century Sweden

Eriksson, Björn

2015

Björn Eriksson

Dynamic Decades

A micro perspective on late nineteenth century Sweden

This dissertation examines various aspects of mobility during the closing decades of the nineteenth century in Sweden. In terms of the pace and magnitude of the changes taking place during this industrial breakthrough, no other period in Swedish history compares. By studying three aspects of mobility: firm entry and exit, migration and social mobility, this dissertation provides new insights on the choices and outcomes of men, women and firms during a formative period in which the geographic, social and economic landscapes of Sweden were fundamentally transformed. In order to study processes determined at the micro level, novel data sets is employed which tracks individuals and firms across space and over time.

Distribueras av
Media-Tryck, Lunds universitet
Box 117, 221 00 Lund

bookorder@se.lu.se

ISSN: 1400-4860
ISBN: 978-91-87793-16-5

Dynamic Decades

Björn Eriksson

72

Dynamic Decades

A micro perspective on late nineteenth century Sweden

Björn Eriksson

LUND STUDIES IN ECONOMIC HISTORY 72

LUND UNIVERSITY

Dynamic Decades

A micro perspective on late nineteenth century Sweden

Björn Eriksson

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Scool of Economics and Management,

Lund University, Sweden.

To be defended at EC3:207. Date 2015-09-26 and time 10.15.

Faculty opponent

Andrew Seltzer

Organization	Document name
LUND UNIVERSITY	DOCTORAL DISSERTATION
	Date of issue 2015-09-04
Author(s)	Sponsoring organization
Björn Eriksson	

Title and subtitle: Dynamic Decades: A micro perspective on late nineteenth century Sweden

Abstract

This dissertation examines various aspects of mobility during the closing decades of the nineteenth century in Sweden. In terms of the pace and magnitude of the changes taking place during this industrial breakthrough, no other period in Swedish history compares. By studying three aspects of mobility: firm entry and exit, migration and social mobility, this dissertation provides new insights on the choices and outcomes of men, women and firms during a formative period in which the geographic, social and economic landscapes of Sweden were fundamentally transformed. In order to study processes determined at the micro level, novel data sets is employed which tracks individuals and firms across space and over time.

Key words

migration, social mobility, firm survival, competition, female employment, Industrialization, microeconomic data

Classification system and/or index terms (if any)

Supplementary bibliographical information | Language: English

ISSN and key title 1400-4860 | ISBN 978-91-87793-16-5

Recipient's notes | Number of pages 204 | Price

Security classification

Dynamic Decades

A micro perspective on late nineteenth century Sweden

Björn Eriksson

LUND
UNIVERSITY

School of Economics and Management, Department of Economic History
ISBN 978-91-87793-16-5
ISSN 1400-4860

Printed in Sweden by Media-Tryck, Lund University
Lund 2015

Acknowledgements

The completion of this dissertation was made possible through funding provided by the following research projects: *Lön efter möda eller lön efter kön? Könsskillnader före den moderna arbetsmarknadens genombrott*. Vetenskapsrådet (Dnr 421-2013-671); *Lön efter möda eller lön efter kön? Om betydelsen av produktivitet, konkurrens och kontrakt för lönediskriminering.* Crafoordska stiftelsen (Dnr 2011569) and *Banken som arbetsmarknad. Feminisering av svenskt bankväsende under 1900-talet.* Handelsbankens Forskningsstiftelser (P2008-0013:1).

Around a year into my PhD studies I decided to lease an allotment. The idea was that manual labour in the form of digging, weeding, sowing and harvesting would provide a breather from academia. Before long I found out that growing vegetables is not that different from writing a PhD dissertation. In fact, it is in many regards very similar.

First, you need the advice of more experienced gardeners. Maria Stanfors, my main supervisor has been invaluable by providing support, patience and interest in my progress from the first day. Maria's advice resulted in seedlings destined to produce stunted fruit and vegetables being reconsidered, while those more suited to the climate were carefully fertilised and nurtured. Christer Lundh and Anders Ögren have been my assistant supervisors. Both have contributed with valued feedback and suggestions at different stages of completion.

Leasing an allotment comes with having a landlord. In writing this dissertation I have been lucky to have very competent and generous academic landlords in the forms of Anders Nilsson and Mats Olsson as heads of the Department of Economic History, and Tommy Bengtsson and Martin Dribe as directors of the Centre of Economic Demograhy. Tina Wueggertz, Birgit Olsson, Madeleine Jarl and Kristin Fransson have helped me on many occasions with various administrative needs. For this I thank you all.

Growing vegetables is not a solitary pursuit. Implicitly you become part of a larger community of neighbouring gardeners with whom you swap tools and tips. Similarly, I would like to thank numerous colleagues, conference and seminar participants and anonymous reviewers for their suggestions and comments. In my final seminar Anders Nilsson and Thor Berger provided extensive and valuable comments. I am also indebted to many other people that have taken their time to discuss, read drafts of

papers and listen to presentations. In no specific order I would in particular like to extend my gratitude to Joseph Molitoris, Patrik Svensson, Jonas Helgertz, Kirk Scott, Mats Berggren and Joyce Burnette. Monica French and Jeff Neilson helped me with the final weeding by proofreading different parts of the manuscript. I am of course solely responsible myself for any remaining thistles and misshapen carrots.

Finally, I would like to thank Magda, Ivar and Keefe, whom have shown me nothing but love and support, and reminds me every day that there are other things to life than struggling with growing the sweetest garden peas or the crispest lettuce.

Contents

Introduction

In 1870, Sweden was still an undeveloped country on the European periphery. In the decades that followed Sweden went through a period of astonishingly high and persistent economic growth. Less than fifty years later, on the eve of the First World War, Sweden had left its humble origins and emerged as a modern economy. In terms of the pace and magnitude of the societal changes that took place during this industrial breakthrough, no other period in Swedish history compares. This dissertation seeks to contribute to the understanding of this revolution by considering the actions and outcomes of individuals and firms during a formative period in which the geographic, social and economic landscapes were fundamentally transformed.

A revolution implies a sharp break with the past and a radically different future. The late but rapid industrialisation of Sweden meant the realisation of a very different society to that which had been. Material manifestations of this shift include rising output, falling fertility and accelerated urbanisation. Other aspects are less apparent and more intangible but nonetheless as important. The industrialisation of Sweden resulted in further entrenchment of commercialisation, the creation of new markets, and the expansion of old ones. New technology combined with market liberalisation meant that both information and people were able to travel more quickly than ever before while new products and industries changed both the supply of goods and demands for labour. These changes fundamentally affected those that lived through the period.

This dissertation deals with different aspects of individual and firm mobility: migration, social mobility, and firm entry and exit. These topics are not only interesting in their own regard, but are also intimately related to development and the pace at which it proceeds. Studying these topics increases our knowledge about historically important determinants of individual welfare and living standards. The first aim of this dissertation is thus, with the help of these three topics, to further our understanding of the

societal changes gripping Sweden at the end of the nineteenth and beginning of the twentieth century and the impact it had on the people living in this period. Throughout the dissertation a micro perspective is applied. The historical importance of choices made by individuals and firms are thus brought to the forefront. The second aim is to further advance research boundaries within specific topics in economic history by purposefully creating and employing new and novel data. More specifically, this dissertation presents new evidence for determinants of firm survival in an historical setting and provides estimates of how migrants fared in terms of earnings and social mobility. Finally, and perhaps most importantly, this dissertation consciously strives to be as comprehensive as possible in terms of telling a history which includes both men and women.

Migration, social mobility, and firm entry and exit are intimately related to societal transformation and the pace at which it proceeds. The structural change accompanying economic growth creates incentives for migration, provides new opportunities for employment and occupational mobility and displaces established firms' competitive advantages. Correspondingly, for growth to be sustained, migration is required in order to meet new demands for labour. Through migration, labour is reallocated to areas where human capital is scarce and most highly valued. Similarly, a more socially mobile society makes for efficient allocation of individuals into occupations and positions that are consummate with individual ability and aspirations. The market also serves to enhance efficiency through the elimination of unproductive and uncompetitive firms. The net result is a more dynamic and fluid society in which actors responds to market signals and changing incentives.

The first topic and form of mobility considered is the entry and exit of firms into and out of the economy. A distinguishing characteristic of an industrialising society is the transition from cottage based production to large scale manufacturing, operating as some kind of legally defined firm (Landes 1969:2). Firms are subject to rules and regulations, governing their creation, functioning and eventual liquidation (Chandler 1992). Because of these characteristics, the firm presents an easily defined and identifiable unit of analysis which has left behind a wealth of archival sources. The emergence of the firm during industrialisation as the predominant unit of production thus provides the quantitative historian with new forms of data and tools for understanding the functioning of markets, competition and economic transformation. Firms survive and prosper by acquiring capabilities, ceasing

opportunities and adapting to their environment (Hannan and Freeman 1977). In the first analytical chapter of this dissertation, firms and competition are given starring roles. By using the firm as the unit of analysis and survival as the outcome, the Paper shows how an industry may transform from having a workforce dominated by men to becoming predominated by women as a result of a competitive process that favoured firms with a more female workforce.

The second theme of the dissertation is migration. Across industrialising countries during the nineteenth century transportation costs were falling and new opportunities were emerging outside of agriculture. This had an impact on the geographic mobility of men and women. Migration was becoming easier and more worthwhile, leading to a transition from the circular life course migration pattern typical of pre-industrial societies to a more modern type of migration characterised by an increase in long range migration directed towards urban areas (Parish 1973; Pryor 1975; Zelinsky 1971:222). As the century was drawing to an end, Ravenstein (1889:288) concluded that 'an increase in the means of locomotion and a development of manufactures and commerce have led to an increase of migration. In fact you need only seek out those provinces of a country within which migration is proceeding most actively, and you will either find yourself in the great centres of human industry, or in a part of the country whose resources have only recently become available. Migration means life and progress; a sedentary population stagnation'. Ravenstein's optimistic view on migration has found support in more recent scholarship which confirms the importance of migration in the industrialisation of Northern Europe (see for example Baines 1985, 1994; Friedlander 1992; Moch 1992; van der Woude 1992; Williamson 1990). It is very much in this context that we may place Sweden at the end of the nineteenth century, a country in the midst of a rapid industrialisation phase that both depended on and fuelled migration. Internal migration was increasingly taking place over longer distances (The Institute for Social Sciences 1941:42) in response to relatively higher wages in industrialising regions (Jörberg 1972: 348). Simultaneously, large numbers of Swedes emigrated to the United States (Sundbärg 1910). This dissertation furthers the knowledge about migration by considering how internal migrants fared. In Paper 2 the earnings of male and female migrants employed in manufacturing are assessed. The analysis presented in the final Paper show that migrants experienced considerable social mobility compared to those that stayed in the

vicinity of their origin. Although rural migrants also realised gains in terms of upward mobility, urban migrants fared better.

The final central theme of the dissertation concerns the movement of individuals between economic and social positions. Industrialisation resulted in a significant shift of the occupational structure during the nineteenth century. A steady decline of the share employed in agriculture, paired with a change in the type and level of skills demanded in manufacturing and a nascent but growing white collar sector, were radically changing the distribution of occupations and the nature of work (Kuznets 1957: Appendix, Table 4; Goldin and Katz 1998; Carter and Carter 1985; Howlett 2004; Seltzer and Frank 2007). Whether changes in the composition of the workforce or opportunities offered translated into an increase in social mobility is much debated (see for example Lipset and Bendix 1959; Featherman and Hauser 1978; Bourdieu, Ferrie and Kesztenbaum. 2009). Proponents of the "Thesis of industrialism" emphasize the positive impact that industrialisation had on social mobility (Goldthorpe 1985; van Leeuwen and Maas 2010). More specifically, the factors argued to affect social mobility include the expansion of education, a decline in the possibility of transferring human capital between generations, and a change in employers' attitudes and values towards recruitment based on achievement rather than ascription (Treiman 1970; Blau and Duncan 1967). Critics of this optimistic view of industrialisation have argued that the effect of industrialisation on social mobility was counteracted by the superior capacity of elites to invest in human capital and thereby maintain their status (Grusky 1983). In this dissertation social and occupational mobility is addressed from two sides, that of the individual and that of the firm. Firms want to hire the most productive workers in order to improve their profitability and competiveness. Workers on the one hand want to improve their economic situation by moving up the occupational ladder. In the first chapter firms' choice of whether to employ women or men is considered. In line with the argument that industrialisation brings about ascription-based recruitment, the analysis shows that firms that broke the prevailing norm and hired women rather than men faced a considerably lower probability of failing. In the final chapter social mobility is considered from an individual perspective. The results show the existence of a strong association between migration and social mobility, thereby connecting two facets of mobility intimately related to industrialisation.

A micro perspective on economic transformation

Economic history has traditionally been concerned with describing and explaining large-scale processes. This approach is increasingly being complemented by the collection and compilation of high quality micro-data into longitudinal samples. Examples demonstrating the scientific value of such sources include studies based on censuses (Abramitzky, Platt Boustan and Eriksson 2012) personnel records (Seltzer 2013) and business registers (Thompson 2005). These new data sources have allowed economic historians to address old and important questions using the tool box of economics while simultaneously enriching history by considering the trajectories and fates of individual life courses. In historical demography, as a result of the labourious collection of longitudinal micro-data, a now well-established practice of deriving explanations for historical macro trends from carefully considered micro evidence exists (e.g. Alter 1988 Bengtsson Campbell and Lee 2004). For business historians, the preferred approach has been, and still remains the case study analysis. While most often a micro approach, its focus on the actions of single firms or entrepreneurs in practice precludes statistical testing of hypothesis (de Jong, Higgins and van Driel 2015). By retaining a micro perspective but extending the analysis beyond that of the single firm and towards samples of multiple firms, it is possible to not only conduct more explicit hypothesis testing, but also more easily relate the actions of firms to wider societal changes.

The analysis presented in the following chapters is firmly rooted in an economic, demographic and historical practice, concerning itself with choices at a micro level. The units of analysis are either the individuals or firms that had to contend with the challenges and vie for the opportunities offered in a rapidly changing setting. Although the dissertation focuses on actions and outcomes determined at the micro level, aggregate measures and changes are explicitly taken into consideration. In all papers an effort have been made to actively motivate and relate findings at the micro level to aggregate trends and changes. The results show that by analysing the individual choices made by men, women and firms, and the resulting outcomes, the mechanisms underlying aggregate trends and changes may be better understood. By analysing firm's choices regarding whom to employ, and differences in survival rates, the results in Paper 1 show how feminization of occupations and industries may be driven by market competition. By analysing migration, the results from Paper 1 and Paper 4 furthers our understanding of how

historical labour markets worked while simultaneously providing explanations for aggregate trends in geographic mobility.

Because of the high costs of collecting and processing historical micro data, samples almost always have to be limited to certain groups, geographic areas or sectors of the economy. This calls into question the generalizability of results and to what extent evidence derived from analysis based on restricted micro data may be related to aggregate developments and trends. This dissertation, to the extent that it is feasible, uses data that is as comprehensive as possible. The most obvious example is the construction of a panel based on the complete Swedish population in 1880, 1890 and 1900 described in Paper 3 and utilised in the last paper of the dissertation. Even when the focus in terms of data is narrower, an effort has been made to be as encompassing as possible. Paper 1 thus considers the complete population of firms within an industry over an extended period of time (1863-1915). Similarly, although based on cross-sectional data, Paper 2 includes three different industries which cover hundreds of firms and several thousand workers geographically dispersed throughout Sweden around 1900.

Women and men during industrialisation

Through its use of micro data, this dissertation addresses the circumstances facing both men and women in Sweden at the end of the nineteenth century. Women's economic activity was an instrumental part of industrialisation (Goldin and Sokoloff 1982, 1984; Berg 1993). By studying the careers of women and the importance of female labour it is possible to not only provide a richer and more complete historical account, but also a better understanding of the nature of the economic processes which took place during industrialisation (Humphries 1991).

Despite recognition of women's importance in the industrialisation process and an increasing amount of quantitative evidence, we still know much less historically about women than we do about men. Although important, the economic activity of women is harder to capture. Compared to men, historical evidence with regards to not only earnings, but also occupations is much scarcer. In those cases were information exists about what women did for a living and what they earned, data tend to be fragmentary, comprising few observations and often include little information about important individual characteristics. As a consequence our

understanding of several economic processes is almost entirely grounded in evidence exclusively based on men. Migration, a main topic of this dissertation, is a good case in point. The prevailing view is that historically, both internal and international migrants realised considerable returns as a result of their moves (Long 2005; Abramitzky et al. 2012; Collins and Wanamaker 2014). However, this view is based entirely on evidence derived from male samples. This is particularly unsatisfying given the high migration rates of women during industrialisation. By bringing women migrants into the fold, a much more complete debate regarding the relative merits of migration for both the economy as a whole and individual welfare may be pursued.

Apart from telling a more complete story of industrialisation by studying both sexes, this dissertation addresses to what extent and why women and men differed. Understanding nineteenth century differences in what work men and women did and what they earned doing it is important for our understanding of the closing of the gender gap during the twentieth century (Goldin 1992). By both addressing similarities and differences between the sexes, it is possible to more precisely address the eventual causes behind the narrowing of the gender gap. Economic historians have contributed considerably to our understanding of how and why men's and women's circumstances differed historically by addressing differences in norms and customs (Humphries 1987; Rose 1992) and productivity (Cox and Nye 1989; Burnette 2008). The more negative view holds that industrialisation resulted in a marginalisation of women's position within society through segregation, discrimination and a further entrenchment of gender roles (Horrell and Humphries 1995; Humphries and Weisdorf 2015). Other scholars have emphasised the benign effects that market forces had on equalising inequality between the sexes (Pinchbeck 1930; Goldin 1992; Burnette 1997; Burnette 2008).

The data assembled and used in this dissertation is notable by virtue of its inclusion of both men and women. The approach followed is to study men and women side by side and not in isolation wherein each sex occupies a separate sphere of the economy. This is certainly not the only way in which to approach gender in history, but it does have distinct advantages. Importantly, it allows for women's choices and outcomes to be contrasted to those of men. By doing so, it is possible more directly address what factors caused the gender gap to expand or narrow.

An historical comparison of men and women is complicated by the segregation of men and women between industries and occupations. However, there are numerous examples of industries and jobs in which both men and women worked. In Paper 2, two such industries are considered, printing and tobacco. Studying two industries that employed both men and women begs the question to what extent the results are generalizable, or specific for industries that were not segregated by sex. The inclusion of a third male exclusive industry, mechanical engineering, shows that the results found within industries in which men and women worked side by side also applies to a completely segregated industry. Paper 4 takes a more comprehensive view of men and women, using a sample based on the complete population of Sweden between 1880 and 1900. Finally, as shown in Paper 1, when segregation existed it was in some instances eroded away through competition.

Setting the scene: The economic and institutional context

In order to put the main finding of this dissertation into perspective, it is necessary to lay out the particular setting in which the described processes took place. I do this by first briefly discussing the basic facts and explanations concerning Swedish industrialisation. Thereafter, the institutional and economic setting as it relates to the themes of the thesis is considered.

In 1870, Sweden was in comparison to the rest of Europe a distinct laggard. GDP per capita was less than half the level of leading industrialised European countries such as England, Belgium and the Netherlands, two-thirds of France, Germany, Austria and Denmark, and more or less on par with southern countries such as Greece and Spain and its Nordic neighbours Norway and Finland (The Maddison Project 2013). In the following two decades growth progressed at a modest yearly rate of around 1 per cent before doubling to more than 2 per cent between 1890 and 1910 (Schön and Krantz 2012). On the eve of the First World War Sweden had outperformed its Nordic neighbours and the Southern economies, caught up with France and Germany, and closed in on Europe's leading industrialised countries (The Maddison Project 2013).

The high growth rate around the turn of the century was accompanied and matched by several important societal and demographic changes. Much of the growth, particularly after 1890, may be attributed to the profound structural change that gripped the economy. The expansion of the manufacturing and service sectors largely came at the expense of agriculture. By 1910, both manufacturing and the service sector had overtaken agriculture in terms of output (Schön 2010: 193-195). The decline of agriculture also meant that the engine of growth and opportunities were shifting away from the countryside and into cities. The net result was a steady urbanisation of Sweden. These changes all contributed to a rapid decline in fertility. In a period covering less than two generations, marital fertility rates went from about 8 in 1880 to around 3.5 in 1930 (Dribe 2009). This new fertility regime had important human capital implications. For children, less siblings and higher incomes meant more investments in their human capital. For women, fewer children meant a stronger attachment to the labour market and more freedom to participate in economic life.

The rapid development of Sweden during the closing decades of the nineteenth century has both intrigued and puzzled economists and historians. How could a country as backward as Sweden in a time period covering just over a generation sidle up next to nations that for a century had been the forerunners of industrialisation? A number of scholars have emphasised the role played by exports, arguing that Sweden's industrialisation was a natural response to international market integration and an associated increase in foreign demand for natural resources and agricultural products (Fridlizius 1960; O'Rourke and Williamson 1995; Williamson 1996). This externally driven history of Swedish industrialisation is supplemented by a more inward looking perspective centred on the benign domestic conditions particular to nineteenth century Sweden (Jörberg 1961; Schön 1997; Lundh 2010). Accordingly, the reason for the capacious growth of the Swedish economy is to be found in the institutional setting which posed few barriers to private enterprise and enabled rather than constrained geographic and social mobility. Both explanations are useful for understanding different facets of Swedish industrialisation and the transformation of the economy and further acceleration of growth after 1890.

Institutional liberalisation

The institutional setting in which Swedish industrialisation took place was largely defined by the introduction of several markedly liberal acts. The abolition of the guild system in 1846 may be considered to signify a break with the old order and a first step towards a more liberal and modern economy. Soon thereafter, in 1848, the Companies Act introduced the provisions for business owners to operate under limited liability. The exemption or lowering of duties and the suspension of any remaining export and import prohibitions between 1856-1858, together with the near abolishment of the passport requirement in 1860, meant that both labour and capital mobility within Sweden and across its borders faced few constraints (Schön 2010:130-131). The final liberalisation of trade in Sweden was achieved with the introduction of the Freedom of Trade Act in 1864, which served to remove the last barriers to free enterprise, effectively enabling any citizens the freedom to trade as long as the principal was a major and registered the business with the relevant authority (Rabenius 1888).

The liberalisation of trade served to increase the dynamism of the Swedish economy by stimulating the entry and expansion of both small and large enterprises. Between 1870 and 1900 the number of shopkeepers almost doubled (Ericsson 1993:167).At the other end of the spectrum, the Companies Act made it feasible for capital to be invested into larger ventures and new technology without excessive risk (Schön 2010:131, 219). In total, the number of firms registered in official statistics quadrupled during the last three decades of the nineteenth century (Jörberg 1961: 412-420). This wave of new enterprises and the liberal legal setting that caused it did not only differ from Sweden in the past, but also compared favourably to more developed countries (Hannah 2015; Foreman-Peck and Hannah 2015).

Just as the deregulation of the economy stimulated the continuous entry of new firms, it also resulted in an increase in exits. The removal of legal barriers to entry meant that incumbent firms were increasingly challenged by new entrants. Beyond having to vie with new domestic contenders, the lowering of trade barriers also meant that firms were facing competition from foreign firms in the form of imported goods. The net effect of these changes was an increase in bankruptcies which increased the turnover of firms (Ericsson 1993:168).

The laissez faire permeating business ownership and management also extended to individual workers. There were no legal obstacles to internal

migration as long as certain administrative requirements were met. Before moving, a prospective migrant was required to notify the ministers of both the home and destination parish. Permission to settle in a new parish was given as long as it was not suspected that the migrant would have difficulty supporting him- or herself. Refusal of permission to move was exceedingly rare, with less than 1 per cent of applications denied (Eriksson & Rogers 1978:180-181). One institutional barrier to rural labour mobility did however exist, the Servants Act, which hampered migration and occupational mobility for agricultural workers. The act mandated that yearly employment contracts for farmhands and maids must begin on the 1st of November and run until the 24th of October the following year. This resulted in little down time between employment contracts, which made it difficult for agricultural workers to find employment anywhere beyond the vicinity of their last place of work (Lundh 1999:61; Lundh 2003). Outside of agriculture, transactions between employers and workers were more equal. A more liquid labour market and workers' greater bargaining power resulted in more balanced relations and flexible contracts negotiated on an ad hoc basis (Lundh 2010:69-72).

Infrastructure

The institutional liberalization of the economy meant that the Swedish state in many regards removed its hand from interfering in private enterprise. In contrast, a much more active approach was pursued with regards to improvements in infrastructure. The rapid industrialization and transformation of the Swedish economy in the late nineteenth century was underpinned by a range of investment and improvements in infrastructure taking place throughout the nineteenth century. A canal construction boom between 1780 and 1830 served to connect the hinterland of central Sweden with the Baltic and North Seas. This was followed by the opening of stagecoach lines connecting Stockholm with Gothenburg and Scania in the 1830s. The building of canals and stagecoach lines primarily served to lower the cost of transporting bulky goods and improving the flow of information through the post (Schön 2010:101-102). The real breakthrough in terms of affordable passenger transport followed the construction of railways which commenced in 1855. The arrival and expansion of the railway markedly reduced both costs and transit times for passengers. As a result the pace of urbanization began to accelerate as migrants started flocking to towns

connected to the railway (Berger and Enflo 2013). The development and funding of infrastructure was paralleled by a necessary expansion and modernization of the financial sector. Between the mid-1850s and early 1870s Sweden underwent a financial revolution which resulted in a marked increase in the provision of liquidity and financial services (Ögren 2009).

Work and skills

A final important explanation for the dynamism of the period under study is to be found in the nature of work and skills in late nineteenth and early twentieth century. A number of factors relating to the skill level of the general population, production processes and the mobility of labour together resulted in a particularly fluid labour market.

The first point to note is that most work required few advanced skills. Out of the Swedish workforce in 1900, 31 per cent of male workers and 27 per cent of female workers held occupations that may be classified as unskilled, and a further 22 per cent of men and 53 per cent of women held low skilled occupations[1]. For their livelihood most of the labour force thus depended on supplying basic manual labour. Combined with the high geographic mobility of workers, most labour were thus a commodity easily substituted and transferred between places or firms.

In those instances when expertise mattered, skills tended to be of a general nature (within occupation that is) and not firm specific. Moreover, full proficiency was typically acquired at an early age as evidenced by age-earnings and experience-earnings profiles (Burnette and Stanfors 2015). In manufacturing, production methods and technology had still not become complex processes specialized within firms. As a result new workers were easily integrated into a firm's workforce and required little training (Magnusson 1987; Owen 1995). Common white collar occupations such as teaching, also primarily relied on a standard set of general skills (Owen 2004:47-49). Workers and their skill were thus easily transferrable between different employers (Jacoby 1983). For workers, because firm specific human

[1] Calculations based on the Swedish census of 1900, using the HISCLASS-scheme (van Leeuwen and Maas 2011)

capital had not yet become important, changing employer did not have a negative impact on earnings (Burnette and Stanfors 2012a).

The net result was a labour market characterized by high worker turnover and mobility (Elmquist 1899: 94; Slichter 1919). This flexibility enabled in particular women to combine work with motherhood through breaks or quitting and rehiring (Hareven, 1982: 77; Burnette and Stanfors 2012b). Firms were also affected. Little or no resources had to be set aside to teach workers firm specific skills, while the need for screening job applicants was minimal as a consequence of the prevalence of piece rates and the ease of firing unsuitable employees (Leunig 2003).

This flexibility did however not last for long. During the first decades of the 20th century labour markets again became less flexible and workers became increasingly attached to geographic places and employers. The growth of large scale enterprises led to the establishment of policies aimed at staff retention such as tenure based promotion (Sundstrom 1988). The development of internal labour markets changed the incentives of workers. Quitting and short tenures with employers became detrimental for career prospects, resulting in a more loyal and less mobile labour force (Owen 1995). These changes affected women in particular who were finding it increasingly difficult to combine the responsibilities of family life with a more rigid labour market which penalised absence and career interruptions more severely (Goldin 1992: 116). The increasing importance of unions and collective bargaining after 1900 and the introduction of the modern welfare state in post-war Sweden further changed the dynamics of the labour market (Lundh 2010).

Summary

In summary, the second half of the nineteenth century may be considered a distinctly liberal and vigorous period in Swedish history. The take off in growth towards the end of the century was preceded by a number of reforms and investments which increased market efficiency. A wave of market friendly reforms and a demand for general rather than firm specific skills promoted the flexibility of the labour market and reinforced competition between firms. Simultaneously, improvements in infrastructure served to lower costs and remove barriers to migration and trade. The result was a dynamic environment temporarily created by the confluence of a benign

institutional setting, developed infrastructure and a flexible and rapidly changing economy. Put into a longer historical context, the exceptionality of the end of the nineteenth century and beginning of the twentieth becomes even more apparent. The period occupies a distinct historical space that was preceded by a more rigid mercantile regime and followed by the introduction of the modern welfare state.

Theoretical considerations

In the following section the basic premises of the employed models are laid out and the merits and limitations of their application discussed. The model underlying the analysis in this dissertation follow a neo-classical approach based on rational choice theory. Narrowly defined, rational choice theory implies that decisions are made in order to maximize utility under the assumption of perfect information and consistent preferences. No society is or has been perfectly rational, or irrational for that matter, since there are numerous reasons as to why the basic assumptions may not be fulfilled. In practice, decision making is a complex task and information is seldom complete, resulting in choices that satisfy rather than optimize (Simon 1957).

Migration

The models considered in this dissertation have their basis in human capital theory (see Becker 1993). The major strength of the human capital theory lies in its wide applicability theoretically and empirically. The departure point of the model is that choices may be regarded in terms of prospective investments in human capital, each yielding some future stream of benefits which is discounted to their present value. Each choice also entails some cost, often incurred upfront. The rational utility maximising outcome is thus given by the choice resulting in the highest return in the form of the net present value of future benefits minus any costs incurred. In Paper 2 and 4 investment in human capital in the form of migration is considered. Migration may be conceptualized in a straight forward manner by applying the framework laid out above in which each possible move is associated with certain benefits and costs. If the net return from moving is positive, migration subsequently takes place to the chosen destination (Lee 1966). While

theoretically neat, the migration model poses a number of challenges when applied empirically.

The first problem is the comprehensiveness of what may constitute a benefit or cost. Leaving other utilities aside and only focusing on work and career related benefits, we may include income gains caused by regional wage differences or occupational mobility and improvements in the probability or security of employment as pertinent examples of benefits realisable from migration (Sjastaad 1962). In practice, particularly when taking into account the limitations of historical data, only some of all realised benefits are observed and measurable. Comparing benefits is further complicated when making allowance for individual differences. Some benefits to migration are unconditional on ability and thus potentially realizable by any migrant. Parts of the realized return to migration are, however, determined by individual abilities. Building on work by Roy (1951), Borjas (1987) showed how ability determines returns to migration by sorting individuals into destinations in which returns to skills are consummate with individual ability. Moreover, if the more able are "more efficient in migration" costs will decline as ability increases (Chiswick 1999: 182). Even when costs are fixed, as in the case of a train or boat ticket, migration is still relatively more expensive for the less able because fewer hours of work are required on the part of the more able to cover expenses associated with a move. That benefits and costs may vary according to individual ability brings us to the second problem; sample selection bias. Selection bias is pervasive in any empirical estimation involving choices made at the individual level. Hence, in Paper 2, the extent and explanations of selection and how it differs between men and women is considered. In Paper 4, returns to migration are considered primarily in the form of occupational mobility. Differences in costs and unobserved ability are primarily addressed through the use of geographic and family fixed effects.

Discrimination

Discrimination theory seeks to explain instances when an individual's human capital is not objectively valued in the market place but influenced by some other characteristic which does not affect productivity (Becker 1957). Formally, an employer is defined as discriminatory when refusing to hire someone with a marginal value product that is greater than cost. The rationale

behind the choice is provided by what Becker terms as "a taste for discrimination" (Ibid. 1957:39), which is factored into the profit maximizing calculus as a non-monetary cost to the employer, reflecting the negative utility derived from having to interact with the group discriminated against. Employers that have a taste for discrimination against women thus act as though female wages are higher than the actual market wage when deciding whom to hire.

An important implication of Becker's model is that prejudice carries a real cost since employers with a taste for discrimination are willing to forego some profits in order to satisfy their discriminatory preferences. If markets are efficient and employers taste for discrimination is heterogeneous, the expectation is that competition will drive the most discriminatory employees out of the market. In the long run, only the least discriminatory firms are able to survive, resulting in the eventual elimination of discrimination.

In the preceding sections it has been argued that industrialization resulted in a transition towards labour recruitment based on ability rather than ascription and more competitive markets, factors which should have benefited groups discriminated against. The prediction that the existence of discrimination is subject to competition and not a stable equilibrium in the long run have been extensively tested in a contemporary setting (Ashenfelter and Hannan 1986; Hellerstein, Neumark and Troske 2002; Weber and Zulehner 2014) Although there is a voluminous historical literature directed at estimating the extent of discrimination against women and the effect it had on earnings (i.e. Goldin 1992; Burnette 2008), only anecdotal evidence exists concerning the long term implications of Becker's discrimination model historically. In Paper 1, a historical perspective is added to the contemporary literature by analyzing the short run impact of employing women rather than men (which is interpretable as a proxy for taste discrimination) on individual firms and the long run implications for industry and work force feminization.

Delimitations

The preceding sections have outlined the motivation, contribution and setting of this dissertation. By necessity, the papers presented do not provide an exhaustive account of the topic at hand. Two aspects related to the period of study and the transformation of Sweden, education and emigration, falls

outside of the scope of this dissertation. This is not to say that schooling and emigration were historically insignificant factors. The interactions and causal links between education, industrialisation and growth are clearly important topics (Goldin and Katz 1998). In relation to the analysis presented in this dissertation, it should be noted that basic literacy was near universal in Sweden at the end of the nineteenth century (Sandberg 1979). However, as shown by Ljungberg and Nilsson (2009) the stock of human capital in the form of education also increased in this period, particularly through increases in compulsory schooling. However, because the provision of schooling was the responsibility of local parishes and not the state, quality varied considerably. Between 1870 and 1900, the average years of schooling in the population more than doubled from around one to more than two and a half years. This is however, in particular compared to the great expansion of educational opportunities a relatively small amount of schooling. Instead family resources and on the job training rather than public schooling may be regarded as the primary sources of differences in human capital between individuals.

Similarly, by just considering the magnitudes, emigration to the United States looms large in Swedish history. Between 1860 and 1908 more than a million Swedes left their native country for the U.S., a significant number for a country with a population numbering just over 5 million in 1900 (Sundbärg 1910: 128). The most intensive period was in the 1880's when almost 400,000 Swedes emigrated. By 1910, emigration had declined substantially as a result of the decline in the difference between real Swedish and US wages (Bohlin and Eurenius 2010)

In the 1850s and 1860s the migration of agricultural workers and their children dominated the migration flow. These migrants differed from later cohorts both in terms of their marital status and their intention to remain in agriculture once they reached the United States. As the nineteenth century progressed the composition of Swedish migrants to the US changed. By the end of the century the typical migrant was a young unmarried man or woman whose aim was not to continue farming in the United States. Instead they mostly settled in one of the urban centres in the eastern parts of the United States, men taking jobs in industry and women in urban household as domestic servants (Lindberg 1930:188-201). In the last decades of the nineteenth century unmarried women (34.65 percent) and men (35.33 per cent) constituted equal parts of the outflow of emigrants from Sweden (Sundbärg 1910: 166)

How do the results in this dissertation relate to emigration? Firstly, emigration affected those that stayed behind. In an essay written at the height of Swedish emigration to the US, Wicksell (1882) argued that emigration was a boon for Sweden, solving the problem of surplus labour in agriculture and poverty in the countryside, a claim which has been bolstered by quantitative evidence showing that emigration impacted those that stayed behind by raising real wages (Taylor and Williamson 1997).

Secondly, emigration was clearly an important option next to migrating internally within Sweden. Because of data limitations, emigration remains an unseen option in the analysis presented in this dissertation. However, it may be argued that emigrants were in many regards similar to internal urban migrants. The development of emigration to the United States to a great extent mirror the internal migration within Sweden, initially being mainly rural to rural, and gradually shifting it locus towards urban areas. Moreover, by the end of the nineteenth century, just as in the case of internal migration, women were emigrating to the same extent as men. Because most late nineteenth century male and female emigrants choose to primarily settle in urban areas in the United States, it is likely that they would also have moved to a city had they stayed in Sweden. In fact, often eventual emigrants moved to cities for an extended period prior to emigration as part of a stage migration pattern (Nilsson 1970).

Data

This dissertation uses three novel historical micro-level data sets. Each one is based on different data sources which have their own merits and limitations in terms of coverage and application. This section provides a brief presentation of the sources, a description of the scope in terms of coverage and the level of detail of each data set.

Firm register

The first dataset encompasses all tobacco firms active in Sweden between 1863 and 1885, followed until their eventual demise or 1915, the year in which the Swedish tobacco industry nationalised. The main body of the data has been sourced from annual ledgers that were compiled by the Swedish

Board of Commerce as an intermittent step in the production of aggregated official statistics covering factories and manufacturers (Kommerskollegium kammarkontoret, 1863–1885). The archive comprises complete scans of annual ledgers from 1740-1890. Each ledger is organized geographically and lists all businesses operating in each town or city, or for those located in the countryside, by county. This study is to my knowledge the first to utilise the archive for building a longitudinal panel of firms. Hopefully it may serve to draw other scholar's attention to this rich source. The potential of the archive for business and economic historians is indeed tremendous, effectively allowing for the reconstruction of firm level panels with annual observations for any Swedish manufacturing industry covering 150 years.

The choice of industry and periodization was motivated by a number of separate reasons. Firstly, the year 1863 was chosen as a starting point since this was the first year in which a distinction was made between men and women workers by the Swedish Board of Commerce in the source material. The choice of the tobacco industry specifically was essentially motivated by four factors. Firstly, the industry employed sufficient numbers of women to make it viable to study the effect that employing women had on the survival prospects of individual firms. Secondly, a complimentary source, Hantverkare och Fabriker inom Svensk Tobakshantering 1686–1915, authored by Erik Angelin (1950) an archivist at the Swedish Tobacco Monopoly, made it possible to add the date of founding and demise (and the reason for demise) for all firms that entered or exited the industry before or after our window of observation between 1863-1885. Thirdly, the time period covered was one in which no important innovations were introduced in Swedish tobacco manufacturing, allowing me to discount unobserved heterogeneity caused by differences in technology employed. The nationalisation of the whole industry in 1915 serves as the final motivation by constituting a natural end date to the study.

For each firm, on an annual basis, there exists information about number of workers by sex, the number of foremen employed, output for each product category by volume and value, machinery utilised and whether it was powered by steam, electricity or animals and the estimated effect thereof measured in horse powers, and the legal status (limited liability or not) of the firm.

Industry surveys

The second source of data is the HILMA-database, which contains individual level data covering more than 24 000 workers employed in three manufacturing industries in Sweden around the turn of the century and close to 3000 workers employed in bakeries. The data has been sourced from investigations conducted on behalf of the Swedish Board of Commerce. The database is hosted by The Department of Economic History at Lund University and administered by Maria Stanfors.

For the purpose of this dissertation, I restrict myself to the three manufacturing industries: tobacco, printing and mechanical engineering. The industries were surveyed in 1898 (tobacco), 1899 (mechanical engineering) and 1903 (printing) respectively. The tobacco and printing industries were surveyed in their entirety, while the investigation of the mechanical engineering industry included what were deemed to be the 32 most important firms. In 1901, a second survey was conducted of the mechanical engineering industry. This survey included smaller firms. The HILMA-database does not include this information.

The surveys were led by Henning Elmquist, whom together with travelling agents collected information from 4,380 workers employed in the tobacco industry, 7,855 workers in the printing industry, and 12,060 workers employed in mechanical engineering. The Swedish Board of Commerce had several motivations for why these industries were chosen specifically. All three industries were deemed to be important manufacturing industries, both in terms of employment and output. The motivation behind surveying the tobacco and printing industries were to a large extent driven by concerns regarding working conditions. Tobacco, and to some extent also printing relied on female and underage labour, which in combination with general hygienic conditions being "particularly adverse" in both industries were given as the primary reasons as to why these two industries were chosen (Elmquist 1909). The reason for surveying workers employed in mechanical engineering was quite different, motivated by the future potential and importance of the industry as a result of very much being at the vanguard Swedish industrialization (Elmquist 1901).

The workers employed in the surveyed industries were interviewed by travelling agents that conducted interviews in accordance with a pre-printed, detailed questionnaire. A number of questions were asked relating to both workers' professional careers and their private life. We thus have information

about each worker's age, experience, health, occupation, tenure, union and friendly society membership, place of birth and details about marital status and children. Importantly, the questionnaires asked for both weekly earnings and work hours, enabling hourly earnings to be calculated. Moreover, for each worker the employer and work place is known.

Censuses

The final sources of data are the digitalized Swedish censuses of 1880, 1890 and 1900. The three censuses have been digitized and coded by The Swedish National Archives (*Riksarkivet*) and Stockholm City Archives (*Stadsarkivet*). The coding of the censuses was funded by a Swedish Research Council grant and the North Atlantic Population Project at the Minnesota Population Centre. (The Swedish National Archives 2012).

Paper 3 describes how these three sources were turned into a panel through the linking of individuals between the censuses. The result is a panel covering more than 1.8 million men and women observed every ten years. The digitalization and coding of the data was undertaken by The Swedish National Archives. The Swedish censuses differ from the U.S. and British censuses by not being the product of census taking by enumerators visiting and counting the populace. Instead, with one exception (the city of Stockholm) the Swedish censuses were the result of a compilation of excerpts from continuous parish registers which were kept by the Swedish Lutheran church and maintained by the parish priest. For Stockholm, the sources of the censuses were excerpts from the Roteman register, an administrative register supervised by Mantalsnämnden (Geschwind & Fogelvik 2000:207-208). Because the Swedish census is no more than an excerpt from a continuous and consistent source rather than a recreation of a population register as in the case of the U.S. and British censuses, the raw data is of comparatively better quality. The lack of age heaping in the data is direct evidence of this. Moreover the under-enumeration of the population as a whole and of specific groups is less of a problem than is normally expected from censuses.

Compared to the industry surveys described above the census data includes little information regarding work, experience and income. The primary economic variable is recorded occupation, which has been coded into HISCO (van Leeuwen, Maas and Miles 2002). Other basic information about individual characteristics in the censuses includes sex, age, religious

affinity, any recorded disabilities, birth place and current residency. The strength of the data is rather to be found in the detailed references which allows for the identification of different kinds of relationships between individuals, something which I exploit Paper 4 in which the identification strategy relies on comparisons between siblings. A final advantage of the data is that it may fairly easily be expanded by adding for example military, school and death records to individual observations.

In the Swedish censuses of 1880, 1890 and 1900, married women as a rule appear with their maiden name intact, rather than as in most other historical sources sharing the last name of their spouse. This allows for a final linked sample that is broadly representative of the population as a whole, including both sexes. The underlying quality of the data meant that an exceptionally high number of individuals could be linked between each census. The final three census panel covers 1832260 individuals, equivalent to 60.2 per cent of all Swedes born in 1880 or before and still alive and residing in Sweden in 1900. About half of all the linked individuals are women, making it, to my knowledge, the largest historical panel of women currently in existence. These numbers compares favourably to the typical linkage rates between censuses which tends to range from 20-30 per cent (Collins and Wanamaker 2014; Long and Ferrie 2013; Abramitzky et al. 2012).

Summary of papers and main results

Paper 1: A winning strategy? The employment of women and firm longevity during industrialization

The feminization of occupations, industries and the workforce is an enduring aspect of labour market history. However, the employment of women has historically not followed a uniform pattern. Although work in many industries was increasingly being put into the hands of women during the nineteenth and twentieth century, some occupations remained male dominated, while others in contrast experienced periods of de-feminization. Received explanations have emphasized the role of mechanization, the transition from the artisan shops to factory production, and the associated increased division of labour and deskilling as explanations for workforce

feminization. The first paper of the dissertation add an alternative and complimentary explanation by showing how feminization can result from a competitive process in which firms employing more men are continuously selected out of the population.

The paper uses unique data comprising all the firms operating in the Swedish tobacco industry from 1863 to 1885. All firms are followed until their eventual demise or the nationalization of the industry in 1915. The independent variable of interest is the extent to which a firm's workforce was feminized relative to its competitors. A series of proportional hazard models are estimated in order to evaluate the impact of employing women on firm's survival prospects. Factors internal and external to the firm as well as the different forms in which a firm may exit the industry are taken into account. The primary result is that the least feminized firms in the industry faced a highly elevated hazard of failing, and were thus continuously selected out of the population.

