Testing the universality of historical occupational stratification structures across time and space

Paul Lambert Stirling University, UK, paul.lambert@stirling.ac.uk
Richard Zijdeman Utrecht University, Netherlands, R.L.Zijdeman@fss.uu.nl
Ineke Maas Utrecht University, Netherlands, W.A.F.Maas@fss.uu.nl
Ken Prandy Cardiff University, UK, prandyk@cardiff.ac.uk
Marco van Leeuwen International Institute for Social History, Netherlands, mle@iisg.nl

Paper presented to the ISA RC28 Social Stratification and Mobility Spring 2006 meeting, Nijmegen, May 11-14th.

DRAFT VERSION OF **5/5/06**, PLEASE DO NOTE CITE OR QUOTE WITHOUT THE AUTHORS' PERMISSION

Acknowledgements

This research was supported by an NWO research grant. We are extremely grateful to the coordinators of the research projects (cited in text) who allowed us access to the marriage record datasets used in this paper. We also wish to thank Jereon Vermunt for his provision of the lEM software package, used extensively in the work of this paper. We gratefully acknowledge Vincent Buskens and Stephanie Rosenkranz for the use of facilities of the Experimental Laboratory for Sociology and Economics (ELSE) at Utrecht University, The Netherlands.

Abstract:

In this paper we report on the estimation of a variety of 'HIS-CAM' scales of occupational stratification. The scales are estimated from marriage records collected over the period 1800-1938 in six countries: Britain, Canada, France, Germany, the Netherlands and Sweden. A dataset of approximately 400,000 marriages is collated, from which the occupation of the groom, bride, and parents of the groom and bride are given HISCO codes. The analysis proceeds by conducting 'RC-II' association models on cross-tabulations of child-to-parent occupations. These models may be estimated at a 'universal' level (a single occupational scale on the entire dataset), but they may also be estimated on subsets of data from different time periods and/or countries (a series of 'specific' occupational scales, for different countries). We find that specific occupational scale estimations are statistically favoured, and offer some revealing insights into the structure of occupational stratification in the period. Nevertheless, we also find evidence that universal occupational stratification scales are substantively parsimonious, insofar as they do a fair job of incorporating most, if not all, of the important differences in occupational stratification positions.

1. Introduction

> *"..what do the marriage patterns observed tell us about the social distance between classes?" (van Leeuwen and Maas 2005, p23)*

The question posed above embodies a research interest shared by contemporary and historical sociologists alike. The patterns of social associations exhibited between people from different occupational or class positions are felt to be revealing about the overall structure of social distance and stratification (cf. Stewart et al 1980; Bottero 2005), and to be indicative of numerous micro-social processes influenced by social structures and inequalities (Kalmijn 1998; McPherson et al 2001). Historical analyses of occupational structures – the focus of this paper – have increasingly been able to use large scale occupational data resources in order to examine patterns of social associations between occupations, ordinarily in terms of marriage record databases. Most commonly, occupational marriage patterns are analysed in terms of the binary association between the occupation of the groom's father and the bride's father (e.g. van Leeuwen and Maas 2005; Miles and Vincent 1993). Here we use the HISCO standardised occupational coding scheme (van Leeuwen et al 2002) to compare similar occupational data patterns using large samples from six societies over the period 1800-1938 (Britain, Canada, France, Germany, the Netherlands, and Sweden).

However this paper marks a departure from – and assessment of - some of the assumptions implicit in the quotation above. Firstly, we adopt a slightly more liberal approach to the interpretation of occupational social associations through marriage, analysing the patterns between any two occupational positions which are related through intergenerational marriage connections (for instance, groom to bride's father; bride to groom's mother; etc). Table 1 gives an overview of the familial relationships encompassed in the data resources used. Moreover, whilst marriage records provide our data source, we also utilise conventional intergenerational mobility records (e.g. groom-groom's father), as well as non-consanguineous intergenerational marital connections, in order to explore the patterns of relations between occupational locations. In fact, contemporary research on occupational social associations has increasingly suggested that *any* measures which indicate a social connection between occupational positions – for instance, whether based upon records of marriage, friendship, and inter-generational and intra-generational mobility - is likely to indicate the same empirical structures of social distance between occupations (cf Prandy and Lambert 2003; Rytina 2000; we return to this topic in section 2 below).

Secondly, we adopt a critical perspective to the occupational structure of 'classes' implied above. By conceptualising the structure of social stratification as primarily defined by social interaction relationships (Bottero 2005; Prandy and Lambert 2003; Stewart et al 1980), we are led to the belief that occupational positions can serve themselves as indicators of relative stratification position (rather than economic determinants of the same, the implicit assumption of many uses of the term 'class'). This leads to an open – but important – question about how categorical measures of occupational difference can best be translated into indicators of relative stratification position. On the basis of what, for us, seems overwhelming evidence that the primary (if not only) structure of stratification is one-dimensional hierarchical (cf. Rytina 2000; Prandy 1990), we support a strategy of assigning scale scores to categorical measures of occupational difference, on the basis of analyses of the frequency of social interaction positions. This strategy has had numerous implementations on contemporary survey datasets (cf Lambert and Prandy 2006; Chan and Goldthorpe 2004; Rytina 2000; Prandy 1990; Stewart et al 1980), and we are aware of a few similar activities with historical occupational records (Schumacher and Lorenzetti 2005; Maas and van Leeuwen 2005).

A neglected question in such research concerns how we should decide how many categorical measures of occupational difference we wish to scale, and how many different scales we should implement. Two *a priori* positions used during the numerous contemporary 'CAMSIS' scaling projects summarised by Lambert and Prandy (2006) have been that we should work with as fine a level of occupational disaggregation as possible; and that we should be prepared to estimate different relative locations for any different context to the occupational social associations (such as different time periods or different countries). Existing occupational classification systems have tended to represent a variety of locations in the dimensions polarised by these two positions. A selection of these are summarised schematically in figure 1. Aside from the contemporary CAMSIS scaling systems, most other social-interaction based measures of occupational scaling have, for a variety of practical and theoretical reasons, worked with a relatively small number of occupational locations (Chan and Goldthorpe 2004; Schumacher and Lorenzetti 2005, though cf. Rytina 2000).

The latter position has been referred to as the principle of 'specificity' in occupation-based social classifications (Lambert et al 2005). It may be contrasted with positions of 'universality', the claim that occupational locations are for most purposes fixed in their social meaning across time and space (cf Treiman 1977; Ganzeboom and Treiman 1996; Ganzeboom 2005). A major advantage of universalist schemes is the *ex post* comparability of terms between different eras and regions. However perhaps the majority of occupational schemes incorporate some degree of 'specificity' in their measurement, whether inherent to their design (as for instance in the CAMSIS approach), or, more commonly, in terms of their adaptation to limited specific features of the occupational structure of the given society (cf. Breen 2004). In this paper we present empirical evidence from historical data on the likely degree of universality that characterises occupational measurements of stratification structures over the nineteenth and early twentieth centuries.

