Micro-level aspects of Swedish development

ANDREW SELTZER[*]   *Royal Holloway, University of London*

Björn Eriksson, *Dynamic Decades: A micro perspective on late nineteenth century Sweden*, Lund studies in economic history 72 (Lund: Ekonomihögskolan, Lunds universitet 2015).

Introduction

Björn Eriksson's dissertation examines various micro-level aspects of Swedish economic development in the late-nineteenth and early-twentieth centuries. The period was characterised by rapid growth, urbanisation, and industrialisation. Unlike today, during this time period Sweden had a *laissez faire* economy with little government regulation of the business environment. Although the overall level of government economic intervention was low, one area in which the Swedish government was particularly activist was in the collection of data. The Swedish Census and industry surveys are considerably more detailed and higher quality than those collected in the USA, UK, or other Western European countries at the same time. This dissertation uses micro-level data from these sources to examine various micro-level aspects of Swedish industry, labour markets, and migration patterns in a time of rapid industrialisation. The nature of the data allows a more nuanced and complete analysis than can be done for other countries. The analytical approach of this dissertation is to use the excellent Swedish data along with modern micro-economic and econometric methodologies to examine several aspects of Swedish development. The contributions of this dissertation are both methodological and empirical, and I summarise the specific contributions of the individual chapters below.

[*] Professor of Economics and Economic History; fakultetsopponent

Results

The first substantive chapter of the dissertation, "A Winning Strategy? The employment of women and firm longevity during industrialization", examines the relationship between female employment and firm survival in the tobacco industry in the late-19th century. The tobacco industry was competitive and largely unregulated during this period. The work was labour intensive, much of it unskilled. The main hypothesis tested in this chapter is whether, all else equal, firms that hired more women were more likely to survive. The thesis outlines a strong theoretical reason to believe that this would be the case. Gary Becker's model of employer discrimination implies that if labour markets are competitive, discrimination will ultimately hurt discriminatory employers by increasing their costs relative to non-discriminatory employers. In other words, if employers' "tastes" for discrimination are unrelated to productivity, they will lower their profits by hiring men. If the product market was competitive, as was almost certainly the case for a fairly homogeneous product such as tobacco, firms that voluntarily accepted the additional costs associated with not employing women would not have been profitable and thus more likely to fail.

To test this hypothesis, Eriksson uses data collected from annual ledgers compiled by the Swedish Board of Commerce. As with all data used in this thesis, the Board of Commerce data is remarkably comprehensive, comprising *all* firms in the industry over a 20+ year period. The data contain firm-level information on the size of the workforce, the machinery used, the amount of different tobacco outputs, and geographic location. The chapter estimates survival probabilities of these firms using Cox regression models. These probabilities are based on firms' observed characteristics, in particular on the employment of women. The main finding of the regressions in this chapter is that more feminized firms were less likely to fail. This finding is statistically significant, quantitatively large, and fairly robust to alternative specifications.

The chapter then offers a theoretical explanation for this finding based on Becker's model of discrimination. While the Becker "taste" model is plausible and intuitively compelling, it is not the only possible explanation for the observed relationship between female employment and firm failure. The Cox regressions reported in this chapter show a relationship between female share and firm failure. By their nature, they do not show causation. Causation is assumed with the Becker "taste for discrimination" model. A possible alternative explanation that is consistent with these findings is a model of managerial failure. Good managers are constantly looking for new ways of keeping costs down. Bad managers may simply stick with existing practice. Thus good managers implement many new practices, including employment

of women. In this model, the reason that some firms are less likely to fail than others is the quality of management. Individual practices that good managers bring in may only be of second order importance. However, the model predicts that failure probability will be correlated with feminization, as this is correlated with management quality. The reason for the estimated relationship between feminization rates and failure probability is omitted variable bias, i.e. managerial quality is not directly observed or controlled for in a regression, but is correlated with feminization. One (weak) piece of evidence that managerial failure is at least part of the explanation for the statistical significance of the coefficient on female share is that this coefficient declines when some managerial variables are included in the regression.