Beyond putting forward and testing an alternative explanation for feminization the paper adds to a small but growing literature which takes a quantitative approach to firm survival and its determinants in an historical setting. The secondary results are entirely consistent with not modern findings on determinants of firm survival, supporting the generalizability of a number of stylized facts derived from modern data.

Paper 2: Industrious migrants: Earnings and migrant selection in turn of the century manufacturing

The second paper estimates and seeks to explain earnings premiums of migrants employed in manufacturing around the year 1900. We contribute to the research literature by focusing not only on men but also considering women. By addressing the gender dimension, we shed light on a neglected aspect of the literature on historical labour markets as well as the literature on migrant selection. Female labour migrants, despite being an instrumental part of the industrialization process, have received comparatively little attention from scholars. This is understandable given the available sources. Data based on censuses and parish registers is notoriously unreliable when it comes to identifying what women participated in the labour force and more specifically what work they did. This paper seeks to remedy this shortcoming

by exploiting data based on detailed surveys of three manufacturing industries conducted around the turn of the century.

We find that migrants' earnings premiums, irrespective of gender, were consistent with migrants on average being positively selected. The extended analysis shows that the average positive estimate of selection masks considerable heterogeneity in terms of the extent of and explanations for selection and migrant ability. For both men and women, there is evidence of sorting based on skills consistent with a Roy selection model. An increase in work on both the extensive and intensive margins further explains the earnings premium of migrants.

By exploring the mechanisms behind selection, we establish that while men and women faced different positions and opportunities in the labour market, markets forces allowed many women to cope with migration costs and low pay by working longer and harder. Although women may have occupied a marginal position relative to men in the formal economy around the year 1900, women migrants were by no means marginal individuals, being just like men positively selected. However, that we find costs to be a relevant factor in the selection of women migrants does suggest that women's lower earnings constrained migration and affected selection. This finding is consistent with the observed increase in effort expended by women migrants working in the tobacco industry; indicating that working was a possible strategy used by this relatively low paid group to cope with costs associated with migration.

Paper 3: A Link to the past: Linking and evaluating Swedish historical censuses

The third paper of this dissertation is a methodological paper devoted to linking of historical censuses. This paper outlines a linking process which is transparent, easily replicable, dependent on a minimum of judgement calls, not in itself a cause of bias, and simultaneously results in a high number of accurate links. To this end three Swedish historical censuses enumerated in 1880, 1890 and 1900 are linked and evaluated.

Digitized censuses provide a potentially rich source material for historians. Thanks to increased computational power and methodological advances, the possibility of creating large panel data sets from historical censuses has become a reality. The continuous release of further complete

count censuses means that for the first time ever life courses of full count historical populations over long time spans can be studied. The nature of census data also lends itself to be combined with other supplementary sources such as death records and geographic information, something which will undoubtedly result in many new research questions being asked and insights gained.

Turning cross sectional data into panel form increases its applicability substantially. The first step on the path from raw data to final research result is the linking of individuals between censuses. While striving for a minimum of bias and false positive links and a high rate of linked observations is important when creating linked samples, it is even more so when linking multiple censuses. If a linkage method results in biased and inaccurate links between two censuses, we can expect bias to increase and quality of links to decrease further as additional censuses are linked together. What may at first be a small bias or low rate of false positives may thus with each additional linked census aggregate into a significant problem.

Although linkage rates are normally given secondary concern after bias and accuracy, a too low linkage rate is also problematic if it leads to an unacceptable depreciation of the final sample when multiple linked censuses are joined together. Thus, while the creation of panels based on multiple censuses brings with it much promise, it also means that the linking method itself and the effect it has on link quality and quantity becomes even more important. The paper demonstrates that by carefully considering different thresholds for deeming a link true or false and evaluating the resulting linkage rate and accuracy, it is possible to increase linkage rates and accuracy considerably. Moreover, the method is wholly replicable and applicable to other data sources.

Paper 4: Onwards and upwards? Migration and intergenerational mobility in Sweden 1880-1900

The last paper of this dissertation evaluates returns to internal migration in Sweden in the form of intergenerational social mobility between 1880 and 1900. The societal changes associated with industrialisation have been hypothesised to not only denote a shift in migration patterns but also herald a break from a socially stagnant past. An increase in social mobility is generally viewed as something normatively good, implying fairness and

equality for individuals and a more open society. Migration is often stressed as an important route to social advancement. Historical studies have shown that relocation to urban areas yielded particular benefits in terms of career prospects, pulling individuals away from stagnant rural regions with limited opportunities and towards growing and dynamic towns and cities. Short range rural migration commonly associated with pre-industrial societies have in contrast been characterized as a push phenomenon, with few associated gains for movers. This paper contrasts these of migration and the gains realized thereof.

The paper uses a subsample of a cohort of young men and women drawn from linked census data created following the methods described in Paper 3. In 1880 I observe both the geographic and social origin of each individual. I then proceed by considering each individuals own social status in 1900 as the outcome variable together with migration status as the main independent variable of interest. This paper seeks to place the relative importance of internal migration for social attainment in a greater context by considering multiple destinations, comparing the outcome of migrants both to urban and rural areas and by distance moved. In its nature migration is self-selected. In order to control for this, the main identification strategy rests on the comparison of brothers and sisters. The results show that there were substantial returns to migration in the form of upward social mobility for men. These gains were however not restricted to urban migrants but also to a large extent realised by rural migrants. For women, the results are more mixed. Gains in the form of upward social mobility were considerably smaller in particular for migrants moving to an urban area. When instead using the social status of husbands as a measure of returns, a clearer pattern emerges. This suggests that in terms of social mobility, women primarily benefited through either the occupational upgrade of their spouse, or better prospects on the marriage market in their chosen destination.

Discussion

Findings at the micro level have implications for understanding and explaining aggregate change. The actions of millions of people responding to incentives that pushed and pulled men and women away from their social and geographic origin did not only have individual consequences but also altered

the social and geographic landscape of Sweden. Likewise, choices made by firms about whom to employ did not only change the prospects of individual workers and enterprises but also transformed an entire industry through the competitive churning of firms. These forms of mobility aided the rapid industrialisation of Sweden by contributing to a more efficient and flexible economy. By considering choices made at the micro level and the consequences thereof the results presented in this dissertation furthers our understanding of aggregate change in a transformative and dynamic period that is not only a vital part of Swedish history, but also have wider implications beyond its specific geographic and temporal context. When considered together, the findings of the above papers may be synthetized into more general points with broader implications. People are and have always been influenced by their surrounding environment. This is something which late nineteenth century Sweden together with numerous industrialising countries pay testament to. Economic change is however not mechanical and deterministic. The rate and direction of change is very much determined by the individual actions of people and organisations, each acting according to their own preferences and abilities in response to various incentives. In line with explanations that emphasise the importance of domestic conditions for explaining Sweden's remarkable economic growth at the end of the nineteenth century, the results in this dissertation points to some important internal factors observable at the micro level. The results in all papers provide evidence that market forces played an important part in driving these changes within Sweden. Increasing geographic mobility in the period was accompanied and rewarded by sizable returns to migration. Similarly, firms that were willing and able to transform their workforce by hiring women rather than men were rewarded through significantly better prospects of survival. This implicit threat of failure, in turn, would have spurred surviving firms into employing even more women, thereby affecting the average share of women employed in individual firms and the industry as a whole positively.

Turning to the topic of women's position within society, the results have implications for how we view the role and circumstances of women during industrialisation. Akin to how the rapid development in the period studied informs us about the history and conditions underpinning Sweden's subsequent growth and progress during the twentieth century, an analysis of the standing of women within society and the processes which determined it also provides a historical background and point of departure for

understanding the relatively high levels of gender equality prevailing in modern Sweden.

The results from the papers presented in this dissertation points towards positive effects associated with market efficiency and competition. The analysis in Paper 1 demonstrates how competition between firms worked in the favour of women, by favouring the survival of more feminized firms and by extension encouraging the employment of women. These processes show that even in the total absence of state intervention, market forces may serve to narrow the gap between men's and women's opportunities and relative position in society. This is not to say that state interventions are inconsequential or that traditions and norms were unimportant. Some firms clearly resisted the transformation of the industry and continued to predominantly employ men. As a consequence, the process of feminization, even in an industry as competitive as tobacco in the nineteenth century, was fairly slow. Resistance was however not a tenable position in the long run. By the end of the nineteenth century tobacco had completed its transformation from a male to a female dominated industry.

The analysis in Paper 2 supports the same broad conclusion. Estimates from the Roy selection model show that women, just as men, were able to sort into destinations in which the returns to skills in the form of earnings were consummate with individual ability. This result supports the notion that the labour market in this regard worked to allocate both men and women into ideal geographic destination based on ability and returns to skills.

The results from Paper 4 serve to nuance the picture further. In terms of upward social mobility as measured by own occupation, the returns realized by women are compared to men smaller. This is primarily explained by occupational segmentation, which prohibited female migrants from realizing the same gains in terms of upward mobility as men. That women, despite this, were as migratory as men may be explained by the fact that other important gains apart from upward occupational mobility were realized by women as a result of migration.

Taking a wider geographic perspective, the findings of this dissertation have implications for our understanding of the industrialisation and transformation of Sweden in the aggregate. The importance of migration for eroding regional wage differences has been emphasised as an explanation for historical income convergence both between and within countries (Boyer and Hatton 1997; Taylor and Williamson 1997). Recent studies have shown that aggregate migration flows was also an important factor in driving regional

convergence in Sweden, in particular in the period leading up to 1910 (Enflo, Lundh and Prado 2014; Enflo and Roses 2015). Paper 4 gives evidence in favour of urban and long distance migrants being positively selected in terms of social background while the results in Paper 2 indicate that migrants in manufacturing also tended towards positive selection. That migration was primarily undertaken by the most able would have worked to counteract convergence by dampening growth in sending regions and promoting it in receiving regions. The results from the Roy model on the other hand paints a picture that is more compatible with convergence: the allocation of low skilled migrants to destinations with low returns to skill and high skilled migrants to destinations with high returns to skill is a process which over time would have served to equalize skill premiums across regions.

In both Paper 2 and 4 the results clearly demonstrate that migration was in itself a highly selective process, typically undertaken by the most able and resource rich part of the population. It is thus unclear to what extent migration had an impact on the overall rate of social mobility, given that migration was pre-dominantly undertaken by the privileged. A central issue pertaining to industrialisation is inequality, both of opportunities and outcome. Much of recent focus have been on aggregate measure of inequality (i.e. Long and Ferrie 2013; Piketty 2014). By considering evidence at the micro level it is possible to consider more subtle aspects of inequality through the comparison of different groups. An important form of inequality relates to what individuals or groups are able to capitalize on the opportunities offered by industrialisation. As an example, the increased provision of education around 1900 has been shown to not be "the great leveller of opportunity" since more socially favoured groups seemed to have been better positioned to take advantage of educational opportunities offered and profit thereof (Carlsson 1958:127-137; Bourdieu and Passeron 1977; Parman 2011). The results presented in this dissertation points towards a similar conclusion with regards to migration, which, because of selection, probably further entrenched inequality instead of equalising opportunities.

References

Abramitzky, R., Platt Boustan, L. and Eriksson, K. (2012). and Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, 102: 1832–1856.

Alter, G. (1988). *Family and the Female Life Course: The Women of Verviers, Belgium, 1849-1880*. Madison: The University of Wisconsin Press.

Angelin, E. (1950). *Hantverkare och Fabriker inom Svensk Tobakshantering 1686–1915*. Nordisk Rotogravyr: Stockholm.

Ashenfelter, O. and T. Hannan (1986). Sex Discrimination and Product Market Competition: The Case of the Banking Industry. *The Quarterly Journal of Economics*, 101, 149–174.

Baines, D. (1985). *Migration in a Mature Economy: Emigration and Internal Migration in England and Wales 1861-1900.* Cambridge: University Press.

Baines, D. (1994). *European labour markets, emigration and internal migration, 1850-1913. In Hatton, T. and Williamson, J.G. eds., Migration and the International Labour Market, 1850-1900.* London: Routledge.

Becker, G. (1957). *The Economics of Discrimination*. Chicago: The University of Chicago Press.

Becker, G. (1993). *Human Capital*. Chicago: The University of Chicago Press.

Bengtsson, T, Campbell, C. and Lee, J.Z. (2004). *Life under Pressure: Mortality and Living Standards in Europe and Asia, 1700-1900.* Cambridge: The MIT Press.

Berg, M. (1993). What difference did women's work to the industrial revolution? *History Workshop Journal*. 35: 22-44.

Berger, T. and Enflo, K. (2013). *Locomotives of local growth: the short- and long-term impact of railroads in Sweden*. EHES Working Papers in Economic History, 42.

Blau, P. and Duncan, O.D. (1967). *The American Occupational Structure.* New York: Wiley.

Bohlin, J. and Eurenius, A_M. (2010). *Why they moved - Emigration from the Swedish countryside to the United States, 1881–1910. Explorations in Economic History*, 47: 533-551.

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

Bourdieu, J., Ferrie, J. and Kesztenbaum, L. (2009). Vive la difference? Intergenerational Mobility in France and the Unites states during the Nineteenth and Twentieth Centuries. *Journal of Interdisciplinary History*, 39: 523-557.

Bourdieu, P. and Passeron, J. (1977). *Reproduction in Education, Society and Culture*. London: Sage.

Boyer, G.R. and Hatton, T.J. (1997). *Migration and labour market integration in late nineteenth-century England and Wales*. *Economic History Review*, 50: 697–734

Burnette, J. (1997). An investigation of the female-male wage gap during the industrial revolution in Britain. *The Economic History Review*, 50: 257-281.

Burnette, J. (2008). Gender, work and wages in industrial revolution Britain. University Press: Cambridge.

Burnette, J. and Stanfors, M. (2012a). *Before Careers: Experiences of Wage Growth among Late Nineteenth-Century Swedish Cigar Workers*. Lund Papers in Economic History, 121.

Burnette, J. and Stanfors, M. (2012b). Was there a family gap in late nineteenth century manufacturing? Evidence from Sweden. *The History of the Family*, 17: 31-50.

Carlsson, G. (1958). *Social Mobility and Class structure*. Lund: Gleerup

Carter, M.J. and Carter, S.B. (1985). Internal labor markets in retailing: The early years. *Industrial and Labor Relations Review*, 38: 586-598.

Chandler, A.D. (1992). What is a firm? A historical perspective, *European Economic review*, 36: 483-492.

Chiswick, B. (1999). Are immigrants favorably self-selected? *The American Economic Review,* 89: 181-185.

Collins,W.J. and Wanamaker, M.H. (2014). Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data. *American Economic Journal: Applied Economics*, 6: 220-252.

Cox, D. and Nye, J. V. (1989). Male-female wage discrimination in Nineteenth-century France. *Journal of Economic History*, 49: 903–920.

De Jong, A., Higgins, D.M. and van Driel, H. (2015). *Towards a new Business History?* Business History, 57:5-29.

Dribe, M. (2009). Demand and Supply factors in the fertility transition: a county-level analysis of age-specific marital fertility in Sweden, 1880-1930. *European Review of Economic History*, 13:65-94.

Elmquist, H. (1899). *Undersökning af tobaksindustrien i Sverige*. Stockholm: Kommerskollegium.

Elmquist, H. (1901). *Undersökning af den mekaniska verkstadsindustrien i Sverige*. Stockholm: Kommerskollegium.

Elmquist, H. (1909). *Undersökning af tryckerier och därmed förenad handtering i Sverige.* Stockholm: Kommerskollegium.

Enflo, K. and Roses, J.R. (2015). Coping with regional inequality in Sweden: structural change, migrations, and policy, 1860-2000. *The Economic History Review*, 1: 191-217.

Enflo, K., Lundh, C. and Prado, S. (2014). The role of migration in regional wage convergence: Evidence from Sweden 1860-1940. *Explorations in Economic History*, 52: 93-110.

Ericsson, T. (1993). Social Mobility and the Urban petite bourgeoisie: Sweden in European Perspective, 165-181, in Miles A. and Vincent, D., *Building European society: Occupational change and social mobility in Europe, 1840-1940*. Manchester: Manchester University Press.

Featherman D.L. and Hauser R.M. (1978). *Opportunity and Change.* New York: Academic Press.

Foreman-Peck, J. and Hannah, L. (2015). The diffusion and impact of the corporation in 1910. *Economic History Review,* 3: 962-985.

Fridlizius, G. (1960). The Crimean War and the Swedish Economy. *Economy and History*, 3: 56-103.

Friedlander, D. (1992). Occupational structure, wages, and migration in late nineteenth-century England and Wales. *Economic Development and Cultural Change*, 40: 295-318.

Geschwind, A. and Fogelvik, S. (2000). The Stockholm Historical Database, 207 – 231, in Hall, P.K., McCaa, R. and Thorvaldsen, G. *Handbook of International Historical Microdata for Population Research*. Minneapolis: Minnesota Population Center.

Goldin, C. (1992). Understanding the Gender Gap: An Economic History of American Women. Oxford: Oxford University Press.

Goldin, C. and K. Sokoloff (1982). Women, Children, and Industrialization in the Early Republic: Evidence from the Manufacturing Censuses. *The Journal of Economic History,* 42, 741–774.

Goldin, C. and K. Sokoloff (1984), The Relative Productivity Hypothesis of Industrialization: The American Case, 1820 to 1850, *The Quarterly Journal of Economics*, 99, 461–487.

Goldin, C. and Katz, L. (1998). The origins of technology-skill complementarity. *The Quarterly Journal of Economics*, 113: 693-732.

Goldthorpe, J.H. (1985). On Economic Development and Social Mobility. *The British Journal of Sociology*, 36: 549-573.

Grusky, D.B. (1983). Industrialization and the Status Attainment Process: The Thesis of Industrialism Reconsidered. *American Sociological Review*, 48: 494-506.

Hannah, L. (2015). A global census: publicly traded and close companies in 1910, *Economic History Review*, 68: 548-573.

Hannan, M. T. and J. H. Freeman (1977). The population ecology of organisations. *American Journal of Sociology*, 82, 929–964.

Hareven, T. K. (1982). *Family time and industrial time: The relationship between the family and work in a New England industrial community*. New York: Cambridge University Press

Hellerstein, J. K., D. Neumark and K. R. Troske (2002). Market Forces and Sex Discrimination. *The Journal of Human Resources*, 37: 353–380.

Horrell, S. and Humphries, J. (1995). Women's Labour Force Participation and the Transition to the Male-Breadwinner Family, 1790-1865, *The Economic History Review,* 48: 89-117.

Howlett, P. (2004). The Internal Labour Dynamics of the Great Eastern Railway Company, 1870-1913. Economic History Review, 57: 396–422.

Humphries, J. (1987). "…The Most Free From Objection…" The Sexual Division of Labor and Women's Work in Nineteenth-Century England. *The Journal of Economic History*, 47: 929-949.

Humphries, J. (1991). 'Lurking in the Wings. . .': Women in the Historiography of the Industrial Revolution, *Business and Economic History*, 20: 32-44

Humphries, J. and Weisdorf, J. (2015). The Wages of Women in England, 1260-1850. *Journal of Economic History*, 75: 405-447.

Jacoby, S. M. (1983). Industrial labor mobility in historical perspective. *Industrial Relations,* 22: 261-282.

Jörberg, L. (1961). *Growth and Fluctuations of Swedish Industry, 1869-1912*. Lund: Almqvist & Wiksell.

Jörberg, L. (1972). *A History of Prices in Sweden 1732-1914.* Vol. 2, Lund: CWK Gleerup.

Kommerskollegium kammarkontoret (1863-1885). *Årsberättelser fabriker serie 4, Da4, vol S05823–44, vol S02054–61 and vol S05806–19*. Kommerskollegiets arkiv, National Archives (Riksarkivet), Stockholm.

Kuznets, S. (1957). Quantitative Aspects of the Economic Growth of Nations: II: Industrial Distribution of National Product and Labor Force. *Economic Development and Cultural Change*, 5:1-111.

Landes, D.S. (1969). *The Unbound Prometheus.* Cambridge: Cambridge University Press.

Lee, E.S. (1966). A Theory of Migration. *Demography*, 3: 47-57.

Leunig, T. (2003). Piece rates and learning: understanding work and production in the New England textile industry a century ago. *Economic History Working Papers*, 72/03. Department of Economic History, London School of Economics.

Lindberg, J.S. (1930). *The Background of Swedish Emigration to the United States*. Minneapolis: The University of Minnesota Press.

Lipset S.M. and Bendix, R. (1959), [1967]. *Social Mobility in Industrial Society*. Berkeley: University of California Press.

Ljungberg, J. and Nilsson, A. (2009). Human Capital and Economic Growth: Sweden 1870-2000. *Cliometrica*, 3: 71-95.

Long, J. (2005). Rural-Urban Migration and Socioeconomic Mobility in Victorian Britain. *The Journal of Economic History*, 65: 1-35.

Long, J. and Ferrie, J. (2013). Intergenerational Occupational Mobility in Great Britain and the United States since 1850. *American Economic Review*, 103: 1109-37.

Lundh, C. (1999). Servant Migration in Sweden in the Early Nineteenth Century. *Journal of Family History*, 24: 53-74.

Lundh, C. (2003). Life Cycle Servants in Nineteenth Century Sweden – Norms and Practice. *Lund Papers in Economic History*, No. 84.

Lundh, C. (2010). *Spelet regler: Institutioner och lönebildning på den svenska arbetsmarknaden 1850-2010*. Stockholm: SNS.

Magnusson, L. (1987). *Arbetet vid en svensk verkstad: Munktells 1900-1920*. Lund: Arkiv förlag.

Moch, L.P. (1992). *Moving Europeans: Migration in Western Europe since 1650*. Bloomington: Indiana University Press.

Nilsson, F. (1970). *Emigrationen från Stockholm till Nordamerika 1880-1893: en studie I urban utvandring*. Stockholm: Nordstedt.

O'Rourke, K.H. and Williamson, J.G. (1995). Open Economy Forces and Late Nineteenth Century Swedish Catch-Up: A Quantitative Accounting. *Scandinavian Economic History Review*, 43: 171-203.

Ögren, A. (2009). Financial revolution and economic modernisation in Sweden. *Financial History Review*, 16: 47-72.

Owen, L. (2004). An economic Perspective on Career Formation, 42-55, in Mitch, D., Brown, J. and van Leeuwen M.H.D. *Origins of the Modern Career*. Aldershot: Ashgate.

Owen, L. J. (1995). Worker Turnover in the 1920s: The Role of Changing Employment Policies. *Industrial and Corporate Change*, 4: 499–530.

Parish, W.L. (1973). Internal Migration and Modernization: The European Case. *Economic Development and Cultural Change,* 21: 591-609.

Parman, J. (2011). American Mobility and the Expansion of Public Education. *The Journal of Economic History*, 71: 105-132.

Piketty, T. (2014). *Capital in the Twenty-First Century*. Cambridge: Harvard University Press.

Pinchbeck, I. (1930). *Women Workers and the Industrial Revolution, 1750-1850.* London: George Routledge & Sons.

Pryor, R.J. (1975). Migration and the process of modernization, 23-38, in Kosinski, L.A. & Prothero, R.M. *People on the Move*. London: Methuen & Co.

Rabenius, T. (1888). Näringsfrihet. *Nordisk Familjebok*, 12, 19–22.

Ravenstein, E.G. (1889). The Laws of Migration: Second Paper. *Journal of the Royal Statistical Society*, 52: 241-305.

Rose, S.O. (1992). *Limited Livelihoods: Gender and Class in Nineteenth-Century England*. Berkeley: University of California Press.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3: 135-146.

Sandberg, L.G. (1979). The Case of the Impoverished Sophisticate: Human Capital and Swedish Economic Growth before World War I. *The Journal of Economic History*, 39: 225-241.

Schön, L and Krantz, O. (2012). Swedish Historical National Accounts 1560-2010. *Lund Papers in Economic History*, 123.

Schön, L. (1997). External and Internal Factors in Swedish Industrialization. *Scandinavian Economic History Review*, 45: 209-223.

Schön, L. (2010). *Sweden's Road to Modernity: An Economic History*. Stockholm: SNS.

Seltzer, A. (2013). The impact of female employment on male salaries and careers: evidence from the English banking industry, 1890–1941. *The Economic History Review*, 66: 1039–1062.

Seltzer, A.J. and Frank, J. (2007). Promotion tournaments and white collar careers: evidence from William Deacon's Bank, 1890-1941. *Oxford Economic Papers*, 59: 49-72.

Simon, H. (1957). *Models of Man*. New York: Wiley.

Sjastaad, L.A. (1962). The Costs and Returns of Human Migration. *Journal of Political Economy*, 70: 80-93.

Slichter, S. (1919). *The turnover of factory labour*. New York: D. Appleton & Co.

Stanfors, M. and Burnette, J. (2015). Estimating Historical Wage Profiles. *Historical Method*s, 48: 35-51.

Statistics Sweden (1999). Befolkningsutvecklingen under 250 år: Historisk statistik för Sverige. *Demografiska rapporter*, 2.

Sundbärg, G. (1910). *Den svenska och europeiska folköknings- och omflyttningsstatistiken i sina hufvuddrag,* Emigrationsutredningen bilaga 4. Stockholm: Norstedt & söner.

Sundstrom, W. (1988). Internal Labor Markets before World War I: On-the-Job Training and Employee Promotion. *Explorations in Economic History*, 25: 424–445.

Taylor, A.M. and Williamson, J.G. (1997). Convergence in the age of mass migration. *European Review of Economic History*, 1: 27-63.

The Institute for Social Sciences, Stockholm University, (1941). *Population Movements and Industrialization*. London: P.S. King & Son.

The Maddison-Project (2013). http://www.ggdc.net/maddison/maddison-project/home.htm, [Accessed 2015-04-26]

The Swedish National Archives. (2012). *The SweCens NAPP Variable Documentation for the Swedish censuses 1890 and 1900.*

Thompson, P. (2005). Selection and Firm survival: Evidence from the Shipbuilding Industry, 1825-1914. *The Review of Economics and Statistics,* 87, 26–36.

Treiman, D.J. (1970). Industrialization and Social Stratification. *Sociological Inquiry*, 40: 207-34.

van der Woude, , A., Hayami, A. and Vries, J. (1990). *Urbanization in History*. Oxford: Clarendon Press.

van Leeuwen, M. and Maas, I. (2010). Historical studies of social mobility and stratification. *Annual Review of Sociology,* 36: 429-451.

van Leeuwen, M. and Maas, I. (2011). *HISCLASS: A Historical International Social Class Scheme.* Leuven: Leuven University Press.

van Leeuwen, M., Maas, I. and Miles, A. (2002). *HISCO: Historical International Standard Classification of Occupations.* Leuven: Leuven University Press.

Weber, A. and Zulehner, C. (2014). Competition and gender prejudice: Are discriminatory employers doomed to fail? *Journal of the European Economic Association*, 12: 492-521.

Wicksell, K. (1882). *Om utvandringen: Dess betydelse och orsaker*. Stockholm.

Williamson, J.G. (1990). *Coping with City Growth during the British Industrial Revolution*. Cambridge: University Press.

Williamson, J.G. (1996). Globalization, Convergence and History. *The Journal of Economic History*, 56: 277-306.

Zelinsky, W. (1971). The Hypothesis of the Mobility Transition. *Geographical Review*, 61: 219-49.

Paper 1

A Winning Strategy? The employment of women and firm longevity during industrialization[2]

Björn Eriksson
Department of Economic History, Lund University

Maria Stanfors
Department of Economic History, Lund University

Abstract
Why do certain firms prosper and grow old while other firms fail? Established knowledge tells us that it is related to the firm's ability to adapt to market conditions, for example through product diversification, learning-by-doing, and through the adoption of new strategies regarding technology, human resources, and management practices. This paper argues that the employment of women constituted an important competitive advantage for firms in nineteenth-century manufacturing. By using new data covering the entire Swedish tobacco industry, estimating duration models, we find that firms which employed more women were considerably less likely to fail than firms which employed men. The strategy of hiring women in order to reduce costs was a winning strategy among firms in a labor-intensive industry in competitive markets. Thus the adopters of this strategy lived on. The extended longevity of more feminized firms, in turn, reshaped the whole industry. Industry feminization may thus be seen as result of a competitive process in which more feminized firms through longevity came to dominate the industry.

JEL classification: L10, L66, C41

Key words: Firm survival, longevity, competing risks, competition, female employment

---

[2] A shorter version of this paper with the same title is forthcoming in *Business History*.

Introduction

The question of why some firms survive and grow old while others fail has attracted attention from scholars in various disciplines. Many efforts have been made to describe and explain firm survival and industry dynamics from an economics perspective (e.g., Dunne, Roberts and Samuelson 1988; Klepper and Simons 1997, 2000; Agarwal and Gort 2002; Klepper and Thompson 2006). Business history literature is replete with examples of companies, which, to varying degrees, failed or succeeded in adapting to a shock or to an evolving environment, especially technological change and market competition (see Jones and Zeitlin 2007: parts I-II) that threatened their survival. The fact that some companies managed to survive and even prosper despite change in their environment leads us to the fundamental question: What differentiates the companies that successfully adapt to change from those that do not? This article provides an answer to this question while contributing to the literature on firm survival and deepening the understanding of firm longevity in an historical context. Using unique data comprising all the firms operating in the Swedish tobacco industry in the second half of the nineteenth century, we focus on how the employment of one particular input – women's labour – affected firm longevity. The evidence presented in this paper shows that firms which employed more women than their competitors faced a significantly lower risk of failure, and thus they lived on and came to dominate the industry.

Firm longevity poses a complex challenge to the historian. On the one hand, old firms differ from young by having acquired experience and productive resources. On the other hand, firms survive and grow old because they are or were fundamentally different from firms that fail. A long tradition exists in business history of addressing these critical issues through case studies describing how firms stay competitive given a specific context (e.g., Nenadic 1993; McGovern 2007; Kipping and Cailluet 2010). This paper remains true to this tradition by studying firms in a specific historical setting but departs from the standard case study approach by considering a particular industry as a whole. While precluding us from considering the most intimate inner workings of firms provided by case studies, our approach has a number of advantages given the topic at hand. First, we avoid any survivorship bias, since the many firms that failed are as much part of our story as the ones that thrived and grew old. Second, our approach allows us to explicitly model longevity and statistically assess the relative importance of firm

characteristics for survival. Third, by considering a complete industry population of firms, we can explore the wider implications of firm longevity on an industry and society at large. In the long run, given an efficient market, under-performing firms will fail and competitive firms survive. If the cause of survival and subsequent longevity among firms can be found in a set of specific capabilities, the characteristics of successful firms will, through natural selection, eventually come to typify the industry as a whole. By exploring and testing how one such characteristic, female employment, affected firm survival, we show how the longevity of the firms that employed more women shaped the whole industry by gradually transforming its workforce from a male- to female-dominated one over a fifty-year period.

In the remainder of this paper, the viability of the employment of women as a winning strategy is discussed and tested by statistical modelling of firm longevity using proportional hazard models. Our case is tobacco manufacturing in nineteenth-century Sweden. For this industry we have collected very rich data covering all firms that were active between 1863 and 1885. We follow firms until 1915, the year in which the industry nationalized.

Firm longevity and female employment as a competitive advantage

Firm longevity may be considered from two perspectives. Firstly, age has implications for performance and viability. With age, firms acquire knowledge and capital that make them less vulnerable. Jovanovic (1982) explains the dependence between age and survival as the result of a passive learning regime in which a firm's true capabilities, while initially unknown, are discovered over the time it spends in the market. Once the firm is better informed about its capabilities, decisions as to whether or not it should exit the market are taken. The prediction is that a few exits will take place soon after establishment, followed by early exits of less capable firms and the continuation of more competitive firms. Learning also lies at the heart of the model proposed by Ericson and Pakes (1995) in which firms actively learn by doing and invest accordingly in order to remain viable. This results in high failure rates during the initial learning phase with improved survival prospects in later periods as a result of resource accumulation. Secondly, firm

longevity is in itself an outcome. As argued by Friedman (1953) survival and longevity in a competitive market setting results from a process favouring efficient and profitable firms. Being efficient and profitable does in this case not necessarily imply that firms are profit maximizing (Alchian 1950). Indeed, not even the most competitive firms are required to be profit maximizers. Rather, what matters for survival and longevity is a firm's efficiency and profitability relative to its competitors.

For most firms, wages constitute a major share of costs, making hiring and retaining staff important managerial decisions (Wright, McMahan and McWilliams 1994; Becker and Gerhart 1996; cf. Gospel 2007). When choosing whom to hire, employers care primarily about the productivity of workers and their associated wages, and seek employees who offer the best combinations of productivity and costs. If men's higher wages relative to women's are not entirely motivated by productivity differences, men's labour is more expensive than women's. Hence, if women can be employed doing the same job as men but at a lower cost, then hiring women will be an efficient strategy and constitute a competitive advantage for firms adopting this strategy. But given that one labor input is relatively less expensive than another, why do not all firms choose to hire the cheaper input?

One answer to this question is provided by Becker's (1957) seminal model in which an employer who refuses to hire an employee with a productivity that equals or exceeds his/her cost is presumed to reveal discriminatory preferences (own, those of workers and/or customers). Discriminating firms will therefore, all else held constant, face higher unit costs of production than their competitors and will be likely to find their long-term position in the market untenable and are therefore expected to exit the industry.

Throughout history women have always earned less than men. To what extent the difference between men's and women's earnings can be explained by productivity differences or discrimination is highly dependent on period and context. During the nineteenth century, earnings were often gender-neutral since work was simple, individual productivity easily measured, and piece rates were widely used (Cox and Nye 1989; Goldin 1990; Burnette 2008). Workers in the nineteenth-century Swedish tobacco industry were paid either by the piece or hour. While piece rates were gender-neutral, hourly wages were not, providing an exploitable opportunity for firms to employ women instead of men and cut costs (Stanfors et al. 2013).

A large body of anecdotal evidence supports the idea that nineteenth-century firms sought to employ women in place of men to increase competitiveness. In the late nineteenth-century English hosiery industry it was seen to benefit employers' if they hired women in place of men as long as women could be paid less for the same amount and quality of work, and this resulted in employers attempting to hire women workers in place of men (Rose 1987: 170-171). The process of substituting women for men was, however, not limited to hosiery factories, but common in several manufacturing industries. In both umbrella making and the manufacturing of tins for preserved food, lower-paid women were replacing higher-paid men (Webb 1891: 647-648). In Glasgow in 1833, the managers of a cotton mill decided to hire women mule spinners "in the expectation that they [the company] would be able to reduce their wages lower than the rates paid to male spinners" (British Parliamentary Papers, 1833, *The First Report of the Central Board*, (XX), 84-85, quoted in Valverde 1988: 623). Moreover, Anthony Austin, reporting to the Handloom Weavers Commission, concluded that in South West England "women are employed, who will readily undertake it, at a lower wage than men receive [...]. By this process (unless the men consent to take the lower rate of wage) the whole of the weaving is gradually put into the hands of women [...] and the men are compelled to seek other work" (British Parliamentary Papers, 1840, *Handloom Weavers Commission* (XXIII) 282, quoted in Burnette 1997). Historical accounts clearly show that nineteenth-century firms were actively adapting their workforce by hiring women instead of men in jobs where both genders were equally useful and productive, as a way to reduce costs and thereby increase competiveness.

In a modern setting, the employment of women has been shown to affect both the profitability and survival of firms. Hellerstein, Neumark and Troske (2002) find that firms which employ a high share of women, and have some market power, are more profitable. For start-up firms, which by nature are predominantly small and inexperienced, and therefore more sensitive to managerial decisions, personnel policies are especially critical. Firms that choose to hire more women in the start-up phase are subsequently rewarded by a lower risk of failure (Weber and Zulehner 2010). First hires are likely to be responsible for important management decisions during start-up and in later periods, thus highlighting the role of gender diversity among firm management (Weber and Zulehner 2009). The effect of the gender composition of the workforce is, however, complicated by crowding and

segregation. Highly segregated firms (measured by having either a strongly male-dominated or female-dominated workforce) perform significantly worse compared to firms with a more gender-balanced workforce (Persson and Sjögren Lindquist 2010).

The incentives for employing women and the impacts thereof are, however, not independent of the market but affected by the level of competition. This is exemplified by Ashenfelter and Hannan (1986) who show that when markets are more competitive, firms are forced to act rationally and change their hiring practices accordingly (that is, in a less discriminating manner). The impetus to hire more women may also come from within the firm as the result of a change of management. By modeling changes in workforce composition following takeovers, Heyman, Svaleryd and Vlachos (2013) find that firms that operate in weakly competitive markets (in which the pressure to minimize costs was less prior to the takeover) increase their share of female employees under new ownership.

While the economic decision made by a firm to employ women instead of men is based on preferences, productivity and wages, this choice is affected by conditions which can either accelerate or constrain the process. An employer who chooses to employ women in order to cut costs feels pressure , on the one hand, from the level of competition in the market while being simultaneously forced, on the other hand, to contend with norms and practices governing the recruitment of labour imposed by society at large (Stinchcombe 1965). In this context female employment may be considered one form of cost-cutting process innovation (cf. Klepper 1996; Cefis and Marsili 2005). That improvements in processes, such as the employment of a particular group of workers, would prove decisive is very plausible given that the firms in our study were part of a well-established and mature industry experiencing little technological progress, few product innovations, competition in labour and product markets, and a work process allowing both men and women to perform most jobs.

The Swedish tobacco industry, 1863–1915

Tobacco manufacturing in nineteenth-century Sweden encompassed a simple non-mechanized work process organized in both small workshops and large factories housing hundreds of workers. Even by 1897, the manufacturing

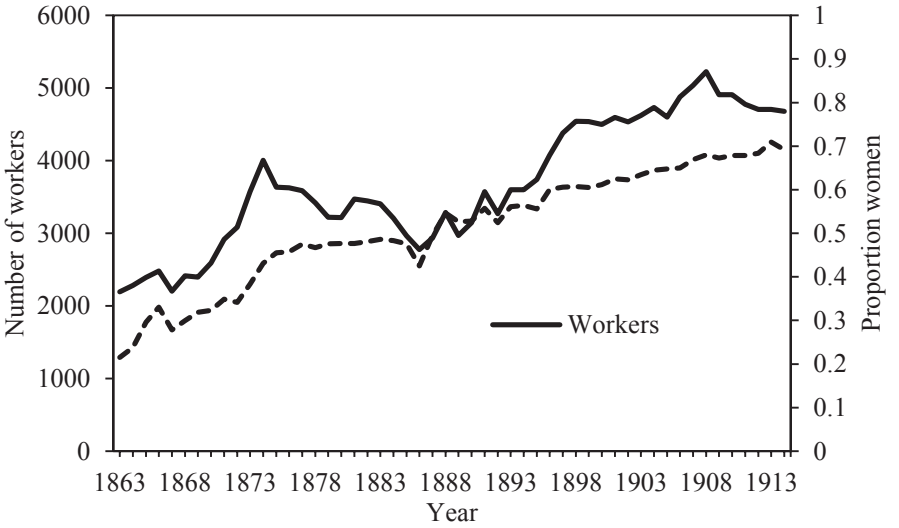
process was considered to be identical to the crafts industry despite being factory-based. In this regard the industry shared similarities with textile manufacturing, which also featured elements of both old-style artisanal and modern factory production (Gospel 2007). At the beginning of the nineteenth century snuff, chewing tobacco and smoking tobacco dominated output, while cigars became increasingly popular after 1850 (Kommerskollegii 1863–1910). Cigars were rolled by hand (sometimes using simple wooden implements) while snuff, chewing tobacco and smoking tobacco were shredded using mills that were often man-powered, but in some cases powered by water, animals or steam. Although the first cigarette machine was patented in the US in 1881 (Bonsack 1881), it was not until after the industry's nationalization in 1915 that mechanization led to a rationalization and transformation of the Swedish tobacco industry.[3] Since mechanization and returns to scale were not decisive factors in production, entry into the industry did not require large initial capital outlays on specialized machinery, making sunk costs very low.

In 1863, the year in which the Swedish Board of Commerce first recorded separately the employment of men and women in manufacturing, only 21 percent of the employees in the tobacco industry were women. Half a century later, on the eve of the industry's eventual nationalization in 1915, women had come to dominate the workforce (see Figure 1). While the feminization that took place was in a fairly persistent, industry growth was more erratic; the industry experienced periods of expansion in the late 1860s and early 1870s, and from the mid-1880s until the years prior to nationalization.

While the overall output of the Swedish tobacco industry increased during the second half of the nineteenth century, the number of active firms remained around 100 in each year throughout the period, in each year, new entrants replaced exiting firms. The industry's rates of entry and exit (on average 8 and 7.5 percent, respectively) are in line with modern rates which typically range between 5 and 15 percent (Dunne, Roberts and Samuelson 1988, 1989; Caves, 1998; Bartelsman, Scarpetta and Schivardi 2005; Lotti 2007). The industry was thus not subject to a "shake-out" followed by

[3] Thus, cigarettes remained a marginal product until after nationalization in 1915, which marks the end of our study.

Figure 1. Industry employment and share of women in the Swedish tobacco industry, 1863–1913



Sources: Kommerskollegi, *Bidrag till Sveriges Officiella Statistik: Fabriker och handtverk* (Stockholm, 1863–1912), and Kommerskollegium, *Industri* (Stockholm, 1913–1914).

a concentration of production in a handful of firms, as happened in the American car industry prior to the Second World War (Geroski and Mazzucato 2001; Klepper 2002), and nor did the growth in output lead to an increase in the number of firms, as happened in the case of the nineteenth-century American shipbuilding industry (Thompson 2005).

Turning to labour market conditions, the late nineteenth-century labour market for tobacco workers is a good example of the liberal labour market regime which dominated the period after 1850. It is best described as lightly regulated by both modern and historical standards. Apart from restrictions on the use of child labour, firms were free to hire as they pleased, and they were free to hire and fire as they pleased; no minimum wage existed, and there were no regulations prohibiting the employment of women and no requirements regarding formal qualifications or guild membership. Swedish tobacco workers were unionized in 1889 but, unlike its American equivalent, the Swedish Tobacco Workers' Union admitted women as members and was not opposed to their employment (Lindbom and Kuhm 1940; Cooper 1987).

The market for tobacco products was characterized by a competitive environment driven by a large number of firms producing relatively cheap,

homogenous, and easily transportable products that were predominantly manufactured from the same imported tobacco. Although the market was mainly domestic, firms, in addition to competition from rivals in the domestic market, and despite the existence of tariffs on manufactured tobacco products, also had to contend with competition from imported tobacco goods (Kommerskollegii 1863–1885).

The beginning of our period of study coincides with the introduction of the Freedom of Trade Act in Sweden. Enacted in 1864, this reform effectively lowered the legal barriers to entry by removing several previous requirements. Prior to 1864, in order to start a business, the principal had to have prior experience of the trade in question, be literate, considered a person "of good standing", and, if business were to be conducted in a city or within a five kilometers thereof, burghership was required. After its enactment, citizens were free to start an enterprise as long as they were of full legal age and registered the business with the relevant authority (Rabenius 1888). Our study ends with the nationalization of the tobacco industry in 1915. The decision by the state to nationalize the tobacco industry was driven by a need to increase government revenue in order to finance an expanding public sector. Tobacco tax was seen as one such potential source of revenue. Taxing an industry, as fragmented as the tobacco industry was, however, deemed unfeasible unless it was consolidated through nationalization. Moreover, the threat of the establishment of British American Tobacco – a company which had become a significant actor in neighbouring countries – in Sweden served as a further impetus to nationalize the tobacco industry (Karlsson 2008: 53-55).

Given the purpose of this paper, the characteristics of the Swedish tobacco industry at the end of the nineteenth century serve us well. A large number of firms were active in every year throughout the period and firm turnover was high enough to provide sufficient variation in the data for modelling survival. Heterogeneity stemming from differences in the technology used may be largely extent be discounted since no important innovations were introduced or adapted during the period studied. Firms produced similar products with no apparent economies of scale in a market where barriers to entry and exit were low and government intervention was non-existent. Raw tobacco leaves were mostly imported and there was no vertical integration into tobacco farming, making the cost of raw material for firms almost identical for all the firms. Finally, since consumers did not interact directly with producers, nobody knew the gender of the individual

who had manufactured the product, and thus there was no need to consider consumer preferences when hiring workers, granting firms more discretion when deciding whom to employ. In sum, our case discounts a number of factors associated with the employment of women that may otherwise complicate the analysis through unobserved heterogeneity.