The HISCO scheme (Historical International Standard Classification of Occupations, van Leeuwen et al 2002) was created to allow sociologists and historians to compare historical occupational titles from different languages, countries and time periods. It categorizes occupational titles from a growing number of countries and languages from the 18th to the early 20th centuries.

Several previous research applications have generated alternative 'univeralist' stratification schemes appropriate to the HISCO classification for the period of the nineteenth century (see also Figure 1). These include the HISCLASS (Maas and van

Leeuwen 2005) and SOCPO (van de Putte and Miles 2005) social class classifications; the linkage of HISCO with standardised routines for coding ISEI and SIOPS scales (using adaptations of the macros provided by Ganzeboom 2006); and a recent presentation on an universalist version of a social interaction based scale (HISCAM v0.1, see Maas et al 2006; the data on this scale may be downloaded from www.camsis.stir.ac.uk/hiscam). Below, we report on an effort to develop and compare alternative specific occupationally based social classification scales for HISCO codes.

2. Data

We work with a database of social associations between HISCO-coded occupational records spanning 6 countries (see Table 1).

In the analyses below, we estimate models on the two-way cross-tabulation between combinations of the child's and parent's occupation. We then consider a sequence of permutations of data structure as they interact with this basic cross-tabulation. As shorthand, we use the following representations:

C – Child's HISCO occupational classification (7, 10, 72, 224 or 581 categories)
P – Parent's HISCO occupational classification (7, 10, 72, 224 or 581 categories)

N – Country of origin of data (6 categories)
M – Gender of occupational combination (2 categories)
T – Time period of marriage (2 categories)

The alternative number of occupational categorisations associated with C and P is used to test the value of alternative levels of occupational detail. In all cases, we start with 5-digit HISCO occupational codes, then recode into smaller numbers of cases. The 7-, 10-, 72-, and 224- category versions, denoted respectively o1, o2, o3, o4, represent the number of categories emerging from the inbuilt HISCO structure of Major groups (o1 and o2), minor groups (o3) and 3-digit unit groups (o4) (van Leewen et al 2002). The 581- category version, denoted o5, is obtained by manually recoding the marital records whilst preserving to the maximum feasible occupational detail. This uses a working principle, favoured during contemporary CAMSIS estimations, that representation of an occupational unit group should exceed 20 cases, and if not, that group should be merged with another OUG with similar occupational properties (see Lambert and Prandy 2006).

The gender of occupational combination variable M is coded to two categories, indicating that both occupational holders are male, or that either of the two are female. This is a collapsed version of the full 4-category detail indicated in Table 1. The classification is used in order to reduce the computational demands of the model estimation process. However in further work we hope to consider alternative classifications of this measure, and also to incorporate information on the consanguinity of the occupational combinations, which are at this stage ignored.

The time period variate was used to try to approximate a single cut-point between more pre-industrialised, and more industrialised, economies. It was chosen after a descriptive review of the changing occupational distributions of the six countries over time. As measures of industrialisation, indicators that are thought to be congruent with the rise of industrialisation are often used, such as time and the declining percentage of labourers in the agricultural sector. We tried to identify a point in time when the occupational distribution shifted to one more typical of an industrialised economy. Therefore, in order to find a cut-point we also included the rise of occupation typically related to industrialisation. The cut-point between the estimated decline in the agricultural sector and the estimated increase in typically industrial occupations was used as a base to separate 'early' (less industrialised) from 'late' (more industrialised) periods. Unfortunately, this method did not work for all countries, for example due to a scarce number of typically industrial occupations.

A different division of periods for different countries was chosen: for the Netherlands and Germany, the periods were 1800-1890, and 1891-1938; for Sweden, 1800-1890 only (there was no more recent data available for Sweden); for France, 1800-1910 and 1911-1938; for Britain, 1800-1850 and 1851-1938; for Canada, 1800-1900 and 1901-1938. We do not wish to claim that these cut-points should be interpreted as definitive cut-points of industrialisation for each country, particularly, because for some countries in our analysis, the number of occupational titles is relatively small or stems from a single region. The next paragraph elaborates these and other difficulties with the data. However, we do note that there is a considerable time gap between the decline of the percentage of agricultural labourers and the increase in the percentage of labourers in typical industrial occupations. Therefore, our cut-points between 'early' and 'late' are typically later than reported in (historical) literature.

The HISCO data embraces a wealth of occupational detail, reflecting the methodological principles adopted during the development of the HISCO scheme (van Leeuwen et al 2002, p25ff). Equally, there are many problems associated with accurately understanding historical occupational data from such a range of countries and time periods (cf van Leeuwen et al 2002). As described below, our experience in working with this data has been that the most significant data problems on our resource involve:

- the uneven historical coverage of the German and Swedish records, which may conflate the understanding of national and time period differences within these two countries’ data resources
- the concentration of the Canadian data into a number of particularly populous occupational categories (agricultural), leading to low representation of Canadian occupations from other positions
- the recognition that although HISCO is translated in several languages, differences in applying a code between coders from different countries may appear. For a large part of the data such differences are controlled for through extensive cross-checking between different researchers, but the process is not yet completed

As noted above, we use data from a variety of types of marital social association (consanguineous and non-consanguineous, and combining different gender relations). Moreover, we make the claim that these patterns will generalise to the social interaction structure as measured by *any* indicator of a social connection between occupational positions. Bottero (2005) summarises theoretical justifications why such diverse relations should embody the same empirical structure of stratification, whilst Prandy and Lambert (2003) and Prandy (1994) have summarised empirical evidence for this in the British example.

It should be noted that this assumption often leads to a mistaken critique, of the apparent endogeneity of any scale estimations derived from such social association data. It may seem that scales derived from data on marriage, friendship, inter- or intra-generational mobility may not reasonably be then applied to subsequent analyses involving any of same concepts. In fact, this position indicates a fundamental misunderstanding of the nature of scale representations of categorical data structures. When a summary of metric information is mapped to a category identity, there is no logically necessary endogeneity to whatever metric source contributed to the information. That is, the scores assigned to categorical relations are overwhelmingly constrained by the information defining the categorical boundaries, rather than by the information contributing to an aggregate metric value which may be assigned to each category. There is of course a potential collinearity between such category scores and the origin metric. However, because of the empirical dominance of the categorical structures in applications such as described above, we are not aware of any previous example of scale representations of occupational data where there was evidence that such collinearity approached a problematic level of tolerance.

