

Chapter 11

Conclusion

The basis for this thesis has been the desire to make efficient use of very large datasets in general and one in particular, the NESPD. For most of its life, the importance of the NES to UK economists was largely due to use of the published NES rather than analyses of the underlying data. Since the micro-data in the NESPD was made available to researchers at the end of the 1980s, the amount of analysis carried out on the data has increased considerably, although this is still relatively small compared to the potential of such a large and comprehensive dataset and utilisation of other surveys like the FES, GHS and WIRS.

This thesis has described the way in which effective use can be made of the data. The core of the work has been to show how linear estimation methods can be made fast, practical, and free from the constraints of having to work at arm's length with data only available at the DE. The validity of this approach is indicated by the fact that, in recent years, a score of papers using OLS regressions on the NESPD have been produced at the Universities of Stirling and Manchester.

The value of cross-sectional studies on the NES is high. In terms of numbers surveyed and the range of work variables defined, it far exceeds any other survey of the UK labour force. It also allows for estimation over a long period (sixteen years worth of data was available for this thesis, and a further five years is due shortly) and the construction of time-series of coefficients. This is a feature of the data which has seen little interest outside the confines of the Universities involved in this project.

While the ability to perform cross-sectional analyses, with time-varying coefficients, quickly and efficiently is a significant step forward, linear regressions have been run by other researchers. With respect to cross-sectional analysis, the NES is similar to many other surveys apart from its size. However, the ability to construct fixed-effects estimators which

allow for individual heterogeneity is almost unique amongst the large UK surveys. The allowance for individual heterogeneity goes some way to making up for the lack of personal and educational variables in the NES; indeed, the fixed-effects assumption may be a better way to treat education than to employ categorical variables which lump different types of education together. Moreover, the fixed-effects assumption also takes into account characteristics which are constant over time but essentially unmeasurable, such as "motivation", "preference", "ability", "attitude", and so on. This is only feasible on a panel dataset, and the NESPD is the only major UK survey constructed as a panel with several years worth of data¹. The importance of allowing for individual heterogeneity was shown in chapter nine, where the results from the TVCS and TVFE models were compared. Although the qualitative results were similar, the scale of the estimates produced by the two models can differ dramatically.

Chapter two described the theoretical advantages of using panel models, and the practical aspects of panel estimation. The basic literature has been well established for some years now, the significant work being Hsiao's (1986) monograph. However, one potential advantage of using panel models has been almost completely overlooked: the ability to let coefficients vary over time and so to create time-series estimates of coefficients. The usefulness of time-varying coefficients is apparent from chapter ten, where a time-series of the decomposition of wage differentials was created. Bell and Ritchie (1995c) have used this to show how the various components of the differential have changed over time. This gives a dynamic aspect to the analysis of male-female wage differentials.

The models presented in this thesis are unusual in that the basic specification of all models (save to some extent the unbalanced differenced model) includes the assumption that coefficients are not constant over time. A brief consideration of the UK labour market should

¹ There are two relatively young panel datasets available: the British Household Panel Survey and the Cohort Study of Youth Training. The latter is limited in its scope but the former is a general survey and has now been running for three years.

suggest that this is a sensible basis for modelling; and from a theoretical position the general-to-specific school of econometric methodology would argue that the "restricted" models of this thesis (the standard "fixed-effects" models of most applied work) should be tested as special cases of a more general hypothesis. The validity of the restricted models was tested to some extent in chapter nine, and it fails fairly comprehensively. This is not a very surprising result but has been overlooked by a large number of researchers.

Chapter five described how cross-section, fixed-effects, and differencing estimators could be created, and the mathematics underlying the efficient analytical routines. This is the crux of the practical aspect of the thesis. Although some other programs (such as STATA) can run regressions on cross-product matrices, the intention to use the cross-product matrix as the basic data type means that both extraction and analysis programs can be tailored to produce particular results - the fixed effects estimator being one such outcome. The basic assumption of coefficients varying over time will also affect the calculation of cross-product matrices, and it was shown that the construction and manipulation of matrices to achieve this requires some forethought.

The end result of this has been the first (indeed, only - so far) estimates to make use of the NESPD's distinguishing feature - its panel nature. Chapter nine showed that the effect of using panel estimators may be considerable, and suggests a significant area for further study. One such possibility is the implementation of minimum-distance estimators to allow for flexible definitions of the error term - another estimator with theoretical advantages which is rare in applied work but whose practical problems are by no means insurmountable in the cross-product framework.

Chapters nine and ten produce a number of results on male and female earnings, largely using the TVFE estimator. Whilst most of the results are in broad agreement with other applied work, there are some surprises; most notably the very small coefficients on collective agreement, and their path over time. Several of the coefficients illustrate the importance of

constructing time-series of the coefficients, where possible: estimates of the effects of public sector working, unions, and region, for example, are very dependent on when the estimates are made.

Finally, one other aspect of collecting and using data efficiently has been described, although it has had a marginal effect on this thesis. This is the observation history, an extremely compact way of storing information about states. This is currently receiving some interest in the construction of multi-destination transition matrices, but in this thesis it has been used largely to provide an alternative view of the labour market by comparison with the more usual employment cohort. The descriptive analysis of the labour market presented in chapter eight generally ties in with other studies of the labour market. However, the employment and data cohorts illustrated that the dynamic effects are both important in explaining the characteristics of the labour market and the differences between the sexes.

In summary, the final contribution of this thesis has been twofold. Firstly, to provide some efficient techniques for retrieving and analysing data from one of the largest datasets of its kind in the world; secondly, to provide the first analyses on this dataset allowing for individual heterogeneity and a flexible parameter structure. And as a side effect, it is hoped that the benefits of allowing for time-varying coefficients when estimating on repeated micro-data have been demonstrated convincingly.