Econometric analysis

Data and variables

Our analysis is based on data of all firms registered as active in the Swedish tobacco industry between 1863 and 1885. We have constructed the panel by collecting information from annual ledgers compiled by the *Swedish Board of Commerce*[4] as an intermittent step in the production of aggregated official statistics covering factories and manufacturers (Kommerskollegium kammarkontoret 1863–1885). Firms are followed yearly between 1863 and 1885. Identification and linking of firms between years is based on the name of the firm. The ledgers helpfully recognize firms that change names by including both the old and new name of the firm in their listings, thus enabling us to accurately and consistently track firms over time. The detailed source material includes annual firm-level information on the firm's workforce (number of men, women, girls, boys and foremen employed), the legal form of the business, the types of machinery used and its power output measured in horse powers, the amount of tobacco produced by product category (cigars, cigarettes, snuff, chewing tobacco and smoking tobacco by weight and value) and the firm's geographical location. The registered firms range from small owner-operated firms to large factories employing hundreds of workers. The source material also includes firms that were only registered with the local authorities but never in operation. All the firms were single plant units and no firms owned subsidiaries. Because some firms enter before the first year of observation while others exit after the last year of observation, the year of establishment and demise for firms established prior

[4] Authors translation, original name: *Kommerskollegii.*

to 1863 or exiting the industry after 1885 was obtained from a comprehensive directory over the Swedish tobacco industry (Angelin 1950). To account for the Swedish industries competitiveness relative to the rest of the world, annual information on the volume of manufactured tobacco exports and imports is sourced from official trade statistics (Kommerskollegii 1863–1885). These measures are yearly industry indicators and, hence, identical for all firms in a given year.

In total, 268 firms were registered between 1863 and 1885. A majority of the firms were small and only employed a handful of workers. Such firms would often cease for periods before restarting production, resulting in intermittent appearances in the register. Furthermore, since most small firms were essentially workshops rather than factories, operations tended to cease once the proprietor retired or died. For modelling purposes, we only include firms that reached a minimum size of ten employees at any time in our dataset. This restriction and the exclusion of observations with missing information for any variable of relevance, yields a sample of 109 firms. Although restricting the sample by more than half may sound severe, in practice it is not so: the eliminated firms only contributed a small share to total industry output, producing less than eight percent of total output during the period of investigation. Descriptive statistics of both the full and restricted samples are presented in Table 1. The restrictions result in the average firm in the final sample being significantly larger in terms of workers employed, machinery utilized and output produced when compared to the total population of firms. Moreover, the firms in the restricted sample were more likely to operate as a limited liability company and employ a foreman but were only slightly more feminized than the total population of firms.

We assess firm longevity by considering firms that survive compared to those that exit the industry, controlling for a number of factors. A firm could cease to exist in a number of ways. We identified different reasons by supplementing the register-based data with information on the circumstances of the failure of each firm. Reasons included their being the subject of an acquisition, voluntary liquidation or bankruptcy, and finally - specific to our case - being nationalized as part of the Swedish Tobacco Monopoly in 1915. Because the source material uses the terms liquidation and bankruptcy interchangeably, we are not able to differentiate between these two types of failure. Although the underlying causes of takeovers are debatable, voluntary liquidation and bankruptcy are both in a similar way related to the viability of the firm, which is what we are ultimately interested in. We therefore treat

Table 1. Descriptive statistics of firms explored in the empirical analysis

	All observations (268 firms)					Restricted sample (109 firms)				
	Mean	Std. Dev.	Min.	Max.	Obs.	Mean	Std. Dev.	Min.	Max.	Obs.
Workforce										
Women	8.97	26.90	0.00	299.00	2,051	15.43	34.24	0.00	299.00	1,168
Employees	30.96	58.31	0.00	401.00	2,051	51.93	70.25	1.00	401.00	1,168
Output ('000 SEK)										
Cigars	28.45	66.52	0.00	444.88	2,051	48.05	82.50	0.00	444.88	1,168
Cigarettes	0.05	0.85	0.00	33.02	2,051	0.08	1.13	0.00	33.02	1,168
Snuff	33.36	88.71	0.00	970.09	2,051	50.18	111.27	0.00	970.09	1,168
Tobacco	23.70	53.83	0.00	542.78	2,051	40.84	66.30	0.00	542.78	1,168
Management										
Foreman	0.35	0.48	0.00	1.00	2,051	0.59	0.49	0.00	1.00	1,168
Limited liability	0.26	0.44	0.00	1.00	2,051	0.37	0.48	0.00	1.00	1,168
Machinery										
Horse powers	2.34	6.84	0.00	100.00	2,048	3.78	8.64	0.00	100.00	1,168

Sources: Kommerskollegium kammarkontoret (1863-1885), *Årsberättelser fabriker serie 4*, Da4, vol S05823–44, vol S02054–61 and vol S05806–19, Kommerskollegiets arkiv, National Archives (*Riksarkivet*), Stockholm

liquidation and bankruptcy identically. Firms in our sample may thus have ceased to exist as a result of one of three possible outcomes:

1. Firm taken over by state as part of the nationalization of the tobacco industry in 1915.
2. Firm acquired by another firm.[5]
3. Firm exiting from industry because of voluntary liquidation or

Our main explanatory variable of interest is the firm's employment of women. The most simple and straightforward measure of firm feminization is the share of women in the total number of employees in each firm. Using raw ratios does, however, raise concerns about spurious relationships. As shown in Figure 1, the average share of women employed by firms increased almost monotonically from 1863 until the nationalization of the industry in 1915. Because the firms and the Swedish tobacco industry as a whole became gradually more feminized during the period covered, raw ratios are highly time-dependent. If other variables that affected the survival of firms (such as disposable income, female labor force participation or tobacco consumption)

[5] One single firm, which entered the industry in 1883, was recorded by Angelin (1950) to have merged with another firm (that entered after 1885). We treat this firm as having been the subject of an acquisition. Whether we treat the merger as an acquisition, a liquidation or bankruptcy, or consider the new firm that resulted from the merger as a continuation of the original firm founded in 1893, makes no difference to the results.

increased accordingly over time any correlation between the firms' employment of women and survival may be spurious. To address these issues we construct a measure of female employment that is adjusted by the local level of feminization in each year[6]:

$$\dot{r}_{it} = \frac{r_{it} - \bar{r}_{it} + 1}{2} \quad (1)$$

where r_{it} denotes the share of women of firm *i* at time *t*, and $\bar{r}_{it}$ is the average share of female employees in tobacco manufacturing within a 50 km radius off each firm. A further concern is functional form, because the impact of employing women may be non-linear, something that has been shown both theoretically, when assuming friction in the labor market (Rosén 2003), and empirically (Weber and Zulehner 2009: Figure 4). Here, we allow for the impact on firm survival of employing women to vary in the distribution by coding the adjusted measure of feminization into four ordinal groups of firms delimited by quartiles (respectively labelled least feminized, moderately feminized, highly feminized and most feminized).The least feminized group of firms, with a level of feminization falling below the first quartile, is used as the reference category in all models.

Naturally, the success or failure of a firm is not solely determined by its hiring strategies pertaining to whom it chooses to employ but also contingent on a number of factors internal and external to the firm. Little evidence exists about general patterns of firm survival and its determinants in historical settings. This lack of knowledge may be attributed to a scarcity of historical data on industries in their entirety, that is populations of firms. Fewbut nonetheless interesting exceptions include Thompson's (2005) study of the survival of firms in the shipbuilding industry from 1825 to 1914, Box's (2008) exploration of the impact of the economic environment on survival among seven birth cohorts of firms during the first half of the twentieth century, and Mackie's (2001) study of the survival prospects of family firms in Kirckaldy, Scotland, 1870–1970.

As a consequence of better data availability, including data from modern business registers, coupled with the applicability of increasingly

[6] Our approach is thus similar to that advocated by Weber and Zulehner (2009) but instead of adjusting by an industry average, we adjust by the local level of feminization.

sophisticated econometrics, the literature is more extensive on firm survival, and its determinants, in contemporary settings. A number of stylized correlations between survival and factors internal and external to the firm may be identified. The most generally observed determinants of firm survival are size and age, exemplified by high rates of failure among recently established and small firms, followed by declining rates in later periods (see e.g., Evans 1987; Dunne, Roberts and Samuelson 1988, 1989; Mata and Portugal 1994). We control for firm size by including a variable indicating the number of workers employed by the firm along with a variable indicating output measured in 1,000 SEK (in fixed prices adjusted for inflation). Regarding output, a distinction is made between cigars, snuff, and cigarettes. Because of inconsistencies in the source material regarding the definition of chewing tobacco and smoking tobacco, the two categories have been merged into one (labelled 'other tobacco'). Based on the commonly observed relationship between firm size, age and survival, the number of employees is expected to have a negative effect on the hazard of failure. Although the general trend in the impact of output is predicted to be positive, the impact of individual categories of output is *a priori* more difficult to hypothesize about because some branches of tobacco manufacturing may have been more competitive than others.

The legal structure of the firm is related to its age and size. Firms typically start out as sole proprietors or partnerships before changing their legal status when reaching a more mature stage. Whether firms that operate under limited liability are more vulnerable than those under unlimited liability (or vice versa) seems, however, not unidirectional but highly dependent on context (Brüderl, Preisendörfer and Ziegler 1992; Harhoff, Stahl and Woywode 1998; Esteve-Peréz and Mañez-Castillejo 2008). To account for differences in the legal form of the firm, a dummy variable is included that indicate whether the firm was registered as a limited liability company (the reference category being sole proprietor or partnership). Possible differences in management practices are controlled for by the inclusion of a dummy variable denoting whether the firm employed a foreman or not. We interpret the reference category (no foreman employed) as an indication of the firm being managed by its owner. Although mechanization was limited, animals, steam engines or water were used in some cases to power mills for shredding tobacco. To alleviate concerns about mechanization driving the results, we include a variable indicating the number of horse powers employed by the factory.

Besides internal factors, the survival of firms is affected by environmental variables. Industry-specific characteristics such as competition and economies of scale together with the geographic location of the firm all result in differences in survival rates between firms (Manjón-Antolín and Arauzo-Carod 2008: 18-20). The risk of failure is moreover correlated with the business cycle, being lower in times of industry expansion and elevated during downturns (Audretsch and Mahmood 1995; Geroski 1995; Caves 1998; Disney, Kaskel and Heden 2003). We consider these factors by including a number of control variables. The impact of general and location-specific environmental conditions is incorporated through the inclusion of the yearly growth rate in tobacco industry employment along with an identifier of the province in which the firm was located. Finally, the impact of competition from foreign firms is measured by a variable indicating imports and exports of manufactured tobacco products as a percentage of total domestic industry output.

Empirical model and results

To test whether the employment of women affected firm longevity, and, to account for the interdependence between the firm's age and its survival discussed at the beginning of this paper, duration models are employed. We begin by estimating standard Cox (1972)[7] proportional hazards models, treating both being subject of an acquisition or the industry's eventual nationalization as censoring events:

$$h(t|x) = h_0(t)*exp(\boldsymbol{\beta}\mathbf{X}) \qquad (2)$$

[7] The model requires two key assumptions: non-informative censoring and proportional hazards. It is plausible that being the subject of an acquisition does not constitute a case of non-informative censoring. This concern is explicitly dealt with when we proceed to modeling competing risks. The proportional hazards assumption requires that all covariates are time-independent, that is, that the effect (β) of a covariate is constant throughout a firm's life. In contrast to the non-informative censoring assumption, the fulfillment of the proportional hazards assumption may be evaluated empirically. Tests of the proportional hazards assumption based on Schoenfeld residuals confirm that the assumption of proportionality is not violated.

The first term in Cox' proportional hazards model, $h_0(t)$, denotes the baseline hazard of failure and corresponds to the probability that a firm whose explanatory covariates are all equal to zero will fail. The estimation of the model necessitates leaving the baseline hazard un-estimated. This, in turn, permits $h_0(t)$, the firms' baseline hazard of failure to increase, decrease, remain unchanged or vary with time. The model thereby allow for any theoretically motivated survival pattern with regards to firm age. We thus remain agnostic about the specific relationship between age and survival and instead focus on the estimated parameters ($\boldsymbol{\beta}$) for covariates ($\mathbf{X}$) on survival. For ease of interpretation, we report all parameters as hazard ratios ($exp(\boldsymbol{\beta})$). A hazard ratio above one means an increase in the risk of failure, while a hazard ratio less than one denotes a decrease. For example, an estimated hazard ratio of 0.75 for a dummy covariate means that a firm fulfilling the dummy criterion at any point in time is 25 percent less likely to fail compared to a firm in the reference category. If the covariate in question is continuous, each unit change in the covariate will lead to reduction in the risk of failure by the corresponding hazard ratio. A five unit increase in a covariate with an estimated hazard ratio of 1.10 would thus imply a change in increased risk of failure by 61 percent ($= 1.10^8$).

The results from model estimations are presented in Table 2. We begin by estimating a very parsimonious model (I) with firm-level feminization as the only explanatory variable. Thereafter, we add control variables in a step-wise manner. With no controls, a the substantial reduction in the hazard of failure for more feminized firms is notable. When compared to the least feminized group of firms (i.e., the reference category), the risk of failure is reduced by 52 percent for moderately feminized firms; 71 percent for highly feminized firms; and 47 percent for the most feminized group of firms. It is likely that this effect may be attributable in part to excluded variables covarying with the employment of women. Hence, in Model II, we account for firm size by including the number of employees and output by value as covariates. The estimated hazards associated with these covariates are in line

[8] A sub-hazard differs from the regular hazard in the Cox model by being the instantaneous probability of failure from a specific cause at a point in time given no failure from such cause prior or having failed but from an alternative cause prior. The Fine and Gray (1999) approach to estimating sub-hazards accounts for competing risks by keeping subjects that fail from a competing cause in the risk population using a time dependent weighing function.

Table 2. Estimations of hazard ratios of failure from Cox's proportional hazards models

	I	II	III	IV	V
Level of feminization					
Moderate	0.473 **	0.672	0.450 **	0.420 **	0.342 ***
	(0.163)	(0.224)	(0.181)	(0.169)	(0.141)
High	0.291 ***	0.400 ***	0.264 ***	0.254 ***	0.223 ***
	(0.112)	(0.131)	(0.119)	(0.113)	(0.101)
Most feminized	0.532 **	0.610	0.458 **	0.430 *	0.378 **
	(0.169)	(0.210)	(0.177)	(0.186)	(0.170)
Workforce					
Employees		0.973 ***	0.973 **	0.960 ***	0.964 ***
		(0.010)	(0.011)	(0.012)	(0.014)
Output ('000 SEK)					
Cigars		1.003	1.003	1.002	1
		(0.004)	(0.005)	(0.006)	(0.008)
Cigarettes		1.036	1.06	1.034	1.021
		(0.048)	(0.052)	(0.048)	(0.047)
Snuff		0.993	0.986	0.986 *	0.988
		(0.005)	(0.010)	(0.008)	(0.008)
Tobacco		0.980 **	0.975 **	0.97 **	0.971 **
		(0.008)	(0.012)	(0.013)	(0.014)
Management					
Foreman				3.392 ***	3.098 **
				(1.328)	(1.364)
Limited liability				1.203	1.051
				(0.392)	(0.376)
Machinery					
Horse powers				0.999	0.992
				(0.008)	(0.008)
Macro variables					
Industry growth					0.961 **
					(0.016)
Exports					0.745
					(0.164)
Imports					1.032
					(0.020)
Province dummy	No	No	Yes	Yes	Yes
Firms	109	109	109	109	109
Firm years	1,168	1,168	1,168	1,168	1,168
No of failures	65	65	65	65	65
No of censored firms	44	44	44	44	44
$\chi 2$	11.72 ***	45.00 ***	140.8 ***	2064.0 ***	185.1 ***

Notes: Robust standard errors clustered by firm in parenthesis. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

Sources: Kommerskollegium kammarkontoret (1863-1885), *Årsberättelser fabriker serie 4*, Da4, vol S05823–44, vol S02054–61 and vol S05806–19, Kommerskollegiets arkiv, National Archives (*Riksarkivet*), Stockholm, and Kommerskollegi, *Bidrag till Sveriges Officiella Statistik F, Utrikes handel och sjöfart* (Stockholm, 1863–1885)

with expectations that both more employees and higher firm output lead to a reduction in the risk of failure. After including controls for firm size, the impact of employing women is somewhat weakened, indicating that more feminized firms also tended to be bigger. In Model III we add geographic controls in the form of province dummies to account for differences in firm survival based on location. In Model IV, we proceed by controlling for whether the firm was managed by a foreman, and registered as a limited liability company, and for the extent to which machinery was used in the production process. We find no significant relationship the firm's legal form and survival. Furthermore, machinery had no impact on firm survival. This result is in line with the characterization of the Swedish tobacco industry as being essentially artisanal in form and organization. Being managed by a foreman implies a highly elevated hazard of exit. Owner-managed firms were less likely to fail than those employing a foreman. The final model (V) is the most extensive model, incorporating all control variables of previous models together with macro-level variables of the firm, potentially affecting its survival. The estimated effect of macro variables on firms are as expected: the hazard of failure is positively related to the business cycle (i.e., industry growth) and, although not statistically significant, imports are negatively associated with survival while exports are positively associated with survival.

Turning to the main variable of interest, feminization, we find that firms whose share of women was above the bottom quartile faced a significantly lower risk of failure than those whose share fell below that level. This is the case in all estimated models. In the most extensive model (V), relatively more feminized firms' hazards of exit are between 64 and 81 percent lower compared to the least feminized group of firms. In order to get a sense of the relative magnitude of these numbers, a comparison with the estimated hazards of other covariates is illustrative. The difference in the hazard of failure between the least feminized group of firms and the moderately feminized group (a difference of 64 percent) is equivalent to a decrease in hazard of failure associated with an increase in firm size by about 30 employees. An equivalent decrease in hazard associated with industry growth implies a year-by-year expansion of the tobacco industry by approximately 25 percent. Given that the average firm in the sample employed 52 workers (see Table 1) and that a yearly industry growth rate of 25 percent must be considered very favourable, it is clear that the importance of employing women relative to other factors was not only statistically but also economically significant. Interestingly, we find that the impact of

feminization is non-linear: while employing more women than the least feminized category of firms resulted in a considerably lower risk of failure, no discernible further gain can be identified for firms feminizing beyond this threshold in terms of survival. The dynamics of the selection process is thus not to favor the forerunners of feminization in terms of survival but rather to eliminate extreme laggards among of firms.

A shortcoming of the standard Cox model is its failure to accommodate different forms of failure. In order to account for the different ways in which a firm in our sample could exit the industry and the effect this could have on the results, we extend the analysis by treating each form of exit as a distinct event. In our case the additional event in question is the firm being subject to an acquisition, constituting a competing risk in that it prevents the occurrence of eventual failure through bankruptcy or liquidation. Employing a competing risks model acknowledges that firms may cease to exist from causes other than liquidation or bankruptcy. We estimate the following competing risks model suggested by Fine and Gray (1999):

$$h_j(t|x) = h_{j,0}(t)*exp(\boldsymbol{\beta}\mathbf{X}) \quad (3)$$

where $h_{j,0}(t)$, the sub-hazard of experiencing failure from one of j events in year t is again left unspecified (in our case $j=1$: exit due to liquidation or bankruptcy and $j=2$: exit due to acquisition). Alternative types of failure are thus treated not in terms of censoring but as separate events (acquisitions) preventing the outcome of interest (liquidation or bankruptcy) from occurring. As in the Cox model, our interest lay in the parameters ($\boldsymbol{\beta}$) which are reported as sub-hazard ratios ($exp(\boldsymbol{\beta})$).

The results from the competing risk analysis are presented in Table 3. Of the 109 firms in our sample, 10 were the subject of an acquisition, 65 failed and 34 were censored as a result of the industry's nationalization. The step-wise inclusion of control variables exactly mirrors what was presented in Table 2. None of our results are substantially altered after taking into account the multiple ways in which a firm may fail. When compared to the results estimated using the standard Cox model, the impact of the employment of women is slightly stronger and more precisely specified. In the most extensive model (V), the hazard of bankruptcy or liquidation among the most feminized firms is between 66 and 81 percent lower than the reference category.The results of the estimated models are highly robust.

Table 3. Estimations of sub-hazard ratios of failure from proportional hazards model with competing risks

	I	II	III	IV	V
Level of feminization					
Moderate	0.462 **	0.667	0.441 **	0.428 **	0.338 ***
	(0.139)	(0.221)	(0.177)	(0.171)	(0.138)
High	0.258 ***	0.348 ***	0.219 ***	0.219 ***	0.191 ***
	(0.089)	(0.116)	(0.099)	(0.099)	(0.087)
Most feminized	0.472 **	0.535 *	0.394 **	0.377 **	0.329 **
	(0.143)	(0.189)	(0.154)	(0.165)	(0.152)
Workforce					
Employees		0.969 ***	0.970 **	0.958 ***	0.960 ***
		(0.011)	(0.012)	(0.012)	(0.014)
Output ('000 SEK)					
Cigars		1.005	1.004	1.004	1.002
		(0.004)	(0.005)	(0.006)	(0.008)
Cigarettes		1.038	1.061	1.036	1.025
		(0.049)	(0.052)	(0.048)	(0.047)
Snuff		0.995	0.988	0.988	0.990
		(0.004)	(0.009)	(0.008)	(0.007)
Tobacco		0.983 **	0.978 *	0.975 *	0.976 *
		(0.008)	(0.012)	(0.014)	(0.014)
Management					
Foreman				3.228 ***	2.926 **
				(1.249)	(1.288)
Limited liability				1.084	0.932
				(0.370)	(0.352)
Machinery					
Horse powers				1.004	0.995
				(0.008)	(0.008)
Macro variables					
Industry growth					0.964 **
					(0.016)
Exports					0.756
					(0.166)
Imports					1.039 *
					(0.021)
Location					
Province dummy	No	No	Yes	Yes	Yes
Firms	109	109	109	109	109
Firm years	1,168	1,168	1,168	1,168	1,168
No of failures	65	65	65	65	65
No of competing	10	10	10	10	10
No of censored	34	34	34	34	34
$\chi 2$	18.97 ***	44.93 ***	523.56 ***	1094.49 ***	761.00 ***

Notes: Robust standard errors clustered by firm in parenthesis. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.
Sources: See Table 2

We undertake a number of sensitivity tests. When replacing the measures of the Swedish tobacco industry's competitiveness relative to the rest of the world (i.e., annual information about the volume of manufactured tobacco exports and imports) with year dummies, our results hold up to the extent that the impact of the employment of women actually gets stronger. Moreover, the robustness of our results is tested by a re-estimation of the most extensive competing risks model (model V in Table 3) after imposing a number of restrictions. The results of various robustness checks are presented in Table 4. We begin by investigating whether the results are driven by segregated firms by excluding firms that exclusively employed either women or men (see models I-III in Table 4). A further concern is that the estimated results may be driven by well-established firms that entered the industry before our first year of observation. To ensure that this is not the case we proceed by re-estimating the model only for those firms that entered during our observation period, thereby excluding all firms that entered the industry before 1863. The results are presented in column IV. Finally, to test to what extent the results are sensitive to following firms up to 30 years after the last year of observation, 1890 is used as an alternative year of censoring (in model V). It should be noted that, after applying the above restrictions, our results change little. Although some of the coefficients are estimated with somewhat less precision after applying the restrictions, the main results of the analysis remain consistent: the impact of female employment (i.e., firm-level feminization) on survival remains large, non-linear, and statistically significant.

Firm longevity and industry feminization

Our study has implications for how we regard firm longevity and industry feminization. Feminization of the workforce has alongside industrialization profoundly transformed the economy. However, the employment of women has not followed a uniform pattern historically. Although work in many industries was put increasingly into the hands of women during American and European industrialization, some occupations remained dominated by men while others experienced periods of de-feminization (e.g., Abbott 1910; Bradley 1989). A prominent explanation for why particular firms and industries feminized during the nineteenth century rests upon the assumption

Table 4. Sensitivity tests of estimations of sub-hazard ratios of failure from proportional hazards model with competing risks

	I	II	III	IV	V
Level of feminization					
Moderate	0.513	0.314 ***	0.472	0.303 **	0.466 *
	(0.242)	(0.135)	(0.240)	(0.150)	(0.209)
High	0.217 **	0.184 ***	0.215 **	0.103 ***	0.152 ***
	(0.142)	(0.087)	(0.151)	(0.061)	(0.091)
Most feminized	0.377 *	0.295 **	0.338 *	0.336 **	0.394 *
	(0.191)	(0.145)	(0.189)	(0.172)	(0.194)
Workforce					
Employees	0.946 ***	0.962 **	0.947 ***	0.956 ***	0.945 **
	(0.013)	(0.015)	(0.013)	(0.014)	(0.022)
Output ('000 SEK)					
Cigars	1.004	1.001	1.004	1.003	1.009
	(0.006)	(0.008)	(0.006)	(0.006)	(0.012)
Cigarettes	1.03	1.022	1.028	1.029	1.034
	(0.043)	(0.045)	(0.041)	(0.048)	(0.047)
Snuff	0.992	0.989	0.992	0.99	0.99
	(0.009)	(0.007)	(0.009)	(0.019)	(0.013)
Tobacco	0.987	0.976 *	0.986	0.981	0.976
	(0.013)	(0.014)	(0.013)	(0.023)	(0.025)
Management					
Foreman	2.772 *	3.58 ***	3.281 *	2.442 *	2.444 *
	(1.465)	(1.683)	(2.018)	(1.168)	(1.198)
Limited liability	1.021	0.913	0.963	0.996	1.197
	(0.432)	(0.369)	(0.448)	(0.478)	(0.486)
Machinery					
Horse powers	1	0.996	1.001	0.987	0.974
	(0.011)	(0.009)	(0.012)	(0.011)	(0.023)
Macro variables					
Industry growth	0.969 *	0.958 **	0.959 **	0.985	0.968 **
	(0.017)	(0.018)	(0.020)	(0.013)	(0.016)
Exports	0.874	0.764	0.873	0.878	0.731
	(0.207)	(0.174)	(0.207)	(0.210)	(0.167)

Imports	1.015	1.032	1.009	1.023	1.023
	(0.028)	(0.021)	(0.030)	(0.019)	(0.024)
Location					
Province dummy	Yes	Yes	Yes	Yes	Yes
Censoring year	1915	1915	1915	1915	1890
Sample restrictions					
Excluding all-male firms	Yes	No	Yes	No	No
Excluding all-female firms	No	Yes	Yes	No	No
Excluding pre-1863 entrants	No	No	No	Yes	No
Firms	95	108	94	65	109
Firm years	893	1,136	867	437	1,168
No of failures	50	61	47	48	53
No of competing	10	10	10	7	5
No of censored	35	37	37	10	51
$\chi 2$	1048.50 ***	1399.32 ***	1400.30 ***	504.65 ***	1003.57 ***

Notes: Robust standard errors clustered by firm in parenthesis. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.
Sources: See Table 2.

that female labour was complementary to technological change in the form of mechanization and the increasing division of labour and deskilling. Marx (1970 [1867]: 420) declared the onset of feminization of a workforce to begin when "machinery dispenses with muscular power, [at which point] it becomes a means of employing labourers of slight muscular strength, and those whose bodily development is incomplete, but whose limbs are all the more supple. The labor of women and children was, therefore, the first thing sought for by capitalists who used machinery". Goldin and Sokoloff (1982; 1984) provide a similar argument, but emphasize the role of increased demand for unskilled labour in the nineteenth-century manufacturing sector. Coupled with women's and children's inferior labour productivity and lower wages in the agricultural sector, feminization is attributed to the relative cheapness of women's and children's labour coupled with de-skilling and the increasing division of labour resulting from the transition from artisan shops to factory production (cf. Gospel 2007).

In this article, we do not directly dispute the reasons for feminization given by previous researchers. Instead, we argue that our results provide a basis for an alternative and complementary explanation for feminization based on competition and firm longevity. The extended longevity of more feminized firms coupled with the eventual feminization of the tobacco industry may be considered a case of *isomorphism,* a process that shapes a population to become increasingly compatible with its environment (Hawley 1968). Isomorphism may be described as a constraining process that forces entities in a population to tend towards a resemblance of each other in line with survival of the fittest. Within an industry, competition provides the constraining mechanism that induces change and eventual homogenization. By favouring competitive firms and causing unprofitable firms to exit, competitive firms will, through their extended longevity, increasingly dominate an industry (Hannan and Freeman 1977). It thus follows that an industry will with time increasingly come to display the attributes of its most competitive firms. Based on our analysis of female employment as being a competitive advantage, the difference in longevity between more and less feminized firms provides a previously neglected explanation for industry feminization. Our results show that the least feminized firms in the Swedish tobacco industry failed to a significantly greater degree than more feminized firms. In terms of survival it was highly detrimental not to employ women. Firms that did not employ women were continuously selected out of the population. No additional premium was, realized, however, by the

forerunners of feminization. The elimination of the least feminized firms with no apparent further gain realized by firms that were very highly feminized is consistent with the fairly slow and gradual replacement of men by women over the period depicted in Figure 1. While markets were clearly efficient enough to favour firms that feminized, the effect of the differences in survival between more and less feminized firms on the composition of the industry workforce was not radical. Instead, a persistent but slow movement towards equilibrium over a relatively long period of time can thus be observed.

Concluding discussion

In this article, we have explored whether the employment of women was a strategic decision among employers during industrialization. Our results are robust against the inclusion of a number of relevant control variables and hold both when employing a standard Cox model and when considering the role of competing risks. The results are not diminished by subjecting the models to several sensitivity tests. We focused on tobacco firms' employment decisions during the second half of the nineteenth century and how it affected the gender composition of the industry as a whole. This study demonstrates that during nineteenth-century industrialization firms which employed relatively more women than their competitors faced a considerably lower risk of failure. Since firms that were more feminized than their competitors were more likely to survive, the least feminized firms' long-term position in the market were untenable.

It is important to emphasize that the impact of employing women above men is large: firms that failed to feminize thus faced severe repercussions. Firstly, this tells us something about firm survival and nineteenth-century markets. There are numerous *a priori* reasons why nineteenth-century markets should be less efficient than what may be the case in a modern setting. While significant inefficiencies were likely to exist because of limited means of transportation and high costs of information, such obstacles to competition were remedied by liberal labor laws and low material and regulatory barriers to entry and exit. This contributed to make markets efficient enough to reward firms that acted in accordance with profit maximization and adapted to market competition. Secondly, because the

survival prospects of the least feminized firms were substantially poorer relative to more feminized firms, the incentives for capable firms to feminize were high enough to act upon. Industry feminization may thus be seen as a result of two related processes: the continuous elimination of the least feminized firms in the distribution which directly affected the share of men employed in the industry negatively. This implicit threat of failure, in turn, served to incentivize surviving firms to employ even more women, thereby affecting the average share of women employed in individual firms and the industry as a whole positively.

The period considered in this paper stands in sharp contrast to that which followed. After the nationalization of the Swedish tobacco industry in 1915, mechanization brought a second transformation of the industry's workforce driven by rationalization and personnel reductions, a common development mirrored by, for example, the American tobacco industry (Cooper 1987). The transformation of the Swedish tobacco industry prior to 1915 – from a male-dominated to a female-dominated workforce – was much more subtle by nature. Prevailing explanations for nineteenth-century feminization processes typically attribute the increasing employment of women to changes in production processes associated with the Industrial Revolution. In contrast to this, the present study emphasizes competition between firms as a powerful mechanism behind feminization. As exemplified by the Swedish tobacco industry, technological and organizational changes were not prerequisites for feminization. Instead, given sufficiently efficient markets, feminization resulted from a competitive process that unfolded during several decades, a process distinctly different from the industry-wide technologically induced transformation of an industry's workforce.

With regards to the literature reviewed, the evidence presented here corroborates both historical witness accounts of nineteenth-century hiring practices and contemporary studies of the impact of employing women in a modern setting. That the least feminized firms in our sample were not viable in the long run lends credibility to the cited claims made by nineteenth-century observers that women were hired in place of men in order to increase a firm's competitiveness. Moreover, the results are consistent with modern findings showing that more feminized firms generate higher profits and, by extension, are less likely to fail. Given the consistency of our results with modern quantitative studies and qualitative historical evidence, we have good reason to surmise that our conclusions have bearings on firm survival, firm longevity, and the feminization of labour in a wider historical setting.

Finally, from a methodological perspective this study demonstrates the value of considering complete populations by using individual firms as the unit of analysis when seeking to explain not only the inner workings of competition within industries but also changes to society at large. We therefore argue that studies attempting to explain overall patterns should, to a greater extent, explicitly consider how processes at the micro-level shape long-term aggregate trends.

References

Abbott, E. (1910). *Women in Industry: A study in American economic history*. D. Appleton and Company: New York.

Agarwal, R. and Gort, M. (2002). 'Firm and product life cycles and firm survival.

American Economic Review, 92: 184–190.

Alchian, A. (1950). Uncertainty, Evolution and Economic Theory. *Journal of Political Economy*, 58: 211–221.

Angelin, E. (1950). *Hantverkare och Fabriker inom Svensk Tobakshantering 1686–1915*. Nordisk Rotogravyr: Stockholm.

Ashenfelter, O. and Hannan, T. (1986). Sex Discrimination and Product Market Competition: The Case of the Banking Industry. *The Quarterly Journal of Economics,* 101: 149–174.

Audretsch, D. B. and T. Mahmood (1995). New firm survival: new results using a hazard function. *Review of Economics and Statistics*, 77: 97–103.

Bartelsman, E., S. Scarpetta and F. Schivardi (2005). Comparative analysis of firm demographics and survival: evidence from micro-level sources in OECD countries. *Industrial and Corporate Change*, 14: 365–391.

Becker, B. and Gerhart, B. (1996). The Impact of Human Resource Management on Organizational Performance: Progress and Prospects. *The Academy of Management Journal*, 39: 779–801.

Becker, G. S. (1957). *The Economics of Discrimination*. The University of Chicago Press: Chicago.

Bonsack, J. A. (1881). *Cigarette Machine*, No. 238,640, United States Patent and Trademark Office.

Box, M. (2008). The death of firms: exploring the effects of environment and birth cohort on firm survival in Sweden. *Small Business Economics*, 31: 379–393.

Bradley, H. (1989). *Men's Work, Women's Work*. Basil Blackwell: Oxford.

Brüderl, J., Preisendörfer, P. and Ziegler, R. (1992). Survival Chances of Newly Founded Business Organizations. *American Sociological Review*, 57: 227–242.

Burnette, J. (1997). An Investigation of the Female-Male Wage Gap during the Industrial Revolution in Britain. *The Economic History Review*, 50: 257–281.

Burnette, J. (2008). *Gender, work and wages in industrial revolution Britain*. University Press: Cambridge.

Caves, R.E. (1998). Industrial organization and new findings on the turnover and mobility of firms. *Journal of Economic Literature*, 36: 1947–1982.

Cefis, E. and Marsili, O. (2005). A matter of life and death: innovation and firm survival. *Industrial and Corporate change*, 14: 1167–1192.

Cooper, P.A. (1987). *Once a cigar maker: men, women, and work culture in American cigar factories, 1900-1919.* University of Illinois Press: Urbana.

Cox, D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society*, 34: 187–220.

Cox, D. and Nye, J.V. (1989). Male-female wage discrimination in nineteenth-century France. *The Journal of Economic History,* 49: 903–920.

Disney, R., J. Kaskel, and Y. Heden (2003). Entry, exit and establishment in UK manufacturing. *The Journal of Industrial Economics*, 51: 91–112.

Dunne, T., M.J. Roberts, and L. Samuelson (1989). Patterns of firm entry and exit in U.S. manufacturing industries. *Rand Journal of Economics*, 19: 495–515.

Dunne, T., Roberts, M.J. and Samuelson, L. (1989). The growth and failure of US manufacturing plants. *Quarterly Journal of Economics*, 104: 671–698.

Ericson, R. and Pakes, A. (1995). Markov-perfect industry dynamics: a framework for empirical work. *Review of Economic Studies*, 62: 53–82.

Esteve-Pérez, S. and J. Mañez-Castillejo (2008). The resource-based theory of the firm and firm survival. *Small Business Economics*, 3: 231–249.

Evans, D.S. (1987). The relationship between firm growth, size, and age: estimates for 100 manufacturing industries. *Journal of Industrial Economics*, 35: 567–581.

Fine, J.P. and Gray, R.J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94, 496–509.

Friedman, M. (1953). *Essays in Positive Economics.* University of Chicago Press: Chicago.

Geroski, P.A. (1995). What do we know about entry? *International Journal of Industrial Organization*, 13: 421–440.

Geroski, P.A. and Mazzucato, M. (2001). Modelling the dynamics of industry populations. *International Journal of Industrial Organization*, 19: 1003–1022.

Goldin, C. (1990). *Understanding the gender gap: an economic history of American women*. Oxford University Press: New York.

Goldin, C. and Sokoloff, K. (1982). Women, Children, and Industrialization in the Early Republic: Evidence from the Manufacturing Censuses. *The Journal of Economic History,* 42: 741–774.

Goldin, C. and Sokoloff, K. (1984). The Relative Productivity Hypothesis of Industrialization: The American Case, 1820 to 1850. *The Quarterly Journal of Economics*, 99: 461–487.

Gospel, H. (2007). The management of labor and human resources, in G. Jones and J. Zeitlin, eds., *The Oxford Handbook of Business History.* Oxford University Press: Oxford, 420–446.

Hawley, A. (1968). Human ecology, in D.L. Sills, ed., *International Encyclopedia of the Social Sciences.* MacMillan: New York.

Hannan, M. T. and Freeman, J. H. (1977). 'The population ecology of organisations,' *American Journal of Sociology*, 82: 929–964.

Harhoff, D., Stahl K. and Woywode, M. (1998). Legal Form, Growth and Exit of West German Firms - Empirical Results for Manufacturing, Construction, Trade and Service Industries. *The Journal of Industrial Economics*, 46: 453–488.

Hellerstein, J. K., Neumark, D and Troske, K. R. (2002). Market Forces and Sex Discrimination. *The Journal of Human Resources,* 37: 353–380.

Heyman, F., Svaleryd, H., and Vlachos, J. (2013). Competition, Takeovers and Gender Discrimination. *Industrial and Labor Relations Review*, 66: 409–432.

Jones, G. and Zeitlin, J. eds. (2007). *The Oxford Handbook of Business History.* Oxford University Press: Oxford.

Jovanovic, B. (1982). Selection and the Evolution of Industry. *Econometrica*, 50: 649–670.

Karlsson, T. (2008). Downsizing: Personnel Reductions at the Swedish Tobacco Monopoly, 1915-1939. *Lund Studies in Economic History*, 47.

Kipping, M. and Cailluet, L. (2010). Mintzberg's emergent and deliberate strategies: Tracking Alcan's activities in Europe, 1928-2007. *Business History Review*, 84: 79–104.

Klepper, S. (1996). Entry, exit, growth and innovation over the product life cycle. *American Economic Review*, 86: 562–583.

Klepper, S. (2002). The capabilities of new firms and the evolution of the US automobile industry. *Industrial and Corporate Change*, 11: 645–666.

Klepper, S. and Simons, K. (1997). Technological extinctions of industrial firms: an inquiry into their nature and causes. *Industrial and Corporate Change,* 6: 379–460.

Klepper, S. and Simons, K. (2000). The making of an oligopoly: firm survival and technological change in the evolution of the US tire industry. *Journal of Political Economy,* 108: 728–760.

Klepper, S. and Thompson, P. (2006). Submarkets and the evolution of market Structure. *RAND Journal of Economics,* 37: 861–886.

Kommerskollegii (1863-1885). *Bidrag till Sveriges Officiella Statistik F, Utrikes handel och sjöfart 1863–1885.* Stockholm.

Kommerskollegii (1863-1910). *Bidrag till Sveriges officiella statistik. D: Kommerskollegii underdåniga berättelse för år 1863-1910*. Stockholm.

Kommerskollegium kammarkontoret (1863-1885). *Årsberättelser fabriker serie 4*, Da4, vol. S05823–44, vol. S02054-61 and vol. S05806-19. Kommerskollegiets arkiv, National Archives (*Riksarkivet*): Stockholm.

Lindbom, T. and Kuhm, E. (1940). *Tobaksabetarnas förbund i Sverige 1889-1939*. Tiden: Stockholm.

Lotti, F. (2007). Firm dynamics in manufacturing and services: a broken mirror? *Industrial and Corporate Change*, 16: 347–369.

Mackie, R. (2001). Family Ownership and Business Survival: Kircaldy, 1870-1970. *Business History*, 43: 1–32.

Manjón-Antolín, M. C. and Arauzo-Carod, J-M.. (2008). Firm survival: methods and evidence. *Empirica*, 35: 1–24.

Marx, K. (1970 [1867]). *Capital: A Critique of Political Economy.* 1, International Publishers: London.

Mata, J. and Portugal, P. (1994). Life duration of new firms. *The Journal of Industrial Economics*, 42: 227–245.

McGovern, T. (2007). Why do Successful companies fail? A case study of the decline of Dunlop. *Business History,* 49: 886–907.

Nenadic, S. (1993). The small family firm in Victorian Britain, *Business History*, 35: 86–114.

Persson, H. and Sjögren Lindquist, G. (2010). The survival and growth of establishments: does gender segregation matter? *Research in Labor Economics,* 30: 253–282.

Rabenius, T. (1888), Näringsfrihet, in *Nordisk Familjebok*, 12: 19–22.

Rose, S.O. (1987). Gender Segregation in the Transition to the Factory: The English Hosiery Industry, 1850-1910. *Feminist Studies*, 13: 163–184.

Rosén, Å. (2003). Search, bargaining and employer discrimination. *Journal of Labour Economics,* 21: 807–829.

Stanfors, M., Leunig, T. Eriksson, B. and Karlsson, T. (2013). Gender, productivity and the nature of work and pay: evidence from the late nineteenth-century tobacco industry. *The Economic History Review*, 67; 48–65.

Stinchcombe, A. (1965). Social structures and organizations, in J.G. March ed., *Handbook of Organizations*. Rand McNally: Chicago.

Thompson, P. (2005). Selection and Firm survival: Evidence from the Shipbuilding Industry, 1825-1914. *The Review of Economics and Statistics*, 87, 26–36.

Valverde, M. (1988). Giving the Female a Domestic Turn: The Social, Legal and Moral Regulation of Women's Work in British cotton Mills, 1820-1850. *Journal of Social History*, 21: 619–34.

Webb, S. (1891). The Alleged Differences in the Wages Paid to men and to Women for similar Work. *The Economic Journal*, 1; 635–662.

Weber, A. and Zulehner, C. (2009). Competition and Gender Predjudice: Are Discriminatory Employers Doomed to Fail? *NRN working papers,* 0926, The Austrian Center for Labor Economics and the Analysis of the Welfare State.

Weber, A. and Zulehner, C. (2010). Female Hires and the Success of Start-up Firms. *American Economic Review Papers & Proceedings*, 100: 358–361.

Wright, P. M., McMahan, G. C. and McWilliams, A. (1994). Human resources and sustained competitive advantage: a resource-based perspective. *International Journal of Human Resource Management*, 5: 310–26.

Paper 2

Industrious Migrants: The earnings of male and female migrants in Swedish manufacturing around 1900

Björn Eriksson
Department of Economic History, Lund University

Maria Stanfors
Department of Economic History, Lund University

Abstract
Migration played a central role in facilitating industrialization by reallocating labour to where it was most valued while at the same time improving life chances for many migrants. In this paper we consider the labour market performance of internal migrants in Sweden around the turn of the last century. Using detailed individual data on both men and women employed in three manufacturing industries, we find evidence that migrants, irrespective of gender, performed well in terms of earnings. For both men and women, there is evidence of sorting based on skills with earnings premiums consistent with a Roy selection model. An increase in work on both the extensive and intensive margins further explains the earnings premium among female migrants.

JEL classifications: J24, J61, N33, N63

Keywords: Migration, Wage premium, Selection, Industrialization

Introduction

During the late nineteenth and early twentieth century, labour mobility increased dramatically across Western Europe (Baines 1985; 1994). The frequent yet short-range migration patterns of the pre-industrial era were gradually being supplanted by longer-range migration directed toward industrial centers and cities; a shift in mobility driven by both decreasing migration costs and new incentives for migration (Parish 1973). This movement of workers had important implications for individuals, firms and the economy as a whole. By leaving their place of origin, many individuals' life chances and living standards improved as a function of their moves. Employers gained from the mobility of job-searchers by having a large pool of workers to choose from in a labour market where turnover was generally high (Elmquist 1899: 94; Hareven 1982; Slichter 1919). Overall, migration greased the wheels of the industrialization process through a reallocation of labour.

Previous historical studies on internal migrants have focused on the correlations between geographic and occupational mobility. By comparing the occupational status of migrants to different reference groups, questions relating to returns to migration, selection into migration streams and the assimilation of migrants into local labour market have been addressed (Long 2005; Stewart 2006; Salisbury 2015; Silvestre et al. 2015). This paper contributes to the existing literature on internal migrants by closely examining earnings instead of occupations. By focusing on earnings, we are able to more precisely quantify and explain factors determining the labour market performance of migrants around the turn of the twentieth century.