3. Model Fit Statistics

Our key question is whether the same social classifications may reasonably be applied to the same occupational categories across the range of countries and time periods over which the HISCO scheme offers coverage. In this section we report upon the statistical modelling issues involved in undertaking numerous alternative specific scale classifications. In section 4 we then evaluate the substantive properties of such alternative measures, in comparison with other alternative measures.

The various HISCAM scales reported here were estimated using implementations of widely used RC-II association statistical models (Goodman 1979), as implemented in the package lEM (Vermunt 1997). The theoretical ideas and practical implementations of this approach to measuring occupational structures of social interaction through statistical models estimated in lEM are described in detail by Lambert and Prandy (2006). The core idea is that a scaled association model can be used to estimate a set of numeric dimension scores which indicate the relative social interaction distance between different occupational unit groups, but which may reasonably be interpreted as a measure of the typical social stratification position of each occupation.

The most basic model is a universal model in which all data is combined in one table, and an association model run on the relationship between the Child's and Parent's occupations. In the shorthand of loglinear modelling (borrowing the terminology of Vermunt 1997), this may be denoted as the model C,P,ass(CP), described as model (1) in Table 2. Here, the C,P terms indicate that marginal terms account for the total distribution of childs and parents occupations, and the ass(CP) expression indicates that scaled association values are assigned to help predict the number of occurrences of each C-P permutation.

This model may then be extended in several ways. Firstly it may incorporate alternative permutations of the N,M,T factors within the marginal and/or scaled association parts of the model. For the former, the terms act as a control for any disproportionate representation of occupations between different countries, genders or time periods. For the latter, the terms act to estimate a separate scale for different countries, genders or time periods – i.e., a series of specific scales.

The model may also be extended by incorporating further dimensional structures to the CP association (for instance, it is often informative to estimate a second dimension which is constrained to coincide with gender segregation data on the proportion of women in each occupation. Another extension to the model concerns the possibility of estimating particular parameters for specific combinations of occupations, such as diagonal (occupational inheritance) or structurally related occupations. This extension incorporates quite an extended range of model choices. Here, we present models which use a large volume of such parameters and exclude a considerable number of marital records from the analysis (61% of all records) on the grounds of being ‘diagonals’ or ‘pseudo-diagonals’. The large volume of these exclusions concern the intergenerational farming-farming combination, but also other types of intergenerational linkage. We do not expand further here on the details of this strategy, and although the total exclusion is considerable, we would note that it is standard practice in log-linear modelling of social association patterns; and point out that the percentage here is higher than in contemporary applications almost entirely because of the disproportionate role of agricultural occupations in historical datasets. Further information on this aspect of the work is offered on email request to the authors. Finally, one further modelling possibility is the option of whether to constrain the scores estimated for children and parents to be equal on the same occupations, or whether the scores for each may be different. Table 2 lists the array of model permutations which are theoretically possible given this dataset.

Whilst Table 2 points to a vast range of potential models for our dataset, there are several practical constraints to these permutations. Firstly, the model estimations for larger numbers of occupational unit groups require substantial amount of computing time on a desktop pc (approximately 5 days per model). Secondly, data sparsity means that many of the theoretical model permutations may not be successfully identified in practice. Thus, in this paper we report upon a selected subset of the models of Table 2.

Figure 2 reports trends in the model fit statistics for selected models spanning five possible levels of occupational detail (o1, o2, o3, o4 and o5). Indications of the Log-likelihood and BIC statistics are presented, in both cases, standardised around the highest BIC and lowest log-likelihood obtained from the different models. Both measures summarise the extent to which the model structures improve our description of the data relations (the latter statistic taking some account of the relative parsimony of the explanation). The conventional interpretation of both measures is that the lower the BIC statistics, and the higher the log-likelihood statistics, the better fitting the model. However, it is important to realise that only the models within each level of occupational detail should be compared as nested (for instance, different o5 measures may be compared with each other, and different o1 measures with each other, but o1 and o5 may not be directly compared in this presentation).

When values within a series are compared, Figure 2 leads to a clear conclusion: in statistical terms, models with greater levels of control for specific contexts always offer greater explanatory power. Moreover, although it is not easily seen within the figure, the specific models when compared to the non-specific equivalents (such as 7n, 7m compared to 7), always offer a slightly improved fit on both measures. This is a revealing conclusion, which may not necessarily have emerged were the social interaction occupational structures identical across different countries or time periods. Nevertheless, the finding should be tempered with the consideration that, with a large sample such as used here, it is widely recognised that any meaningful model parameter is likely to lead to a reduction in BIC, regardless of the true substantive importance of the parameter (cf Breen 2004).

As indicated above, these model outputs are a selection from a number of estimations, but those estimations themselves are still at this stage some way from a complete review of all theoretical permutations. There are a number of complexities in alterative modelling specifications and convergence strategies that may be explored more fully in this evaluation. Of particular substantive interest will be to explore more examples of 'cross-classified' specific structures (ie, T*N, T*N*M, etc) which are not yet reported upon in this research. Additionally, there is much fine-tuning of scale estimations which may be attempted, particularly concentrating upon the specification of particular occupational combinations as 'diagonals' or structurally related 'pseudo-diagonals', described in section 2. Below, we report on some of the finer details of the scales encompassed by Figure 2, and we note at several points that an later, improved model estimation may be expected to bring more convincing scale scores for the relevant occupational database.

4. Substantive differences between specific occupational scales

Thus whilst there is clear evidence that there is statistical efficiency in estimating specific models, the more substantively interesting question concerns the nature of specific differences in occupational locations. In this section we examine the empirical patterns of the alternative universal and specific HIS-CAM scales. Given the large volume of alternative scales and considerations, we present selectively a number of indicative outputs. They suggest that there are many revealing differences between occupational positions in specific scales. However, they also suggest two patterns which may be taken as favouring a universalist approach: there is evidence of probable measurement error in many of the specific scale positions; and it is important to remember that much of the primary structure of social interaction patterns is shared between most specific schemes, whereas many of the specific differences are of a more particular nature.

Our strategy of enquiry which involves asking how the values derived from specific scales map on to alternative universal social classifications, whether a universal version of a HIS-CAM scale, or an alternative historical universal occupational measure (we operationalise the SIOPS prestige scale by adapting the macros of Ganzeboom 2006; and the HIS-CLASS scheme using a macro provided by two of the authors). Figures 3-5 and Tables 3-5 all indicate alternative aspects of this comparison.