The second substantive chapter "Industrious migrants: The earnings of migrants in Swedish manufacturing around 1900" examines the earnings of internal migrants from other parts of Sweden relative to locals in three manufacturing industries: tobacco, printing, and mechanical engineering. For the tobacco and printing industries, this analysis is done separately for men and women. Mechanical engineering was exclusively male, and thus gender-based comparisons are not possible.

The data for this chapter are drawn from Swedish Board of Commerce ledgers. As with the other data used in this dissertation, these data are remarkably comprehensive containing information on both firm and worker characteristics. The data contains information on workers' earnings and on a range of their socio-economic characteristics such as age, place of birth, experience, father's social status, health, number of children, and migrant status. The analysis of migrant earnings requires a definition of who is considered to be an internal migrant. The definition used in this chapter is based on the distance between each worker's place of birth and their place of employment. The Board of Commerce data contains each employee's birth parish and the location of each firm. These locations are then mapped out on a 25-kilometre by 25-kilometre grid mapping the whole of Sweden. An individual whose birthplace and workplace are located on different grids is classified as a migrant. One comment that I have on this approach is that a 25x25 grid is an *ad hoc* mapping to determine immigrant status. This, in itself, is not a criticism as there is no theory to guide us as to who is local and who is an immigrant. However, it would be useful to know whether any of the main results of this chapter depend on the definition of an immigrant. A priori it is not clear that a 25x25 grid is better than 10x10, 50x50, or 100x100 grids, and I would like to see some robustness checks to assure the reader that the choice of definition is not driving the results described below.

The main analytical approach used by Eriksson in this chapter is Mincer-type earnings regressions. The log of hourly earnings of individual workers

is estimated as a function of their socio-economic characteristics and of firm and location fixed effects. Separate regressions are run for men and women where possible. The results consistently show that migrants earned more than local workers, with the premium ranging from about 5 percent to about 12 percent across gender and industry. The coefficients on immigrant are strongly significant for both men and women in all three industries. The chapter then turns to explaining why migrants earned more than locals after controlling for observable characteristics, examining two explanations that are prevalent in the migration literature: migrant selection and effort level. The selection hypothesis is drawn from the Roy model of migration. The Roy model states that if there is greater wage dispersion in the receiving area than in the sending area, migrants will tend to have above-average abilities or skill levels. The test for sorting involves estimating the returns to skills at the destination location, the returns to skills at the location of birth, and the interaction of these returns to migrant status. The chapter finds strong evidence for positive sorting; migrants' returns are highest in regions where the returns to skills are highest. The effort hypothesis says that migrants will tend to work harder than native-born workers. Using data on hours worked, the chapter finds some evidence for this effect for female tobacco workers, but not for other workers.

The third substantive chapter "A link to the past: Linking and evaluating Swedish historical censuses" is methodological, rather than historical, in nature. It develops a new approach to linking historical censuses and uses this approach to create a new data set from the linked Swedish censuses. Linkage of censuses has been an important problem for economic historians as the data collected in individual censuses is cross-sectional by nature, but combined data from successive censuses is longitudinal. Cross-sectional data from decennial census provides an array of useful information at a single point in time, but is not useful for examining economic changes over time. However, longitudinal data from linked censuses can be used to examine changes over time.

Economic historians have used linked census data to address questions ranging from the impact of childhood exposure to lead on subsequent health to, as in the fourth paper, the impact of migration on subsequent earnings. However, linking census faces several fundamental difficulties that any matching procedure needs to overcome. Information may be recorded incorrectly in a Census. For example, enumerators may not spell names correctly or individuals may not remember their exact age. In addition, individuals' names can change over time, an issue that has made it particularly difficult to match women across subsequent censuses. Because of these inherent difficulties, previous studies linking individuals across US and UK censu-

ses have been able to attain match rates only slightly over 20 percent. The huge sample sizes of these censuses may mean that these linkage rates are tolerable when linking any pair of adjacent censuses, however, the linkage rate is likely to decline exponentially as more census are introduced. For this reason, improvements in linking procedures result in larger and better longitudinal data sets.