Despite their importance in the industrialization process, female labour migrants remain a largely neglected group in the historical migration literature. This is understandable considering the scarcity of information about women's occupations and income in historical records, yet insufficient and undesirable when taking women's high geographic mobility and the importance of migration for both individual welfare and the overall economy into account. Our second contribution is to extend the analysis beyond the prevailing focus on men by including female migrants in the analysis. Studying female migrants alongside their male counterparts provides us with important evidence about differences and similarities between men and women in historical labour markets.

Our focus is on migrants employed in manufacturing in Sweden around 1900; a period characterized by rapid industrialization, high economic growth, rising migration rates and increasing involvement of women in the formal economy. By using detailed matched employer-employee data covering three manufacturing industries around 1900, we quantify the earnings premium of internal migrants compared to both the native population in a specific location and to coworkers within the same firm. In all three industries, migrants earned more than both natives and coworkers. This pattern holds for both men and women. Moreover, the premium remains after accounting for differences in human capital by including a rich set of control variables. We therefore proceed by considering explanations for why migrants performed so well in terms of earnings by testing for differences in ability and returns to skills, and effort expended on both the intensive and extensive margins by migrants.

Migrant earnings

Migration constitutes a choice between moving and staying. In a simplified form, migration may be understood as an investment decision. By weighing the benefits of moving against costs, potential migrants decide whether migration is expected to be profitable or not and act accordingly (Lee 1966). Ignoring any possible non-monetary benefits or costs, three distinct reasons why migrants earn more than those that remain in their place of origin may be considered. First, migrants may earn more as a result of wage levels being higher at their chosen destination than at their origin. Historically, this was a result of migration from mainly agricultural and rural areas with plentiful labour and low wages to urban and industrialized areas where labour was in demand and wages accordingly were higher (Boyer 1997; Boyer and Hatton 1997). Secondly, differences in earnings may reflect differences in occupational structure between the destination and origin of migrants. By moving to a location with better prospects for upward occupational mobility earnings may increase as a result (Sjaastad 1962; Long 2005). These two reasons are to some extent unconditional on ability and therefore potentially realizable by any migrant. Thirdly, part of the observed earnings premium of migrants is, however, not realizable by all but the result of migrants being inherently different from non-migrants in their motivations, preferences and

capacity to exploit opportunities presented. Migrants are thus a self-selected group of the population, which differ in terms of abilities and motivations for moving, a factor that affect earnings apart from migration itself (Marshall 1920: 199; Chiswick 1999).

In a series of articles building on Roy (1951), Borjas (e.g., 1987; 1991; 1994) shows how migrants' earnings depend on differences in individual ability and returns to skills between locations. The model's point of departure is an initial mismatch between the workers' skills and place of origin which is resolved through migration. The logic of the model is straightforward if we consider two stylized regions, A and B, in which skills are valued differently. In the first region (A), returns to skills are high, and income is therefore highly correlated with ability. In the second region (B), return to skills are lower than in A, and income is subsequently less dependent on ability. This makes region A (i.e., high returns to skills) more attractive to high-ability individuals while region B (i.e., low returns to skills) offers greater benefit to low-ability individuals. In order to improve their lot, high-ability individuals will move from Region A to Region B while low-ability individuals will move in the opposite direction.

In terms of migrant's earnings relative to the native population, the Roy-model (Borjas 1987: 533-534) predicts that when returns to skills are lower at origin than at destination, migrants will constitute the most able persons and will outperform the destination's native population in terms of earnings. Conversely, if returns to skills are high at origin compared to the destination, migrants will be drawn from the bottom of the ability distribution and will underperform in terms of earnings when compared with the destination's native population.[9]

Migrant's earnings are also affected by their ability to assimilate and become incorporated in the local labour market. Although the importance of assimilation is not as pressing as when studying international migrants (see for example Abramitzky et al 2014; Hatton 1997; Minns 2000) or more heterogeneous societies than nineteenth century Sweden, concerns about internal migrant's labour market incorporation should not be dismissed

[9] Borjas (1987: 534) defines a third case, that of 'refugee sorting', under the condition that a negative correlation exists between the way skills are valued at origin and destination. Given that this paper concerns migration within industries in a single country, this is highly unlikely to apply and is therefore not considered further.

outright (Hatton and Bailey 2002; Silvestre et al. 2015; Williamson 1990). A lack of specific skills or information are examples of factors that may still have inhibited internal migrants from achieving their full earnings potential, resulting in migrants occupying a marginal position and underperforming relative to the native population.

Although one can hypothesize about direction and causes, the question of how migrants fared in historical labour markets is ultimately an empirical one. With a few exceptions the historical evidence generally supports the notion that male internal migrants fared well, realizing both important gains from migration while simultaneously outperforming the native population in the destination. Attempts to assess the selectivity and labour market performance of migrants in historical populations include diverse and innovative approaches to measuring observed and unobserved characteristics. Studies using census data have typically relied on measures derived from recorded occupations (Salisbury 2014) or wealth indicators (Abramitzky et al. 2013). Alternative approaches have estimated differences in migrants' human capital by using heights (Humphries and Leunig 2009; Kosack and Ward 2014) literacy (Feliciano 2001) and age heaping (Stolz and Baten 2012) as measures thereof. In the case of the US, there is evidence that American frontier migrants tended to be negatively selected from the source population (Stewart 2006, 2012), a notion also apparent in Thernstrom's description (1973: 42) of US internal migrants being a "permanent floating proletariat made up of men ever on the move spatially but rarely wining economic gains". In contrast, Collins and Wanamaker (2014) find that black migrants moving from the American South to the North at the beginning of the twentieth century were positively selected. In the case of Great Britain, Long (2005) finds that rural-to-urban migrants in the Victorian period were positively selected. Humphries and Leunig (2009) show that nineteenth-century English and Welsh seaman who moved to London were positively selected in terms of height.

Comparisons between natives and migrants have generally painted a positive picture of how internal migrants fared. Hatton and Bailey (2002) find that migrants into London around 1930 were not marginal workers, outperforming the native population both in terms of earnings and lower rates of unemployment. Silvestre et al. (2015) show that migrants in Barcelona in

1930 who arrived prior to 1900 tended to do as well as the native population, while those that moved to the city later did relatively worse[10]. In the case of Sweden, Maas and van Leeuwen's (2004) study of the Sundsvall region indicates that migrants seemed to fare somewhat better than the native population in terms of upward career mobility during the mid-nineteenth century.

Gender and Migration

The high geographic mobility of women during industrialization did not go unnoticed by contemporary observers. As the nineteenth century was drawing to an end, Ravenstein (1885: 196) declared that "woman is a greater migrant than man" and that "they [women] migrate quite as frequently into certain manufacturing districts, and the workshop is a formidable rival of the kitchen and scullery". While it may be debated whether Ravenstein was correct in arguing that women's mobility exceeded that of men, his observations indicate that internal migration was prevalent and important among women as well as among men (Sharpe 2001; Trent and Steidl 2012). Industrialization increased the demand for female labour in manufacturing centers and urban areas, which in turn prompted the migration of large numbers of women (Boyer and Hatton 1997: 706; Moch 1992: 130). Many employers made special arrangements to attract and accommodate the many female migrants required to meet the labour demands of the expanding manufacturing sector (Dublin 1994; Gordon 2002).

Theoretically, migrants' earnings premiums is typically thought of as positive in terms of high returns to skills at destination, regional wage differences and more favorable to migrants with economic motives. If men and women differ in these respects, we should also expect migrants' earnings to vary according to sex. If there were discrimination and glass ceilings barring women from skilled positions, putting a downward pressure on skill premiums and earnings, the scope for sorting across different locations

[10] The authors argue that this result is driven by both a change in the composition of migrants and a failure on their behalf to assimilate into a changing labour market.

according to skill will be more limited for women than for men. On the other hand, if discrimination and glass ceilings are not universal but only local, differences in returns to skills between different locations for women may, on the contrary, be greater resulting in larger effects for women than for men.

Another potentially significant difference between women and men concerns the motivation for moving. As pointed out by Lee (1966: 51): "not all persons that migrate reach that decision themselves". Many, especially women, have been migrants as a result of being tied to a moving spouse (Mincer 1978), married women's migration decision thus being to a lesser degree autonomous than men's decision. But even for unmarried women the reasons for moving may have been different from those of men. Most benefits from migration are not realized instantaneously after moving but rather at some time in the future. Since marriage in the late nineteenth and early twentieth century typically meant that women withdrew from the labour market, work-related monetary returns to migration and ability may have served as a weaker pull factor for women, resulting in smaller earnings premiums for female migrants than for male migrants.

Context

Sweden circa 1900

Sweden was not among the earliest industrializers; modern economic growth took off in the mid-nineteenth century, and the period after 1870 marks the most successful part of industrialization and growth. Swedish development lagged behind continental North Western Europe but was ahead of Southern Europe (Crafts 1985: 57-58). Sweden benefited from favorable resource endowments and from latecomers' advantages with respect to technological and organizational advances made elsewhere. Innovations, new technologies, new machines and materials appeared throughout the last half of the nineteenth century and culminated with electrification, capital-intensive industry, mass production and the emergence of big business in the early twentieth century (Chandler 1977; Landes 1969). By 1900, 20 percent of the total labour force was employed in manufacturing. 19 percent of working-age women, most of them young and single, were employed in the formal economy, principally in domestic service and textile production. Female

employment was on a par with the United States, but below the British rate (Mitchell 1981; Richards 1974; Sawyer 1949).

Concern about economic issues and the labour market conditions of the growing working classes led governments in Europe and the United States to undertake surveys and censuses from around 1880. The *Swedish Board of Commerce* instigated statistical surveys of a number of industries, including mechanical engineering, printing and tobacco.[11] Statistician Henning Elmquist was the person in charge. Together with a number of travelling agents he asked detailed questions of all employers and employees in the surveyed industries.

The industries surveyed

At the time when they were surveyed (between the years 1898-1903), the industries considered in this paper - tobacco, printing, and mechanical engineering - may all be categorized as factory-based manufacturing industries. They were nonetheless distinguished by differences in skill requirements, organization, technology and the employment of women. Tobacco manufacturing was still considered at the turn of the last century to be a crafts industry, thus in many ways resembling more closely Sweden's pre-industrial economy (Lindbom and Kuhm 1940: 82). The printing industry was also a trade with historical roots, but was at the time of the survey going through a period of modernization and rationalization. The introduction of the Linotype and Monotype machines was gradually rendering typesetting by hand obsolete, changing both the nature of the tasks performed by workers and the overall industry. The progress made by the mechanical engineering industry was in contrast to the backwardness of the tobacco industry and the

[11] There were several motivations why these industries were surveyed. The tobacco industry was known for its conflicts between workers and employers; it had a considerable share of female workers; and a previous public inquiry on the issue of labour insurances had showed that the work environment in the tobacco industry was particularly unhealthy. Largely the same motivations applied for the printing industry, yet the share of women workers was not as high as in tobacco. Printing was, however, becoming an increasing important industry at the time of survey. Mechanical engineering was signifying the new and highly important industries. The geographical location of the industries (a relative concentration to cities) was thought to simplify the realization of the surveys.

gradual changes in the printing industry in essence a product of the second industrial revolution. It was highly capital-intensive, it often made use of electricity power, and both employed and produced technologically advanced machinery. Between 1890 and 1910, the Swedish mechanical engineering industry grew on average by more than 10 percent yearly (Schön 2010: 197) and was, according to Elmquist (1901: 12): ”fully modern and by necessity compelled to remain at the vanguard of development”. The survey of the mechanical engineering industry includes factories such as *LM Ericsson* and *Husqvarna*; firms that would go on to become global leaders within their respective fields in the twentieth century.

The formal skill requirements for entering the tobacco industry may be considered to have been minimal. Required skills were acquired through job training during a period of apprenticeship that typically lasted at least two years. The printing industry was, in comparison, more skill-intensive. Compositors obviously had to have a level of literacy beyond basic reading and writing to perform their work. Apprenticeship was also more formalized and lasted a longer time. Formal apprenticeship was, in contrast, uncommon in mechanical engineering. In its place, many workers attended evening classes where they were taught mathematics, mechanics, physics and technical drawing, important skills to have in an industry that was based on high-precision manufacturing of products such as weapons, vehicles, and steam and electric engines (Elmquist 1901).

A striking difference is apparent when comparing the extent to which women were employed in the industries. In tobacco in 1898, women constituted 63 percent, which was a sizable share of the workforce (Kommerskollegium 1899). In the printing industry, women only made up 25 percent of the workforce when surveyed in 1903. The share of women employed in each industry reflects the relative standing of women within the industries. Although women could be found performing skilled work next to men in both tobacco and printing at the time of the surveys, they were more likely to hold a skilled occupation in tobacco. Women in both industries were, however, less skilled than men in either industry (Elmquist 1899, 1909). Unions in both industries allowed women to become members, but the male workforce in printing was considerably more hostile towards women, fearing that the employment of women would lead to an increase in competition and lowering of wages (Ekdahl 1983). The mechanical engineering industry stands in stark contrast to both printing and tobacco: not a single mechanical engineering firm surveyed employed women. The reason

for this may be found in the nature of the work performed. Printing and tobacco were not physically demanding whereas mechanical engineering work often required physical strength, thus providing men with an advantage, creating a barrier to entry for women.

In summary, the industries surveyed span a wide range in terms of degree of modernity, from the manual craft of tobacco manufacturing to what was by the standard of its day a technology-intensive mechanical engineering industry. The industries also differ regarding the skills and physical strength required and the employment and standing of women, and are consequently representative of several facets of the Swedish manufacturing industry.

Data

In order to estimate migrants' earnings premiums and to explain their existence, we exploit cross-sectional data covering the three industries described above. The data has been sourced from investigations of the Swedish labour force conducted by the *Swedish Board of Commerce* around the turn of the last century. The industries were surveyed in 1898 (tobacco), 1899 (mechanical engineering) and 1903 (printing) respectively. The tobacco and printing industries were surveyed in their entirety, while the investigation of the mechanical engineering industry included what were deemed to be the 32 most important firms.[12] Henning Elmquist and travelling agents collected information from 4,380 workers employed in the tobacco industry, 7,855 workers in the printing industry, and 12,060 workers employed in mechanical engineering. All the workers employed in the industries surveyed were interviewed by one of the travelling agents in accordance with a pre-printed, detailed questionnaire that covered the same background information and indicators on earnings and experience, while also including a set of industry-specific indicators of working conditions. The agents also collected information from the employers, which included the number of employees, the wage bill, machinery, working hours, employment contracts, fringe benefits (which were trivial), and experiences of strikes and lock-outs. We are able to link this information in an employer-employee matched dataset.

[12] In 1901, a second survey was conducted of the mechanical engineering industry. This survey included smaller firms. Our data does not include this information.

A study of migration requires a definition of who is to be considered a migrant. The simplest and most common way to define migration is to consider individuals who have crossed a border between administrative units such as countries, counties or parishes as migrants. While this is intuitively a straightforward measure, it makes little sense in our case. As in most countries, Swedish counties are large while parishes are small, and either definition expresses migration distances, be they relatively long or very short, with a great deal of inaccuracy. To introduce greater subtlety and precision, we proceed by considering the actual distance moved by each worker in terms of kilometers, calculating the distance (as the crow flies) from the individual's parish of birth to the parish in which their current place of work is located. In the process of identifying an individual's birth parish, our sample is reduced since we lose those individuals who had given too imprecise information about their place of birth (e.g., had stated a country, county or province instead of a specific parish as their place of birth).[13] In the remainder of the paper we define any individual working more than 50 kilometers away from his or her parish of birth as a migrant.

Migration entails moving from one place to another with each setting having a different impact on earnings, a factor which we take into account by employing fixed effects, and considering location-specific returns to skills. To make each setting more consistent in terms of size we define a set of geographic locations by dividing Sweden into a grid consisting of squares measuring 25 by 25 kilometers, and by assigning to each individual an origin location based on the coordinates of their birth parish and a destination location based on the coordinates of their current workplace. In the analysis that follows, we also will consider a setting that is even more subtle than a geographic location: that of the firm. By comparing migrants to their native coworkers within the same firm, we are able to discount any differences in earnings attributable to factors fixed at the firm level such as pay-scales, hiring practices, management and productivity.

Importantly, and in contrast to most previous historical studies of migration, we observe actual recorded earnings of individual workers. Since the workers reported both weekly earnings and hours worked, we are able to

[13] The numbers lost because of imprecise information regarding birthplace are as follows: 387 men and 469 women in tobacco; 1,063 men and 320 women in printing; and 4,910 men in mechanical engineering.

use hourly earnings as the dependent variable, which is rare but highly important not least from a gender perspective, since women may be paid less because they work shorter hours. Moreover, the use of hourly earnings allay concerns about a potential bias arising as a result of differences in the number of hours worked by migrants, a point to which we will return.

Thanks to the meticulous surveys carried out by the *Swedish Board of Commerce*, additional controls not normally available when studying historical populations are included for confounders. A number of questions were asked relating to both workers' professional careers and their private life. We thus have information about each worker's age, experience, earnings, working hours, father's occupation[14], health, own occupation, employer, tenure, union and friendly society membership, place of birth and details about marital status and children. We exclude any individuals with missing variables as well as all workers aged younger than 15, leaving us with a total sample of 1,032 men and 2,101 women in tobacco, 4,295 men and 1,444 women in printing, and 6,540 men in mechanical engineering. Summary statistics disaggregated by sex and migration status for each industry are presented in Table 1.

Among male tobacco workers, 31 percent of men and 20 percent of women were migrants according to our definition of working in a firm located at least 50 km away from their parish of birth. The corresponding shares in printing were 29 percent for men and 14 percent for women. The share of migrants in mechanical engineering (30 percent) was similar to those of men in tobacco and printing.

In terms of workers' average hourly earnings, men were better paid than women, while workers in the printing industry were better paid (men 35 öre per hour; women 19 öre) than were the workers in mechanical engineering (men 24 öre) and tobacco manufacturing (men 23 öre; women 16 öre).

[14] Father's occupations were coded using this HISCO scheme and thereafter classified into twelve HISCLASS categories (see van Leeuwen et al. 2002 and van Leeuwen et al. 2011). In the analysis we use four aggregate groups: Managers and professionals (encompassing HISCLASS categories 1-5), skilled workers (HISCLASS 6-8), lower skilled workers (HISCLASS 9-10) and unskilled workers (11-12).

Table 1. Summary statistics: means (with standard deviations in parentheses) and proportions

	Tobacco						Printing						Mechanical Engineering		
	Men			Women			Men			Women			Men		
	All	Migrant	Non-migrant	All	Migrant	Non-migrant	All	Migrant	Non-migrant	All	Migrant	Non-migrant	All	Migrant	Non-migrant
Hourly earnings	23.06	27.66	20.98	15.68	17.08	15.35	34.57	42.52	31.29	19.21	24.11	18.39	24.09	26.75	22.96
	(11.15)	(10.03)	(11.02)	(6.38)	(6.72)	(6.25)	(23.17)	(21.15)	(23.17)	(9.94)	(12.62)	(9.17)	(7.12)	(6.75)	(6.98)
Age	33.35	38.46	31.05	27.90	31.39	27.06	27.92	33.15	25.76	24.38	27.46	23.86	31.48	33.81	30.48
	(15.12)	(13.48)	(15.26)	(12.17)	(13.23)	(11.75)	(11.67)	(12.34)	(10.65)	(8.76)	(10.18)	(8.40)	(13.72)	(13.59)	(13.66)
Experience	17.50	22.46	15.27	9.71	10.50	9.52	13.61	18.47	11.61	9.09	10.85	8.80	10.58	11.71	10.10
	(15.40)	(14.40)	(15.33)	(9.90)	(10.17)	(9.83)	(11.17)	(11.89)	(10.21)	(7.92)	(9.29)	(7.63)	(11.11)	(11.63)	(10.84)
Children	1.30	1.51	1.21	0.54	0.68	0.51	0.81	1.24	0.63	0.08	0.16	0.07	1.42	1.46	1.40
	(1.98)	(2.04)	(1.94)	(1.15)	(1.18)	(1.14)	(1.56)	(1.83)	(1.40)	(0.40)	(0.62)	(0.35)	(2.12)	(2.04)	(2.15)
Unmarried	0.54	0.42	0.59	0.77	0.67	0.80	0.64	0.47	0.71	0.95	0.88	0.96	0.51	0.46	0.54
Married	0.41	0.51	0.36	0.17	0.25	0.15	0.35	0.51	0.29	0.04	0.10	0.04	0.46	0.51	0.44
Previously married	0.05	0.07	0.05	0.06	0.08	0.05	0.01	0.02	0.01	0.01	0.02	0.01	0.03	0.04	0.02
Healthy	0.83	0.80	0.84	0.74	0.67	0.76	0.84	0.81	0.86	0.89	0.86	0.90	0.89	0.88	0.89
Apprentice	0.08	0.01	0.11	0.03	0.02	0.03	0.25	0.10	0.32	0.12	0.10	0.12	0.01	0.00	0.01
Temperance society member	0.10	0.10	0.10	0.07	0.05	0.08	0.12	0.14	0.12	0.06	0.08	0.06	0.13	0.11	0.13
Union member	0.47	0.62	0.40	0.26	0.29	0.26	0.69	0.78	0.65	0.38	0.53	0.36	0.34	0.42	0.31
Friendly society member	0.57	0.64	0.53	0.47	0.54	0.46	0.57	0.67	0.53	0.57	0.56	0.57	0.88	0.85	0.89
Father's social status															
Manager or professional	0.09	0.08	0.09	0.07	0.07	0.08	0.13	0.17	0.11	0.12	0.20	0.10	0.05	0.06	0.05
Skilled	0.35	0.42	0.32	0.33	0.45	0.30	0.46	0.49	0.45	0.41	0.42	0.41	0.51	0.52	0.51
Lower skilled	0.21	0.18	0.23	0.22	0.19	0.22	0.18	0.14	0.19	0.19	0.14	0.20	0.17	0.17	0.16
Unskilled	0.22	0.19	0.23	0.24	0.17	0.25	0.15	0.12	0.16	0.12	0.11	0.12	0.15	0.13	0.16
Unknown	0.13	0.13	0.13	0.14	0.12	0.14	0.08	0.08	0.08	0.16	0.13	0.16	0.12	0.12	0.13
N	1032	320	712	2101	408	1693	4295	1254	3041	1444	206	1238	6540	1957	4583

Note: Standard deviations for continuous variables within parentheses.

Sources: Undersökning av tryckerier mm 1903, Avdelningen för arbetsstatistik, HII a:1 volumes 1-6 and HII a:2 volumes 1-12, Kommerskollegiets arkiv, National Archives (Riksarkivet). Stockholm; Specialundersökningar Tobaksindustrien 1898, Statistiska avdelningen, HIII b:1 and HIII b:1 aa volume 1, Kommerskollegiets arkiv, National Archives (Riksarkivet), Stockholm; De större mekaniska verkstäderna 1899, Statistiska avdelningen, HIII c:1, Kommerskollegiets arkiv, National Archives (Riksarkivet), Stockholm.

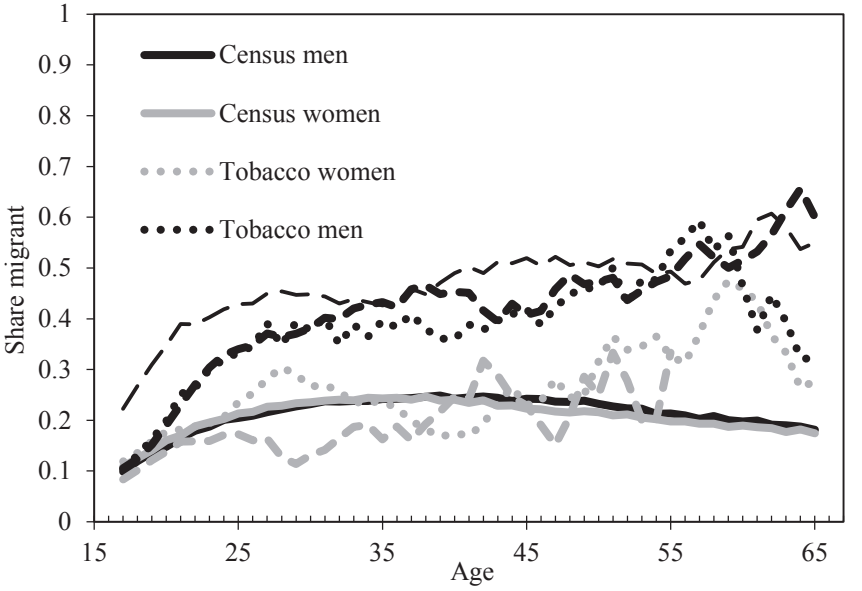
Migrants consistently earned more than non-migrants across all occupations and both sexes. This basic difference in earnings is, however, to great extent spurious since the migration status observed here is the outcome of a cumulative process, with older workers more likely to be migrants than younger workers because a number of younger migrants-to-be had not yet moved. Thus, the average migrant in the sample was older, more experienced, more likely to be married and to have children at home than the non-migrants, all these being differences that are part of the course of life or an outcome of migration itself, and at the same time likely to affect earnings.

One variable, father's social status, does not suffer from the problem of potentially being an outcome of migration. When comparing the social background of migrants and non-migrants by sex and industry some distinct similarities and differences are apparent. First, in tobacco and printing, migrants have a tendency to come from a more privileged social background than non-migrants. The skilled category of father's social status is significantly overrepresented among male (42% vs 32%) and female (45% vs 30%) migrants, while the lower skilled and unskilled categories are underrepresented. In printing the story is similar, with the modification that the highest category, manager or professional is relatively overrepresented among both male (17% vs 11%) and female (20% vs 10%) migrants. In the tobacco and printing there is thus no evidence implying important differences between men and women in how migrants were selected based on father's social status. Moreover, within each industry there are not important differences when considering the selection into being employed in the industry. When considering all men and women in each industry the social backgrounds of the male and female workforce mirror each other in both tobacco and printing. In mechanical engineering migrants and non-migrants does not seem to differ in terms of social background.

How does the prevalence of migration in these industries compare to the general population in Sweden, and how does it differ between men and women when we take age into account? The answer is provided in Figure 1, which plots the share of migrants in each industry against age together with the share of migrants in the population as a whole. In Figure 1, migrant status is defined as residing in a county other than that of birth in order to make possible a comparison to census statistics.

It is clear from Figure 1 that the share of male migrants in all three industries surveyed supersedes that of the overall population by a fair margin. For female tobacco workers the share is similar to that of the overall

Figure 1. Migrant share of population and industries by sex and age



Note: The industry-specific series are five-year moving averages.
Sources: For the industry statistics see Table 1. The census statistics are from *1900 Micro censuses (SweCens),* National Archives (*Riksarkivet*), Stockholm.

population while for female printing workers the share is somewhat lower in the 20 to 40 age band. The overall pattern is consistent with the argument that industrialization in particular exerted a pull on migrants. That migrants constituted such a large share of the work force of the industries surveyed serves to further motivate a more thorough examination of migrants and their performance in terms of earnings.

Our first descriptive evidence thus point towards male and female migrants being on the whole rather similar, with the exception of migrants constituting a larger share among men than women. Both groups' unadjusted earnings superseded that of non-migrants. Moreover, there is no evidence that the selection into migration based on father's social status differed between men and women within the surveyed industries.

Estimates of migrant's earnings premiums

It is clear from the descriptive statistics that migrants earned a significant unadjusted premium relative to non-migrants. However, this simple difference does not take into account other aspects in which migrants differed from non-migrants and provides few explanations for the existence of the premium. To further investigate whether an adjusted earnings premium existed for migrants and, if so, whether it differs between industries and by sex, we begin by estimating the following basic model:

$$ln\ \mathrm{Y} = \beta_0 + \beta_1 \mathrm{Migrant} + \beta_k \mathbf{X} + \varepsilon \quad (1)$$

in which the dependent variable (Y) is the logarithm of hourly earnings, and **X** denotes a vector of individual characteristics. The coefficient β_1 simply denotes the difference in earnings between migrants (those having moved at least 50 km) and non-migrants. Models stratified by sex are estimated for each industry. Controls for age, experience, father's social status, civil status, number of children, health status, and whether the individual was an apprentice, member of a temperance society, union or benefit society are included to allow for the fact that migrants differ from non-migrants on a number of observable variables that affect earnings apart from migrant status. These controls are added in a step-wise manner. Although we have information about occupations, we follow Hatton's (1997) example and deliberately omit this variable to avoid the bad control problem (Angrist and Pischke 2009: 64-68). Table 2 presents the results from an OLS estimation of β_1 from equation (1) estimated separately by industry and sex. The first estimated model, displayed in column (I), includes only a cubic function of age as a control. Father's social status is added as a control in column (II). Column (IV) adds controls for experience (in years as a cubic function), civil status, children and health status. In column (V) the most extensive model is presented which adds a final set of controls in the form of dummies indicating if the worker was an apprentice or a member of a union, benefit society, or temperance society.

In the most basic model (I), controlling for age only, migrants earn between 4.9 and 11.8 log points more than non-migrants. After adding additional father's social status to the model in column (II), estimates change little. The addition of experience as a control in model (III) has little impact with the exception of β_1 for women in tobacco which nearly doubles from 4.7

Table 2. Estimates of OLS regressions of migrant status on log hourly earnings

	I	II	III	IV	V
Panel A: Tobacco					
Men					
Migrant	0.111 ***	0.114 ***	0.109 ***	0.117 ***	0.107 ***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.024)
R^2	0.641	0.645	0.658	0.667	0.690
N	1032	1032	1032	1032	1032
Women					
Migrant	0.053 ***	0.047 **	0.092 ***	0.092 ***	0.083 ***
	(0.020)	(0.020)	(0.018)	(0.018)	(0.017)
R^2	0.326	0.333	0.412	0.421	0.470
N	2101	2101	2101	2101	2101
Panel B: Printing					
Men					
Migrant	0.049 ***	0.048 ***	0.04 ***	0.039 ***	0.034 ***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)
R^2	0.739	0.739	0.762	0.763	0.785
N	4295	4295	4295	4295	4295
Women					
Migrant	0.118 ***	0.109 ***	0.13 ***	0.125 ***	0.090 ***
	(0.028)	(0.029)	(0.027)	(0.027)	(0.027)
R^2	0.566	0.571	0.591	0.596	0.647
N	1444	1444	1444	1444	1444
Panel C: Mechanical Engineering					
Men					
Migrant	0.091 ***	0.091 ***	0.098 ***	0.096 ***	0.091 ***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
R^2	0.517	0.52	0.56	0.563	0.617
N	6546	6546	6546	6540	6540

Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Included controls: (I): age (as a continuous cubic function). (II): (I) + father's social status (categorical: Manager or professional, skilled, low skilled, unskilled, missing). (III) + experience (as a continuous cubic function). (IV): (III) + civil status (categorical: unmarried, married, previously married), children (categorical by each child), health (dummy). (V): (IV) + dummies for apprentices, members in unions, benefit societies, and temperance societies.
Sources: See Table 1.

log points to 9.2, suggesting that female migrants in tobacco possessed less experience than non-migrants. Estimates of β_1 including further sets of controls are presented in column (III)-(V). The effect on the premium of adding these controls is to decrease the effect somewhat. After including a full set of individual controls a positive earnings premium is still evident in all groups. The overall results from Table 2 indicate that an extensive set of individual characteristics only explain a relatively small part of migrants' basic earnings premium, irrespective of gender.

In order to make more clear-cut comparisons we introduce a number of fixed effects that allows for migrants earnings to be compared to different reference groups. By adding a location or firm fixed effect to (1), the resulting β_1 coefficient on migrant status in the model measures the difference in earnings of migrants relative to natives in a specific destination, or alternatively when adding firm fixed effects, the difference in earnings of migrants relative to coworkers within a firm.

In the location fixed effects model, the β_1 coefficient on migrant status tells us to what extent migrants' earnings conform to the general predictions of positive (a) and negative (b) selection as generated by the Roy model (Borjas 1987: 533-534):

(a) *Positive selection*: If returns to skills are lower at origin than at destination, migrants will constitute the most able persons. Migrants will be drawn from the top of the ability distribution at origin and will outperform the destination's native population in terms of earnings.
(b) *Negative selection*: This is the converse case. If returns to skills are high at origin compared to the destination, migrants will be drawn from the bottom of the ability distribution at origin and will underperform in terms of earnings when compared with the destination's native population.

Hence, if $\beta_1 > 0$, migrants outperform the native population at destination in terms of earnings, a finding that corresponds to migrants being on average positively selected. On the other hand, if $\beta_1 < 0$, migrants fare relatively worse than the native population and this supports negative selection. The results are presented in Tables 3 and 4. The estimations proceeds stepwise by gradually including the set of individual controls described for Table 2.

Table 3. Estimates of current location fixed effects OLS regressions of migrant status on log hourly earnings

	I	II	III	IV	V
Panel A: Tobacco					
Men					
Migrant	0.111***	0.114***	0.109***	0.117***	0.107***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.024)
R^2	0.641	0.645	0.658	0.667	0.69
N	1032	1032	1032	1032	1032
Women					
Migrant	0.053***	0.047**	0.092***	0.092***	0.083***
	(0.020)	(0.020)	(0.018)	(0.018)	(0.017)
R^2	0.326	0.333	0.412	0.421	0.47
N	2101	2101	2101	2101	2101
Panel B: Printing					
Men					
Migrant	0.049***	0.048***	0.04***	0.039***	0.034***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)
R^2	0.739	0.739	0.762	0.763	0.785
N	4295	4295	4295	4295	4295
Women					
Migrant	0.118***	0.109***	0.13***	0.125***	0.09***
	(0.028)	(0.029)	(0.027)	(0.027)	(0.027)
R^2	0.566	0.571	0.591	0.596	0.647
N	1444	1444	1444	1444	1444
Panel C: Mechanical Engineering					
Men					
Migrant	0.091***	0.091***	0.098***	0.096***	0.091***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
R^2	0.517	0.52	0.56	0.563	0.617
N	6546	6546	6546	6540	6540

Notes: See Table 2
Sources: See Table 1.

Table 4. Estimates of firm fixed effects OLS regressions of migrant status on log hourly earnings

	I	II	III	IV	V
Panel A: Tobacco					
Men					
Migrant	0.113***	0.116***	0.113***	0.117***	0.115***
	(0.033)	(0.033)	(0.031)	(0.030)	(0.029)
R^2	0.669	0.672	0.686	0.693	0.701
N	1032	1032	1032	1032	1032
Women					
Migrant	0.028	0.028	0.065***	0.062***	0.066***
	(0.018)	(0.018)	(0.015)	(0.016)	(0.014)
R^2	0.313	0.317	0.414	0.423	0.456
N	2101	2101	2101	2101	2101
Panel B: Printing					
Men					
Migrant	0.031***	0.029**	0.022**	0.02**	0.02**
	(0.012)	(0.011)	(0.010)	(0.010)	(0.009)
R^2	0.762	0.762	0.789	0.791	0.807
N	4295	4295	4295	4295	4295
Women					
Migrant	0.037*	0.034	0.057**	0.051**	0.045*
	(0.022)	(0.024)	(0.022)	(0.023)	(0.024)
R^2	0.595	0.599	0.625	0.629	0.657
N	1444	1444	1444	1444	1444
Panel C: Mechanical Engineering					
Men					
Migrant	0.017	0.016	0.026***	0.026***	0.027***
	(0.011)	(0.012)	(0.008)	(0.008)	(0.007)
R^2	0.595	0.596	0.652	0.653	0.681
N	6546	6546	6546	6540	6540

Notes: See Table 2
Sources: See Table 1.

Table 3 includes a location fixed effect, β_1 thus being identified through a comparison of a difference in earnings between migrants residing in the same location. Table 4 makes an even finer comparison by comparing migrants to their coworkers within the same firm. The impact on β_1 of adding location and firm fixed effects varies widely by industry and sex. For men in tobacco, the estimates increase somewhat (from 10.7 log points without fixed effects to 11.5 log points with firm fixed effects) when considering the most extensive models presented in column (V) although this difference is not statistically significant. For the remaining groups a general reduction is observed, one that is coherent with migrants choosing on the whole a better-paying location as their migration destination. For women in tobacco, the remaining within-firm migrant earnings premium after having added firm fixed effects is 6.6 log points, while for male workers in printing the remaining premium is 2.0 log points, and for women in printing it is 4.5 log points. The largest decrease is noted for workers in mechanical engineering, with the estimate falling from 9.1 log points to 2.7 log points. Despite differences in the estimated premiums and the reductions in these when including fixed effects, the main point we take from the results so far is that a positive and statistically significant earnings premium exists in all industries and for both sexes in favor of migrants both when compared to natives in a specific location and when compared to coworkers within a firm.

Although the earnings premiums of migrants are positive in all industries for both men and women, there may still be differences between the sexes in terms of the magnitude. Formal testing for differences between the migrant coefficients of men and women estimated by the most extensive model in Tables 2, 3 and 4 are presented in columns (I), (II) and (III) of Table 5. In tobacco, the migrant coefficient for men exceeds that of women in a statistically significant manner only when migrants are compared to natives in column (II). In printing, the pattern is reversed. The migrant coefficient of women exceeds that of men when no fixed effects are included, or when migrants are compared to natives. When firm fixed effects are included, the difference between men and women becomes statistically indistinguishable.

One concern is that the estimated earnings premium in the above estimations, even after including location and firm fixed effects, may be biased downwards as a result of migrants not realizing their full earnings capacity during a period of assimilation following migration. After location specific human capital has been acquired by migrants, earnings typically grow faster and, when migrants are positively selected, they eventually

Table 5. Test of equality of migrant coefficient between men and women

	I	II	III
Panel A: Tobacco			
Men	0.107	0.122	0.115
	(0.024)	(0.029)	(0.029)
Women	0.083	0.059	0.066
	(0.017)	(0.009)	(0.014)
χ^2	0.7	6.55**	2.61
Panel B:Printing			
Men	0.034	0.021	0.020
	(0.011)	(0.006)	(0.009)
Women	0.09	0.077	0.045
	(0.027)	(0.014)	(0.024)
χ^2	3.79*	11.36***	1.05
Fixed effects	-	Current location	Firm

Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Included controls: age (as a continuous cubic function), experience (as a continuous cubic function), father's social status (categorical: Manager or professional, skilled, low skilled, unskilled, missing). civil status (categorical: unmarried, married, previously married), children at home (categorical by each child), health (dummy) dummies for apprentices, membership in unions, benefit societies, and temperance societies.
Sources: See Table 1.

overtake the earnings of natives (Chiswick 1978, Borjas 1987). Given that this paper is only considering internal migration, we discount the possibility of significant cultural dissimilarities and language differences affecting the earnings of recent migrants, and instead ascribe any possible negative effects to factors such as a lack of knowledge about local labour markets or personal networks, or a lack of location- or firm-specific skills.

An inspection of age-earnings and experience-earnings profiles of migrants and natives in our sample shows no evidence of the effect of failure to assimilate being severe. The age and experience profiles of migrants are similar in shape to those of the natives, and at no point fall below the level of the natives. This is consistent with skills in this period being of a general nature (within occupation, that is) and not firm-specific, and therefore easily transferable between different employers and locations (Jacoby 1985).

In summary, these results indicate that, for both men and women and across all three of the industries surveyed, migrants earned more than natives

and their fellow coworkers. Although the magnitude varies, the regularity of the positive direction shown by the estimates strongly supports the interpretation that migrants were on average positively selected, and that this was the case regardless of industry and sex. The estimated premium is largest for male tobacco workers, followed in descending order by their female coworkers in tobacco, women in printing, men in mechanical engineering and men working in the printing industry.

Explaining the earnings premium of migrants

The preceding section clearly shows that migrants not only earned more than natives in their chosen destination, they also outperformed coworkers within firms. That observable individual characteristics, location and place of work fail to fully explain migrants' earnings premiums is testament to migrants constituting a select group, potentially differing on a number of unobservable variables. The evidence presented so far is in favor of migrants being on average more able than both the native population and co-workers. However, finding that migrants' earnings were higher, net of differences attributable to observable characteristics and premiums associated with specific location s and firms provides little in terms of explanations for why. We seek next to investigate more precisely the nature of the observed earnings premium by first evaluating the role played by sorting according to ability, thereafter considering differences in work intensity as explanations for migrants' higher earnings.

Sorting and skills

Sorting according to ability in the Roy model implies that migrants choose a destination that is ideal in terms of compatibility between one's own skills and the return to skills at destination.[15] The results presented in Table 6 shows that the earnings premium of migrants is consistent with a positive

[15] For historical applications of the Roy model, see for example Salisbury (2014) and Stolz and Baten (2012).

selection on average according to the predictions of the Roy model. To further assess whether sorting was an important factor underlying migrants earnings premiums, we proceed by estimating a Roy selection model on the basis of conditioning earnings on returns to skills at an individual's origin and destination (using the 25 by 25 kilometers grid defined in the data section above). We take the same approach as Borjas et al. (1993), and approximate returns to skill on the basis of wage dispersion in a location, given more precisely by the standard deviation of hourly earnings by industry and sex (see Juhn et al. (1993) for a discussion about the relationship between wage dispersion and returns to skills).

Given the data at hand, this approach has some unavoidable shortcomings. Because some migrants originated from locations in which no firms were operating at the time of the survey, thus prohibiting the calculation of returns to skills for these migrants' origins, our sample is somewhat reduced. Moreover, because the data at hand is cross-sectional, we assume that the current wage dispersion in a location is a viable proxy for the wage dispersion that prevailed at the time of migration. On the upside, our approach is preferable to more commonly used measures of returns to skills since it records actual dispersion within an industry, arguably a better approximation for returns to skills than population-based measures of dispersion such as the Gini coefficient or top-to-bottom income ratios (see e.g., Cobb-Clark 1993 and Feliciano 2005).

We proceed by further assessing migrants' earnings by estimating the following model:

$$ln\ \mathrm{Y} = \beta_0 + \beta_1 \mathrm{Migrant} + \beta_2 \mathrm{R_d} + \beta_3 \mathrm{Migrant}*\mathrm{R_o} + \beta_4 \mathrm{Migrant}*\mathrm{R_d} + \beta_k \mathbf{X} + \varepsilon \quad (2)$$

where R_o denotes return to skill at individual's origin and R_d denotes return to skill at individual's current location. To facilitate comparisons and interpretation, R_o and R_d have been standardized (mean = 0 and standard deviation = 1). Selection in the Roy model implies that migrants leaving origins with low returns to skills for destinations with high returns to skills will be positively selected and vice versa. Moreover, individuals whose skills are complementary to the return to skill in a location (i.e., high skill-high return or low skill-low return) are more likely to remain (for stayers, R_o is naturally equal to R_d). Evidence of selection according to the Roy model is thus consummate with $\beta_2 > 0$, $\beta_3 < 0$ and $\beta_4 > 0$ in equation (2). The results from estimations by OLS are presented in Table 6.

Table 6. Estimates of OLS regressions of migrant status and returns to skill on log hourly earnings

	Men			Women		
	I	II	III	I	II	III
Panel A: Tobacco						
Migrant	0.113***			0.059***		
	(0.024)			(0.022)		
Migrant x R_o	-0.033			-0.067***		
	(0.023)			(0.016)		
Migrant x R_d	0.098***	0.079***	0.089***	0.074***	0.062***	0.070***
	(0.019)	(0.019)	(0.017)	(0.018)	(0.016)	(0.013)
R_d	0.067***			0.079***		
	(0.016)			(0.008)		
R^2	0.711	0.475	0.45	0.499	0.528	0.502
N	979	320	320	1916	408	408
Panel B: Printing						
Migrant	0.041***			0.074**		
	(0.012)			(0.031)		
Migrant x R_o	-0.036***			-0.095***		
	(0.010)			(0.020)		
Migrant x R_d	0.042***	0.039***	0.045***	0.117***	0.116***	0.111***
	(0.014)	(0.013)	(0.015)	(0.024)	(0.028)	(0.021)
R_d	0.046***			0.125***		
	(0.007)			(0.011)		
R^2	0.791	0.619	0.63	0.695	0.659	0.666
N	4226	1254	1254	1376	203	203
Panel C: Mechanical engineering						
Migrant	0.069***					
	(0.008)					
Migrant x R_o	-0.074***					
	(0.007)					
Migrant x R_d	0.020***	0.058***	0.058***			
	(0.007)	(0.005)	(0.014)			
R_d	0.044***					
	(0.003)					
R^2	0.636	0.579	0.572			
N	5558	1957	1957			
Fixed effects	None	None	Birth location	None	None	Birth location

Notes: Robust standard errors, calculated using the Huber White sandwich estimator, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Included controls: age (as a continuous cubic function), experience (as a continuous cubic function), father's social status (categorical: Manager or professional, skilled, low skilled, unskilled, missing). civil status (categorical: unmarried, married, previously married), children at home (categorical by each child), health (dummy) dummies for apprentices, membership in unions, benefit societies, and temperance societies.
Sources: See Table 1.