Firstly, Table 3 indicates the overall magnitude of correlation between occupational scales derived from specific models within the structure specified by Table 2, and an appropriate universal HIS-CAM scale, across three different candidate levels of occupational detail (o1, involving 7 categories; o3, involving 72; and o5, involving 581). As would normally occur, the correlations are usually higher at the lower level of occupational differentiation. Across different specific schemes, the correlations with universal scales vary considerably, from almost perfect relationships to very low associations. The interpretation of these differences is ambiguous. A lack of perfect correlation could be a good thing – indicating a genuinely important substantive difference between the occupational structures in the different contexts. Equally, a low correlation is also likely to indicate some level of measurement error or a related estimation problem (such as inappropriate dominance by pseudo-diagonal effects). Indeed, a variety of literature can be cited as evidence that the broad structure of occupational stratification does not change dramatically between specific contexts over the time period concerned (e.g. van Leeuwen and Maas 2005; Miles and Vincent 1993; Treiman 1977). For this reason, we treat correlations to a universal scale of less than 0.5 as probable evidence of flawed specific estimations. Correlations greater than 0.5, however, might be presented as evidence of genuine specificity, pending further exploration of the scale structure.

Figures 3 and 4, and Table 4, begin to provide some indication of the finer structure of occupational stratification measured by the alternative specific scales. Figure 3 indicates the relative scores assigned to HISCO major groups ('o1') across these scales. In this figure, higher values on the vertical axis indicate occupational scores interpreted as of having less relative advantage (the ordering is of course arbitrary, though in most other examples, HIS-CAM scales are coded so that higher values are associated with greater advantage). The figure shows a clear structure of stratification in social association patterns between HISCO major groups, and is interpreted as evidence that the specific HIS-CAM scales will usually (subject to estimation problems) map onto a stratification system. Equally, Figure 3 highlights some instances where particular specific scales locate occupations differently from the general pattern of other locations. German Clerks, for instance, are rated as less advantaged that Clerks in other countries or scales; German sales workers, as more advantaged.

Figure 4 and Table 4 look at some of these patterns in slightly greater detail, concentrating on the finer level of occupational differentiation of HIS-CAM scales estimated at the 'o5' level of detail (581 different OUG codes). Table 4 picks out, somewhat arbitrarily, the occupations scaled at the extremes of the various HIS-CAM scales. It indicates that these are usually different between different specific versions. However in several instances, particular occupations are located in positions which might appear substantively implausible (for instance, those highlighted with an asterisk). Again this highlights an ambiguity in these estimation strategies : these locations are probably the outcome of model estimations which don't satisfactorily account for a particular combination of structurally related occupations, but there is no obvious criteria, except for a substantive judgement, to tell us that this is the case rather than there being a genuine difference between the occupations in each context.

Figure 4 also highlights this message of possible measurement error, by summarising all the different scores associated with the two historically specific scales derived from model (11n). We see their basic correlation with the universal HIS-CAM scale, and with each other, but we also so various patterns of divergence from this correlation. The key question is whether those divergences are substantively important, or measurement errors. Given that the entire process is essentially concerned with sampling estimations, there is necessarily some measurement error, but it is worth reminding that in conducting a metric estimation of relative stratification position or social distance, it is not theoretically problematic to tolerate some degree of uncertainty in the location of different occupations. Moreover, many of the outlying occupations of Figure 4 are occupational positions for which a theory of historical change in occupational compositions might reasonably be applied (and thus explain discrepancies in location between specific scale versions). Our position is that large gaps in scale score locations might reasonably be treated as evidence of specific patterns, but that greater attention needs to be paid to the sampling representativeness of the contributing occupations.

Lastly, Table 5 and Figure 5 focus on some more defined patterns of difference between selected specific scales. Figure 5 indicates where specific values at the major group level model (o1) indicate greater temporal or gendered differences within HISCO major groups. A ‘positive premium’ implies that the HIS-CAM score estimated for the major group is more advantaged on average for respectively the later time period or for female occupations, with a negative value implying the reverse – so for example, service workers have relatively higher locations in the later period, but relatively lower locations for women then for men. The value of the specific approach lies upon the claim that such differences are substantively plausible, and, moreover, interesting. At the major group level, where the sampling representation of each group within countries is good, we suggest that Figure 5 gives support to a specific approach.

Table 5 is similarly presented as possible support for a specific approach to occupational social classifications. Variations in the magnitude of correlation with universal prestige scale scores (SIOPS) and a class scheme influenced by a Weberian framework (HISCLASS) suggest possible specific differences. Moreover, the fact that the magnitude of the difference between the specific scores and universal scores is usually strongly correlated to SIOPS and HISCLASS measures suggests that at least some of the reason for specific differences has some meaningful structure.

The outputs presented focus upon the question of universality and specificity, though they also provide some evidence on our second research question, concerning the importance or otherwise of detailed occupational disaggregation. We do not present fit statistics on this question, due to the non-nested nature of the model as currently implemented, though it is anticipated that, again given the large sample size, conventional fit statistics would always favour more disaggregated data. The more important question is probably again one of the substantive value of occupational disaggregation. Firstly, Table 4 indicates a negative feature of disaggregate strategies, the possibility that they introduce uncertainty over the estimation of particular occupational locations. Equally, Table 4 and Figure 4 both indicate particular examples where occupational differences are revealed at a disaggregate level which do not coincide with more aggregate occupational classifications and thus would be obscured by them. A theoretically appealing strategy would be to propose as fine a level of occupational disaggregation in social classifications as may reasonably be supported by sample based data (although we recognise that in many research and data projects, the value added may not outstrip the data and potential measurement error costs involved, cf. Ganzeboom 2005).

Conclusions

Universalist strategies might not seem to do justice to the extended sensitivity to occupational variations applied during the development of HISCO. Firstly, they may appear crude and generalising, when historians who exploit HISCO are often interested in the minutiae of differences between occupations in the given historical and national context. Secondly, they offer little adaptation to distributional changes in occupational structures over time and between countries. These are influenced, in particular, by processes of industrialisation, which create new occupational positions, and may alter the circumstances of those in existing occupations. Indeed, there is a substantial literature which suggests evidence that the nature of occupational conditions, and the social associations between occupational positions, are subject to historical and national change (e.g. van Leeuwen and Maas 2005).

