7

Conclusions and Outlook

The aim of this book was twofold: In the first part of the book we wanted to give some practical guidance for researchers who want to evaluate active labour market programmes (or other measures) and have to decide between different available non-experimental estimators. Thereby we concentrated on (propensity score) matching and discussed several issues arising during its implementation. We used this knowledge in the second part of the book to evaluate the employment effects of JCS for the participating individuals in Germany.

In the first chapter we presented an overview of the most relevant microeconometric evaluation methods and gave researchers some guidance on how to choose between different approaches. The discussion in this chapter has shown that each non-experimental estimation strategy relies on identifying assumptions and has to be justified case-by-case. Clearly, in an ideal world, the evaluator is already involved at early stages of the programme design and has influence on the data collected for later evaluation. In that case, one can make sure to collect those data needed to justify either the unconfoundedness assumption or to create an instrument (or an exclusion restriction) that allows to use IV methods or selection models. If the evaluator is instead faced with an ongoing programme, he carefully has to assess which identification strategy works for the situation at hand, taking the design of the programme, the selection process, and the available data into account. With matching one tries to construct a 'comparable' control group, selection models try to model the selection decision completely and IV methods focus on finding a good instrument. As Smith (2004) notes, matching methods make no sense without rich data, IV methods make no sense without a good instrument, and finally, longitudinal methods make no sense when selection into treatment depends on transitory rather than permanent shocks.

Chapter 2 than focussed on the matching approach. We have presented several alternative matching estimators and also have shown, that asymptotically all approaches should yield the same results. However, we have highlighted that in small samples the choice can be important. All matching estimators

contrast the outcome of a treated individual with the outcome of comparison group members. The presented estimators differ not only in the way the neighbourhood for each treated individual is defined and the common support problem is handled, but also with respect to the weights given to these neighbours. Usually a trade-off between bias and variance arises. We have shown that there is no clear winner for all situations and that the choice of the estimator crucially depends on the situation at hand. Pragmatically, it seems sensible to try a number of approaches. Should they give similar results, the choice may be unimportant. Should the results differ, further investigation may be needed in order to reveal more about the source of the disparity. Finally, we have also presented the benefits which can be achieved by combining matching with other methods, e.g. to take time-invariant unobservable factors into account or to exploit the relation between covariates and outcomes.

In chapter 3 we have focussed on some practical issues when implementing propensity score matching. Our discussion has made clear that the researcher faces a lot of decisions during the implementation. The first step of implementation is the estimation of the propensity score. As to that we have discussed not only the choice of the model but also the choice of the variables to be included in the model. Second, our discussion has emphasised that treatment effects can only be estimated in the region of common support and that testing the overlap in the propensity score distribution of participants and non-participants is a pre-condition for applying matching estimators. The third choice, namely the choice between different matching algorithms usually involves a trade-off between bias and variance and has been discussed extensively, too. Since the main goal of the matching procedure is to balance the distribution of covariates in the treatment and control group, one has to check in a fourth step if that goal was achieved. We presented several procedures to do so. Finally, one has to estimate standard errors and to decide when to measure the effects. A last step of matching analysis is to test the sensitivity of the results with respect to 'hidden bias'. We have presented an approach (Rosenbaum bounds) that allows the researcher to determine how strongly an unmeasured variable must influence the selection process in order to undermine the implications of matching analysis.

Based on the methodological discussion in chapters 1 to 3, part II of the book was concerned with the evaluation of the employment effects of JCS on the participating individuals in Germany.

In chapter 4 we have presented the institutional setup of labour market policies in Germany in general and of JCS in particular. We have shown that microeconometric evaluation of JCS in Germany was hampered for a long time by the absence of suitable data. The earlier evaluation studies of JCS mainly concentrated on the East German labour market and were based on survey data, either the labour market monitor of East Germany or the one of Sachsen-Anhalt. The relatively small groups of participants did not allow to take adequately into account effect heterogeneity. The presentation of the dataset we used for the empirical analysis made clear, that this is not a