Overall, the estimated coefficients conform to the expectations yielded by the Roy selection model. The effect of the origin skill premium has a negative effect on migrants' earnings while the destination skill premium has a positive effect. As predicted, the earnings of stayers increase along with theskill premium in the origin. The migration term is positive for all groups, reflecting either gains from migration or the fact that migrants are positively selected through some other mechanism other than sorting. The standardization of R_o and R_d means that the magnitude of the estimates is easily interpretable, a change in one standard deviation in the origin or destination skill premium being consummate with a change in log hourly earnings reported in Table 6. To give an example, a female tobacco migrant worker from a location with a skill premium one standard deviation above the mean and coming to a destination with a return to skill equal to the mean would thus have been negatively selected and would have earned 6.7 log points less than the average migrant. A coworker making a move in the opposite direction, from a location with a return to skill equal to the mean to a destination lying one standard deviation above the mean, would thus have earned 7.4 log points more than the average migrant, illustrating the range of selection outcomes.

An alternative approach to testing the Roy model is to make sorting across locations conditional on migration. Although a more restricted model, it has the advantage of allowing for the inclusion of migrants from locations for which we have no estimate of the skill-premium. Column (II) in Table 6 thus presents results from models using a sample consisting of only migrants. The reported coefficient denotes the effect of the return to skill in the migrants chosen destination on hourly earnings. In column (II) a birth location fixed effect is added. On the whole the results for these restricted models are very similar to the more extensive model presented in column (I).

One possible reason for the relatively stronger results for women is that our measure of the skill premium also captures to some extent the general career opportunities for women in a given location, since the existence of a glass ceiling in that location would have affected the wage dispersion of women negatively. Able female migrants in the case of what we define as a high return to skill location may thus also have gained from better prospects for upward occupational mobility, and also benefited from any gain associated with better matching between individual ability and the return to skill.

Effort

The explanation underlying selection considered so far stem from presumed differences in ability or skills in favor of migrants. However, differences in observed earnings may also be a result of migrants exerting more effort by supplying more labour, on the extensive or intensive margin (Galor and Stark 1991). Simply put, if migrants work longer and/or harder, an earnings premium interpreted as a difference in ability and skills may in reality be the result of differences in motivation and effort. *A priori* there are a number of plausible reasons why migrants may choose to work longer hours and harder. To begin with, both moving to another location and working more diligently may be jointly determined if migrants are innately more ambitious and hard-working. Moreover, costs or changes in preferences associated with moving *per se* can affect earnings through an increase in working hours and intensity. Moving from one's place of origin typically means leaving behind, if not all, then at least most of one's friends and family, resulting in fewer opportunities for socializing when off work, and implying that migrants may value leisure time to a lesser extent than do their native coworkers (Schaeffer 1995). Moreover, migration incurs certain costs such as travelling expenses (associated with not only the initial move but also a recurring expense when visiting one's place of origin) and, for young migrants in particular, the costs associated with setting up a new and independent household. To cover these costs it is plausible that migrants may target a higher income and as a result expend greater effort.

So far differences in overall (that is, daily, weekly or yearly) earnings attributable to working additional hours have been discounted on purpose in this analysis through the use of hourly earnings as the dependent variable. To test whether migrants expended more effort, we begin by considering increases on the extensive margin. To test whether migrants worked longer hours, we re-estimate equation (1) but include time worked as the dependent variable (measured as minutes per week). We only include individuals with self-reported hours of work. This shows to the sample of male tobacco workers in particular being severely reduced. The results from OLS estimations are presented in Table 7 in Columns (I) and (II).

Table 7. Estimates of OLS regressions of relationship between migrant status and effort

Dependent variable:	Minutes worked per week		Log hourly earnings	
	(I)	(II)	(III)	(IV)
Panel A: Tobacco				
Men				
Migrant	45.064	38.652	0.008	0.036
	(33.429)	(24.468)	(0.031)	(0.042)
R^2	0.149	0.093	0.681	0.624
N	276	276	336	336
Women				
Migrant	47.599***	32.349*	0.052**	0.064***
	(18.141)	(16.296)	(0.022)	(0.017)
R^2	0.121	0.102	0.43	0.394
N	1555	1555	870	870
Panel B: Printing				
Men				
Migrant	-8.824	-8.686		
	(8.827)	(6.955)		
R^2	0.019	0.025		
N	4295	4295		
Women				
Migrant	-8.812	2.982		
	(18.065)	(13.802)		
R^2	0.205	0.136		
N	1444	1444		
Panel C: Mechanical engineering				
Men				
Migrant	5.387	-2.28		
	(4.661)	(2.973)		
R^2	0.045	0.012		
N	6540	6540		
Fixed effects	None	Firm	Piece rate	Piece rate and firm

Notes: See Table 6.
Sources: See Table 1.

Only in tobacco do we find a difference in working hours between migrants and non-migrants. Female migrants in tobacco worked on average 48 minutes longer per week than non-migrants, the difference falling to 32 minutes after including firm fixed effects. The estimates for male tobacco workers are almost identical although not statistically significant at conventional levels, possibly due to the small number of observations. In printing and mechanical engineering, no discernible difference is found in the hours of work supplied by migrants. This difference may be explained by the nature of the work in the three industries. In tobacco, workers were predominantly paid by the piece and had more discretion in deciding how many hours to work than did those in printing and mechanical engineering

who to a greater extent relied on hourly wage rates and enforced a more regular work schedule.

For one occupational group, cigar makers, we have information about actual piece rates paid. This enables us to test for an increase in the supply of labour on the intensive margin. To test whether migrants worked harder than natives, we include a piece rate fixed effect in a re-estimation of equation (1), thus comparing workers being paid exactly the same piece rate. The results are presented in Table 7, columns (III) and (IV), which adds firm fixed effects. Again, the only impact we find is on women, with the extra effort expended by female cigar workers resulting in a premium of 6.1 log points relative to their coworkers.

Conclusion

This paper adds to the historical literature on labour markets and migrant earnings. Significantly, we contribute by considering female migrants, which has been a previously understudied group. By addressing the gender dimension, we shed light on a neglected aspect of the literature on historical labour markets as well as the literature on migrant selection.

Our first and most basic finding is that male and female migrants outperformed both the native population and coworkers in terms of earnings in all three industries. These earnings premiums remain even after including an extensive set of controls. This general earnings pattern suggests that migrants differed from the rest of the population in some unobserved way. Our extended analysis shows that the earnings of migrants correspond to sorting according to ability as predicted by the Roy model. The results demonstrate that both female and male migrants' earnings varied a great deal depending on the returns to skills at both the origin and destination of the migrants. Migrants were thus, irrespective of gender, able to capitalize on their abilities and industriousness by moving into locations that rewarded these traits. Our results support the conclusion that workers who migrated rightfully anticipated wage gains as they sorted into locations with returns to skill that were consummate with individual ability. This suggests that workers, irrespective of gender, were well informed about skill premiums in different locations and acted upon this information. This is indicative of an

efficient and well-functioning market characterized by labour market flexibility.

Knowledge about how migrants fared has bearing on our understanding of their circumstances and career prospects and enables us to address questions regarding the supply and recruitment of labour during industrialization. The intensification of economic activity during the nineteenth century meant that the demand for labour was becoming ever more localized and concentrated in urban areas and centers of manufacturing. Business owners and managers were thus increasingly forced to look beyond the native population in order to secure a ready supply of labour. Evaluating the productivity and earnings of migrants informs us as to what extent migrants constituted desirable employees relative to the native population. Our analysis implies that migrants were a valuable resource for employers: Migrants were on average more skilled and able than the native population and their native-born coworkers and, in some cases, also harder working. Expanding firms were thus not hampered by relying on labour supplied by locals, but were also able to tap into a on a stream of able and qualified migrant labour.

Although women may have occupied a marginal position relative to men in the formal economy around the year 1900, women migrants were by no means marginal individuals. When comparing the results for men and women, similarities trump any differences. The results suggest that overall men and women were subject to the same basic economic processes. The fact that men and women were similar in this respect should factor into our understanding of internal migration during industrialization.

These findings also have wider implications. Migration is an important feature of the second industrial revolution, which is reflected by the high incidence of migration in both the Swedish population and in our sample. Evidence in favor of migrants being positively selected implies that migration may have served to exaggerate regional inequality, with the sending regions being stripped of their most able workers to the benefit of the receiving regions. Yet the results of the Roy selection model also indicate a converging process in which migrants acted upon regional differences in returns to skills; a process which over time would have served to equalize skill premiums across regions.

References

Abramitzky, R., Platt Boustan, and L. Eriksson, K. (2012). Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration. *American Economic Review*, 102: 1832-1856.

Abramitzky, R., Platt Boustan, L. and Eriksson, K. (2013). Have the poor always been less likely to migrate? Evidence from inheritance practices during the age of mass migration. *Journal of Development Economics*, 102: 2-14.

Angrist, J. D. and Pischke, S. J. (2009). *Mostly Harmless Econometrics: An Empiricists Companion.* Princeton: University Press.

Baines, D. (1985). Migration in a Mature Economy: Emigration and Internal Migration in England and Wales 1861-1900. Cambridge: Cambridge University Press.

Baines, D. (1994). European labour markets, emigration and internal migration, 1850-1913. In T. J. Hatton and J. G. Williamson (Eds.), *Migration and the International Labour Market, 1850-1900*. London: Routledge.

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

Borjas, G. J. (1991). Immigration and self-selection. In J. M. Abowd and R. B. Freeman (Eds.), *Immigration, Trade, and the Labor Market.* Chicago: University of Chicago Press.

Borjas, G. J. (1994). The economics of immigration. *Journal of Economic Literature*, 32: 1667-1717.

Borjas, G. J., Bronars, S. G. and Trejo, S. J. (1993). Self-selection and internal migration in the United States. *Journal of Urban Economics*, 32: 159-185.

Boyer, G. R. and Hatton T. J. (1997). Migration and labour market integration in late nineteenth-century England and Wales. *Economic History Review*, 50: 697-734.

Burnette, J. (2008). *Gender, Work and Wages in Industrial Revolution Britain.* Cambridge: Cambridge University Press.

Chandler, A. D., Jr. (1977). *The Visible Hand.* Cambridge: The Belknap Press.

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

Chiswick, B. (1999). Are immigrants favorably self-selected? *The American Economic Review*, 89: 181-185.

Cobb-Clark, D. A. (1993). Immigrant selectivity and wages: The evidence for women. *American Economic Review*, 83: 986-993.

Collins, W. J. and Wanamaker, M. H. (2014). Selection and economic gains in the Great Migration of African Americans: New evidence from linked census data. *American Economic Journal: Applied Economics*, 6: 220-252.

Crafts, N. F. R. (1985). British Economic Growth during the Industrial Revolution*. New York: Clarendon Press.*

Deltas, G., Sicotte, R. and Tomczak, P. (2008). Passenger shipping cartels and their effect on transatlantic migration. *Review of Economics and Statistics*, 90: 119-133.

Dribe, M. and Lundh, C. (2005). Gender aspects of inheritance strategies and land transmission in rural Scania, Sweden, 1720–1840. *History of the Family*, 10: 293-308.

Dublin, T. (1994). Transforming Women's Work: New England Lives in the Industrial Revolution. Ithaca: Cornell University Press.

Ekdahl, L. (1983). *Arbete mot kapital.* Lund: Studentlitteratur.

Elmquist, H. (1899). *Undersökning af tobaksindustrien i Sverige*. Stockholm: Kommerskollegium.

Elmquist, H. (1901). Undersökning af den mekaniska verkstadsindustrien i Sverige. Stockholm: Kommerskollegium.

Elmquist, H. (1909). Undersökning af tryckerier och därmed förenad handtering i Sverige. Stockholm: Kommerskollegium.

Feliciano, C. (2001). The skill and economic performance of Mexican immigrants from 1910 to 1990. *Explorations in Economic History*, 38: 386-409.

Feliciano, C. (2005). Educational selectivity in U.S. migration: How do immigrants compare to those left behind? *Demography*, 42: 131-152.

Galor, O. and Stark. O. (1991). The probability of return migration, migrants' work effort, and migrants' performance. *Journal of Development Economics*, 35: 399-405.

Goldin, C. (1990). Understanding the Gender Gap: An Economic History of American Women. New York: Oxford University Press.

Gordon, W. M. (2002). Mill Girls and Strangers: Single Women's Independent Migration in England, Scotland and the United States 1850-1881. Albany: State University of New York Press.

Hareven, T. K. (1982). Family Time and Industrial Time: The Relationship between the Family and Work in a New England Industrial Community. New York: Cambridge University Press.

Hatton, T. J. (1997). The immigrant assimilation puzzle in late nineteenth-century America. *Journal of Economic History*, 57: 34-62.

Hatton, T.J. and Bailey, R.E. (2002). Native and migrants in the London labour market, 1929-1931, *Journal of population Economics*, 15: 59-81.

Hatton, T. J. and Williamson, J. G. (2008). Global Migration and the World Economy. Two Centuries of Policy and Performance. Cambridge: MIT Press.

Humphries, J. and Leunig, T. (2009). Was Dick Whittington taller than those he left behind? Anthropometric measures, migration and the quality of life in early nineteenth century London? *Explorations in Economic History*, 46: 120-131.

Jacoby, S. M. (1985). Industrial labor mobility in historical perspective. *Industrial Relations* 22: 261-282.

Juhn, C., Murphy, K. M. and Brooks, P. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 101: 410-442.

Kanbur, R. and Rapoport, H. (2005). Migration selectivity and the evolution of spatial inequality. *Journal of Economic Geography,* 5: 43-57.

Keeling, D. (1999). The transport revolution and trans-Atlantic migration, 1850-1914. *Research in Economic History*, 19: 39-74.

Kommerskollegium (1863-1902). Bidrag till Sveriges officiella statistik. D, Kommerskollegii underdåniga berättelse 1860–1900. Stockholm.

Kosack, E. and Ward, Z. (2014). Who crossed the border? Self-selection of Mexican migrants in the early twentieth century. *Journal of Economic History*, 74: 1015-1044.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99: 483-499.

Landes, D. S. (1969). *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present*. Cambridge: Cambridge University Press.

Lee, E. S. (1966). A theory of migration. *Demography*, 3: 47-57.

Lindbom, T. and Kuhm, E. (1940). *Tobaksabetarnas förbund i Sverige 1889-1939*. Stockholm: Tiden.

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

Maas, I. and van Leeuwen, M. H. D. (2004). Occupational careers of the total male labour force during industrialization: the example of nineteenth-century Sweden. In D. Mitch, J. Brown and M. H. D. van Leeuwen (Eds.), *Origins of the Modern Career*. Aldershot: Ashgate.

Marshall, A. (1920). *Principles of Economics*, 8th edition. London: Macmillan

McKenzie, D. and Rapoport, D. (2010). Self-selection patterns in Mexico-U.S migration: The role of migration networks. *Review of Economics and Statistics*, 92: 811-821.

Mincer, J. (1978). Family migration decisions. *Journal of Political Economy*, 86: 749-773.

Minns, C. (2000). Income, cohort effect, and occupational mobility: A new look at immigration to the United States at the turn of the 20th century. *Explorations in Economic History*, 37: 326-350.

Mitchell, B. R. (1981). *European Historical Statistics, 1750-1975*. London: Macmillan.

Moch, L. P. (1992). *Moving Europeans: Migration in Western Europe since 1650*. Bloomington: Indiana University Press.

Parish, W. L. (1973). Internal migration and modernization: The European case. *Economic Development and Cultural Change*, 21: 591-609.

Ravenstein, E. G. (1885). The laws of migration. *Journal of the Royal Statistical Society,* 52: 167-235.

Richards, E. (1974). Women in the British economy since about 1700: an interpretation. *History*, 59: 337-357.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3: 135-146.

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

Sawyer, C. (Ed.) (1949). Historical Statistics of the United States 1789–1945: a supplement to the Statistical Abstract of the United States. Washington D.C.: United States Department of Commerce.

Schaeffer, P. (1995). The work effort and consumption of immigrants as a function of their assimilation. *International Economic Review*, 36: 625-642.

Schön, L. (2010). Sweden's Road to Modernity: An Economic History. Stockholm: SNS.

Sharpe, P. (Ed.), (2001). Women, Gender and Labour Migration: Historical and Global perspectives. London: Routledge.

Silvestre, J., Ayuda, M.A. and Vicente, P. (2015). The occupational attainment of migrants and natives in Barcelona, 1930, *The Economic History Review*, forthcoming.

Sjastaad, L. A. (1962). The costs and returns of human migration. *Journal of Political Economy*, 70: 80-93.

Slichter, S. (1919). *The Turnover of Factory Labor*. New York: D. Appleton & Co.

Stewart, J. I. (2006) Migration to the agricultural frontier and economic mobility, 1860-1880. *Explorations in Economic History*, 43: 547-577.

Stewart, J. I. (2012). Migration to US frontier cities and job opportunity, 1860-1880. *Explorations in Economic History*, 49: 528-542.

Stolz, Y. and Baten, J. (2012). Brain drain in the age of mass migration: does relative inequality explain migrant selectivity? *Explorations in Economic History*, 2: 205-220.

Thernstrom, S. (1973). The Other Bostonians. Poverty and Progress in the American Metropolis. Cambridge: Harvard University Press.

Trent, J. A. and Steidl, A. (2012). Gender and the "Laws of Migration": A reconsideration of nineteenth-century patterns. *Social Science History*, 36: 223-241.

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.

Wegge, S. (2002). Occupational self-selection of European emigrants: Evidence from nineteenth-century Hesse Cassel. *European Review of Economic History*, 6: 365-394.

Data sources

De större mekaniska verkstäderna 1899, Statistiska avdelningen, HIII c:1, Kommerskollegiets arkiv, National Archives (*Riksarkivet*), Stockholm.

Specialundersökningar Tobaksindustrien 1898, Statistiska avdelningen, HIII b:1 and HIII b:1 aa volume 1, Kommerskollegiets arkiv, National Archives (*Riksarkivet*), Stockholm.

*Undersökning av tryckerier mm 190*3, Avdelningen för arbetsstatstik, HII a:1 volumes 1-6 and HII a:2 volumes 1-12, Kommerskollegiets arkiv, National Archives (*Riksarkivet*). Stockholm.

Paper 3

A Link to the Past: Linking and evaluating Swedish historical censuses

Björn Eriksson
Department of Economic History, Lund University

Abstract
Digitized historical censuses provide a new and rich source material for social scientists. Thanks to increased computational power and methodological advances, the potential to create extensive panel data sets from historical censuses has become a reality. This paper describes a linking process which emphasizes transparency and relies on an absolute minimum of judgement calls. The methods followed are completely replicable and require a minimum of labour input beyond basic programming. To this end, the process of linking three Swedish historical censuses enumerated in 1880, 1890 and 1900 is described.

JEL classification: C81, C83, C87, C40

Key words: Quantitative methods, Microeconomic data

Introduction

Digitized historical censuses provide a new and rich source material for social scientists. Thanks to increased computational power and methodological advances, the potential to create extensive panel data sets from historical censuses has become a reality. The continuous release of complete count censuses means that, for the first time ever, the life courses of full count historical populations over long time spans can be studied. By tracking individuals over time and space, a number of long-standing but pertinent historical questions are being re-examined. Using linked samples drawn from American and British censuses, Long and Ferrie (2013) contrast the aggregate social mobility of the two leading industrial nations at the end of the nineteenth century. Individual predictors of social mobility such as public education (Parman 2011), family resources (Bleakley and Ferrie 2013) and migration (Long 2005; Stewart 2006; Abramitzky et al. 2012; Collins & Wanamaker 2014) have correspondingly been investigated using samples based on censuses linked both within and between countries. Although important, these studies provide only a glimpse of the future potential of historical census data (see Ruggles 2014). The comprehensive nature of census data also lends itself to combination with other sources such as death records and geographic information, a feature which enhances its usefulness and applicability even further. The creation of panels based on multiple censuses offers much promise, but it also means that the linking method itself and the effect it has on linkage rate and sample quality becomes even more important.

The purpose of this paper is to present a linking process that results in a high quality linked sample. The linking process described emphasizes transparency and relies on an absolute minimum of judgement calls. The methods followed are completely replicable and require a minimum of labour input beyond basic programming. To this end, this paper describes the linking of three Swedish historical censuses enumerated in 1880, 1890 and 1900, and then evaluates the resulting samples. This evaluation is based on the number of links achieved and the quality of the resulting sample in terms of representativeness and the occurrence of false positive links. A distinguishing feature of the Swedish censuses is that the material therein allows for the linking of women throughout their life course, after marriage as well. It is thus possible to consider how the application of methods of record linking affects men and women differently.

The usefulness of any sample is determined by its size and quality. Converting cross-sectional data into panel form greatly increases its applicability. However, it also raises valid concerns about the impact that the linking process has on the quality of the resulting sample. This means that the linking of individuals between censuses is a critical step on the path from raw data to final research results. Because of the increasing popularity and importance of linked census-based samples, the linking process itself and the quality of the resulting samples warrant further scrutiny. The extent of bias and false positives in a linked sample poses serious threats to research validity, while the final size determines its applicability and statistical power. Normally linkage rates, and hence final sample size, are of secondary concern when linking censuses. Methods that result in samples free from bias and representative of the underlying population are the ones favoured instead (Ruggles 2006). This is often a sound strategy, because censuses normally contain enough observations to yield a feasible sample even when linkage rates are low. However, too low a linkage rat may prove problematic if it leads to an unacceptable depreciation of the final sample when multiple-linked censuses are joined together.

While striving to achieve a minimum in terms of bias and false positive links and a high rate of linked observations is important when creating linked samples, it becomes even more important when linking multiple censuses. If a linkage method results in biased and inaccurate links between two censuses, we can expect an increase in bias and a further decrease in the quality of the links as additional censuses are linked together. What will have appeared at first to be a small bias or low rate of false positives may with each additional linked census aggregate to pose a significant problem. Methods are therefore required that result in a high linkage rate which does not come at the expense of a decline in sample quality.

Source material

At present, three complete historical censuses in Sweden (dated 1880, 1890 and 1900) have been digitized and coded by the Swedish National Archives

(*Riksarkivet*) and the Stockholm City Archives (*Stadsarkivet*).[16] The coding of the censuses has been done in accordance with the principles set out by the North Atlantic Population Project (the Swedish National Archives 2012).[17]

The Swedish censuses differ from the US and British censuses in that they are not the product of census-taking by enumerators visiting and counting those who make up the populace. They are instead, with the exception of the city of Stockholm, the result of a compilation of excerpts from continuous parish registers which were kept by the Swedish Lutheran Church and maintained by the parish priest. The census source in the case of Stockholm consists of excerpts from the *Roteman* register, an administrative register supervised by *Mantalsnämnden* (the population and tax registration board) which replaced the church registers in 1878 in order to cope with the rapidly increasing and mobile population of Stockholm at the end of the nineteenth century (Geschwind & Fogelvik 2000:207-208).

Because Swedish censuses are no more than excerpts from a continuous and consistent source rather than the recreation of a population register every ten years, as in the case of the US and British censuses, one can expect the raw data to be of comparatively better quality. Common errors resulting from misreporting or the recording of mistakes on the part of enumerators may thus be largely discounted. Moreover, because the data on Swedes were entered in the parish books at the time of christening and not removed until the time of death or emigration, the under-enumeration of the population as a whole and of specific groups is less of a problem than what might normally expected of historical censuses. Finally, thanks to a peculiarity of the Swedish censuses, it is possible to link not only men but also women between them. Unlike most other censuses, the majority of married women appear with their maiden name recorded as their last name instead of that of their spouse, the exception once again being Stockholm, where many women

[16] The coding of the censuses was funded by a Swedish Research Council grant and the North Atlantic Population Project at the Minnesota Population Centre.

[17] In the near future, these censuses will be supplemented by two more, the 1910 and 1930 censuses, which are currently being processed. Once completed, these digitized censuses will provide full coverage of the entire Swedish population over half a century, starting in 1880 – an important historical period in which Sweden went from being a relatively backward Western European country to a developed industrial nation. By linking individuals between these censuses, it will be possible to study the life course of the generation which experienced this important historical period.

appear with no last name recorded. This allows for a final sample that is broadly representative of the population as a whole and includes both sexes.

The Swedish censuses are, however, not without limitations. One major shortcoming is that a large proportion of individuals appear with no surname recorded. Individuals appearing in a census without a surname fall into one of two categories: married women residing primarily in Stockholm County, and children still living at home with their parents. A child's surname may be derived from the father's name in order to link it to future or past censuses. In the case of married women with no maiden name recorded, there is no straightforward way of linking them to past censuses in which they appeared as unmarried.

Another data quality issue relevant to linking is the variation in the spelling of names. Most spelling variation stem from the source material itself and is the result of inconsistency on the part of the parish priest. The letters *k* and *ch* were often used interchangeably in spelling, as were *t* and *th*, resulting in spelling variations of the same name. For example, the name Kristina was sometimes spelled Christina, and at other times as Kristhina. Apart from these errors in the original sources, further typing and spelling errors were introduced during the digitizing of the censuses as a result of misreading the handwritten source material (see Kuckich 1992 for a discussion of typing and spelling errors). In order to deal with these issues, an approach to linking is required that allows for inconsistencies in the source material while at the same time producing high quality links.

Linking procedure

The linking of individuals between censuses can be either deterministic of probabilistic (Herzog et al. 2007:82-83). A deterministic approach only allows for exact matches when creating links. Modern identity numbers are a prime example, which makes the free linking of individuals easy over time and across various sources. Historical data requires a more forgiving approach that takes into account the inconsistencies and errors often found in censuses and parish and tax registers. These problems were initially dealt with by finding potential matches manually and judging on the basis of the evidence at hand whether the two observations were indeed the same individual (for examples of these see Thernström 1964 and Alter 1988).

Linking a particular individual was thus "a very labour-intensive process" often reliant on "a highly varied and opportunistic linkage strategy" (Alter 1988:52).

Despite early pioneering work in the probabilistic methods of linking (Newcombe el al. 1959; Fellegi and Sunter 1969), technological limitations have meant that the automated linking of large numbers of individuals has only become feasible in the last couple of decades (Ruggles 2006). Increased computing power and the digitalization of historical censuses have seen the creation of a number of linked samples derived from historical populations. Recent examples include linked samples of US censuses (Long and Ferrie 2013; Collins and Wanamaker 2014; Abramitzky et al. 2012) and UK censuses (Long and Ferrie 2013), and the ongoing work of the North Atlantic Population Project (Ruggles et al. 2011). For Sweden, different methods of record linking have been tested for the 1900 and 1890 censuses using the populations of the two northern parishes of Byske and Skellefteå (Wisselgren et al. 2014). Since this paper also uses Swedish historical censuses, the methods employed are in some respects similar to those in Wisselgren et al. (2014), particularly as regards the treatment of missing surnames and patronyms. The scale of the exercise described in this paper and the methods relating to the classification of links as true or false do, however, differ significantly. By linking the complete censuses of 1900, 1890 and 1880, this paper highlights the value of achieving high linkage rates where multiple censuses are linked together. Moreover, this paper employs a method of classifying links as true or false which is completely replicable and does not rely on manually constructed training data.

Because censuses differ in terms of how the population was enumerated and which information was collected, the approach to linking also necessarily differs depending on the source material. The methods used for linking must therefore be adopted in order to meet source-specific challenges. While differences exists in the methods used by researchers when linking census data, general agreement exists on three overall objectives set in order to produce a high quality linked sample (Feigenbaum 2014:3; Mill 2013:85; Ruggles 2006):

(a) The linking process should produce a final linked sample that is representative of the original cross-sectional records. Two sources of bias are possible. The first potential source of bias is inherent in the source material itself and relates to the accuracy with which information was recorded at the time of the census-taking. If certain

groups are susceptible to under-enumeration (e.g. Hacker 2013) and thus only appear intermittently in the source material, the probability of their being linked between censuses naturally decreases. Moreover, even if the enumeration is complete, the accuracy of the recording of information such as name and age will affect the probability of whether a link is deemed true or not. This is problematic where less educated individuals are more likely to be uncertain about their age or the exact spelling of their names (see, for example, A'Hearn et al. 2009). A further bias may be introduced by the researcher if a method is chosen which favours the linking of certain individuals above others. It thus falls on the researcher to follow a linking process which minimizes any further bias beyond that inherent in the source material.

(b) The linking process should result in as few false positives as possible. While an unrepresentative sample may be addressed through weighting, a high occurrence of false positive links is more problematic. A false positive link constitutes an erroneous link between two separate individuals rather than a true link between the same individuals observed at different points in time. Naturally, the heterogeneity in characteristics between the two observations that constitute a false positive link will greatly exceed that of a true link. The severity of the problem may be further emphasized when considering the question of intergenerational mobility, whereby a perfect socially mobile society is one in which parental status has no effect on the status of the offspring, meaning that the distribution of parental occupations is independent of the parents' children. A false positive link will in this case bias the results in favour of high mobility, since false positive links by construction reveal either the absence or only small extent of status correlation between two points in time.

(c) Finally, the linking process should strive for a high linkage rate. The successful linking of a large number of individuals will increase the future usefulness of the eventual sample. To what extent high linkage rates are of primary concern depends on the intended use of the sample. If the initial cross-sections from which links are made are small in size, or if the purpose of future research is to study narrowly defined groups or rare events, linkage rates are important in achieving a final sample size large enough to be usable.

Achieving a high linkage rate is also important where the ultimate aim is to link not only two but multiple censuses together. While a linkage rate of, say, 25 percent may be satisfactory when considering only two time periods, this will most likely be inadequate when linking multiple censuses together. This is because of the gradual attrition of the sample which will result from the addition of each census: assuming that the linkage rate is fixed over time and randomly assigned at the individual level, the initial two census sample sizes of 25 per cent will decline to 6.25 per cent (=0.25^2) after adding a third census, and to 1.56 per cent (=0.25^3) after adding a fourth, and so on. Clearly, if the aim is to be able to study the complete life course of an individual using multiple censuses, the issue of linkage rates may not be ignored.

Potential trade-offs exist between the above objectives. A high linkage rate is achievable at the expense of representativeness by using auxiliary information which may have a biasing effect on the final sample. A low level of false positives is achievable by considering only those individuals with a low probability of being erroneously linked, such as persons with a rare name. This comes at the cost of a decrease in the representativeness of the linked samples and the linkage rate.

In practice, we cannot know whether a link is true or false. A classifier is thus required which will deem links to be either true or false based on the calculated similarity scores. One popular method is to employ machine learning tools. This is the approach followed by Wisselgren et al. (2014) and the Minnesota Population Center (Ruggles et al. 2011), to name a couple of examples. By using pre-existing samples of manually confirmed links, a program is trained to recognise true links from false by assigning a confidence score. The downside of this method is that it requires a data set which has already been linked. Normally the training data employed has been created using manual linking. This aspect unavoidably introduces a feature in the linking process which is based on human judgement and is therefore not replicable. Moreover, this method requires the creation of new training data for each project, a factor which increases costs.

The method advocated in this paper requires no training data. Instead, I rely on auxiliary information not used for linking (second names) to measure

the accuracy of matches between individuals in the censuses. This approach has some distinct advantages. Firstly, it is cheap and easy to implement using standard software.[18] Secondly, it is entirely transparent and replicable, relying on an absolute minimum of judgement calls. Finally, and most importantly, it allows the researcher to make informed decisions based on both the quantity and accuracy of the links made when choosing the level at which to classify links as true or false.

The linking of the Swedish historical censuses consists of four steps. The first involves some very basic standardization of the data, followed by a probabilistic matching between the censuses of select variables using a matching algorithm. When carrying out probabilistic matching, it is not uncommon for an individual to end up with more than one probable match which results in duplicates. The third step therefore involves the removal of any duplicate and ambiguous links. Finally, the links are evaluated in terms of accuracy and representativeness.

Standardization and matching

Any data-linking project starts with identifying which variables are suitable for matching individuals. In order not to 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. The variables available that fulfil the criteria of being fixed over time include birth year, birthplace, sex, and names. Birth year, sex, and birth place do not suffer from the problems of variation in spelling associated with names, and are therefore used to index the data. Some parishes of birth have been aggregated in order to ensure time-consistent boundaries between 1880, 1890 and 1900.[19] In practice, indexing implies that individuals are only compared between censuses if there is an exact match of year and parish of birth and sex between the two censuses. Using three variables to index the data means that

[18] All stages of the linking were implemented using Stata 13.

[19] I am grateful to Sebastian Klusener for being able to use his coding for defining time-consistent parishes of birth.

the computing power required is also much reduced since the number of comparisons always increases exponentially when data is not indexed, while the number of potential matches increases only linearly.

Names (first names and surnames) are thus the only remaining variables used for probabilistic linking. To make sure that the number of names held by a person does not influence the probability of being linked, only the first name and surname recorded are used (whichever name appears as the first entry in the census). An alternative strategy was to extract in alphabetical order the first ordered name. This resulted, however, in an unwanted homogenization of names, particularly in the case of women. After arranging the first names in alphabetical order, a large proportion of women ended up with the common name 'Anna'. This in turn led to an unacceptable loss of links when removing duplicate matches.

A peculiarity of the Swedish censuses is that the majority of children residing with their parents have no surname recorded. Normally this would not present a problem since it may be inferred that a child's surname, although not explicitly recorded, is in fact identical to that of the father. This is not as straightforward in the case of nineteenth-century Sweden because of the existence of three distinct types of surnames. The first type is the traditional family name which was passed down from generation to generation, typically through the paternal line. Examples of such names include *Borg* and *Lindskog*. The second type is the true patronymic name which was constructed by appending *–sson* (son) or *–sdotter* (daughter) to the father's first name, resulting in names such as *Svensson* and *Nilssdotter*. The third type is the name that is patronymic in appearance but is in practice identical to the family name in the sense that it is passed down through the generations in an unchanged manner. If one is to succeed in linking children that appear in one census without a surname and living in their parental home with those appearing in a later census with a recorded surname and after having left home, allowance needs to be made for all the possible surnames a child might have received. I address this problem in two ways. Firstly, all those children with missing surnames who were living in the same household as their fathers have had their father's surname appended, one which may be either a family name or a patronymic name inherited in the same manner. Secondly, to allow for the possibility that a child's surname was a true patronymic name, a patronymic surname was constructed using the father's first name. Each child appearing without a recorded surname thus had two potential surnames which were used for probabilistic matching: a family

name derived from the father's surname and a patronymic name derived from the father's first name. To summarize, an individual can thus be linked using surnames in four possible ways:

(1) Own surname to own surname: This method covers the majority of cases where the individual has a surname recorded in both censuses.
(2) Father's surname to father's surname: This method ensures that individuals without a recorded surname who reside with their parents can be linked in both censuses.
(3) Own surname to father's surname: This method covers those who appear in their paternal home without a recorded surname and in a later census *with* a recorded surname that is a family name or a patronymic name inherited in the same manner.
(4) Own surname to constructed patronymic name: This method covers those who appear in their paternal home without a recorded surname and in a later census *with* a recorded surname that is a true patronymic name.

Prior to linking, names were subjected to some very off-hand and basic standardization. The letter *w*, which in Swedish does not differ in terms of pronunciation from *v*, was thus standardized as *v*. Moreover, any non-alpha characters were removed. In the case of patronymic surnames, the suffixes *–sson* and *–sdotter* and variations thereof were parsed out in order to decrease the higher homogeneity of patronymic surnames relative to family names. Finally, nobiliary particles (e.g. *von* and *af* or variations thereof) were eliminated from the surname string. The likeness of names was evaluated using the Jaro-Winkler algorithm, which is commonly used for name matching. The Jaro-Winkler algorithm produces a similarity score by considering common characters, transpositions and common character pairs and increasing the score if a text string has the same initial characters, and it checks for closer agreement between long strings than for that between short ones and adjusts the score accordingly (see Christen 2006 and Christen 2007:41-52 for a more detailed discussion of matching algorithms). For each potential match a similarity score is produced by the algorithm ranging from 0 (for completely dissimilar records) to 1.0 (for identical records). To illustrate this procedure, Table 2 presents similarity scores calculated by comparing two sets of simulated name observations listed in Table 1, in which record B_i is a more or less altered version of A_i.

Each comparison yields a vector consisting of the calculated similarity scores A_iB_i: $\{w_1, w_2\}$ with the diagonal representing the scores of "true" links

(A_1B_1: {1.00, 1.00}, A_2B_2: {0.87, 1.00}, A_3B_3: {0.94, 0.94}, A_4B_4: {1.00, 0.96}) and the off-diagonal representing the scores of "false" links. These scores provide a useful metric for assessing whether a link is likely to be true or false, and also allows for the identification of links that are ambiguous.

Table 1. Simulated records of names

Record	First name	Surname	Record	First name	Surname
A_1	Thomas	Persson	B_1	Thomas	Persson
A_2	Jan	Eriksson	B_2	Janne	Eriksson
A_3	Algoth	Sköld	B_3	Algot	Skiöld
A_4	Assar	Persson	B_4	Assar	Pehrsson

Table 2. Matching scores from Jaro-Winkler string comparisons

	B_1		B_2		B_3		B_4	
	First name	Surname	First name	Surname	First name	Surname	First name	Surname
A_1	1.00	1.00	0.00	0.87	0.46	0.44	0.46	0.96
A_2	0.00	0.87	0.87	1.00	0.51	0.51	0.51	0.83
A_3	0.44	0.00	0.46	0.44	0.94	0.94	0.46	0.00
A_4	0.46	1.00	0.47	0.87	0.47	0.44	1.00	0.96

Removing duplicate and ambiguous links

The process of probabilistic matching means that some individuals will end up being matched to more than one individual. Naturally, only one at most of the matches made will be true, which will lead to ambiguity as to which of the potential matches is correct. A rule is therefore needed for governing how to deal with duplicates. One way is to try to discern between true and false duplicates by carrying out further evaluation of duplicates according to certain criteria and favouring the top-ranking match (Abramitzy et al. 2012). An alternative and stricter method is to minimize in advance the possibility of false duplicate links by eliminating common names (Ferrie 1996). While this method reduces the number of false positives, it also removes a large portion of the population from the linking process, thus reducing the linkage rate while also threatening the representativeness of the sample. A third method involves the removal of all ambiguous links. In practice this means that links are only retained if an individual appearing in a later census is linked to one, and only one, individual in an earlier census, who in turn is not linked to any other individual in the later census. Given the problematic nature of false positive links, I choose to err on the conservative side by following this approach. No attempt is therefore made to establish which of the ambiguous

links are true or false. Instead, all ambiguous links are viewed as suspicious and therefore discarded. As regards the removal of false links, full-count data has a distinct advantage over samples since it enables the identification and subsequent removal of a relatively larger number of duplicate and potentially false links than that possible given the restrictive nature of samples.

Choosing a threshold

A simple and easily replicable way of classifying a link as true or not is to set a threshold value manually for the calculated similarity score, which a potential link has to exceed in order to be classified as true. How high should this threshold be? It is important to note that setting a higher threshold does not necessarily lead to a reduction in links successfully made or an improvement in link quality. This is because an increase in the threshold means a simultaneous narrowing of the span within which links are compared for duplicates. Less restrictive criteria will thus initially yield more potential matches, but they will also result in an increased proportion of links being lost during the deduplication stage due to an increased number of ambiguous links. A trade-off is inevitable: a high threshold will yield a smaller initial sample of matches, of which a few will be discarded because of ambiguity, while a lower threshold will yield a larger initial sample of matches, of which many will be discarded because of ambiguity. In practice, therefore, a higher threshold is not necessarily a guarantor for a high quality sample in terms of the rate of false positives.

Because of differences between countries in name homogeneity, spelling and the similarities between distinct names, no ideal universal threshold exists. In practice this means that researchers often choose a more or less arbitrary threshold above which a link is classified as true (see Wisselgren et al. 2014:150, note 6; Goeken et al. 2011:9). I diverge from this approach in that I evaluate with care the impact on the final sample by choosing a specific threshold. Two features of samples produced by different thresholds were considered: the number of links made and the accuracy of the links made. The aim of this exercise is to identify an appropriate threshold which results in a maximum number of accurate links.

The accuracy of the links was checked by comparing the similarity of second names for all linked individuals who had a recorded second name (about two thirds of all men and women had this). Links were deemed

accurate if the Jaro-Winkler similarity score of second names exceeded 0.8. It is important to note that a link not confirmed as accurate does not automatically mean that it is a false positive. Manual inspection of links revealed that the names of linked individuals were often recorded in a different order in each census or sometimes dropped and omitted, resulting in failure to confirm a link even if it was true. The share of confirmed links is, however, a useful metric for evaluating differences, if any, between the relative accuracy of links generated at different threshold levels.

In order to identify an appropriate threshold for the similarity score for the Swedish censuses, a range of values was tested. Table 3 presents the results of the linking between censuses by using different Jaro-Winkler thresholds for first and last names ranging from 0.7 to 1.0. The last column presents the share of links with a recorded second name that could be confirmed as being accurate. In terms of linkage rates, the threshold value and the resulting linkage rate form an inverted U-shape. Setting a low threshold results in a large number of initial matches, of which a considerable proportion are removed in the deduplication process. Conversely, setting a high threshold results in a smaller number of initial matches with fewer links removed in the deduplication stage. When based on the share of links confirmed by a comparison of second names, the discernible difference in the quality of links generated at different threshold levels is small. Any negative effect of a lower and more forgiving threshold regarding link accuracy would thus seem to be remedied by a subsequent increase in the removal of ambiguous duplicate links. Given that the accuracy of links does not vary to a great extent according to the threshold level, 0.8 was chosen as an appropriate threshold in order to maximize the total linkage rate. Admittedly, an even higher linkage rate would have been achievable if different thresholds had been chosen for men (0.8) and women (0.85). However, in the interests of consistency and simplicity, 0.8 was chosen as a threshold for both men and women. For all linked individuals the Jaro-Winkler similarity scores have been retained, enabling, if wished, the extraction of a sub-sample of more precisely matched individuals.

Table 3. Linking results for men and women between 1900-1890 and between 1890-1880 at different Jaro-Winkler thresholds

Jaro-Winkler threshold	Matched		Removed after deduplication		Linked		Confirmed
	n	%	n	%	n	%	%
1900-1890							
Men							
0.70	1653940	83.93	539890	32.64	1114050	56.53	98.37
0.75	1622551	82.33	407893	25.14	1214658	61.64	98.47
0.80	1598495	81.11	350568	21.93	1247927	63.32	98.53
0.85	1481058	75.15	269388	18.19	1211670	61.48	98.31
0.90	1408751	71.49	224677	15.95	1184074	60.08	98.39
0.95	1328619	67.42	196210	14.77	1132409	57.46	98.37
1.00	1277166	64.81	188682	14.77	1088484	55.23	98.33
Women							
0.70	1684349	80.77	615041	36.52	1069308	51.28	97.64
0.75	1642149	78.75	448449	27.31	1193700	57.24	97.82
0.80	1609231	77.17	339600	21.10	1269631	60.89	97.96
0.85	1550201	74.34	269539	17.39	1280662	61.41	98.03
0.90	1459616	70.00	226990	15.55	1232626	59.11	98.08
0.95	1334790	64.01	185353	13.89	1149437	55.12	98.09
1.00	1264357	60.63	178271	14.10	1086086	52.08	98.08
1890-1880							
Men							
0.70	1487218	82.89	502088	33.76	985130	49.99	97.95
0.75	1457961	81.26	385230	26.42	1072731	54.43	98.08
0.80	1434325	79.94	334432	23.32	1099893	55.81	98.15
0.85	1317062	73.41	256644	19.49	1060418	53.81	97.87
0.90	1250595	69.70	219687	17.57	1030908	52.31	97.95
0.95	1172522	65.35	189280	16.14	983242	49.89	97.91
1.00	1128667	62.91	183117	16.22	945550	47.98	97.87
Women							
0.70	1570378	80.97	586980	37.38	983398	47.16	96.67
0.75	1530524	78.92	444460	29.04	1086064	52.08	96.85
0.80	1497152	77.20	344057	22.98	1153095	55.30	97.02
0.85	1437017	74.10	275963	19.20	1161054	55.68	97.15
0.90	1336277	68.90	234820	17.57	1101457	52.82	97.21
0.95	1206441	62.21	189421	15.70	1017020	48.77	97.22
1.00	1137880	58.67	181431	15.94	956449	45.87	97.19

Sources: 1900, 1890 and 1880 censuses, Swedish National Archives

Secondary links

After creating an initial sample of primary links, it is possible to improve linkage rates by exploiting the indirect linking of households created by primary links in the first stage. This means that individuals who were not linked in the process described above are given a second chance to be linked if they are recorded in both censuses as residing with one person who was a

primary link. The creation of secondary links mirrors exactly 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, parish of birth and sex), thereby narrowing the initial criteria for matching individuals of the same sex and born in the same parish in the same year, and who resided with a particular individual linked in both censuses. Because the new indexing severely reduces the size of the group within which individuals are compared between the censuses, only first names are used for probabilistic matching, again using a Jaro-Winkler threshold of 0.8. Presuming that the primary link is true, a false positive link would thus require an individual with a primary link in each separate census to have resided with two different people who by coincidence were of the same sex, were born in the same parish in the same year, and in addition shared a first name. By using household information in the creation of secondary links, I deliberately deviate from the practice of using only those variables for matching that are stable over time. The inclusion of household links thus runs the risk of biasing the sample in favour of individuals living in larger and more stable households. In the following sections, the contribution of secondary links to the overall linkage rate and the effect of these links on the representativeness of the sample are scrutinized further.