This paper has shown that multiple specific historical occupational stratification scales may readily be computed, by applying principles from the CAMSIS occupational scaling strategies to large scale historical data on intergenerational marital occupational associations. Moreover, the very process of empirically estimating and then examining these occupational intergenerational social associations provides for a revealing evaluation of the structures of social stratification inequalities in the 19th and early 20th centuries – there are many interesting research questions and theories which might be explored in an evaluation of the difference between an occupations circumstances in different contexts. Methodologically, there is a very important question about the quality of previous attempts of analysing occupational locations as indications of social stratification advantage: if there are genuine differences between contexts, and if there are genuine patterns of difference revealed at the finer level of occupational detail, then there may be inadequacies in previous analyses which have ignored such differences.

Nevertheless our evidence on the importance of specificity and of the importance of occupational disaggregation is mixed. There is clear statistical support for specific schemes, insofar as specific models capture more of the patterning of occupational social associations than universalist scales do. There are clear patterns of substantively plausible difference between different specific schemes, and substantively plausible patterns of difference between particular occupations which would have been obscured by more aggregated levels of occupational analysis.

Equally, there remain some evidence to favour a universalist approach: there is evidence of probable measurement error in many of the specific scale positions; and it is important to remember that much of the primary structure of social interaction patterns is shared between most specific schemes, whereas many of the specific differences are of a more particular nature.

The evidence of possible measurement errors in the more detailed models raises a further interesting question concerned with the standardisation of occupational coding between countries. In the estimation of CAMSIS scales for numerous contemporary national datasets, models have not usually been estimated in a pooled context (pooled between countries), because it was felt that the OUG scheme (whether a national-specific scheme or an international standard code) should best be uniquely coded within the context of the country's dataset. This refers to the stage of research involving manually recoding occupational units so that all have a reasonable level of representation (which generated the 581 categories of the 'o5' version reported in this paper). In this project, the manual recoding was done on an internationally pooled dataset; arguably, more robust results may be obtained by recoding within countries, and only later comparing scale scores for related occupations.

Future analyses may be employed to address a number of interesting questions on this data. These include further efforts in estimating and exploring alternative specific models (for instance alternative time period differences or gender patterns). However the most important further development may be an exploration of the reliability of particular occupational estimations described above, to offer an indication of the extent of possible measurement error.

Table 1: Data Sources on Inter-generational and marital connections

Study	Year range	Median year	N Child-Parent relations (% of which consanguineous) M-M	M-F	F-M	F-F	%cs
Netherlands							
ZA	1800-1923	1874	154251	69570	71756	52661	*49*
HSN	1812-1938	1895	17448	3434	4441	1813	*48*
Germany							
Knodel / Imhof	1800-1849	1827	364				*47*
(regional subsets)	1800-1849	1827	2084				*50*
	1800-1849	1826	490				*50*
	1800-1849	1829	1476				*49*
	1800-1938	1880	7429		419		*69*
France							
TRA	1803-1938	1877	54149	26981	30044	20150	*50*
Sweden							
DDB	1803-1889	1864	11685		7312		*47*
Britain							
Miles / Vincent	1839-1914	1874	19415				*50*
FHS	1800-1938	1873	31815	679	9464	388	*51*
Canada (Quebec)							
BALSAC	1800-1938	1895	488046		11636		*42*

Notes:

- Years refer to the year at which the marriage occurred.
- Cases indicate marriage records where HISCO occupational data was successfully coded for the relevant occupations. It is possible for the same marriage to contribute intergenerational information on more than one child-parent relation: the total sample covers 1,099,389 records of intergenerational marital associations, which are obtained from 475,919 distinct marriages.
- '%cs' refers to the percent of records within each study where the data refers to a parent and child from the same family (e.g. groom – groom's father, compared to groom-bride's father).

Sources:

HSN	Historical Sample of the Netherlands. International Institute for Social History (IISH). Amsterdam, The Netherlands. URL: http://www.iisg.nl/~hsn/database/
ZA	Civil Records of Zeeland. Zeeuws Archief. Middelburg, The Netherlands. URL: http://www.zeeuwsarchief.nl
Knodel / Imhof	Ortssippenbüchern, Germany (personal contact)
TRA	Base TRA Patrimoine. L'institut national de la recherche agronomique (INRA). Paris, France
DDB	Demographic Data Base. University of Umeå, Umeå, Sweden. URL: http://www.ddb.umu.se/index_eng.html
Miles / Vincent	Marriage records - literacy database (see Vincent 1989; Miles 1999)
FHS	Cambridge Family History Study (genealogical database, see Prandy and Bottero 2000)
BALSAC	BALSAC population register. University of Quebec at Chicoutimi, Quebec, Canada. URL: http://www.uqac.ca/balsac

Table 2: Selected alternative RC-II association models for social association patterns

Universal scales		**Specific scales**	
Model	**Shorthand**	**Model**	**Shorthand**
(1)	C,P,ass(CP)	(5n)	(5)+ass(CPN)
(2)	C; P; N; ass(C,P)	(7n)	(7)+ass(CPN)
(3)	C; P; M; ass(C,P)	(11n)	(11)+ass(CPN)
(4)	C; P; N; M; ass(C,P)	(6m)	(5)+ass(CPM)
(5)	C; P; N; CN; PN; ass(C,P)	(7m)	(7)+ass(CPM)
(6)	C; P; M; CM; PM; ass(C,P)	(11m)	(11)+ass(CPM)
(7)	C; P; N; M; CN; PN; CM; PM; ass(C,P)	(8t)	(8)+ass(CPT)
(8)	C; P; T; ass(C,P)	(10t)	(5)+ass(CPT)
(9)	C; P; N; M; T; ass(C,P)	(11t)	(7)+ass(CPT)
(10)	C; P; T; CT; PT; ass(C,P)		
(11)	C; P; N; M; T; CN; PN; CM; PM; CT; PT; ass(C,P)	(7nm)	(7)+ass(CPNM)
		(11nm)	(11)+ass(CPNM)
		(11nt)	(11)+ass(CPNT)
		(11mt)	(11)+ass(CPMT)
		(11nmt)	(11)+ass(CPNMT)

Further permutations:

(i)	(1_o1), (2_o1), (11n_o2), etc	*Occupational units o1, o2, ..,o5*
(ii)	(1a), (1b), etc	*(a) – scores unequal C-P; (b) equal*
(iii)	(1)-[1]; (1)-[2]; (2)-[1], etc	*Alternative structural cell parameters*
(iv)	(1.1), (2.1), etc	*Alternative subsidiary dimension structures, e.g:* *.1 No subsidiary dimension* *.2 Gender segregation dimension* *.3 Time period index dimension*