Underlying any matching procedure are a few guiding principles: the procedure should minimize false positive matches, the procedure should have a high match rate, and the resulting sample should be close to a random draw from the population. The procedure developed in this paper begins with time-invariant information such as year of birth, sex, place of birth, and name. Because of the issue with names not exactly matching over time, individuals can only be matched probabilistically. The paper uses standard algorithms to assign probabilities that two individuals with differently spelled – but similar sounding – names are actually the same person. Further matching can be done on the basis of secondary links, matching other individuals in the same household. The paper then explores the optimal probability threshold to achieve matches, taking into consideration that a high threshold implies eliminating many likely matches and a low threshold means many duplicate matches. Finally, the paper examines properties of the resulting linked sample and concludes that it is fairly representative of the population.

The final substantive chapter "Onwards and upwards? Internal migration and social mobility in Sweden, 1880–1900", uses the linked census sample described in the previous chapter to estimate the returns to internal migration in Sweden. The data used in this chapter is the set of individuals who were children in the 1880 census who can be linked to their adult selves in the 1900 census. This sample comprises a large proportion of Swedish children in 1880, excluding only the fairly random set of individuals who could not be linked across censuses and the likely non-random set of individuals who emigrated from Sweden.

The censuses are a very rich data source containing information on age, sex, location, occupation, whether from a farming household, family wealth, number of siblings, and father's occupational status. One thing that the censuses do not contain is information on earnings, and thus the returns to migration cannot be estimated in the standard manner. In order to estimate the returns to migration, Eriksson uses occupational status as his dependent variable. All occupations in the 1900 census are classified using the Hiscam scale, a standardised approach to measuring social status. Unlike other measures, such as the NRS social grades, the Hiscam scale provides a continuous assessment of social status ranging from 37 (lowest status) to 99 (highest status).

The chapter then estimates a series of Mincer-type regressions on the determinants of Hiscam scores in 1900. Separate regression specifications are run for men, unmarried women, and married women. The independent variables include age, whether from a farming household, family wealth, number of siblings, father's occupational status, and migrant status measured by distance categories and whether the migration was to a rural or urban area. The main findings of these regressions are as follows. Migrants attained higher occupational status than non-migrants. The size of this effect increases with the distance migrated; there seems to be an approximately monotonic increase across the discrete distance categories used in this chapter (0, 0–25, 25–75,>75). The returns to migration to urban areas are larger than to rural areas for men, but not for unmarried women. The returns for married women are difficult to estimate directly, as most married women did not participate in the labour force. The chapter comes up with a clever alternative to using the Hiscam scores of married women, namely using the Hiscam scores of their husbands. Using this approach, Eriksson finds that the impact for married women is of a similar order of magnitude as for men and unmarried women.

Conclusions

What broader conclusions can be drawn from this dissertation? Because Eriksson's dissertation is written in the modern style of separate essays rather than a unified book, there is no single historical theme that emerges throughout the dissertation. However, there are a number of lessons from this dissertation that will be of broad interest to economic historians. First, the Swedish data are exceptional by historical standards, and scholars who are interested in a range of historical themes should pay careful attention to the work using this data – even if they are not historians of Sweden. The higher quality of data mean that it is possible to examine the micro-level mechanisms at work in Swedish development to a greater extent than has been the case for other counties. Second, market forces played an important role in Swedish economic development. The least efficient firms, i.e. those who did not employ women, were most likely to fail. Workers freely migrated in search of higher wages, resulting in increased output and a more urbanised nation. Finally, there are many more related projects that can be done with this data, particularly the linked censuses used in the last two chapters. For example, because the data contains information on fathers' occupations, it could be used to examine intergenerational transmission of inequality. Was having a father in a high status occupation the best way to attain a high status occupation oneself? Was having a father in an occupation requiring literacy or numeracy an important determinant of sons or daughters oc-

cupational status? There are also additional questions concerning migration that these data can be used to address? For example, was education level an important determinant of the returns to migration? Were there important differences in the returns to migration based on the characteristics of the place of origin, e.g. average earnings or employment rate, or the place of destination, e.g. population size or employment rate?