Basic results

Table 4 presents the linkage rates between the 1900 and 1890 censuses and between the 1890 and 1880 censuses. The overall result of the linking procedure is that more than three quarters of the population in each census are linked to the preceding census, the rate being 1.90 per cent lower for 1890 when linked to 1880 than for 1900 when linked to 1890. Primary links provide the majority of links, viz. 62.07 per cent of those between 1900 and 1890 and 60.34 per cent of those between 1890 and 1880, with secondary links adding a further 13.97 per cent and 13.80 per cent respectively to the total linkage rate. When comparing the linkage rates of men and women, it is clear that women are not linked to the same extent as men in terms of either primary and secondary links. This difference is not, however, large. Having counted all the links, the linkage rate for women falls short by 2.97 per cent

(for 1900 to 1890) and 2.66 per cent (for 1890 to 1880) when compared to men.

Table 4. Links and linkage rates between the censuses of 1900 and 1890 and 1890 and 1880 respectively

	1900-1890			1890-1880		
	Men	Women	Total	Men	Women	Total
Census count (n, age > 10 years)	1970676	2085277	4055953	1794231	1939366	3733597
Primary links (n)	1247929	1269631	2517560	1099893	1153095	2252988
Secondary links (n)	280639	286057	566696	255128	259971	515099
Primary linkage rate (%)	63.32	60.89	62.07	61.30	59.46	60.34
Secondary linkage rate (%)	14.24	13.72	13.97	14.22	13.40	13.80
Total linkage rate (%)	77.57	74.60	76.04	75.52	72.86	74.14

Sources: 1900, 1890 and 1880 censuses, Swedish National Archives

By using several methods for creating links, it is of interest to see the relative importance of probabilistic matching on the linkage rates. Closer examination of the name similarity scores produced by the Jaro-Winkler algorithm revealed that of all the primary links made between 1900 and 1890 (and 1890 to 1880), 80.33 per cent (79.10 per cent) of men and 77.89 per cent (75.44 per cent) of women have a match score of 1.00 for first and last names, implying an exact match between names in the two censuses. More than one fifth of all the primary links thus depends on the probabilistic linking of either first, last or both names. Moreover, probabilistic linking increases the number of women linked relative to that of men, which is of value given the overall lower linkage rate of women. Clearly, the contribution of the probabilistic matching of names to the total share of links is not marginal.

Figures 1-4 provides a more detailed picture of how the different methods of matching surnames and adding secondary links contribute to the total linkage rate across different ages. An important aspect of Figures 1-4 is the degree to which the achieved linkage rate depends on age. The linkage rate starts out at above 80 per cent in the 10-15 age groups before falling to around 60 per cent in the 25-30 age groups. Up to the age of 40 there is no discernible difference in the manner or extent to which men and women are linked. The overall lower linkage rate of women may instead be attributed to differences in linkage rates between older men and women. After the age of 40, the total linkage rate of men steadily increases. For women an increase,

Figure 1. Linkage rate by link type for men, 1900-1890

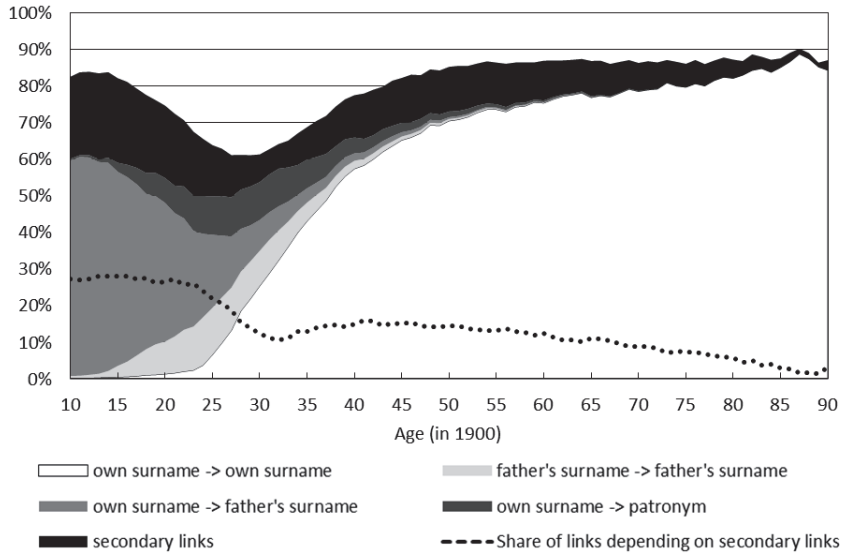


Figure 2. Linkage rate by link type for women, 1900-1890

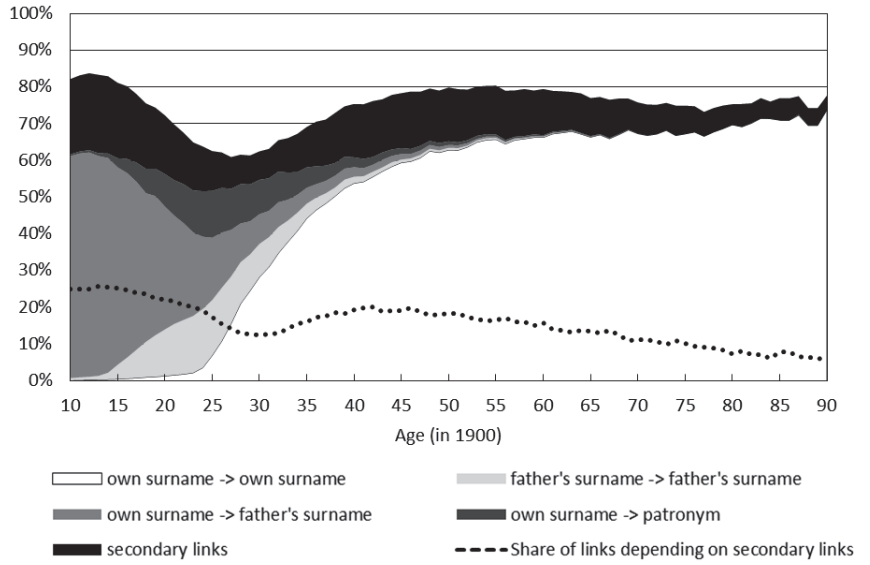


Figure 3. Linkage rate by link type for men, 1890-1880

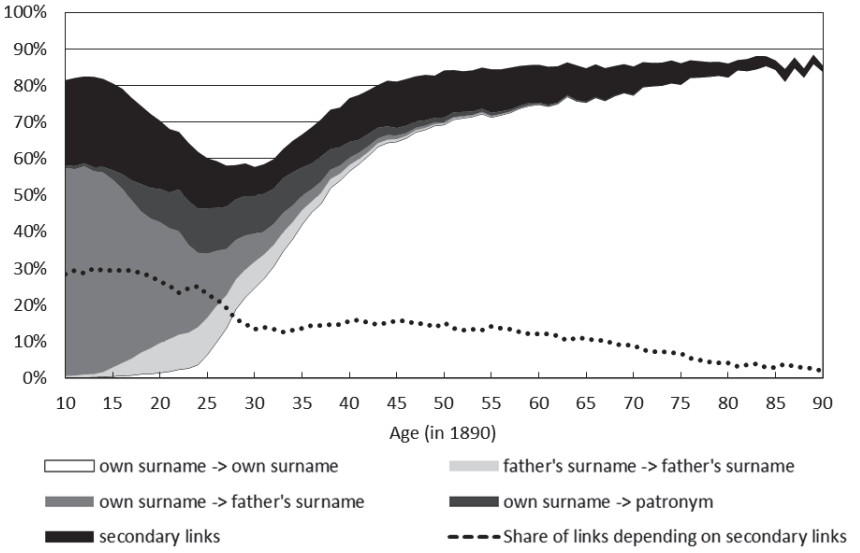
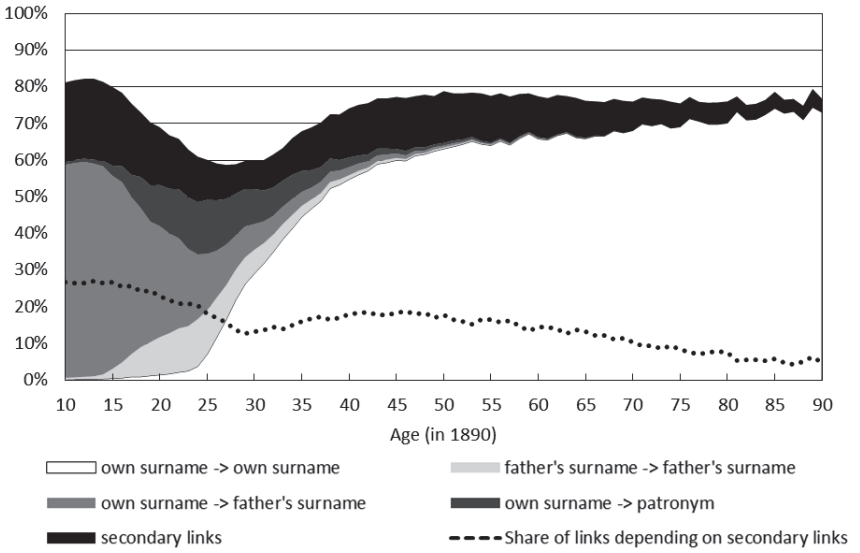


Figure 4. Linkage rate by link type for women, 1890-1880



albeit a slower one, is noticeable until the age of 50 when the total rate levels out.

Age also determines the method by which surnames were matched. If the linking procedure had simply been limited to the recorded surnames in the census, few individuals aged 25 years or less would be linked to the preceding census. In the case of the young, the predominant links are those based on matches between fathers' surnames. Links between the individual's own recorded surname in the later census and the father's surname or a constructed patronymic in the preceding census constitute an important share of the links between individuals aged 20-35. For those of more advanced age, links between the own recorded surname of individuals becomes the predominant link type. Secondary links accentuate the pattern produced by primary links by linking a few individuals in the 25-30 age group and more of those both younger and older.

There are two explanations for the observed decrease in the linkage rate observed from the age of 15. First, the dip coincides with what tends to be a transient period in an individual's life course, involving their leaving home, migration and marriage and family formation. These are events that would have required changes to be made by the priests in parish registers, thus increasing the probability of errors being made when withdrawing and re-entering the information. For women, the depressed linkage rate relative to men after the age of 45 may in some cases be explained by name changes preventing linkage or – as in the case of Stockholm – the lack of a recorded surname following marriage, the net result being a gap in linkage rates between men and women that exceeds 10 per cent at a more advanced age.

The second explanation for the pattern pertains to the construction of surnames. Individuals aged 10-14 in the 1900 or 1890 census had almost always been members of the same household 10 years prior to that. These individuals are therefore almost exclusively linked using $father\ surname_{1900}$ -> $father\ surname_{1890}$ or else these are added as secondary links. Individuals aged 15-30 were, in contrast, likely to be living in the parental household in 1890 (1880) and in a different household in 1900 (1890). Because children living at home have no recorded surnames, members of the 15-30 age group are more often linked on the basis of inferred surnames by using either $own\ surname_{t}$ -> $father\ surname_{t-10}$ or $own\ surname_{t}$ -> $patronym_{t-10}$, this being a method that yields more ambiguous links which are subsequently discarded in the deduplication stage, leading to an attrition of the linked sample in the 15-30 age group. Finally, the older age groups are almost always linked using

own surname$_{1900}$ -> *own surname*$_{1890}$, a method that yields a high number of links and fewer losses in the deduplication stage.

Table 5 displays the linkage rate achieved across three censuses by combining the links created between 1900-1890 and 1890-1880. The three-census linkage rate is considerably lower than the rates achieved when linking between two censuses. Of the more than 3 million individuals present in the 1900 census and born in 1880 or before, 60.30 per cent were linked between both 1900-1890 and 1890-1880, resulting in a panel constituting of more than 1.8 million individuals observed at three points in time. If linking is constrained to the sole reliance on primary links, the total linkage rate would decrease by more than 10 percentage points, resulting in the linking of less than half of all the individuals present in the 1900 census and born in or before 1880.

Table 5. Links and linkage rates across three censuses (1900, 1890 and 1880)

		Total	Men	Women
Census count 1900 (n, age > 20 years)		3038719	1454002	1584717
Linked 1900-1890-1880 (n)		1832260	908973	923287
Thereof:	primary link 1900-1890, primary link 1890-1880 (n)	1501994	746440	755554
	primary link 1900-1890, secondary link 1890-1880 (n)	89822	46970	42852
	secondary link 1900-1890, primary link 1890-1880 (n)	67847	29150	38697
	secondary link 1900-1890, secondary link 1890-1880 (n)	172597	86413	86184
Linkage rate 1900-1890-1880, only primary links (%)		49.43	51.34	47.68
Linkage rate 1900-1890-1880, including secondary links (%)		60.30	62.52	58.26

Sources: 1900, 1890 and 1880 censuses, Swedish National Archives

Figures 5-6 plot the three-census linkage rates by age for men and women. It is clear that the age-dependent pattern apparent in Figures 1-4 becomes even more accentuated when additional censuses are joined together. At a more advanced age the two-census linkage rates do not differ much from the three-census rates. At a lower age, however, the reduction is considerable. In the case of 1900, around 70 per cent of all men and women aged 35-40 may be linked to the 1890 census. When extending the panel back to 1880, the linkage rate for this group falls to approximately 45 per cent. That this group is so severely affected is disconcerting given that the age span covered by the three censuses for this group (10-15 in 1880, 20-25 in 1890 and 35-40 in 1900) is of particular interest to researchers, covering as it does important life course events such as leaving home, migration, marriage and childbearing.

Figure 5. Linkage rate for men using primary and secondary links, 1900-1890-1880

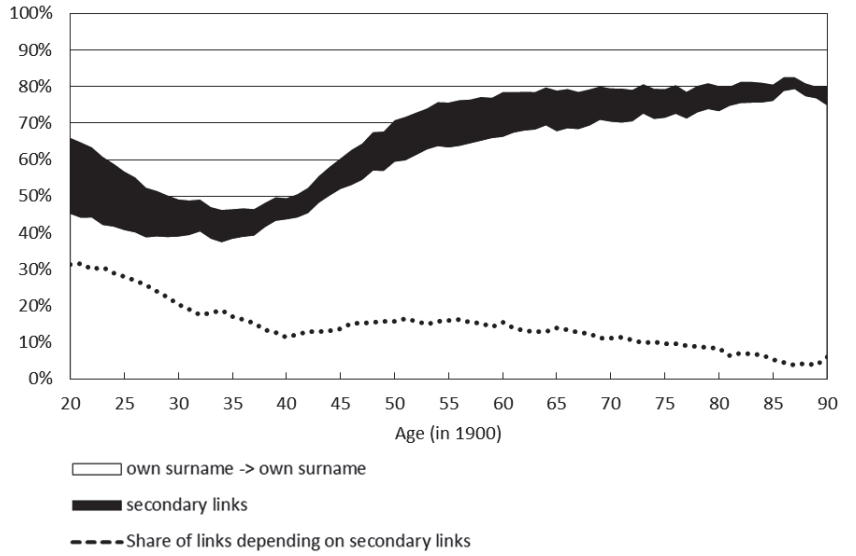
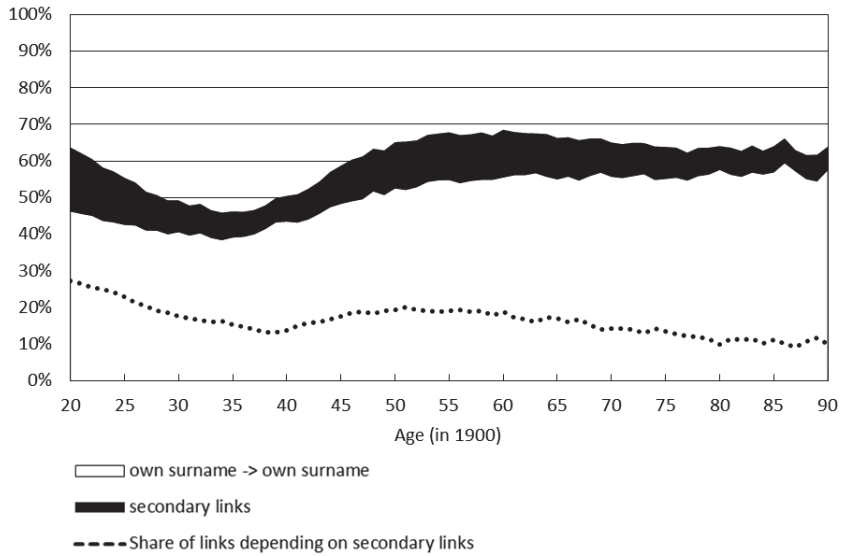


Figure 6. Linkage rate for women using primary and secondary links, 1900-1890-1880



In each one of Figures 1-6, the contribution of secondary links has been plotted with a dashed line. As with primary links, strong age dependence is noticeable, and it favours young individuals in particular. Among the youngest cohorts, more than a quarter of all the links depend on secondary links. The importance of secondary links thereafter declines before increasing once again after age 40 until age 50. The possibility of linking any individual across multiple censuses is thus often dependent on secondary links at some stage in the life course, in particular when the individual is young, but also at a more advanced age. Constraining linking to primary links alone thus comes at the cost of a severe depreciation in sample size when seeking to cover extended periods of an individual's life course. Given the usefulness of secondary links when creating panels extending beyond two censuses, the effect of secondary links on the representativeness of the linked sample is pertinent, this being an issue to which the paper turns next.

Evaluation of links

Bias

Bias in the final sample may be the result of bias inherent in the source material or introduced by the linking process. While the choice of indexing and linking variables is made in order to minimize problems, some potential issues remain. Variations in spelling are perhaps the most problematic for linking purposes. Since the information contained in the censuses stems from the parish church books, any spelling variation is likely to be aggravated by the number of times an individual's name was withdrawn from and entered into a church book. Hence, geographically mobile individuals and those experiencing the transient period in life before and during family formation may exhibit more variation in name spelling between censuses, a factor which affects the likelihood of being linked. The necessary removal of duplicate and ambiguous links is a further potential source of bias. It is more likely that individuals belonging to large birth year cohorts and born in more heavily populated parishes will end up with more than one match, and this will result in their elimination from the final sample. The same holds for individuals with common names.

Because being linked is conditional on surviving between the two censuses, survivor bias is unavoidable in the process of forward linking (i.e. from an older census to a more recent one). To what extent survivor bias is an issue depends on the existence of a mortality gradient across groups (e.g. social, sex, married/unmarried or urban/rural). No social gradient in mortality was seen to emerge in Sweden until the period after World War II (see Bengtsson and Dribe 2011). Any bias in the occupational structure stemming from differences in mortality should therefore be limited. For the purpose of this paper, the issue of mortality bias has been left to one side in considering backward linking, which requires that an individual is present and was alive at the time of the most recent census to be at risk of being linked.

Although the Swedish censuses are in terms of enumeration largely representative of the country's population (or at least arguably more so than are the censuses in other countries from the same period), there is an inherent bias in the prevalence and accuracy of certain variables in the censuses. Variables based on individual characteristics that change over the life course are more affected than stable characteristics. This is most pertinent when it comes to occupational information. A large proportion of individuals appear in the censuses with no occupation recorded, while certain occupational categories are only vaguely specified. Furthermore, the extent to which occupations have been recorded is dependent on geographic location. In general, occupational information is more prevalent and accurate for the countryside than for urban areas (BISOS A 1907: xxxv). Statistics Sweden (*Statistiska Centralbyrån*) were well aware of the problem at the time, and cited increased geographic and occupational mobility coupled with a more diversified economy and a decline in the personal relations between the populace and clergy (who were responsible for updating the parish records used to compile the censuses) as the reasons behind the deficient quality of the occupational statistics (SOS 1926:3). Furthermore, the under-reporting of occupations is particularly marked in the case of women, who have a higher instance of missing occupational information than men.

A basic and descriptive check of the representativeness of the linked samples consists of investigating to what extent the samples mirror the geographic distribution of the Swedish population as enumerated in the censuses. I thus ask to what extent linkage rates differ between different geographic regions in Sweden. Table 6 presents linkage rates (including both primary and secondary links) separated by county of residence. By breaking down the population in terms of county of residence, it becomes clear that

Table 6. Linkage rates by county of residence, primary and secondary links

	1900-1890			1890-1880			1900-1890-1880		
	Men	Women	Total	Men	Women	Total	Men	Women	Total
Stockholm	62.8%	56.4%	59.4%	61.5%	57.8%	59.5%	47.8%	39.4%	43.3%
Uppsala	80.0%	75.1%	77.4%	77.3%	74.7%	76.0%	66.0%	59.8%	62.7%
Södermanland	79.8%	77.3%	78.5%	76.5%	74.9%	75.7%	64.3%	60.7%	62.5%
Östergötland	81.2%	78.6%	79.9%	78.6%	76.9%	77.7%	67.1%	63.7%	65.3%
Jönköping	81.9%	81.5%	81.7%	80.8%	80.2%	80.5%	68.3%	68.1%	68.2%
Kronoberg	83.2%	83.0%	83.1%	81.3%	80.6%	80.9%	69.3%	68.8%	69.1%
Kalmar	79.2%	77.4%	78.3%	77.1%	75.4%	76.2%	64.0%	61.5%	62.7%
Gotland	84.5%	79.3%	81.8%	83.3%	76.8%	79.9%	74.6%	64.8%	69.3%
Blekinge	76.9%	76.1%	76.5%	77.0%	74.8%	75.9%	63.0%	61.0%	62.0%
Kristianstad	82.1%	80.1%	81.1%	80.7%	79.1%	79.9%	70.2%	66.9%	68.5%
Malmöhus	79.0%	76.1%	77.5%	76.6%	74.0%	75.3%	64.6%	60.5%	62.5%
Halland	81.5%	79.8%	80.6%	81.1%	78.0%	79.5%	69.5%	65.9%	67.6%
Göteborg och Bohus	76.2%	74.1%	75.1%	67.2%	63.8%	65.4%	57.1%	53.1%	55.0%
Älvsborg	81.4%	79.7%	80.5%	80.7%	78.5%	79.5%	68.5%	65.9%	67.1%
Skaraborg	81.4%	79.8%	80.6%	78.9%	77.2%	78.0%	67.0%	64.9%	65.9%
Värmland	80.4%	77.2%	78.7%	78.7%	74.8%	76.6%	66.1%	61.3%	63.6%
Örebro	77.8%	74.1%	75.9%	77.7%	75.1%	76.3%	63.4%	58.5%	60.9%
Västmanland	79.8%	76.3%	78.1%	78.0%	74.5%	76.2%	65.4%	59.5%	62.3%
Kopparberg	69.8%	68.6%	69.2%	67.0%	65.9%	66.4%	49.4%	48.4%	48.9%
Gävleborg	77.0%	72.8%	74.9%	74.7%	71.3%	73.0%	61.0%	55.3%	58.1%
Västernorrland	78.6%	74.1%	76.3%	74.5%	71.2%	72.8%	62.5%	56.4%	59.5%
Jämtland	80.8%	79.8%	80.3%	77.6%	77.1%	77.3%	66.4%	65.1%	65.8%
Västerbotten	77.5%	74.1%	75.8%	75.7%	71.1%	73.4%	60.9%	54.8%	57.8%
Norrbotten	72.8%	67.6%	70.3%	71.6%	65.5%	68.6%	55.0%	47.2%	51.2%
Total	77.6%	74.6%	76.0%	75.5%	72.9%	74.1%	62.5%	58.3%	60.3%

Sources: 1900, 1890 and 1880 censuses, Swedish National Archives

linkage rates vary significantly as regards place and period. Linkage rates in a handful of counties exceed 80 per cent. At the other end of the spectrum is Stockholm, its linkage rates being around 60 per cent between 1900 and 1890 and 1890 and 1880. Combining the links between two censuses in a three-census panel, gives an exaggerated overall pattern. Only 43.3 per cent of all the individuals resident in Stockholm in 1900 and born in 1880 or beforehand can be located in both the 1890 and 1880 censuses. In contrast, almost 69.1 per cent of Kronoberg residents can be linked between all three censuses. The lacklustre result for Stockholm is, given the county's circumstances, not surprising. It was by far the most urban of all Swedish counties at the end of the nineteenth century it received more migrants than any other county, and it had the most densely populated parishes, factors which all made more difficult the keeping of an accurate and up-to-date population register.

An important concern is to what extent the probability of linking an individual between censuses is correlated with specific individual characteristics. In order to separate out variables associated with a higher or lower probability of linking, probit regressions were estimated, in which the dependent variable equals 1 for individuals who were linked between the years 1900-1890, 1890-1880 or 1900-1890-1880, and equals 0 for individuals who were not. All variables refer to observed characteristics in the most recent census and are categorical except for age, which is continuous and includes a square and cubic term in order to capture the non-linear relationship between age and linkage rates depicted in Figures 1-6. To test whether there is a social gradient in the linked sample relative to the full population information, occupations were coded in line with the HISCLASS-scheme (see Van Leeuwen, Maas, and Miles 2002). The estimates are presented in the form of marginal effects for the randomly drawn subsamples of 100,000 men and women in Tables 7 and 8[20].

Beginning with the results in the columns labelled I, which consider only primary links, these indicate that the linked samples are very representative of the overall population. Although in statistical terms several individual characteristics are significantly correlated with the probability of being linked, the magnitude of the marginal effects tends to be small. The results regarding social status are particularly encouraging. Although some occupational groups show a tendency towards being associated with higher probabilities of being linked, no social gradient is apparent in the linked samples. The strongest effects are to be found among the variables describing household size and an individual's position within the household. Belonging to a larger household means a higher probability of being linked, while occupying a more peripheral position within the household means a lower probability of being linked. In the case of men, marital status matters little. In the case of women, divorcees and widows are linked to a lower degree than are women who are married or never did marry. The correlation between being either widowed or divorced and failing to be linked is a plausible explanation for the discrepancy apparent in Figures 1-4 between the linkage rates of men and women of a more advanced age. Linking across three

[20] Because the information regarding certain characteristics is missing in the case of some individuals, the estimated samples are somewhat lower than 100,000.

Table 7. Probit marginal effects (∂P/∂X) on linkage for men

	1890-1880		1900-1890		1900-1890-1880	
	I	II	I	II	I	II
Age	0.002***	0.003***	0.001***	0.002***	0.003***	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household position						
Head	ref.	ref.	ref.	ref.	ref.	ref.
Spouse	-0.060	0.007	-0.108	-0.194	-0.009	0.039
	(0.182)	(0.147)	(0.192)	(0.181)	(0.153)	(0.106)
Child	-0.049***	0.072***	-0.034***	0.081***	-0.006	0.125***
	(0.009)	(0.007)	(0.009)	(0.007)	(0.009)	(0.007)
Other relative	-0.207***	-0.136***	-0.188***	-0.100***	-0.125***	0.023***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)	(0.007)
Non-relative	-0.112***	-0.167***	-0.100***	-0.152***	-0.057***	-0.033***
	(0.010)	(0.009)	(0.010)	(0.009)	(0.009)	(0.004)
Unknown	-0.293***	-0.056	-0.420***	-0.167***	-0.164***	0.110***
	(0.033)	(0.032)	(0.028)	(0.032)	(0.038)	(0.030)
Marital status						
Married	ref.	ref.	ref.	ref.	ref.	ref.
Divorced	0.082	0.045	-0.002	-0.033	-0.101**	-0.055*
	(0.048)	(0.036)	(0.044)	(0.035)	(0.039)	(0.024)
Widowed	0.024**	-0.008	0.009	-0.031***	0.017*	-0.022***
	(0.009)	(0.007)	(0.008)	(0.007)	(0.008)	(0.004)
Never married	-0.019*	-0.003	-0.039***	-0.029***	-0.033***	0.021***
	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.004)
HISCLASS						
1. Higher managers	0.049	0.007	0.018	-0.028	0.037	-0.035**
	(0.027)	(0.023)	(0.028)	(0.024)	(0.024)	(0.013)
2. Higher professionals	0.030	0.005	0.010	-0.015	0.062***	-0.022*
	(0.018)	(0.015)	(0.017)	(0.014)	(0.015)	(0.009)
3. Lower managers	-0.007	-0.018	0.016	0.010	-0.026*	-0.003
	(0.015)	(0.013)	(0.014)	(0.011)	(0.013)	(0.008)
4. Lower professionals, clerical and sales personnel	0.013	0.004	0.003	-0.005	0.011	-0.016**
	(0.011)	(0.009)	(0.010)	(0.008)	(0.009)	(0.006)

5. Lower clerical and sales personnel	0.033*	0.027*	0.011	0.017	0.002	0.010
	(0.016)	(0.012)	(0.014)	(0.010)	(0.013)	(0.009)
6. Foremen	-0.015	-0.019	-0.005	-0.015	0.007	-0.006
	(0.015)	(0.013)	(0.014)	(0.012)	(0.013)	(0.008)
7. Medium skilled workers	ref.	ref.	ref.	ref.	ref.	ref.
8. Farmers and fishermen	-0.038***	-0.018**	-0.044***	-0.020***	-0.018**	0.017***
	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)	(0.004)
9. Lower skilled workers	-0.033***	-0.022**	-0.018*	-0.007	-0.041***	0.012**
	(0.009)	(0.007)	(0.008)	(0.006)	(0.007)	(0.005)
10. Lower skilled farm workers	0.016	0.014	-0.012	-0.009	0.009	0.001
	(0.016)	(0.013)	(0.015)	(0.012)	(0.013)	(0.008)
11. Unskilled workers	-0.008	0.005	-0.012	0.000	-0.022***	0.016***
	(0.008)	(0.006)	(0.007)	(0.006)	(0.007)	(0.005)
12. Unskilled farm workers	-0.025**	-0.028***	-0.013	-0.010	-0.020**	-0.004
	(0.008)	(0.006)	(0.007)	(0.006)	(0.007)	(0.004)
13. Unknown	-0.015*	0.009	-0.009	0.008	-0.014*	0.008*
	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)	(0.004)
Household size						
1	ref.	ref.	ref.	ref.	ref.	ref.
2	0.023**	0.073***	0.004	0.045***	0.026***	0.047***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.007)	(0.003)
3	0.062***	0.156***	0.042***	0.139***	0.058***	0.091***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.004)
4	0.096***	0.202***	0.070***	0.183***	0.072***	0.107***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.004)
5 or more	0.128***	0.236***	0.116***	0.222***	0.089***	0.105***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.005)
Rural	0.069***	0.075***	0.074***	0.055***	0.064***	0.026***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)
Not resident in birth parish	-0.030***	-0.053***	-0.016***	-0.041***	-0.016***	-0.031***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
N	99725	99728	99945	99945	99947	99947

Notes: Standard errors within parenthesis. Significance levels: * 10 per cent, ** 5 per cent, *** 1 per cent. I: primary links only. II: primary and secondary links.
Sources: 1900, 1890 and 1880 censuses, Swedish National Archives.

Table 8. Probit marginal effects (∂P/∂X) on linkage for women

	1890-1880		1900-1890		1900-1890-1880	
	I	II	I	II	I	II
Age	0.001***	0.003***	0.001***	0.003***	0.003***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household position						
Head	ref.	ref.	ref.	ref.	ref.	ref.
Spouse	0.055***	0.135***	0.054***	0.157***	0.112***	0.052***
	(0.012)	(0.011)	(0.013)	(0.011)	(0.012)	(0.006)
Child	-0.006	0.163***	0.007	0.181***	0.019	0.132***
	(0.010)	(0.009)	(0.010)	(0.009)	(0.010)	(0.007)
Other relative	-0.138***	-0.029**	-0.134***	-0.002	-0.058***	0.060***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.010)	(0.005)
Non-relative	-0.058***	-0.064***	-0.030**	-0.031**	0.004	-0.013**
	(0.010)	(0.010)	(0.010)	(0.009)	(0.010)	(0.004)
Unknown	-0.161***	0.013	-0.236***	-0.026	-0.082***	0.099***
	(0.026)	(0.025)	(0.024)	(0.024)	(0.023)	(0.016)
Marital status						
Married	ref.	ref.	ref.	ref.	ref.	ref.
Divorced	-0.140**	-0.109**	-0.121***	-0.093**	-0.110**	-0.028
	(0.045)	(0.042)	(0.036)	(0.032)	(0.034)	(0.023)
Widowed	-0.092***	-0.085***	-0.147***	-0.122***	-0.120***	0.007
	(0.012)	(0.011)	(0.012)	(0.011)	(0.011)	(0.007)
Never married	-0.007	0.044***	0.005	0.059***	0.035***	0.046***
	(0.011)	(0.009)	(0.011)	(0.009)	(0.010)	(0.007)
HISCLASS						
1. Higher managers	-0.046	-0.055	0.032	-0.056	-0.021	-0.009
	(0.160)	(0.142)	(0.172)	(0.151)	(0.098)	(0.062)
2. Higher professionals	-0.044	-0.018	-0.004	-0.003	0.003	0.033
	(0.054)	(0.045)	(0.043)	(0.034)	(0.039)	(0.027)
3. Lower managers	-0.086	-0.049	-0.007	-0.033	-0.004	-0.005
	(0.056)	(0.047)	(0.043)	(0.035)	(0.040)	(0.027)
4. Lower professionals, clerical and sales personnel	0.023	0.049	0.060	0.045	0.062	0.002
	(0.045)	(0.037)	(0.035)	(0.028)	(0.033)	(0.023)

5. Lower clerical and	0.028	0.047	-0.056	-0.028	0.007	0.003
sales personnel	(0.051)	(0.042)	(0.039)	(0.031)	(0.036)	(0.024)
6. Foremen	0.054	0.051	0.044	0.006	0.063	-0.014
	(0.046)	(0.038)	(0.036)	(0.028)	(0.033)	(0.022)
7. Medium skilled workers	ref.	ref.	ref.	ref.	ref.	ref.
8. Farmers and fishermen	-0.004	0.019	0.005	0.004	0.017	0.029
	(0.044)	(0.036)	(0.035)	(0.027)	(0.032)	(0.022)
9. Lower skilled workers	0.029	0.051	0.040	0.030	0.029	0.002
	(0.043)	(0.035)	(0.033)	(0.026)	(0.031)	(0.021)
10. Lower skilled farm workers	-0.010	-0.003	-0.056	-0.018	0.045	-0.013
	(0.076)	(0.063)	(0.067)	(0.053)	(0.058)	(0.043)
11. Unskilled workers	0.015	0.009	0.016	-0.017	0.032	-0.017
	(0.041)	(0.034)	(0.032)	(0.025)	(0.029)	(0.020)
12. Unskilled farm workers	-0.076	-0.043	-0.049	-0.057*	-0.038	-0.001
	(0.046)	(0.038)	(0.036)	(0.029)	(0.033)	(0.023)
13. Unknown	-0.028	0.005	-0.007	-0.008	-0.009	0.009
	(0.041)	(0.034)	(0.032)	(0.025)	(0.029)	(0.020)
Household size						
1	ref.	ref.	ref.	ref.	ref.	ref.
2	-0.016	-0.001	-0.029***	-0.014	-0.019*	0.050***
	(0.009)	(0.008)	(0.009)	(0.008)	(0.008)	(0.003)
3	0.005	0.063***	-0.009	0.059***	0.008	0.085***
	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)	(0.004)
4	0.042***	0.112***	0.026**	0.103***	0.025**	0.109***
	(0.010)	(0.009)	(0.009)	(0.008)	(0.009)	(0.004)
5 or more	0.078***	0.156***	0.057***	0.141***	0.036***	0.115***
	(0.009)	(0.008)	(0.009)	(0.008)	(0.009)	(0.005)
Rural	0.108***	0.097***	0.136***	0.082***	0.144***	-0.008***
	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)
Not resident in birth parish	-0.013***	-0.043***	0.006	-0.030***	-0.004	-0.031***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
N	99953	99953	99929	99929	99913	99913

Notes: Standard errors within parenthesis. Significance levels: * 10 percent, ** 5 percent, *** 1 percent. I: primary links only. II: primary and secondary links.
Sources: 1900, 1890 and 1880 censuses, Swedish National Archives.

censuses did not, with the exception of age, seem to result in furthering the bias.

The estimates listed under the columns labelled II extend the analysis by adding secondary links. The addition of secondary links changes the overall result little, with two exceptions. As expected, secondary links result in a further bias of the sample towards individuals belonging to large households. The results are more mixed when it comes to an individual's position within the household. The bias against individuals whose relationship with the household head is unknown is severely reduced, as is the bias against more distant family. In contrast, the existing bias in favour of linking wives is exaggerated.

Summarizing the results shows clearly that individuals who fail to be linked tend to lead a life which is more transient. Migrants, female divorcees and widowers, rural residents, members of single households and those not belonging to a nuclear family are all linked to a lesser degree than are the more settled and stable share of the population. Moreover, apart from biasing the linked sample in favour of the larger households, secondary links do not seem to pose a serious threat to the representativeness of the linked samples.

False positives

Are false positives a major concern in the linked sample? Apart from eliminating all duplicate matches, no further action was taken to actively limit the extent of false positives. Is the complete elimination of duplicate and ambiguous links enough to reduce false positives to an acceptable level? To get a sense of the magnitude of the problem, I have scrutinized some links further. In practice I have compared auxiliary characteristics not used in the linking process of individuals to deduce whether the link is potentially false or true. The challenge has been to find a variable that is sufficiently stable over time to be useful as a check. I propose comparing the characteristics of the linked individuals' spouses. All those men and women from the linked sample who were married in 1890 and 1900 and had a spouse present in the household in both 1890 and 1900 were selected. The spouse's characteristics were then appended to the selected individual and the links evaluated by comparing the similarity of the spouses' first and last names between the 1900 and 1890 censuses. The Jaro-Winkler algorithm was used, and any link where the spouse showed a first name similarity score exceeding 0.8 was

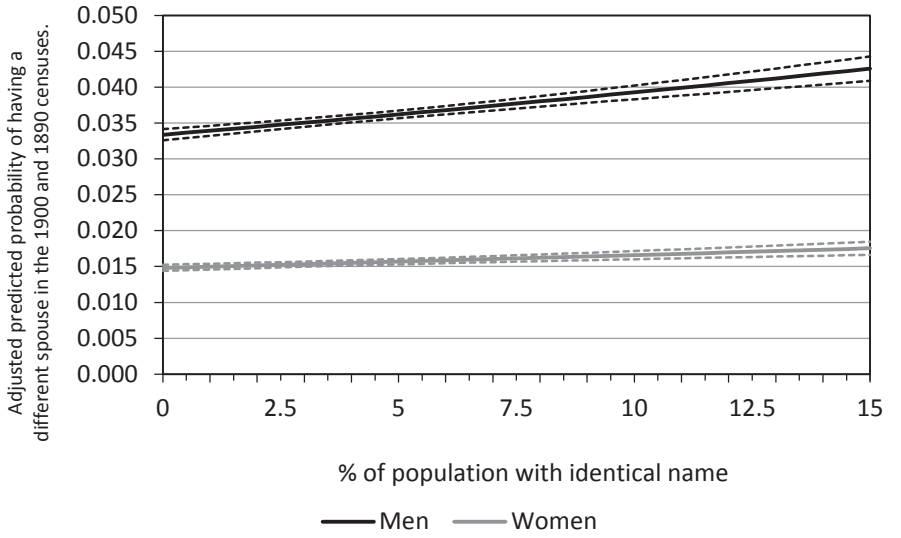
classified as a true link. This method unavoidably deems many links to be false by construction because a significant share of individuals will have different spouses in each census as a result of remarriage rather than being erroneously linked; the basic result should therefore be considered a generous upper bound of the extent of false positives among married individuals. Of the 457,939 men and 445,078 women who met the criteria of being married in both 1890 and 1900 and whose spouse was present, 441,476 men and 438,175 women were verified as being true links after comparison of the spouses' names between the censuses. This yields an upper bound of false positives of 3.60 per cent for husbands and 1.55 per cent for wives.

The likelihood of a link being made when it is in fact false is dependent on the number of potential matches considered for each individual. Because sex, year of birth and parish of birth are used to index the data, and because names are used for probabilistic matching, false positive links are less likely to be made between individuals who hold rare names and were born into small birth year cohorts in less populated parishes. Conversely, the highest rates of false positives are expected among individuals who hold common names and were born into populous parishes in years of high birth rates. The rate of false positives among individuals with very rare or even unique names is thus expected to be negligible given the small probability of making an erroneous match between these individuals. For individuals with very uncommon or unique names, the basic measure of false positives calculated above – i.e. the share of all those individuals with a different spouse in each census - should therefore be almost devoid of false positives and should closely approximate the true share of individuals who remarried between the two censuses. A more precise measure of false positives may thus be calculated by deducting the share of those who actually remarried (approximated by the share of individuals with uncommon names who remarried) from the upper bound rates of false positives calculated above.

Figure 7 plots by sex the adjusted predictions of the probability of having a spouse in 1900 who is different to the one in 1890 and across a measure of the relative commonness of an individual's first name (the percentage of the whole population sharing that individual's name). The adjusted predictions were calculated from the logit estimation of the following equation:

$$\Pr(\text{different spouse} = 1) = F(\beta_0 + \beta_1 \text{woman} + \beta_2 \text{name commonness} + \beta_3 \text{woman*name commonness})$$

Figure 7. Adjusted predictions of men and women at different levels of name commonness, with 95 per cent confidence intervals



It is clear from Figure 7 that the probability of having a different spouse recorded in each census is significantly and positively related to the commonness of an individual's first name. The adjusted prediction of the probability of a man with one of the rarest names having a different spouse equals 3.34 per cent. For women the corresponding number equals 1.48 per cent. Accepting these estimates as the true share of the sample who remarried between 1890 and 1900 yields an adjusted overall rate of false positives of 0.26 per cent (3.60 per cent - 3.34 per cent) for men and 0.07 per cent (1.55 per cent - 1.48 per cent) for women. False positives may thus be considered a marginal problem in the linked sample.

Conclusion

This paper has detailed the linking of three historical censuses. The results show that is possible to construct a linked sample characterized by high linkage rates and a low number of false positive links using a method

that is cost-effective and easily replicated. In total, around three quarters of all individuals were successfully linked to the preceding census. The results are particularly encouraging in terms of women's representation. Although women were not linked to the same extent as men, the linkage rates are close in magnitude. The resulting panel of women (constituting a total of 923,287 women observed at three points in time over 20 years) is to my knowledge the largest linked sample of a historical population of women available at this time. This in itself opens up innumerable opportunities for further research in many areas.

When combining the pairs of linked censuses into a panel covering three censuses and a 20-year period, the linkage rate declined to 60 per cent. Had secondary links been omitted, the overall linkage rate would have declined to less than 50 per cent. The fall in the linkage rate may be explained by differences in the propensity to be linked at different ages. Thus, despite a remarkably high total linkage rate, lower linkage rates among certain age groups result in bottlenecks. When combining this with the compounding effect that low linkage rates can have, one should expect quite severe declines in the overall linkage rate where longer periods of an individual's life course are covered by linking multiple censuses.

The linked samples are not without issues, however. Firstly, the individual's age clearly matters in terms of the probability of being linked. Young adults in particular are linked to a much lesser degree than are children and older individuals. Secondly, a great deal of variation in linkage rates follows geographic patterns. Rural residents were less likely to be linked. Moreover, differences in linkage rates exceeding 20 per cent were observed between different counties. If the purpose is to study specific cohorts or certain geographic areas, these issues are less problematic. However, if it is to make inferences about the population as a whole, appropriate measures such as weighting should be considered.