Table 3: Correlations between Universal and Specific scale values

	Correlations to Universal scale (11b_o1, 11b_03, 11b_o5)					
	Major group (o1 - 7)		Minor group (o3 - 72)		Unit group (o5 – 581)	
	Units	*Weighted*	*Units*	*Weighted*	*Units*	*Weighted*
HISCAM v0.1					*0.97*	*0.96*
SIOPS					*0.68*	*0.75*
	(11nb_o1)		(11nb_o3)		(11nb_o5)	
Netherlands	0.95	0.97	0.37	0.31	0.61	0.58
Germany	0.76	0.87	0.76	0.90	0.42	0.23
France	0.94	0.96	0.83	0.86	0.54	0.78
Sweden	0.82	0.88	0.10	0.15	0.12	0.41
Britain	0.91	0.90	0.93	0.80	0.49	0.32
Canada	0.86	0.90	0.79	0.89	0.70	0.89
	(11tb_o1)		(11tb_o3)		(11tb_o5)	
Early	0.99	0.99	0.37	0.12	0.81	0.97
Late	0.97	0.95	0.99	0.99	0.95	0.98
	(11mb_o1)		(11mb_o3)		(11mb_o5)	
Male	0.94	0.92	0.99	0.97	0.93	0.92
Female	0.92	0.95	0.24	0.32	0.23	0.60

Models refer to names identified in Table 2.
'Weighted' refers to population level correlations within appropriate context; 'units' to correlations at the level of occupational units only.

Table 4: Most advantaged and disadvantaged occupations, by alternative HIS-CAM schemes (unit group occupational scales, o5)

	Most Advantaged	Most disadvantaged	Greatest difference to Universal: Specific more advantaged	Greatest difference to Universal: Specific more disadvantaged
Universal	Chemical engineer Dentist University teacher	House servant Railway brakeman Mine cageman		
Netherlands	Jurist School inspector Lawyer	Cook Vine grower Farmer	Metal grinders Mine cagemen Machinery fitter	Cook Vine grower Farmer
Germany†	Retail salesperson Primary teacher Railway station master	Metal moulder Commercial traveller Supervison / foreman	Fine arts teacher Primary teacher Stenographic secretary	Metal moulder Commercial traveller Plasterer
France†	Clerical supervisors Engineers Jurists	Dentists* Motor vehicle drivers Glass cutters	Clerical supervisor Compositor Metal smelter	Dentist* School inspector Minister of religion
Sweden†	Dentist Teacher Other professional	Cook Tailor Dairy product process	Dentist Teacher Other professional	Tailor Sawyer Cook
Britain†	Primary teacher House servant* Office clerk	Fine arts teacher* Dentist* Glass cutter	Primary teacher Sales manager Jurist	Fine arts teacher* Trapper Saltmaker
Canada	Insurance clerk Author Secretary	Instrument maker Blacksmith Weighing clerk	Glass cutter Electrical repairman Protective services	Instrument maker Governess Chemist
Early	Insurance clerk Secretary Typist	Presser Sales manager* Machine operator	Insurance clerk Secretary Chemical technician	Presser Machine operator House servant
Late	Judge Jurist Medical doctor	Knitting machine oper. Wood boatbuilder Potter	Minister of religion Hot roller Production manager	Knitting machine oper. Wood boatbuilder Insurance clerk
Male	Secretary Chemical engineer Dentist	House servant Railway brakeman Vehicle loader	Winder Paper box maker Sewer	Governess Auxillary nurse House servant
Female†	Pattern maker* Malt cooker* Commercial artist*	Leather cutter Wooden model maker Decorator*	Pattern maker* Malt cooker* Railway brakeman	Decorator* Engineer Biochemist

* Occupational scores with high standardised residuals, anticipated to be the product of estimation errors / 'pseudo-diagonal' relations.
† The specific scale estimations for Germany, Sweden, Britain, France and Females at the 5-digit level are not considered here to require further treatments before being considered to be reliable measures of occupational stratification (see table 3).

Table 5: Correlations between HIS-CAM schemes and SIOPS and HISCLASS

	SIOPS *Pearson's correlation*100*		**HISCLASS** *Eta*100*	
	Weighted by total population			
	Total	Universal-Spec*	Total	Universal-Spec*
Universal	75		76	
Netherlands	51	-28	77	77
Germany	27	60	32	60
France	66	-50	70	59
Sweden	11	-57	62	71
Britain	1	-47	49	63
Canada	67	-2	80	46
Early	74	75	75	76
Late	74	-25	77	33
Male	62	34	71	62
Female	25	-71	45	76

*Relative advantage of occupations in specific versus universal scheme, correlated to SIOPS advantage or HISCLASS.

Figure 1 : Location of selected occupational schemes in dimensions of:
- **Universality –Specificity (horizontal)**
- **Number of occupational unit groups (OUG) (vertical)**

Universal	**Specific**	
Historical schemes *(linked to HISCO)*	**Historical schemes** *(linked to HISCO)*	**Contemporary schemes** *(social distance measures)*
	Low number of OUGs	
SOCPO **HISCLASS** *{HIS-CAM*}*	**Schumacher & Lorenzetti 2005 (Switzerland only)** *{HIS-CAM*}*	
{HIS-CAM}*	*{HIS-CAM*}*	**Chan & Goldthorpe 2004**
{HIS-CAM}* **SIOPS**	*{HIS-CAM*}*	
ISEI		**Rytina 2000**
HISCAM (v0.1)	*{HIS-CAM*}*	**CAMSIS**
	High number of OUGs	

**Scales evaluated in this paper*

Figure 2.

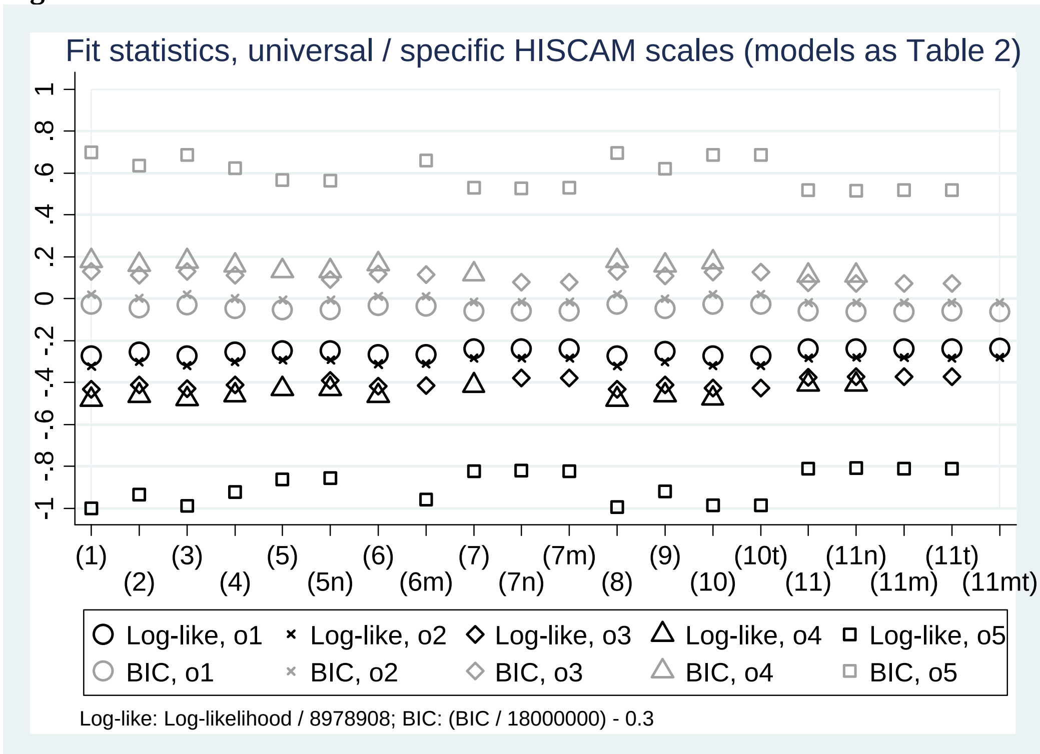


Figure 3: Specific scale score for HISCO Major Groups

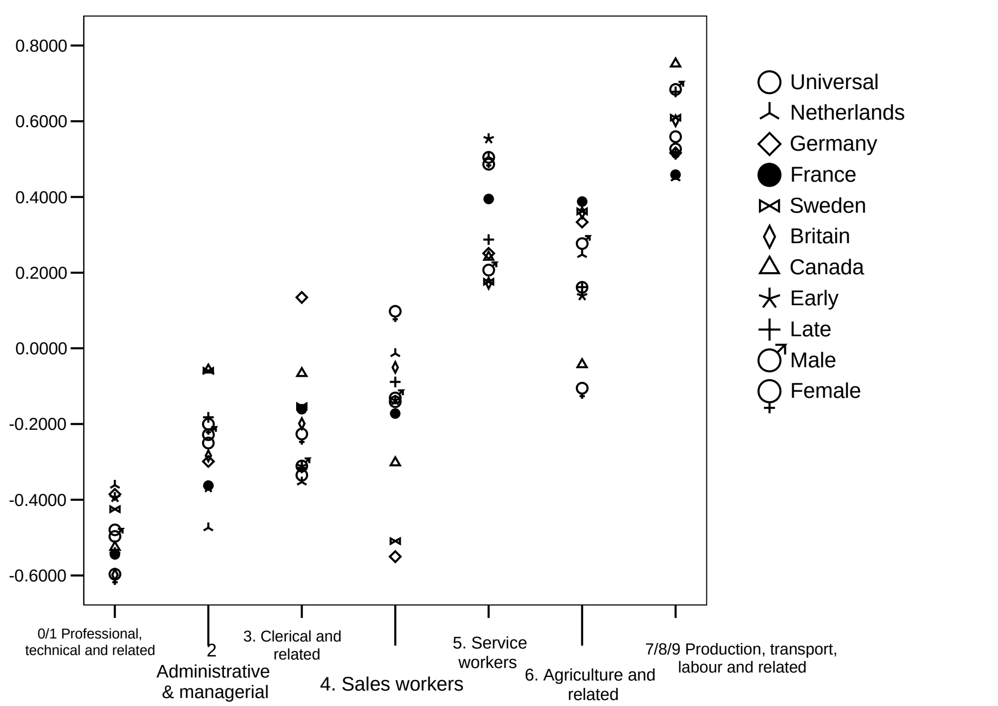


Figure 4: Universal to Historical-specific scale scores, HISCO unit groups

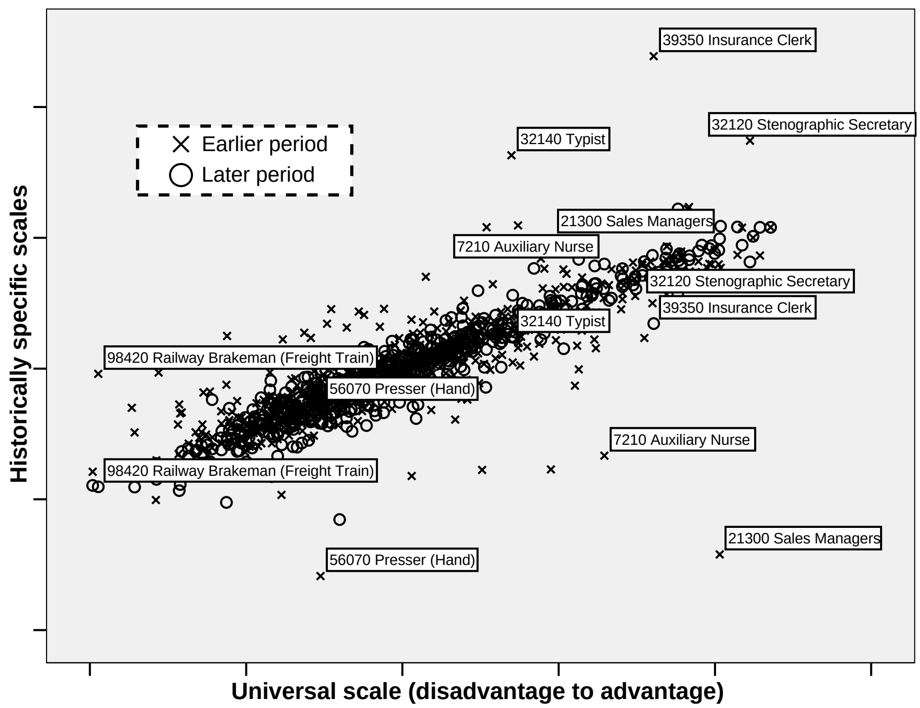
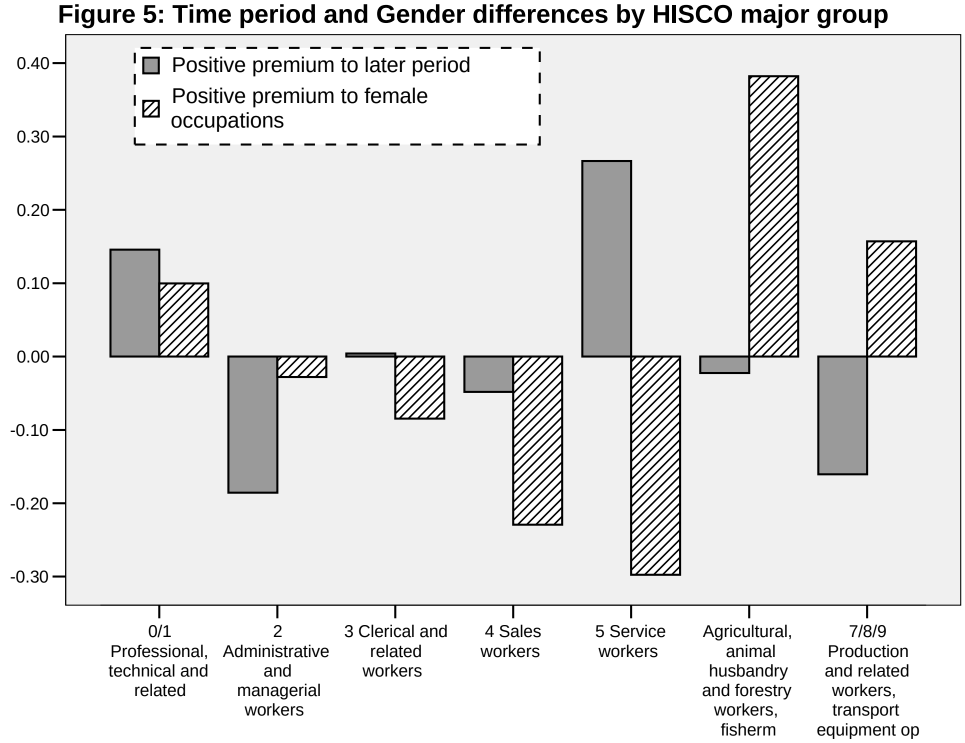