The linkage rates achieved compare favourably to those achieved previously. The two census linkage rates are roughly three times as high as the rates normally achieved. Collins and Wanamaker (2014) achieved a rate of 21 per cent between the 1910 and 1930 censuses, a rate not dissimilar to that of Long and Ferrie (2013) for the UK (20 per cent between the 1851 and 1881 censuses) and the US (22 per cent between the 1850 and 1880 censuses).

What explains the high match rate achieved relative to other attempts at linking historical censuses, and why is the estimated rate of false positives so

low in the resulting sample? Part of this success may be attributed to the source material. The basic quality in terms of the accuracy of variables used in matching certainly plays a role. Secondly, the small administrative units produced the index (the intersection formed by sex, year of birth and parish of birth) within which individuals are compared for possible matches are in most cases very small. The probability of false positives is thus greatly reduced through the elimination of many individuals who have similar names but were born in other parishes that would otherwise have constituted potential (and false) matches.

While it is always beneficial to start with basic data of high quality, the methods and results presented herein show that by carefully considering the linking process it is possible to improve linkage rates without resorting to methods that result in a decline in the quality of the sample. Importantly, the paper shows that it is possible to achieve a high linkage rate by using methods that may be applied generally. Instead of creating training data or carrying out extensive standardization of names specific to the source material, this paper demonstrates that by carefully evaluating the thresholds at which a link is classified as true, significant improvements in linkage rates are possible. More specifically, this method focused on finding an optimal threshold that resulted in a high linkage rate and a high rate of links that could be confirmed as true through a comparison of second names. This exercise revealed that the careful choice of a threshold means that linkage rates may be improved by several percentage points without this resulting in a significantly higher incidence of false positive links. A further approach, one easily adopted, involved the addition of secondary links. The addition of these links raised concerns about the biasing of the sample. After checking the representativeness of this sample, it was found that the addition of secondary links did not seriously bias it further apart from increasing the number of links between individuals belonging to large households.

The results demonstrate that high linkage rates are achievable using basic and transparent methods that depend on a minimal number of judgement calls. This in itself is an advantage since it enables complete replication of research from source material to eventual estimation of model results. To gauge the accuracy of the links in the sample, estimates based on the lower probability of false matches for rare-name individuals were used. The results indicate that the rate of false positives is at most a marginal concern in the linked sample.

References

A'Hearn, B., Baten, J. and Cayen, D. (2009). Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital. *The Journal of Economic History*, 69: 783-808.

Abramitzky, R., Bernheim, B.D., Mahajan, A. and Sin, I. (2011). *A Simple Procedure for Improving the Use of Imperfectly Linked Data.*

Abramitzy, R., Boustan, L.P. and Eriksson, K. (2012). Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, 102: 1832-1856.

Alter, G. (1988). *Family and the Female Life Course: The Women of Verviers, Belgium, 1849-1880*. Madison: The University of Wisconsin Press.

Bengtsson, T and Dribe, M. (2011). The Late Emergence of Socioeconomic Mortality Differentials: A Micro-Level Study of Adult Mortality in Southern Sweden 1815-1968. *Explorations in Economic History*, 48: 389-400.

BISOS A. (1907). *Befolkningsstatistik 1900*, Vol. XLII, No. 3, Statistics Sweden: Stockholm.

Bleakley, H. and Ferrie, J.P. (2013). Shocking Behavior: Random Wealth in Antebellum Georgia and Human Capital across Generations, *NBER Working Paper Series*, Working paper 19348, Cambridge: NBER.

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.

Collins,W.J. and Wanamaker, M.H. (2014). Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data. *American Economic Journal: Applied Economics*, 6: 220-252.

Feigenbaum, J.J. (2014). *Automated Census Record Linking.* Unpublished manuscript.

Fellegi, I.P. and Sunter, A.B. (1969). A Theory for Record Linkage. *Journal of the American Statistical Association*, 64: 1183–1210.

Ferrie, J.P. (1996). A New Sample of Americans Linked from the 1850 Public Use Micro Sample of the Federal Census of Population to the 1860 Federal Census Manuscript. *Historical Methods*, 29: 141-156.

Geschwind, A. and Fogelvik, S. (2000). The Stockholm Historical Database. In P.K. Hall, R. McCaa, G. Thorvaldsen (eds.), *Handbook of International Historical Microdata for Population Research*, 207 – 231. Minneapolis: Minnesota Population Center.

Goeken, R., Huynh, L., Lynch, T.A. and Vick R. (2011). New Methods of Census Record Linking. *Historical Methods*, 44: 7-14.

Hacker, D. (2013). New Estimates of Census Coverage in the United States, 1850-1930. *Social Science History,* 37: 71-101.

Herzog, T.N., Scheuren, F.J. and Winkler, W.E. (2007). Data Quality and Record Linkage Techniques, New York: Springer.

Kuckich, K. (1992). Techniques for Automatically Correcting Words in Text. *ACM Computing Surveys*, 24: 377–439

Long, J. (2005). Rural-Urban Migration and Socioeconomic Mobility in Victorian Britain. *Journal of Economic History*, 65: 1-35.

Long, J. and Ferrie, J. (2013). Intergenerational Occupational Mobility in Great Britain and the United States since 1850. *American Economic Review*, 103: 1109-1137.

Mill, R. (2012). Assessing Individual-Level Record Linkage Between Historical Datasets, in *Inequality and Discrimination in Historical and Modern Labor Markets*, Phd dissertation, pp. 83-104. Stanford University.

Newcombe, H.B., Kennedy, J.M., Axford, S.J. and James, A.P. (1959). Automatic Linkage of Vital Records. *Science*, 130: 954-959.

Parman, J. (2011). American Mobility and the Expansion of Public Education. *Journal of Economic History*, 71: 105-132.

Ruggles, S. (2006). Linking Historical Censuses: A New Approach. *History and Computing*, 14: 213-224.

Ruggles, S. (2014). Big Microdata for Population Research. Demography, 51: 287-297.

Ruggles, S. Roberts, E. Sarskar, S. and Sobek, M. (2011). The North Atlantic Population Project: Progress and Prospects. *Historical Methods*, 44: 1-6.

SOS. (1926). *Folkräkningen 1920*, No. 4, Kungl. Statistiska Centralbyrån: Stockholm.

Stewart, J. (2006). Migration to the Agricultural Frontier and Wealth Accumulation, 1860-1870. *Explorations in Economic History*, 43: 547-577.

The Swedish National Archives. (2012). *The SweCens NAPP Variable Documentation for the Swedish Censuses 1890 and 1900.*

Thernström, S. (1964). Poverty and Progress: Social Mobility in a Nineteenth Century City, Cambridge: Harvard University Press.

van Leeuwen, M. H. D. and Maas, I. (2011). *HISCLASS. A Historical International Social Class Scheme.* Leuven: Leuven University Press.

Wisselgren, M.J., Edvinsson, S., Berggren, M. and Larsson, M. (2014). Testing Methods of Record Linkage on Swedish Censuses. *Historical Methods*, 47: 138-151.

Paper 4

Onwards and Upwards? Internal migration and social mobility in Sweden, 1880-1900

Björn Eriksson
Department of Economic History, Lund University

Abstract
During historical periods of industrialization, migration served as an important route to social advancement. Past studies have shown that relocation to urban areas yielded particular benefits in terms of career prospects, pulling individuals from stagnant rural regions with limited opportunities into towns and cities. Despite the exceptional growth of the economy and cities in this period, the dominating type of migration remained short range and rural, a migration pattern commonly associated with pre-industrial societies and often characterized as a push phenomenon, with few associated gains for movers. This paper contrasts urban and rural migration and the gains realized thereof. The setting is late nineteenth century Sweden, a country going through rapid industrialization and urbanization. Using a new sample of linked men and women from the Swedish complete count censuses of 1880 and 1900, this paper estimates selection into migration and returns realized thereof.

JEL classifications: J10, J61, J62, N33

Keywords: Migration, Selection, Industrialization

Introduction

With the advent of industrialisation, the circular migration pattern emblematic of agricultural societies shifted towards a more modern pattern characterised by longer distance moves to urban, rather than rural, destinations. This shift in demographic behaviour and the consequential transfer of human capital from sending to receiving regions had important implications for both individual welfare and the economy as a whole.

Past studies have found that historically, different forms of migration served as important paths to upward social mobility (Long 2005; Stewart 2006; Abramitzky, Boustan and Eriksson 2012; Salisbury 2014). The existing literature is beset by two limitations. First, estimates of returns to migration have hitherto been limited to those realised by men. In light of women's high rates of mobility, an important aspect of historical migration thus remains untold (Ravenstein 1885; Moch 1992:130; Grant 2005:76). Given the importance of migration, this omission has implications for understanding women's career prospects relative to men and the resulting gender gap (Goldin 1990). Second, most studies only consider specific destinations, thereby neglecting the returns yielded by other potentially important migration streams. By evaluating returns realised across multiple destinations, the relative importance of different forms of migration may be considered in a richer context than that provided by more narrow comparisons.

This paper makes a contribution by examining migration in a comprehensive manner.[21] The analysis includes both male and female migrants and considers multiple destinations. The aim is to place the relative importance of migration for social attainment in a greater context by considering the importance of both short and long range migration to rural and urban destinations for both sexes. I focus on turn of the twentieth century Sweden, a country experiencing rapid industrialization and increasing rates of internal migration. The analysis is based on a rural cohort of men and women born between 1860 and 1875 that transitioned into adulthood during the height of Swedish industrialisation. These men and women are first observed

[21] Because of data limitations, emigration, which primarily took place to the US, has been omitted from the analysis.

as children and adolescents residing with their parents in 1880, and then again twenty years later in 1900, after they left home. Upon leaving their parental home, most moved a relatively short distance, often to a rural area not dissimilar to that in which they were born. By doing so, these migrants were following a well-established pattern dating back to pre-industrial Sweden. Some did however not follow in these well-trodden tracks, instead migrating further away, often choosing to settle in a city rather than remain in the countryside.

Who were these migrants, and how did they fare at this historical juncture between an old and new migration regime? Did a declining agricultural sector lead to push-migration and insecure and low quality urban jobs, or did structural changes in the labour market remove barriers to geographic and social mobility and serve up migrants with new opportunities for upward occupational mobility? I set out to answer these questions by considering both selection into migration, and the returns realised thereof, using linked samples based on full count census data. I find overall that, migrants realised important gains. By moving, migrants improved their chances of moving up the occupational ladder, find steady employment and avoid poverty. However, the findings also show that migration was a highly selective process. A clear social gradient is apparent with regards to both migration distance and the choice between settling in an urban or rural destination.

Related literature

Migration and industrialisation

As the nineteenth century was drawing to an end, Ravenstein (1889:288) concluded that ‘an increase in the means of locomotion and a development of manufactures and commerce have led to an increase of migration. In fact you need only seek out those provinces of a country within which migration is proceeding most actively, and you will either find yourself in the great centres of human industry, or in a part of the country whose resources have only recently become available. Migration means life and progress; a sedentary population stagnation’. The changes in migration patterns described by Ravenstein have been argued to constitute an essential

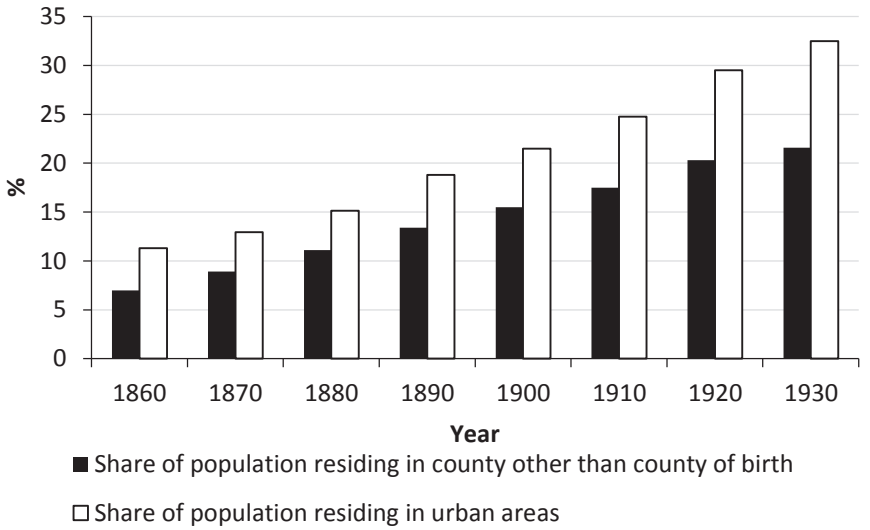
component of the modernization process (Zelinsky 1971:222), conspicuous enough to mark a transition from a pre-industrial to a modern migration pattern (Parish 1973; Pryor 1975). The decline of traditional trades and the emergence of new occupations, faster and cheaper forms of transportation, unbalanced regional development and increasing population growth were some of the factors that served to facilitate this transition (Parish 1973).

This shift does not imply that pre-industrial societies were geographically immobile. Agricultural societies were in fact characterised by high rates of geographic mobility. This is true for Sweden (Gaunt 1977:195, Dribe 2000:5-6, Dribe 2003), and a number of other European countries (Moch 1992; Jackson & Moch 1989), such as Germany (Hochstadt 1983), England (Patten 1976, Clark 1979) and Scotland (Devine 1979). Migration was to a large extent driven by frequent short distance moves between rural areas by young people working as servants (Eriksson and Rogers 1978; Dribe & Lundh 2005). Migration was tied to the seasons and was the means by which the young earned their keep and acquired skills in agricultural work (Moch 1992: 61).

What sets modern geographic mobility apart from the pre-industrial agrarian case is thus not primarily the incidence of migration but rather its form. In contrast to pre-industrial Sweden, geographic mobility during the second half of the nineteenth century display a markedly different pattern (Brändström, Sundin & Tedebrand 2012). Migration was increasingly taking place over longer distances in response to relatively higher wages in industrialising regions (The Institute for Social Sciences 1941:42; Jörberg 1972: 348). As can be garnered from Figure 1, only 7% of the population in 1860 resided in a county different from their county of birth; by 1900 however, the share of county migrants had more than doubled to 15.5% and continued to rise over the following decades (also see Thomas 1941:28). This resulted in most towns growing at a vigorous pace, but the three largest cities, Stockholm, Gothenburg and Malmö grew particularly fast (Statistiska Centralbyrån 1969). The inflow of migrants from the countryside resulted in a doubling of the urban share of the population between 1860 and 1900 (see Figure 1).

This transition was not unique to Sweden, but was reflected in many other European countries (Baines 1985; Moch 1992; van der Woude 1992). With the advent of industrialisation the population of Europe had entered a new phase of geographic mobility. As a result of falling transportation costs, a declining agricultural sector and new opportunities outside of farming, the

Figure 1. Migration and urbanisation, 1860-1930



Sources: Sveriges Officiella Statistik 1930: 84*; Statistics Sweden 1999: 42.

nature of migration was changing. Instead of making a living in agriculture, increasing numbers of people were leaving the countryside in favour of cities.

As long as certain administrative requirements were met there were no legal obstacles to internal migration in Sweden during the nineteenth century. Before moving, a prospective migrant was required to notify the ministers of both the home and destination parish. Permission to settle in a new parish was given as long as it was not suspected that the migrant would have difficulty supporting him- or herself. Refusal of permission to move was exceedingly rare, with less than 1 per cent of applications denied (Eriksson & Rogers 1978:180-181). One institutional barrier to migration did however exist, the Servants Act, which in particular hampered migration for agricultural workers. The act mandated that yearly employment contracts for farmhands and maids must begin on the 1st of November and run until the 24th of October the following year. This resulted in little down time between employment contracts, which made it difficult for farm workers to find employment anywhere beyond the vicinity of their last place of work (Lundh 1999:61; Lundh 2003). Early and mid-nineteenth century migration was also limited by the late arrival of the railway on which construction did not

commence until 1855. The arrival and expansion of the railway meant a considerable reduction in transportation costs and transit times, which resulted in migrants flocking to towns connected to the railway (Berger and Enflo 2013).

Migration and social mobility

The societal changes associated with industrialisation has been hypothesised to not only denote a shift in migration patterns, but also herald a break from a stagnant past in which occupational mobility was primarily determined by social factors and less influenced by individual agency (see Treiman 1970). An increase in social mobility is generally viewed as something normatively good, implying fairness and equality for individuals and a more open society (Featherman and Hauser 1978; Eriksson and Goldthorpe 1992). Whether industrialisation brought about an overall increase in social mobility is debated (see for example van Leeuwen and Maas 2010).

Irrespective of its overall effect on social mobility, industrialisation did change the occupational structure of the workforce and the geographic demand for labour. In rural areas the importance of labour as a factor of production was steadily declining as commercialisation and mechanization took hold in agriculture (Schön 2010). In its place, urban employment in industry and services was gaining in prominence. The associated transition from workshops to factory-based mass production was changing the demand for labour. New production methods and industries meant an increase in mechanization, deskilling of labour in manufacturing and a reduction in the need for artisans (Goldin and Katz 1998). Simultaneously, new innovations and technology required growing cadres of scientists and engineers, leading to an increase in the demand of highly skilled labour (Mokyr 2009). Parallel to this change, the emergence of nascent forms of bureaucracies such as railways, banks and retailers resulted in white collar workers, supervisors and managers increasing in both numbers and importance (Carter and Carter 1985; Howlett 2004; Seltzer and Frank 2007).

How did migrants fare in these choppy waters? Did migrants ascend the occupational ladder by moving into skilled and well paid jobs, or were they pushed into unskilled and low paying jobs at the bottom rungs? Theoretically, migration is often conceptualized as an investment decision in which each possible move is associated with certain monetary and non-monetary benefits

and costs (Lee 1966). We thus expect migration to occur only in those cases where the expectation of upward social mobility is positive. This simple model may be extended by considering the probability of finding employment in a given destination, thereby introducing an element of risk into the migration decision (Sjastaad 1972; Harris and Todaro 1970). A migrant may hence be willing to trade income or occupational status for an increase in security thereof, or vice versa.

The link between internal migration and occupational mobility has been the subject of a number of articles studying individual status attainment during periods of industrialisation in various settings (see Long & Ferrie 2007: C69; Guest, Landale & McCann 1989:376). Long (2005) find that returns to urban migration in Victorian Britain in terms of intergenerational upward mobility was substantial for migrants from all social classes. Migrants to American frontier cities during the same period were likewise rewarded with higher job quality than would have been achieved had they stayed in their place of origin (Stewart 2012). Estimates for black American migrants moving from the south to the north a few decades later paint a similar picture (Collins and Wanamaker 2014). In the Dutch province of Zealand, nineteenth century migrants fared better than non-migrants in terms of social advancement (Zijdeman 2009). For Sweden, Maas and van Leeuwen's (2004) study of the Sundsvall region indicate that migrants seemed to do somewhat better than the stable population in terms of upward career mobility during the middle of the nineteenth century. On the whole, migrants seem to have realised gains in the form of upward occupational mobility across a number of historical settings.

Selection into migration

While some portion of estimated returns to migration may be fixed and realizable by all and sundry, most are not. Expectations, ability, benefits, costs and resources are all characteristics that vary between individuals and simultaneously determine the incidence of migration and the return thereof. Migration is as a result a highly endogenous process undertaken by a certain groups and individuals, each differently selected depending on individual characteristics and circumstances.

Migration cost is an important determinant of selection, increasing the threshold at which migration becomes a profitable and viable option. The

result is selection of the most able, ambitious and entrepreneurial part of the population who are better able to recoup costs in the form of substantial returns (Lee 1966). Similarly, costs may affect selection if cost is a negative function of ability, the able being, in Chiswick's (1999) words "more efficient in migration". Upfront migration costs may also serve as a more direct barrier by preventing the financially constrained from moving.

Borjas (1987) following Roy (1951) showed how migrants sort into different destinations according to ability. Depending on the wage structure in the origin and destination locations, migrants may either be negatively or positively selected with regards to labour market outcomes. If the returns to skills are higher in the destination location relative to the origin location, migrants will constitute the most skilled individuals from the location of origin, resulting in positive selection. If the opposite relationship holds, negative selection should be expected. The notion that some migration streams may be positively and others negatively selected has also been argued to depend on the causes of migration, with migrants reacting to push factors tending towards negative selection while those responding to pull factors tending towards positive selection (Lee 1966:56).

Apart from complicating the estimation of returns to migration, selection is in itself of interest. By analysing selection patterns, the importance of migration for both individual mobility and society at large may be further nuanced. The degree of selection into migration informs how accessible migration was to different social strata and whether migration served as an equalising force by providing a cheap and profitable form of human capital investments, or further entrenched social inequality by being primarily undertaken by the elite. Moreover, the selection of migrants has implications for the economy at large, potentially altering the geographic distribution of human capital by draining sending regions of its most able and productive workers.

Hypothesising about selection is nigh on impossible without knowledge about the specific setting and the characteristics and circumstances of migrants (Jackson and Moch 1989). Ultimately the general direction and degree of migrant selection is a question that has to be answered based on the evidence at hand. Empirical studies of migrants selection have yielded mixed results. In the case of the US, there is evidence that American agricultural frontier migrants in the 1860s tended to be negatively selected from the source population in terms of wealth and literacy (Stewart 2006). This finding contrasts with the positive selection of black migrants moving from

the rural South to the urban North during the beginning of the nineteenth century (Collins and Wanamaker 2014). For nineteenth century England, there is evidence in favour of rural to urban migrants being on average positively selected (Long 2005; Humphries and Leunig 2009). For Sweden, Maas and van Leeuwen's (2004) study of the Sundsvall region instead indicated that internal migrants during the nineteenth century were positively selected. Abramitzky, Boustan and Eriksson (2012; 2013), covering nearly the same period (1865-1900), found that Norwegian emigration and internal migration were discouraged by parental wealth, resulting in negative selection of migrants. Wegge (2002) nuanced the historical picture of migration selection further by showing that nineteenth century migrants to the US from the Hesse-Cassel region of Germany were positively selected in terms of skills, but negatively selected in terms of wealth.

Data

Linked census data

This paper uses two data sources, the complete Swedish censuses of 1880 and 1900. The censuses are the most comprehensive source of individual level data for Sweden around the turn of the century. The Swedish censuses differ from the U.S. and British censuses by not being the product of a census taking done by enumerators actually visiting and counting the populace. Instead, with one exception, that of the city of Stockholm, the Swedish censuses were the result of a compilation of excerpts from continuous parish registers which were kept by the Swedish Lutheran church and maintained by the parish priest. For Stockholm, the source of the census were excerpts from the *Roteman* register, an administrative register supervised by *Mantalsnämnden* (The population and tax registration board), which replaced the Church registers in 1878 in order to cope with the rapidly growing, and increasingly mobile, population of Stockholm at the end of the nineteenth century (Geschwind & Fogelvik 2000:207-208).

Because the Swedish census is no more than an excerpt from a continuous and consistent source rather than a recreation of a population register as in the case of the U.S. and British censuses, the quality of the raw data is comparatively better. Any errors resulting from the misreporting by

the enumerated or recording mistakes by enumerators may thus be largely discounted. Moreover, because Swedes were entered into the parish books at the time of christening and not removed until time of death or emigration, the under-enumeration of the population as whole and specific groups is less of a problem than what is normally expected from historical censuses.

In order to study migration and social mobility it is necessary to be able to ascertain an individual's geographic location and social status before and after migration has taken place. The analysis relies on a new panel sample which has been created by linking individuals between the 1900 and 1880 Swedish complete count censuses.[22] The linking process relies on exact comparisons of sex, birth place and birth year, and probabilistic matching of names for identifying and linking individuals between the censuses. Importantly, and uniquely, women appear with their maiden name, even after marriage, in the Swedish censuses. This enabled women to be linked to nearly the same extent as men between the two censuses.

One significant group has by necessity been excluded from the analysis: emigrants. The reason is that in order to be linked between the censuses, an individual had to reside in Sweden in both 1880 and 1900. Anyone emigrating out of Sweden between the two time points was thus lost in the linking process. Therefore, although the sample used in this analysis is based on the whole population, it is to some extent a self-selected sample by virtue of not having emigrated.

Sample restrictions

From the linked sample of the Swedish censuses of 1880 and 1900, a sub-sample of men and women that were born between 1865 and 1880 in a rural area were selected. I further restrict the analytical sample to those that resided with their father in 1880 and were subsequently observed having left the parental home in 1900. For any analysis relying on own or fathers occupational rank, all individuals with either no information about own or paternal occupation were by necessity excluded. Table 1 presents a comparison of summary statistics for all men and women meeting the sample

[22] For a full description of the linking of the Swedish censuses of 1880, 1890 and 1900 see paper 3 of this dissertation.

Table 1. Means of variables (standard deviation for continuous variables within parenthesis) for men and women in the 1880 census, linked samples, and analytical samples

	Men				Women					
	Census	Linked	Analytical samples		Census	Linked	Analytical samples			
	I	II	IIIa	IIIb	I	II	IIIa	IIIb	IVa	IVb
Own Hiscam-score			52.87	53.12			47.49	48.68		
			(9.34)	(9.64)			(9.34)	(10.07)		
Husband's Hiscam-score									52.43	52.39
									(8.27)	(8.34)
Father's Hiscam-score	50.52	50.74	50.81	51.2	50.58	50.84	50.83	51.85	50.58	50.8
	(6.57)	(6.83)	(7.20)	(7.69)	(6.72)	(7.04)	(7.38)	(8.59)	(6.62)	(6.80)
Distance moved			45.51	46.33			69.82	74.42	30.68	30.45
			(105.95)	(106.53)			(110.50)	(109.38)	(78.17)	(76.18)
Urban migrant			0.17	0.18			0.39	0.44	0.11	0.1
Mother present	0.96	0.97	0.96	0.96	0.96	0.96	0.95	0.95	0.96	0.97
Farm household	0.65	0.67	0.62	0.61	0.65	0.67	0.59	0.59	0.66	0.68
Own age	11.59	11.27	12.12	12.04	11.48	11.27	10.35	10.58	12.22	12.17
	(4.42)	(4.49)	(4.40)	(4.25)	(4.37)	(4.42)	(4.26)	(4.06)	(4.29)	(4.14)
Father's age	46.47	46.42	47.32	47.59	46.42	46.32	46.12	46.94	46.91	46.89
	(9.75)	(9.76)	(10.66)	(11.40)	(10.20)	(9.46)	(10.43)	(7.87)	(8.97)	(7.28)
Siblings	3.36	3.31	3.32	4.2	3.34	3.29	3.43	4.37	3.2	4.15
	(1.88)	(1.89)	(1.90)	(1.76)	(1.89)	(1.89)	(1.84)	(1.75)	(1.92)	(1.80)
Birth order	2.42	2.47	2.4	2.83	2.42	2.45	2.58	3.04	2.29	2.73
	(1.43)	(1.46)	(1.41)	(1.53)	(1.43)	(1.45)	(1.48)	(1.61)	(1.36)	(1.50)
Male servants	0.14	0.18	0.15	0.15	0.17	0.21	0.14	0.18	0.2	0.23
	(0.47)	(0.53)	(0.50)	(0.52)	(0.51)	(0.56)	(0.49)	(0.54)	(0.54)	(0.59)
Female servants	0.21	0.25	0.23	0.26	0.18	0.22	0.16	0.2	0.19	0.19
	(0.57)	(0.63)	(0.62)	(0.68)	(0.56)	(0.62)	(0.53)	(0.61)	(0.57)	(0.61)
N	468298	184527	109411	36556	450875	185772	26874	4197	92609	24124

Sources: SweCens, Swedish National Archives.

restrictions in the 1880 census (I), in the linked sample (II) and in the analytical samples (III and IV). Moreover, each analytical sample includes a sub-sample of sisters and brothers (IIIb, and IVb) used for estimating sibling fixed effects models.

Migration distance and destination

For all men and women in the linked samples the parish of residence in both 1880 and 1900 is known. Using coordinates of the centroid of each parish, the distance between the origin in 1880 and destination in 1900 was calculated using the Euclidian distance between each parish. Or in simpler words, the distance as the crow flies from the middle of one parish to another. Because Swedish parishes were typically small (there were in total around 2500 parishes in Sweden around the turn of the century) the calculated distance is a good approximation of actual distance moved. Apart from distance, the chosen destination is of primary interest. I use the census definition of urban and rural in order to classify migrants accordingly.

To better illustrate the difference in mobility between men and women and to facilitate modelling, I define migration as a categorical variable using distance moved and chosen destination. Any individual that remained in the same parish in both 1880 and 1900 were defined as a non-migrant. This category is used as the reference category throughout. Six more categories were created based on if the destination was rural or urban and whether the distance moved was short (0-25 km) medium (25-75 km) or long (> 75 km). The distribution of men and women between the different categories is presented in Table 2.

Table 2. Distribution of male and female migrants by distance and destination

	Men		Women		Difference	
	N	%	N	%	%	
Non-migrant	49539	45.30%	47014	39.30%	5.90%	***
Rural 0-25 km	22074	20.20%	30271	25.30%	-5.20%	***
Rural 25-75 km	10431	9.50%	13142	11.00%	-1.50%	***
Rural >75 km	8250	7.50%	8490	7.10%	0.40%	***
Urban 0-25 km	5251	4.80%	6574	5.50%	-0.70%	***
Urban 25-75 km	5156	4.70%	6058	5.10%	-0.40%	***
Urban >75 km	8710	8.00%	7934	6.60%	1.30%	***

Sources: see Table 1
Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Among the men and women in our sample, 45 per cent of men and 39 per cent of women still remained in the same parish in 1900 in which they had resided twenty years earlier. Of those that did choose to leave, most moved only a short distance, settling in the nearby countryside. In total, around 17 per cent of all men and women moved to an urban area between 1880 and 1900. Among urban movers, long distance moves were more common. There are differences between men and women in terms of propensity to move and destination chosen. Firstly, the share of non-migrants was significantly higher among men than women, differing by about 6 percentage points. There are also differences between women in terms of distance moved and destination chosen. Overall, women were more likely than men to make a short or medium distance move, while men were more likely to move more than 75 kilometres than women. It should however be noted that although these differences are statistically significant, they are not great in magnitude. On the whole, the general pattern in terms of distance moved and whether the destination was rural or urban is similar between men and women.

Although the Swedish population were transitioning towards a more modern migration pattern during the second half of the nineteenth century, the pre-industrial pattern of rural short distance migration clearly still remained the most common type of migration for the majority of men and women. These basic patterns further motivate a comprehensive analysis of migration which considers for multiple destinations rather than one migration stream.

Measuring social mobility

In this paper social status enters the analysis as both an explanatory and an outcome variable. When analysing selection into migration, father's social status observed in 1880 prior to migration is the explanatory variable of interest. Turning to the second question, in which I seek to estimate the returns to different forms of migration, the focus is instead on the outcome of sons and daughters in the form of their own observed social status in 1900.

Because of data limitations, the prevailing method for measuring social mobility in historical populations typically relies on transitions between occupations. Returns to migration in the form of social mobility are thus typically estimated as the probability of entering a particular social class

(Long 2005). An alternative approach is to transform occupational categories into a continuous measure by assigning an income score to each occupation, which yields a monetized estimate of returns to migration (see for example Abramitzky et al. 2012; Collins and Wanamaker 2014). Transforming occupational information into a continuous measure has distinct advantages, allowing for increased flexibility in terms of modelling and providing simple and more intuitive estimates of returns.

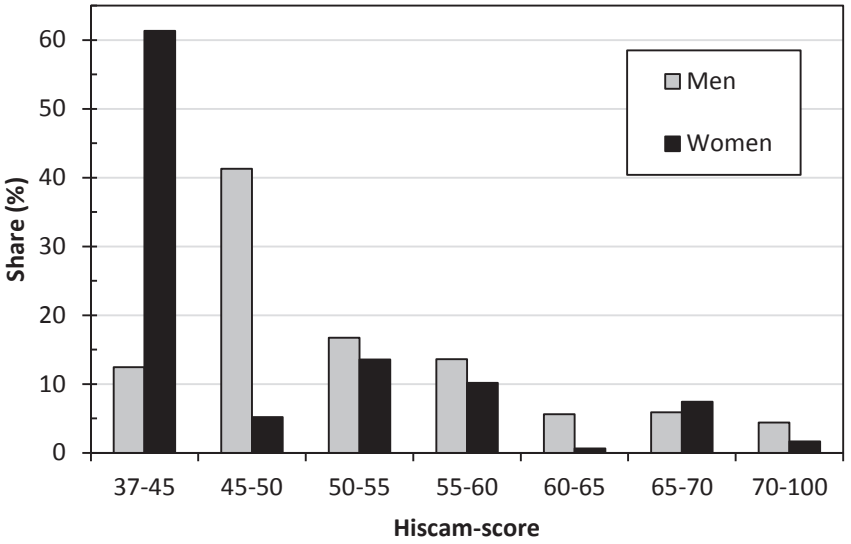
The Swedish censuses include no information about wages or wealth. I therefore rely on occupations to measure the social mobility of migrants. All occupations were coded using the Hiscam-scale (Lambert et al. 2013). Hiscam is a continuous measure of social status denoting an individual's relative position in the social hierarchy in the form of a score ranging from 37 to 99. Because of the small sample size used for constructing the Swedish scale I use the universal scale for both women and men.[23]

Although the linked sample includes as many women as men, only a fraction of all women remain after the sample is restricted to individuals with occupational information in 1900. This is because almost all married women lack a recorded occupation in the censuses. This effectively means that a sizable share of all women has been self-selected out of the sample as a result of marrying between 1880 and 1900. To address this shortcoming married women's social status is measured using the occupational information of their husbands. Unmarried women's returns to migration are thus estimated using individual social status in 1900 (using samples IIIa and IIIb in Table 2) while married women's returns are estimated using the social status of their husbands (using samples IVa and IVb).

The distribution of individual Hiscam-scores for men and unmarried women is shown in Figure 2. Most women fall into the bottom category while men are primarily found in the category above. Men are also more prominent in the upper tail of the distribution. The difference between men and women at the bottom of the distribution is driven by two groups that dominate the occupational distribution, female servants and male farmers. Of all women with occupational information in 1900, 55 per cent were domestic servants (Hiscam-score 41.44). The most common male occupation is that of

[23] The scales are available to download at: http://www.camsis.stir.ac.uk/hiscam.

Figure 2. Distribution of men's and women's Hiscam-scores, 1900



Sources: SweCens, Swedish National Archives.

general farmer (Hiscam-score 49.92), held by 21 per cent of the men in the sample.

Alternative outcomes

Using information derived from occupations suffers from a distinct disadvantage by requiring individuals to have a recorded occupation to be included in any analytical sample. The problem is endemic to all analyses relying on occupational information but becomes particularly pertinent when studying migration. Occupational mobility is only one benefit from migration which migrants may respond to and realise as a return. This is particularly true for economically disadvantaged migrants who may value security of employment above job quality or prospects for upward mobility. To address this shortcoming the analysis is extended by considering two alternative

benefits from migration; the likelihood of employment and of escaping poverty.

The censuses include information about whether an individual was working or not. The census definition of being gainfully employed does not include students. Given the purpose of the analysis I recoded any students as being gainfully employed. The poverty variable was created by defining any individual as poor if the word *fattig* (poor) appears either in their own or their spouse's occupational description, or if it appears in the residence name, as in for example *fattighus* (poorhouse).

Official statistics on the numbers receiving poor relief in 1900 indicate important differences between rural and urban areas. In the countryside, 3.94 per cent of the population received some kind of poor relief. In cities the corresponding share was higher; 6.86 per cent, with the three largest cities being particularly beset by poverty[24] (Statistiska Centralbyrån 1900).

Other covariates

In addition to information about migration and occupations, a number of individual covariates expected to affect social mobility independently from migration are considered. All control variables refer to 1880 and are observed prior to migration and are thus not a realisation thereof. For each individual, own age and father's age are included. These are important variables since we generally observe sons' occupations in 1900 at a younger age than that of their fathers in 1880, which may result in a lifecycle bias. Based on the father´s occupation, a dummy variable was also constructed which indicates whether the household was a farm or not. The effect of siblings (see Becker and Lewis 1973) is accounted for by both the number of siblings present in the household in 1880 and birth order approximated by the age position of the individual within the sibship. The effect of mothers is captured through a dummy which denotes whether an individual's mother was present in the household in 1880. Finally, household wealth was approximated by including variables accounting for the number of male and female servants employed in the household in 1880.

[24] The share of the population receiving poor relief: Stockholm, 7.85 per cent; Gothenburg, 9.32 per cent, Malmö 7.87 per cent.

Migration and social mobility: Empirical results

Migration and selection on father's social status

Before proceeding with estimating returns to different forms of migration I address the issue of selection into migration. Specifically I analyse the relationship between the social status of fathers in 1880 and children's subsequent migration between 1880 and 1900 in terms of distance moved and the choice between settling in an urban or rural destination. To illustrate the relationship between social status and migration distance, Figures 3A and 3B plot the mean distance moved by sons and daughters against fathers' Hiscam-score. The size of the plots denotes the relative size of each group. The Figures reveal a clear selection pattern. Children of fathers from the upper end of the status distribution were susceptible to moving considerably further than those starting out in the bottom of the distribution which on average only moved around 30 kilometres.

The positive selection pattern also extends to destination choice. As shown by Figures 4A and 4B, when they did choose to move, children of high status fathers were more likely to choose an urban destination rather than remaining in the countryside. Only about a third of migrants from a low status household moved to an urban area, a number which doubles for migrants from the top of the income distribution.

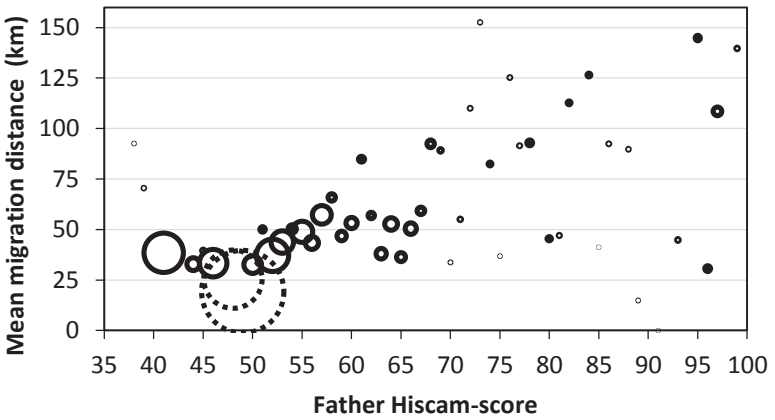
Two groups, denoted by dashed outlines, stand out by virtue of both their size and migration pattern when compared to other groups. The two circles represent the most common agricultural occupations in the census, 'subsistence farmer'[25] (the smaller circle) and 'general farmer'[26] (the larger circle). Sons and daughters of these two groups on average moved a considerably shorter distance than both the overall population and the population as a whole. When sons and daughters of farmers did move, they were also less inclined to settle in an urban area. The children of farmers, perhaps not surprisingly, were thus very much still adhering to a

[25] HISCO-code 61115.

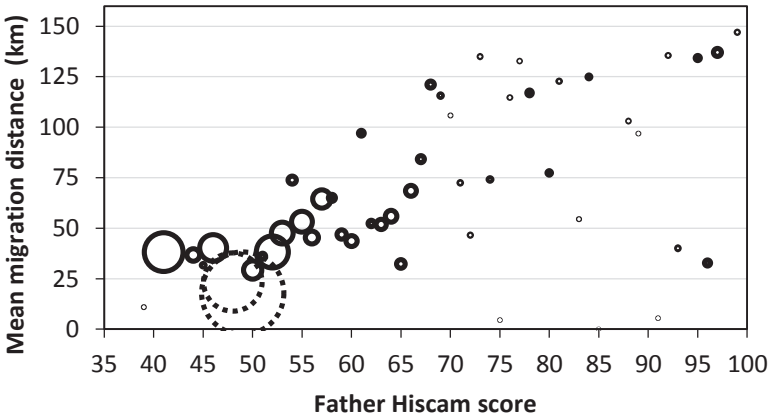
[26] HISCO-code 61110.

Figure 3. Association between father's Hiscam-score and distance moved, 1880-1900

A. Men



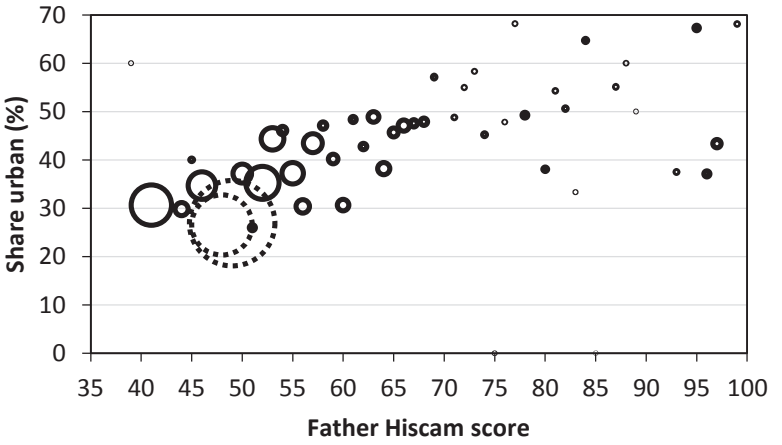
B. Women



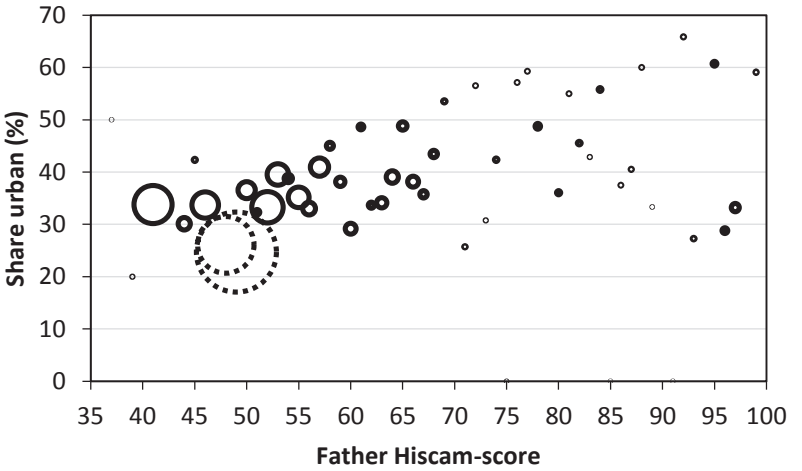
Sources: See Figure 2.

Figure 4. Association between father's Hiscam-score and migrants' choice between an urban or rural destination, 1880-1900

A. Men



B. Women



Sources: See Figure 2

pre-industrial migration pattern dominated by short distance moves in the countryside.

It is possible that the selection patterns in Figures 3 and 4 are caused by differences in local conditions, which in turn is reflected in the social status of fathers. To confirm that the observed pattern is determined at the household rather than local level, I estimated OLS (for distance) and logit (for destination) models including parish of origin fixed effects. Any differences attributable to local conditions are thus differenced out, resulting in a within parish of origin estimate of the correlation between social status and migration. The results are presented in the appendix in Table A1. The coefficient of fathers' Hiscam-score confirms that the positive social gradient for both distance and destination remains after controlling for local conditions in the origin for both men and women.

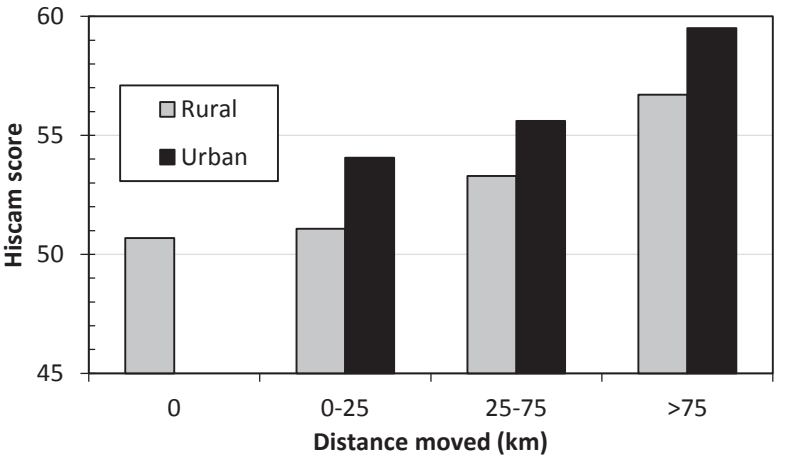
Both graphic evidence and a more precise analysis controlling for local conditions in migrants' geographic origin reveal a clear selection pattern in migration. Long distance migrants were positively selected in terms of social background, while short distance migrants tended to be drawn from the bottom of the social strata. Urban migrants were positively selected and rural migrants were negatively selected. Although the data does not permit us to pinpoint whether migration was restricted by financial resources, a lack of skills or information, it is clear that factors related to the social background of migrants were important determinants of migration for both men and women. Beyond being an important finding in itself, the social gradient in migration informs the identification strategy employed when estimating the returns realised from migration.