Figure 5: Time period and Gender differences by HISCO major group



References

Bottero, W. (2005) *Stratification: Social Division and Inequality*. London: Routledge.

Breen, R. (ed) (2004). Social Mobility in Europe. Oxford: Oxford Univeristy Press.

Chan, T.W. and Goldthorpe, J.H. (2004) "Is There a Status Order in Contemporary British Society?" *European Sociological Review*. 20(5), 383-401.

Ganzeboom H.B.G. (2006) *Tools for Deriving Status Measures from ISK0-88 and isco68*. http://home.fsw.vu.nl/~ganzeboom/PISA/ Utrecht: Utrecht University.

Ganzeboom, H.B.G. (2005) "On the Cost of Being Crude: A comparison of detailed and coarse occupational coding on the ISSP 1987 data" In Hoffmeyer-Zlotnik, J.H-P. and Harkness, J. *Methodological Aspects in Cross-National Research*. Manheim: ZUMA Nachrichten Spezial.

Ganzeboom, H.B.G. and Treiman, D.J. (1996) "Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations" *Social Sciences Research.* 25,201-235.

Goodman, L.A. (1979) "Simple models for the analysis of association in cross-classifications having ordered categories". *Journal of the American Statistical Association* 74, 537-552.

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

Lambert, P.S., Prandy, K. and Bergman, M.M. (2005) "Specificity and Universality in Occupation-Based Social Classifications." Paper presented to the first meeting of the *European Association for Survey Research,* Barcelona, 18-22 July 2005. (http://www.easr.upf.edu/) .

Lambert, P.S. and Prandy, K.. (2006) CAMSIS project webpages. , http://www.camsis.stir.ac.uk/ Stirling: Stirling University.

Maas, I. and van Leeuwen, M.H.D. (2005) "Total and Relative Endogamy by Social Origin: A first international comparison of changes in marriage choices during the nineteenth century". In van Leeuwen, M.H.D. et al (eds) *Marriage Choices and Class Barriers: Social Endogamy in History.* Cambridge: Cambridge University Press (special Edition of *International Review of Social History Supplements).*

Maas, I., Lambert, P.S., Zidjeman, R., Prandy, K. and van Leeuwen, M. (2006) "HIS-CAM - The derivation then implementation of an historical occupational stratification scale." Paper presented to the *Sixth European Social Science History Conference,* RAI, Amsterdam, 22-25 March 2006.

McPherson, M., Smith-Lovin, L. and Cook, J.M. (2001) Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27, 415-444.

Miles, A. (1999) *Social Mobility in Nineteenth- and Early Twentieth-Century England.* Basingstoke: Palgrave.

Miles, A. and Vincent, D. (1993) *Building European Society: Occupational Change and Social mobility in Europe, 1840-1940.* Manchester: Manchester University Press.

Prandy, K. (1990) "The Revised Cambridge Scale of Occupations". *Sociology*. 24(4)629-655.

Prandy, K. (1994) "Similarities of Life-Style and the Occupations of Women." In Crompton, R. and Mann, M. *Gender and Stratification, 2nd edition.* Cambridge: Polity Press.

Prandy, K. and Bottero, W. (2000) “Social Reproduction and Mobility in Britain and Ireland in the Nineteenth and Early Twentieth Centuries” *Sociology*. 34(2)265-281.

Prandy, K. and P.S. Lambert (2003) "Marriage, Social Distance and the Social Space: An alternative derivation and validation of the Cambridge Scale," *Sociology* 37 (3): 397-411.

Rytina, S. (2000) "Is occupational mobility declining in the United States?," *Social Forces* 78 (4): 1227-1276.

Schumacher, R. and Lorenzetti, L. (2005) “’We have no proletariat’: Social Stratification and Occupational Homogamy in Industrial Switzerland, Winterthur 1909/10-1928”. In van Leeuwen, M.H.D. et al (eds) *Marriage Choices and Class Barriers: Social Endogamy in History*. Cambridge: Cambridge University Press (special Edition of *International Review of Social History Supplements).*

Stewart, A., Prandy, K. and Blackburn, R.M. (1980) *Social Stratification and Occupations.* London: MacMillan.

Treiman, D.J. (1977). *Occupational Prestige in Comparative Perspective*. New York: Academic Press.

Van de Putte, B. and Miles, A. (2005) “A class scheme for historical occupational data. The analysis of marital mobility in industrial cities in 19th Century Flanders and England”. *Historical Methods.* 38(2)61-92.

van Leeuwen, M.H.D. and Maas, I. (2005) “Endogamy and Social Class in History: An Overview”. In van Leeuwen, M.H.D. et al (eds) *Marriage Choices and Class Barriers: Social Endogamy in History*. Cambridge: Cambridge University Press (special Edition of *International Review of Social History Supplements).*

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

Vermunt, J.K. (1997) *lEM : A general program for the analysis of categorical data*, Tilburg, Netherlands: Tilburg University.

Vincent, D. (1989) *Literacy and Popular Culture: England 1750-1914.* Cambridge: Cambridge University Press.