Returns to migration

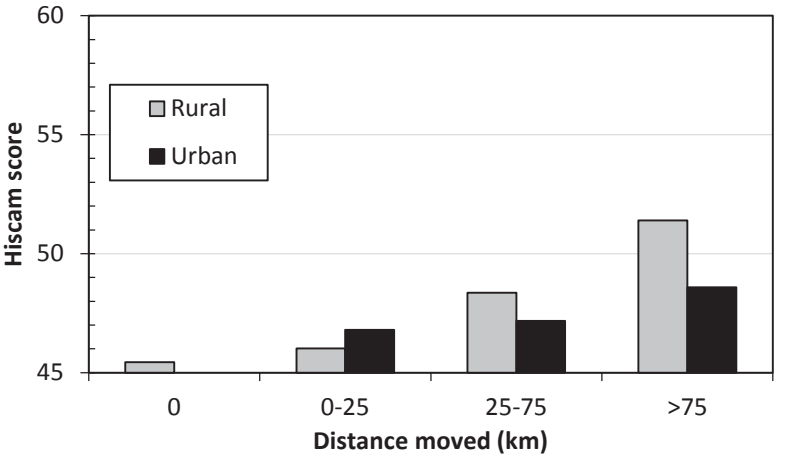
This section seeks to estimate returns realised by migrants in the form of upward social mobility. A simple comparison presented in Figure 5 of mean Hiscam-scores in 1900 according to distance moved and urban/rural destination reveals some clear patterns. For men there is a positive relationship between distance moved and own social status in 1900. In addition, at each given distance, urban migrants seem to have realised a further premium of about 2-3 points. For women, the patterns are less straightforward. Rural female migrants, just as their male equivalents, display

Figure 5. Mean Hiscam-score in 1900 by distance moved and rural/urban destination

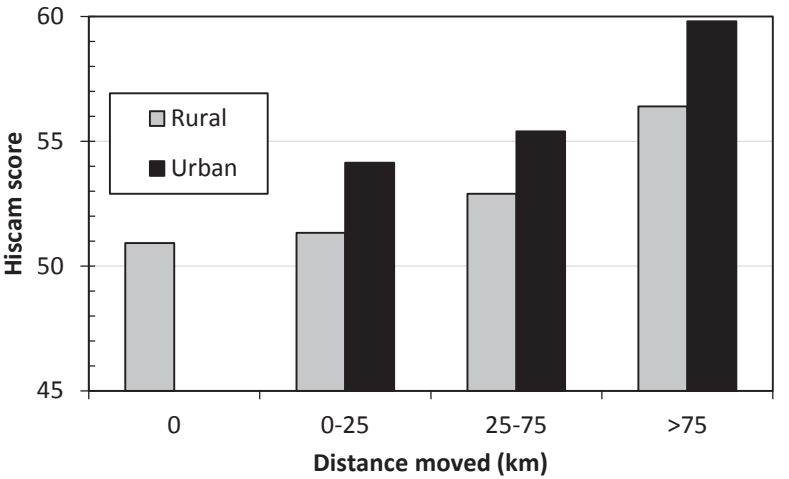
A. Men, individual Hiscam-score



B. Unmarried women, individual Hiscam-score



C. Married women, spouse's Hiscam-score



Sources: See Figure 2.

a clear positive correlation between distance and own social status. The positive urban premium for men is however not present among female migrants. In fact, for medium and long distance migrants an urban penalty is apparent. When we instead consider the status of married women measured by the status of their spouse, the results mirror that of men.

Although illustrative, the above comparisons do however provide little beyond circumstantial evidence. Moreover, the finding that migration was a highly selective process further complicates the issue.

To test whether the correlations between migration and upward occupational mobility stands up to further scrutiny and to strengthen causal claims the following linear model is estimated by ordinary least squares:

$$\ln(Hiscam_i) = \beta + \beta \boldsymbol{M} + \beta \ln(Hiscam_{father}) + \beta \boldsymbol{X} + u \qquad (1)$$

in which the dependent variable, $Hiscam_i$ denotes the Hiscam-score of a son or daughter in 1900. $\boldsymbol{M}$ is a vector consisting of different forms of migration defined above (short, medium and long distance migration to a rural or urban area respectively) with non-migrants as the reference category. The analysis of selection into different migration streams clearly show that fathers social status is a strong predictor of migration behaviour. Beyond this, father's

social status is related to both resources and ability, which affects son's occupational status independent of migration. Father's social status is thus accounted for by $Hiscam_{father}$ which denotes the Hiscam-score of fathers in 1880 prior to migration taking place. $\boldsymbol{X}$ is a vector of individual characteristics, again observed in 1880. The error term is denoted by u.

An important concern is that the error term, u, may be correlated with migration, leading to bias of the estimates as a result of selection into migration. Apart from including the social status of fathers in all models, I address selection in two ways, beginning with omitted place of origin effects. The point of departure in the basic cost-benefit model of migrations is a comparison of expected benefits between ones place of residence and possible destinations. Migrants are thus expected to exhibit a tendency to leave places were opportunities are scarce for destinations with better prospects. By corollary individuals born in more prosperous areas will tend to stay. By adding parish of origin fixed effects, the resulting estimates are interpreted as a within parish of origin comparison of migrants' and non-migrants' outcomes net of observed individual characteristics contained in $\boldsymbol{X}$.

As a final step, migration gains are identified by comparing the migration and outcomes of siblings (for recent historical articles employing a similar strategy see for example Abramitzky et al. 2012; Collins and Wanamaker 2014). Running an OLS regression of equation (1) with sibling fixed effects absorbs the portion of the error term which is identical across siblings. The estimation will eliminate bias derived from aspects such as family resources and genetics shared between siblings that affect both the probability of migration and labour market outcomes. In this case, the coefficients on $\boldsymbol{M}$ measure the return to migration, net of characteristics fixed at the household level that were not captured by the social status of fathers. Moreover, the difference between the within-siblings estimate and the more basic models is informative regarding selection across families. If the coefficients for migration are reduced, it suggests that there is positive selection into migration of sons and daughters belonging to more high achieving household.

I begin by estimating equation (1) for men (using sample IIIa described in Table 2). The results are presented in Table 3. I begin by estimating a parsimonious model (column I) which only includes own age, father's age and father's Hiscam-score as control variables. In column (II), a further set of individual characteristics are added. The coefficients for the control variables reveal nothing abnormal. The Hiscam-score of fathers is positive in relation

Table 3. Ordinary least squares estimations of individual Hiscam-score for men

	I	II	III	IV
Dependent variable: ln(own Hiscam-score)				
Migration (Ref: 0 km)				
Rural 0-25 km	0.013 ***	0.014 ***	0.013 ***	0.016 ***
	(0.001)	(0.001)	(0.002)	(0.003)
Rural 25-75 km	0.049 ***	0.047 ***	0.052 ***	0.046 ***
	(0.002)	(0.002)	(0.002)	(0.004)
Rural >75 km	0.098 ***	0.092 ***	0.097 ***	0.076 ***
	(0.002)	(0.002)	(0.003)	(0.005)
Urban 0-25 km	0.062 ***	0.060 ***	0.059 ***	0.044 ***
	(0.002)	(0.002)	(0.003)	(0.006)
Urban 25-75 km	0.083 ***	0.080 ***	0.083 ***	0.068 ***
	(0.003)	(0.003)	(0.003)	(0.006)
Urban >75 km	0.137 ***	0.128 ***	0.128 ***	0.094 ***
	(0.002)	(0.002)	(0.003)	(0.005)
Father's Hiscam-score	0.353 ***	0.264 ***	0.258 ***	
	(0.005)	(0.005)	(0.006)	
Age	0.003 ***	0.004 ***	0.004 ***	0.004 ***
	(0.001)	(0.001)	(0.001)	(0.001)
Age^2	0.000 ***	0.000 ***	0.000 ***	0.000 ***
	(0.000)	(0.000)	(0.000)	(0.000)
Father age	0.000	0.000	0.000	
	(0.000)	(0.000)	(0.000)	
Father age^2	0.000	0.000	0.000	
	(0.000)	(0.000)	(0.000)	
Farm household		-0.031 ***	-0.023 ***	
		(0.001)	(0.001)	
Mother present		0.006 ***	0.006 ***	
		(0.002)	(0.002)	
Siblings		-0.001 *	0.000	
		(0.000)	(0.000)	
Birth order		0.002 ***	0.002 ***	-0.004 **
		(0.000)	(0.000)	(0.002)
Male servants		0.001	0.003 *	
		(0.001)	(0.001)	
Female servants		0.041 ***	0.041 ***	
		(0.001)	(0.001)	
Constant	2.513 ***	2.857 ***	2.875 ***	3.922 ***
	(0.020)	(0.020)	(0.025)	(0.015)
Fixed effects	-	-	Parish	Brothers
R^2	0.172	0.201	0.192	0.039
N	*109411*	*109411*	*109411*	*36556*

Sources: see Table 1.

Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

to that of theirs sons. Social status increases with age, but at a declining rate. Sons with a family background in farming did somewhat worse than the rest of the cohort. The presence of a mother is positive, as is the number of female servants employed in the household.

With regards to the effect of migration on social status, the results in columns (I) and (II) closely mirrors that observed in Figure 5A: Social status increases both with distance moved and when the destination changes from rural to urban. In column (III) the results show that accounting for differences in parish of origin characteristics through fixed effects changes the estimates of returns to migration little. The realised returns across all types of migration are very similar in magnitudes to those estimated in column (II) which does not take geographic origin into account. This result indicates that the bias stemming from unobserved location-specific characteristics is limited. Estimates from the most extensive model based on differences in migration and realized social status between brothers are presented in column (IV). Because identification in this model requires each individual to have a brother included in the analytical sample that also differ in terms of migration status, the sample size is reduced by about two-thirds. Nevertheless, the remaining brothers still constitute a sample which is large enough to make the exercise feasible. With one exception, rural short range migration, the inclusion of sibling fixed effects reduces the realised returns. The reduction in returns is larger for migration across longer distances than for shorter, and for urban migration than for rural. This pattern makes sense by suggesting positive selection on unobserved household resources which is similar to the relationship between father's social status and subsequent migration distance and destination.

Table 4 presents results from estimations identical to those in Table 3 but for women holding an occupation in 1900. For rural migrants the returns closely mirror those estimated for men in column (I) to (III) and the mean Hiscam-scores displayed in Figure 5A. When it comes to the returns realised by urban migrants, there is an obvious difference between men and women. In contrast to men, urban female migrants' returns are lower than those realised by rural migrants for comparable distances. While there is an urban premium for men, there seems to be an urban penalty for women. In column (IV) the sister fixed effects estimates are presented. Because these estimates are based on a considerably smaller sample (4197 observations) than the brother fixed effects estimates (36556 observations), the results should be interpreted with care. Taken at face value, the sister fixed effects estimates

Table 4. Ordinary least squares estimations of individual Hiscam-score for women

	I	II	III	IV
Dependent variable: ln(own Hiscam-score)				
Migration (Ref: 0 km)				
Rural 0-25 km	0.019 ***	0.020 ***	0.014 ***	0.013
	(0.003)	(0.003)	(0.003)	(0.010)
Rural 25-75 km	0.058 ***	0.058 ***	0.055 ***	0.023 *
	(0.004)	(0.004)	(0.004)	(0.014)
Rural >75 km	0.107 ***	0.101 ***	0.103 ***	0.041 **
	(0.005)	(0.004)	(0.005)	(0.016)
Urban 0-25 km	0.030 ***	0.028 ***	0.005	0.010
	(0.004)	(0.003)	(0.005)	(0.014)
Urban 25-75 km	0.032 ***	0.031 ***	0.025 ***	-0.007
	(0.003)	(0.003)	(0.004)	(0.014)
Urban >75 km	0.048 ***	0.043 ***	0.043 ***	0.009
	(0.003)	(0.003)	(0.004)	(0.013)
Father's Hiscam-score	0.359 ***	0.270 ***	0.266 ***	
	(0.009)	(0.009)	(0.011)	
Age	0.006 ***	0.009 ***	0.008 ***	-0.003
	(0.001)	(0.001)	(0.001)	(0.005)
Age^2	0.000	0.000 *	0.000 *	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Father age	0.001 ***	0.000	0.000	
	(0.000)	(0.000)	(0.000)	
Father age^2	0.000 ***	0.000 *	0.000 **	
	(0.000)	(0.000)	(0.000)	
Farm household		-0.020 ***	-0.015 ***	
		(0.002)	(0.002)	
Mother present		0.010 **	0.011 **	
		(0.005)	(0.005)	
Siblings		-0.003 ***	-0.002 **	
		(0.001)	(0.001)	
Birth order		0.009 ***	0.010 ***	-0.015 **
		(0.001)	(0.001)	(0.007)
Male servants		0.009 ***	0.009 ***	
		(0.003)	(0.003)	
Female servants		0.057 ***	0.054 ***	
		(0.003)	(0.003)	
Constant	2.314 ***	2.651 ***	2.667 ***	3.940 ***
	(0.036)	(0.037)	(0.042)	(0.053)
Fixed effects	-	-	Parish	Sisters
R^2	0.123	0.158	0.154	0.009
N	26874	26874	26874	4197

Sources: see Table 1.

Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

suggest that urban female migrants did not realise any returns in the form of upward social mobility. The returns from medium and long distance rural migration remain, but are reduced by more than half.

The results presented above clearly shows that gains in the form of upward social mobility was not restricted to urban migrants. Although male urban migrants earned the highest premiums, important gains were also realized by rural migrants. This is an important result in light of the high incidence of rural migration in the period.

The observed differences between men and women beg two questions. First, why did women not benefit from urban migration to the same extent as men? The explanation for women's lacklustre returns from urban migration can be found in differences in the occupational structure of urban and rural labour markets. Figure 6 plots the urban and rural cumulative distribution of Hiscam-scores for all men and women in Sweden in 1900. In the rural labour market there were distinct differences between the status of occupation held by men and women. At the lower end of the distribution this difference is primarily driven by the fact that most women employed in rural areas worked as servants, while most men were farmers. However, at the top of the distribution, the difference between men and women is much smaller. Only about 5 per cent of men and women held occupations with a Hiscam-score higher than 65. The chances of moving into a high status job in the countryside were thus equally limited for both men and women.

The urban distribution of occupational status tells a different story. Overall, jobs in urban areas were clearly more prestigious than in the countryside. However, the difference in the kind of jobs held by men and women were more pronounced in towns and cities than in the countryside. Among urban men, around 20 per cent held high ranking jobs with a Hiscam-score above 65. The equivalent number for women was just over 10 per cent. Similarly, less than 25 per cent of all men in urban areas held low-ranking jobs with a Hiscam-score of 50 or less. In the countryside the equivalent number for men was 65 per cent. For women, the difference between rural and urban areas in the number of women holding low-ranking jobs is much smaller, 80 per cent versus 65 per cent. To summarize these numbers suggests that the urban labour market was particularly beneficial to men's careers by providing few low status jobs and more opportunities for upward mobility. For women, the relative differences between opportunities in the rural and urban labour markets were much smaller.

Figure 6. Cumulative distribution of Hiscam-scores of the Swedish population in 1900, rural and urban men and women

Sources: See Figure 2.

The second question is why, given their low or non-existing prospects of upward social mobility, did women still flock to urban areas rather than remain in the countryside? An obvious answer is that estimating returns by considering occupational mobility only captures a small part of the benefits realised by female migrants. Regional and rural/urban wage differences is one important omitted factor which may have driven the migration of women. Given the data at hand I am unable to deal with this issue directly. Other potentially important causes which may have incentivized women' may however be addressed. Part of the answer is provided by considering the social status of female migrants' husbands in place of their individual occupational status as an outcome. I proceed by re-estimating all models for women using the Hiscam-score of husbands as the dependent variable. The results are presented in Table 5. On the whole, the estimated returns are very similar to those presented in Table 3 for the male samples. This finding

Table 5. Ordinary least squares estimations of spouse's Hiscam-score for women

	I	II	III	IV
Dependent variable: ln(spouse Hiscam-score)				
Migration (Ref: 0 km)				
Rural 0-25 km	0.010 ***	0.010 ***	0.008 ***	0.009 ***
	(0.001)	(0.001)	(0.001)	(0.003)
Rural 25-75 km	0.034 ***	0.033 ***	0.036 ***	0.032 ***
	(0.002)	(0.002)	(0.002)	(0.004)
Rural >75 km	0.086 ***	0.080 ***	0.081 ***	0.068 ***
	(0.002)	(0.002)	(0.003)	(0.006)
Urban 0-25 km	0.057 ***	0.054 ***	0.052 ***	0.046 ***
	(0.003)	(0.003)	(0.003)	(0.007)
Urban 25-75 km	0.078 ***	0.074 ***	0.071 ***	0.053 ***
	(0.003)	(0.003)	(0.004)	(0.009)
Urban >75 km	0.142 ***	0.130 ***	0.126 ***	0.087 ***
	(0.004)	(0.004)	(0.004)	(0.009)
Father's Hiscam-score	0.295 ***	0.217 ***	0.214 ***	
	(0.006)	(0.006)	(0.006)	
Age	0.001	0.002 ***	0.002 ***	0.002
	(0.001)	(0.001)	(0.001)	(0.002)
Age^2	0.000	0.000	0.000	0.000 *
	(0.000)	(0.000)	(0.000)	(0.000)
Father age	0.000 ***	0.000	0.000	
	(0.000)	(0.000)	(0.000)	
Father age^2	0.000 ***	0.000 ***	0.000 ***	
	(0.000)	(0.000)	(0.000)	
Farm household		-0.024 ***	-0.017 ***	
		(0.001)	(0.001)	
Mother present		0.006 ***	0.007 ***	
		(0.002)	(0.002)	
Siblings		-0.001 ***	-0.001 *	
		(0.000)	(0.000)	
Birth order		0.003 ***	0.003 ***	-0.003
		(0.000)	(0.000)	(0.002)
Male servants		0.000	0.001	
		(0.001)	(0.001)	
Female servants		0.042 ***	0.041 ***	
		(0.001)	(0.002)	
Constant	2.754 ***	3.051 ***	3.058 ***	3.937 ***
	(0.024)	(0.023)	(0.026)	(0.017)
Fixed effects	-	-	Parish	Sisters
R^2	0.129	0.158	0.148	0.029
N	92609	92609	92609	24124

Sources: see Table 1.

Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

suggests that the migration of married women may have been driven by husband's chances of upward social mobility, while unmarried women may have been motivated by better prospects in the marriage market.

A shortcoming of the above analysis is that it requires an individual to be employed to be included in any analytical sample, or having a husband that is so. The results so far thus neglect potential differences between migrants and non-migrants and men and women in terms of being steadily employed or subject to poverty. I proceed by extending the analysis by considering the impact of migration on the probability of employment or poverty respectively using the following model:

$$Prob(Y = 1) = F(\beta \boldsymbol{M} + \beta \ln(Hiscam_{father}) + \beta \boldsymbol{X} + u) \quad (2)$$

where Y, depending on the outcome considered denotes either being gainfully employed or poor. The rest of the notation follows from (1). In order to not confuse the issue of employment with marriage and subsequent labour market exits, the female sample used for estimating the probability of employment is restricted to unmarried women only. The results are presented in Table 6. Across all distances and destinations there is a significant increase in employment, with migrants being at least twice as likely to be gainfully employed as non-migrants. The estimates of migration on being poor support the results for employment. Across all types of migration, poverty decreases substantially. For women, the strong effect of migration on employment nuances the previous results and serves to further explain why women's mobility was as high as that of men.

Table 6. Alternative outcomes: Odds rations from logit estimations of employment status and poverty

	Dependent variable: Employed		Dependent variable: Poor	
	Men	Women	Men	Women
Migration (Ref: 0 km)				
Rural 0-25 km	2.151 ***	3.556 ***	0.487 ***	0.509 ***
	(0.086)	(0.164)	(0.049)	(0.032)
Rural 25-75 km	2.186 ***	4.135 ***	0.344 ***	0.427 ***
	(0.123)	(0.261)	(0.057)	(0.041)
Rural >75 km	2.901 ***	4.481 ***	0.341 ***	0.369 ***
	(0.208)	(0.319)	(0.064)	(0.047)
Urban 0-25 km	3.129 ***	3.722 ***	0.053 ***	0.069 ***
	(0.282)	(0.231)	(0.031)	(0.022)
Urban 25-75 km	2.947 ***	4.433 ***	0.018 ***	0.111 ***
	(0.261)	(0.273)	(0.018)	(0.029)
Urban >75 km	2.473 ***	5.140 ***	0.011 ***	0.029 ***
	(0.160)	(0.262)	(0.011)	(0.013)
Father's Hiscam-score	0.224 ***	0.584 ***	0.384 **	0.295 ***
	(0.030)	(0.073)	(0.159)	(0.077)
Age	0.982	0.896 ***	0.844 ***	0.921 **
	(0.019)	(0.018)	(0.041)	(0.032)
Age^2	1.001 *	1.002 **	1.007 ***	1.003 ***
	(0.001)	(0.001)	(0.002)	(0.001)
Father age	0.967 **	1.006 **	1.012 *	1.023 ***
	(0.014)	(0.003)	(0.007)	(0.004)
Father age^2	1.000 ***	1.000 **	1.000	1.000 **
	(0.000)	(0.000)	(0.000)	(0.000)
Farm household	0.601 ***	1.113 ***	1.174 *	0.811 ***
	(0.021)	(0.036)	(0.096)	(0.044)
Mother present	1.063	0.883 *	0.696 **	0.636 ***
	(0.067)	(0.057)	(0.107)	(0.065)
Siblings	1.158 ***	1.130 ***	1.110 ***	1.098 ***
	(0.012)	(0.013)	(0.026)	(0.018)
Birth order	0.906 ***	0.824 ***	1.000	0.916 ***
	(0.014)	(0.013)	(0.037)	(0.024)
Male servants	1.000	0.994	0.688 ***	0.761 ***
	(0.028)	(0.031)	(0.084)	(0.057)
Female servants	0.867 ***	0.743 ***	0.829 **	0.694 ***
	(0.020)	(0.021)	(0.076)	(0.057)
χ^2	1749.68	3407.98	487.612	851.671
N	117193	33281	117193	132314

Sources: see Table 1.

Notes: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Conclusion

In the closing decades of the nineteenth century Sweden was going through a mobility transition. Migrants increasingly began seeking their fortune in more distant urban areas. The net result was an increase in overall mobility and urbanisation. The majority of the men and women that left home during this period did however not move to a town or city. Instead they continued to conform to the well-established short range and rural migration pattern of *l'ancien régime*.

The results show that migration played an instrumental role in determining individual social mobility in Sweden at the end of the nineteenth century. This process can however not be said to have worked to equalize opportunities. Men seem to have gained more from moving to a town or city than women did as a result of segmentation of the urban labour market. In terms of selection, the strongest tendency is towards positive selection, both when considering distance migrated and destination choice. Urban long distance migration was, despite falling costs of transportation, to a large extent still an elite phenomenon in late nineteenth century Sweden.

By approaching migration in a comprehensive manner, the results show that rural migrants also realized important gains from moving. This finding provides an explanation for why rural migration, despite rapid industrialization and urbanization, remained the dominating type of migration in Sweden at the end of the nineteenth century.

Moreover this paper adds to our understanding of the incentives that drove migration historically. By extending the prospective benefits from migration by considering employment and poverty, the results show that as an outcome of moving, migrants realised important gains in the form of having a stronger attachment to the labour market, and avoiding poverty. This aspect of migration has been largely neglected in recent work which has almost exclusively focused on the employed.

References

Abramitzky, R., Platt Boustan, L. Eriksson, K. (2012). and Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, 102: 1832–1856.

Abramitzky, R., Platt Boustan, L. Eriksson, K. (2013). Have the poor always been less likely to migrate? Evidence from inheritance practices during the age of mass migration. *Journal of Development Economic*, 102: 2-14.

Baines, D. (1985). *Migration in a Mature Economy: Emigration and Internal Migration in England and Wales 1861-1900*. Cambridge: University Press.

Becker, G.S. and Lewis, H.G. (1973). On the Interaction between the quantity and quality of children, *The Journal of Political Economy*, 81: 279-288.

Berger, T. and Enflo, K. (2013). Locomotives of local growth: the short- and long-term impact of railroads in Sweden, *EHES Working Papers in Economic History*, 42.

BISOS U.

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

Brändström, A., Sundin, J. and Tedebrand, L-G. (2012). Two Cities: Urban Migration and Settlement in Nineteenth-century Sweden. *The History of the Family*, 5: 415-29.

Carter, M.J. and Carter, S.B. (1985). Internal labor markets in retailing: The early years, *Industrial and Labor Relations Review*, 38: 586-598.

Chiswick, B. (1999). Are immigrants favorably self-selected? *The American Economic Review*, 89: 181-185.

Clark, P. (1979). Migration in England during the Late Seventeenth and Early Eighteenth Centuries. *Past & Present*, 83: 57-90.

Collins, W.J. and Wanamaker, M.H. (2014). Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data. *American Economic Journal*, 6: 220-52.

Devine, T.M. (1979). Temporary Migration and the Scottish Highlands in the Nineteenth Century. *The Economic History Review*. 32: 344-59.

Dribe, M. (2000). *Leaving home in a peasant society. Economic fluctuations, household dynamics and youth migration in southern Sweden, 1829-1866.* Södertälje: Almqvist & Wiksell International.

Dribe, M. (2003). Migration of rural families in 19th century southern Sweden. A longitudinal analysis of local migration patterns. *The History of the Family,* 8: 247-65.

Dribe, M. and Lundh, C. (2005). People on the move: determinants of servant Migration in nineteenth-century Sweden. *Continuity and Change*, 20: 53-91.

Dribe, M. and Svensson, P. (2008). Social Mobility in Nineteenth Century Rural Sweden – A Micro Level Analysis. *Scandinavian Economic History Review*, 56: 122-141.

Erikson, R. and Goldthorpe, J.H. (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies.* Oxford: Clarendon Press.

Eriksson, I. and Rogers, J. (1978). *Rural Labor and Population Change*. Stockholm: Almqvist & Wiksell.

Featherman D.L. and Hauser R.M. (1978). *Opportunity and Change*. New York: Academic Press.

Gaunt, D. (1977). Pre-industrial Economy and Population Structure. *Scandinavian Journal of History*, 2: 183-210.

Geschwind, A. and Fogelvik, S. (2000). The Stockholm Historical Database. In P.K. Hall, R. McCaa, G. Thorvaldsen (eds.), *Handbook of International Historical Microdata for Population Research,* pp. 207 – 231. Minneapolis: Minnesota Population Center.

Goldin, C. (1990). *Understanding the Gender Gap: An Economic History of American Women*. New York: Oxford University Press.

Goldin, C. and Katz, L. (1998). The origins of technology-skill complementarity. *Quarterly Journal of Economics*, 113: 693-732.

Grant, O. (2005). *Migration and Inequality in Germany 1870-1913*. Oxford: Oxford university Press.

Guest, A.M., Landale, N.S. and McCann, J.C. (1989). Intergenerational Occupational Mobility in the Late 19th Century United States. *Social Forces*, 68: 351-78.

Harris, J.R. and Todaro, M.P. (1970). Migration, Unemployment: A Two-sector Analysis. The American Economic review, 1: 126-142.

Hochstadt, S. (1983). Migration in Preindustrial Germany. *Central European History*, 16: 195-224.

Howlett, P. (2004). The Internal Labour Dynamics of the Great Eastern Railway Company. 1870-1913. *Economic History Review*, 57: 396–422.

Humphries, J. and Leunig, T. (2009). Was Dick Whittington taller than those he left behind? Anthropometric measures, migration and the quality of life in early nineteenth century London. *Explorations in Economic History*, 46: 120-131.

Jackson, J.H. and Moch, L.P. (1989). Migration and the Social History of Modern Europe. *Historical Methods*, 22: 27-36.

Jörberg, L. (1972). *A History of Prices in Sweden 1732-1914.* Vol. 2, Lund: CWK Gleerup.

Lambert, P. S., Zijdeman, R. L., Van Leeuwen, M. H. D., Maas, I., & Prandy, K. (2013). The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Comparative Research. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 46: 77–89.

Lee, E.S. (1966) A Theory of Migration, *Demography*, 3: 47-57.

Long, J. (2005). Rural-Urban Migration and Socioeconomic Mobility in Victorian Britain. *The Journal of Economic History*, 65: 1-35.

Long, J. and Ferrie, J. (2007). The path to convergence: Intergenerational occupational mobility in Britain and the US in three eras. *The Economic Journal*, 117: C61-71.

Lundh, C. (1999). Servant Migration in Sweden in the Early Nineteenth Century. *Journal of Family History,* 24: 53-74.

Lundh, C. (2003). Life Cycle Servants in Nineteenth Century Sweden – Norms and Practice. *Lund Papers in Economic History*, No. 84.

Maas, I. and van Leeuwen, M.H.D. (2004). Occupational careers of the total male labour force during industrialization: the example of nineteenth-century Sweden: 227-58, in Mitch, D., Brown, J. and van Leeuwen, M.H.D. eds. *Origins of the Modern Career*. Aldershot: Ashgate.

Moch, L.P. (1992). *Moving Europeans: Migration in Western Europe since 1650*. Bloomington: Indiana University Press.

Moky, J. (2009). *The enlightened economy: An economic history of Britain, 1700-1870*. New Haven: Yale University Press.

Parish, W.L. (1973). Internal Migration and Modernization: The European Case. *Economic Development and Cultural Change*, 21: 591-609.

Patten, J. (1976). Patterns of migration and movement of labour of three preindustrial East Anglian towns. *Journal of Historical Geography*. 2: 111-29.

Pryor, R.J. (1975). Migration and the process of modernization. pp. 23-38, in Kosinski, L.A. & Prothero, R.M. *People on the Move*. London: Methuen & Co.

Ravenstein, E. G. (1885). The laws of migration. *Journal of the Royal Statistical Society,* 52: 167-235.

Ravenstein, E.G. (1889). The Laws of Migration: Second Paper. *Journal of the Royal Statistical Society*, 52: 241-305.

Ravenstein, E.G. (1889). The Laws of Migration: Second Paper. *Journal of the Royal Statistical Society*, 52: 241-305.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3: 135-146.

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

Schön, L. (2010). *Sweden's Road to Modernity: An Economic History*, Stockholm: SNS.

Seltzer, A.J. and Frank, J. (2007). Promotion tournaments and white collar careers: evidence from William Deacon's Bank, 1890-1941. *Oxford Economic Papers* 59:49-72.

Sjastaad, L.A. (1972). The Costs and Returns of Human Migration. *Journal of Political Economy*, 70: 80-93.

Statistiska Centralbyrån (1900). *Bidrag till Sveriges officiella statistik. U: Kommunerns fattigvård och finanser,* Stockholm: P.A. Nordstedt & Söner.

Statistiska Centralbyrån (1969). *Historisk statistik för Sverige*, 1, Stockholm.

Stewart, J.I. (2006). Migration to the Agricultural Frontier and Economic Mobility, 1860-1880, *Explorations in Economic History*, 43: 547-577.

Stewart, J.I. (2012). Migration to U.S. frontier cities and job opportunity, 1860-1880. *Explorations in Economic History*, 49: 528-42.

The Institute for Social Sciences, Stockholm University, (1941). *Population Movements and Industrialization*. London: P.S. King & Son.

Thomas, D.S. (1941). *Social and Economic Aspects of the Swedish Population Movement, 1750-1933*. New York: The Macmillan Company.

Treiman, D.J. (1970). Industrialization and Social Stratification. *Sociological Inquiry*, 40: 207-34.

van der Woude, A., Hayami, A. and Vries, J. (1990). *Urbanization in History*. Oxford: Clarendon Press.

van Leeuwen, M. H. D. and Maas, I. (2010). Historical studies of social mobility and stratification. *Annual Review of Sociology* 36: 429-451.

Wegge, S. (2002). Occupational self-selection of European emigrants: Evidence from nineteenth-century Hesse-Cassel. *European Review of Economic History*, 6: 365-394.

Zelinsky, W. (1971). The Hypothesis of the Mobility Transition. *Geographical Review*, 61: 219-49.

Zijdeman, R.L. (2009). Like my father before me: intergenerational occupational status transfer during industrialization (Zeeland 1811-1915). *Continuity and Change*, 24: 455-86.

Appendix

Table A1. Estimates of father's social status on migration distance and destination

Estimation method:	OLS		Logit	
Dependent variable:	Distance moved (km)		Urban destination (=1)	
	Men	Women	Men	Women
Father's Hiscam-score	2.029 ***	1.660 ***	1.028 ***	1.016 ***
	(0.073)	(0.065)	(0.001)	(0.001)
Constant	-57.597 ***	-44.663 ***		
	(3.702)	(3.279)		
Fixed effects	Parish	Parish	Parish	Parish
R^2	0.019	0.017		
χ^2			603.423	196.576
N	109411	119084	58566	70875

Sources: see Table 1.
Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

LUND STUDIES IN ECONOMIC HISTORY

72 Eriksson, Björn
Dynamic Decades: A micro perspective on late nineteenth century Sweden, 2015

71 Persson, Mats R
In good times and in bad: Immigrants, Self-employment and Social Insurance, 2015.

70 Borowiecki, Karol Jan
"Mournful and yet grand is the destiny of the artist." Some Economic History of Music Composers and their Creativity, 2015.

69 López Jerez, Montserrat
Deltas Apart. Factor Endowments, Colonial Extraction and Pathways of Agricultural Development in Vietnam, 2014.

68 Xiao, Jing
Entrepreneurial Dynamics and Acquisitions of New Technology-Based Firms, 2014.

67 Taalbi, Josef
Innovation as Creative Response. Determinants of Innovation in the Swedish Manufacturing Industry, 1970-2007, 2014.

66 Bergenfeldt, Fredrik
Roads to Market. Transport and Agricultural Change – the case of southern Sweden, 1750-1850, 2014.

65 Sjöö, Karolin
Innovation and Transformation in the Swedish Manufacturing Sector, 1970-2007, 2014.

64 Lind, Daniel
Value Creation and Structural Change During the Third Industrial Revolution. The Swedish Economy from a Vertical Perspective, 2014.

63 Silva, Ernesto
Essays on Income and Fertility. Sweden 1968-2009, 2014.

62 Dekhtyar, Serhiy
Early Life Origins of Health and Well-Being in Modern Sweden, 2013.

61 Holmberg, Kajsa
Behind the Counter. Female Inroads, Obstacles and Careers in the Swedish Commercial Bank Sector 1885-1937, 2013.

60 Palacio, Andrés
Mind the Gap. Inequalities in Infant and Child Mortality: the Case of Colombia 1967-2010, 2013.

59 Quaranta, Luciana
Scarred for life. How conditions in early life affect socioeconomic status, reproduction and mortality in southern Sweden, 1813-1968, 2013.

58 Hannemann, Tina
It Breaks a Man's Heart. Socioeconomic Differences in the Onset of Cardiovascular Disease in Contemporary Sweden, 2012.

57 Faran, Turaj S.
The Industrialization Process in the Chinese Mirror: Continuity and change in transition from collective agriculture to market economy in China, 2011.

56 Weisdorf, Jacob L.
Malthusian Progress, 2011.

55 Håkansson, Peter
Ungdomsarbetslösheten – om övergångsregimer, institutionell förändring och socialt kapital (Youth unemployment – transition regimes, institutional change and social capital), 2011.

54 Henriques, Sofia Teives
Energy transitions, economic growth and structural change. Portugal in a long-run comparative perspective, 2011.

53 Helgertz, Jonas
Immigrant Careers – Why Country of Origin Matters, 2010

52 Bohman, Magnus
Bonden, bygden och bördigheten. Produktionsmönster och utvecklingsvägar under jordbruksomvandlingen i Skåne ca 1700-1870 (Peasants, production patterns and paths of development during the agricultural transformation in Southern Sweden (Scania) ca 1700-1870), 2010.

51 Waldenström, Daniel
Lifting All Boats? The Evolution of Income and Wealth Inequality over the Path of Development, 2009.

50 Roubert, Carl-Johan
Småländsk världsindustri - en studie om entreprenörer och kluster 1950-2002, 2009.

49 Örnberg, Lena
Bondens val - System, samspel och stödjande strukturer under moderniseringsprocessen i norra Thailand (Farmer's Choice - Systems, interaction and supporting structures during the modernisation process in Northern Thailand), 2008.

48 Rönnqvist, Sofia
Från diversity management till mångfaldsplaner? Om mångfaldsidéns spridning i Sverige och Malmö (From Diversity Management to Diversity Plans? On the Spread of the Diversity Idea in Sweden and Malmö), 2008.

47 Karlsson, Tobias
Downsizing: Personnel Reductions at the Swedish Tobacco Monopoly, 1915-1939, 2008.

46 Enflo, Kerstin
Regional and Industrial Growth Patterns in 20th Century Western Europe, 2008.

45 Axelsson, Tobias
Peasants and Policymakers, Agricultural transformation in Java under Suharto, 2008.

44 Gustafsson, Fredrik
The Visible Palm: Market Failures, Industrial Policy and the Malaysian Palm Oil Industry, 2007.

43 Forsberg, Le Thanh
Defining Strong Ownership: Institutional Determinants and Stakeholder Interests in Vietnamese Development Planning, 2007.

42 Andera, Jan
Driving under the Influence – Strategic Trade Policy and Market Integration in the European Car Industry, 2007.

41 Krantz, Olle & Schön, Lennart
Swedish Historical National Accounts 1800—2000, 2007.

40 Lundh Nilsson, Fay
Lönande lärande. Teknologisk förändring, yrkesskicklighet och lön i svensk verkstadsindustri omkring 1900 (Profitable Learning – Technological Change, Skills and Wages in the Swedish Engineering Industry around 1900), 2007.

39 Abou-Zeinab, Ali
Commercial Banking and Economic Development, The Lebanese Experience 1856-1974, 2006.

38 Jönsson, Per Ingvar
Chaebolföretagen - en"omöjlig" konfiguration i den sydkoreanska industrialiseringsprocessen (The Chaebol Conglomerate - An "Impossible" Configuration in the South Korean Industrialization Process), 2006.

37 Josephson, Camilla
Growth and Business Cycles -Swedish Manufacturing Industry 1952-2001, 2005.

36 Appelquist, Joakim
Informationsteknik och organisatorisk förändring - Teknik, organization och produktivitet i svensk banksektor 1975-2003 (Information Technology and Organizational Change – Technology, Organization and Productivity in Swedish Banking Sector 1975-2003), 2005.

35 Green, Erik
Peasant production and limits to labour: Thyolo and Mzimba Districts in Malawi, mid-1930s to late-1970s, 2005.

34 Saarinen, Jani
Innovations and Industrial Performance in Finland 1945 – 1998, 2005.

33 Lindgren, Mattias
Does schooling make sense? A household perspective on the returns to schooling for self-employed, farmers and employees in Egypt, 2005.

32 Lundh, Christer, Jonas Olofsson, Lennart Schön & Lars Svensson (eds.)
Wage Formation, Labour Market Institutions and Economic Transformation in Sweden 1860-2000, 2004.

31 Rytkönen, Paulina
Fruits of Capitalism, Modernisation of Chilean Agriculture 1950-2000,

2004.

30 Johansson, Kent
Child Mortality during the Demographic Transition. A Longitudinal Analysis of a Rural Population in Southern Sweden, 1766-1894, 2004.

29 Holmquist, Svante
Kapitalbildning i svensk industri 1870-1930 (Capital Formation in Swedish Industry 1870-1930), 2004.

28 Carlsson, Ellen
To Have and to Hold: Continuity and change in property rights governing water resources among the Meru of Tanzania and the BaKgatla of Botswana; 1925-2000, 2003.

27 Abdulsomad, Kamaruding
Building technological capabilities of local auto parts firms under contrasting industrial policies: a comparative study of Malaysia and Thailand 1960-2000, 2003.

26 Bengtsson, Henrik
Ålder och ekonomisk omvandling. Sveriges tillväxthistoria i ett demografiskt perspektiv, 1890-1995 (Age and Economic Transformation.Sweden's History of Economic Growth in a Demographic Perspective, 1890-1995), 2003.

25 Andersson, Martin
Bending the Kuznets Curve - Wrenching and Levelling Forces during Agricultural Transformation in Taiwan and Chile, 2003.

24 Sandal, Jan-U.
Jakten på Entreprenøren. Kan Joseph A. Schumpeters teori benyttes til å identifisere og differensiere entreprenører i det 21.århundrede? (In Search of the Entrepreneur. Can Joseph A. Schumpeter's theory be used to identify and differentiate between entrepreneurs in the 21st century), 2003.

23 Öberg, Klas
Pharmacy Regulation in Sweden. A New Institutional Economic Perspective, 2003.

22 Stanfors, Maria

Education, Labor Force Participation and Changing Fertility Patterns. A Study of Women and Socioeconomic change in Twentieth Century Sweden, 2003.

21 Klinthäll, Martin
Return Migration from Sweden 1968-1996, 2003.

20 Olsson, Mats
Storgodsdrift. Godsekonomi och arbetsorganisation i Skåne från dansk tid till mitten av 1800-talet (Manorial Economy and Corvée Labour in Early Modern Scania), 2002.

19 Kander, Astrid
Economic growth, energy consumption and CO2 emissions in Sweden 1800-2000, 2002.

18 Rauhut, Daniel
Fattigvård, socialbidrag och synen på fattigdom i Sverige 1918-1997 (Poor Relief, Social Assistance and the View on Poverty in Sweden 1918-1997), 2002.

17 Schånberg, Ingela
Genus och utbildning. Ekonomisk-historiska studier i kvinnors utbildning ca 1870-1970 (Gender and Education.Economic History Studies in Women´s Education ca. 1870 to 1970), 2002.

16 Svensson, Patrick
Agrara entreprenörer. Böndernas roll i omvandlingen av jordbruket i Skåne ca 1800-1870 (Agrarian Entrepreneurs.The Role of the Peasants in the Transformation of Agriculture in Scania, 1800-1870), 2001.

15 Bevelander, Pieter
Immigrant Employment Integration and Structural Change in Sweden, 1970 – 1995, 2000.

14 Lobell, Håkan
Växelkurspolitik och marknadsintegration. De utländska växelkurserna i Sverige 1834–1880 (Exchange Rate Policy and Market Integration.Foreign Exchange Rates in Sweden 1834-1880), 2000.

13 Dribe, Martin
Leaving Home in a Peasant Society. Economic Fluctuations, Household Dynamics and Youth Migration in Southern Sweden, 1829-1866, 2000.

12 Thorburn, Thomas
Economic of Transport: the Swedish Case 1780-1980, 2000.

11 Ankarloo, Daniel
"Institutions", What Is In a Word? A Critique of the New Institutional Economics, 1999.

10 Iglesias, Edgar
Welfare, Security and the Level of Fertility: A Global Analysis, 1985 and 1995, 1999.

9 Scott, Kirk
The Immigrant Experience: Changing Employment and Income Patterns in Sweden, 1970 – 1993, 1999.

8 Siriprachai, Somboon
Control and Rent-Seeking, The Role of the State in the Thai Cassava Industry, 1998.

7 Pei, Xiaolin
The Institutional Root of China's Rural Industry and Gradual Reform, 1998.

6 Olofsson, Jonas
Arbetslöshetsfrågan i historisk belysning. En diskussion om arbetslöshet och social politik i Sverige 1830-1920 (The Unemployment Question in Historical Perspective. A Discussion on Unemployment and Social Policy in Sweden, 1830-1920), 1996.

5 Ahlström, Göran
Technological Development and Industrial Exhibitions 1850-1914. Sweden in an International Perspective, 1995.

4 Arlebäck, Sven Olof
Från ägarmakt under ansvar till företagsledarmakt upa. En studie över maktförskjutningen i tre svenska verkstadsaktiebolag (The Transition of Power from Shareholders to Top-Management. A Study of Three Swedish Engineering Firms), 1995.

3 Staffansson, Jan-Åke
Svenskt smör. Produktion, konsumtion och utrikeshandel 1861-1913 (Swedish Butter. Productivity, Consumption, and Foreign Trade, 1861-1913), 1995.

2 Bergström, Asta

Åtstramning och Expansion. Den ekonomiska politiken i Sverige 1971-1982 (Retrenchment and Expansion. Swedish Economic Policy 1971-1982), 1995.

1 Lundh, Christer (ed.)
Population, Economy and Welfare. Scandinavian Population Studies, vol 10, 1995.

0 Ahlström, Göran
Technological Development and Industrial Exhibitions 1850-1914. Sweden in an International Perspective, 1995.