Accurate statistical analysis in dynamic panel data models

Bun, M.J.G.

Citation for published version (APA):
Chapter 6

Summary and conclusions

In this thesis we have analysed statistical inference methods for dynamic regression models for panel data. We have focused on the analysis of panels with a limited number of observations in both the time and cross-sectional dimensions. In this case first-order asymptotic theory developed either for the finite $T$, large $N$ panel or large $T$, large $N$ panel may be misleading and yield inaccurate inference in finite samples. Hence we have pursued two goals, viz. establishing higher-order asymptotic properties and exploiting these in order to improve on existing first-order asymptotic inference methods. Regarding the former, using asymptotic expansion techniques we have derived several analytical results on the higher-order properties of asymptotic inference techniques for a variety of models. In addition we have examined the actual performance of these methods in small samples by a number of simulation studies. Regarding the latter goal, we have examined to what extent the results from asymptotic expansions may be used to improve on the accuracy of existing techniques. Both bias corrections based on analytical expressions and empirical bootstrap methods have been exploited to enhance the accuracy and efficiency of asymptotic inference methods.

Starting point for the analysis is the first-order stable dynamic panel data model with a scalar disturbance covariance matrix. The inclusion of a lagged dependent variable regressor in the model complicates estimation of the unknown parameters, especially when $T$ but also when $N$ is small. In Chapter 2 we focus on the case where the explanatory variables other than the lagged dependent variable regressor are all strongly exogenous. We examine by simulation the bias and mean squared error of various coefficient estimators, the bias in related estimators for the disturbance variance and estimators of the coefficient standard errors. Also we examine the actual size of simple coefficient tests. As we are interested in panels with both dimensions small, we compare LSDV, which is consistent for large $T$, with various IV and GMM techniques, which are consistent for large
We find that for these typical dimensions of the panel all techniques show substantial distortions. However, a particular bias corrected LSDV estimator (LSDVc) as proposed by Kiviet (1995, 1999) performs relatively well. The main theoretical result in Chapter 2 is the construction of an appropriate estimator for the asymptotic standard error of the LSDVc estimator, which proves to be non-trivial for the small $T$, large $N$ case.

In Chapter 3 we continue to focus on higher-order approximations of the bias of LSDV and other estimators in empirically more relevant models. We consider models which in addition to a lagged dependent variable regressor also have a dynamic feedback mechanism from the dependent variable to the explanatory variables. In Chapter 3 the focus is on lagged feedback mechanisms, i.e. weakly exogenous regressors, and their effects on the finite sample properties of LSDV, IV and GMM estimators. By comparison of asymptotic variances we show that, although less biased, the efficiency losses of simple IV estimators are substantial as compared with LSDV (and GMM). Using asymptotic expansion techniques we find several noteworthy results on the higher-order properties of GMM estimators. We consider GMM estimators using instruments in levels for the equation in first differences (or orthogonal deviations) and vice versa. In general we find that the order of magnitude of the finite sample bias increases with the number of moment conditions used in estimation. In the simulations we compare these GMM estimators with a particular bias corrected LSDV estimator. The bias corrected LSDV estimator has been developed for strongly exogenous explanatory variables and by simulation the impact of weakly exogenous regressors on its finite sample properties has been examined. From the simulations we find that, although a weakly exogenous regressor is included, bias corrected LSDV can still compete with GMM estimators.

We use the results of Chapters 2 and 3 to analyse the determinants driving local economic activity in Moroccan cities. Using annual data for 6 urban areas and 18 industrial sectors indicators for specialisation, diversity and competition of firms within a particular region are constructed for the years 1985-1995. The effects of these and other explanatory variables on local economic activity are estimated using the model and inference methods analysed in Chapters 2 and 3. The estimation results suggest several empirical phenomena. First, we find significant positive specialisation and diversity effects on local economic activity and significant negative competition effects. Second, significant heterogeneity is found reasserting the necessity of analysing small panels in which heterogeneity is mitigated.

In Chapter 4 we continue to analyse the LSDV estimator, but now in higher-order dynamic panel data models with a non-scalar disturbance covariance structure. As the covariance matrix is non-scalar now, also the generalised LSDV estimator is analysed.
Using asymptotic expansion techniques bias approximations are derived for these estimators. The focus is on panel data models with the number of cross-sectional units $N$ relatively small with respect to $T$, hence we derive the bias approximations from a large $T$ perspective. Attention is paid to estimation of both short- and long-run multipliers, to estimation under restrictions and also to variance estimators.

In Chapter 5 we focus on a fundamental assumption of panel data models, viz. homogeneity of parameters across cross-sectional units. Of course, panel data models allow for individual specific effects, but the assumption of equal slope coefficients for all cross-sectional units is a rather strong one. Whenever $T$ is large enough compared to $N$ it is feasible to test for this type of heterogeneity in the parameter vector within the framework of a system of regression equations. Using both simulation evidence and bounds analysis we find that the robustness of the standard $F$ test for single equation models is very weak in case of dynamic regression models and models with non-spherical disturbances. In the latter case, use of the standard $F$ statistic is not appropriate and we examine a generalised version of the $F$ statistic (Roy, 1957; Zellner, 1962), for which approximate distribution theory exists. By simulation it is shown that the finite sample performance of asymptotic tests is rather poor. However, employing the original test statistics with bootstrapped critical values leads to much more accurate inference in finite samples.

With the econometric theory developed in Chapters 4 and 5 the money demand data from Vlaar and Schuberth (1998) are analysed. The possibility of exploiting panel data techniques to analyse these cross-country data is explored in Chapter 4. Significant spillovers between countries are found reflecting the dependence of domestic money demand on foreign developments. The empirical results show that in general plausible long-run effects are obtained by the bias corrected estimators. Moreover, bias correction can be substantial underlining the importance of more refined estimation techniques. Also the efficiency gains by exploiting the heteroscedasticity and cross-correlation patterns between countries are sometimes considerable. Of course, panel data techniques require to some extent slope homogeneity across countries. Hence the results of Chapter 5 are applied to the money demand data to examine the degree of heterogeneity in the slope coefficients across countries. It is shown that classical asymptotic tests and bootstrap procedures may lead to conflicting test outcomes. Given the superior finite sample performance of bootstrap tests over classical procedures these empirical results indicate that aggregating or pooling seems to some extent acceptable in the period after the German reunification.

Based on the results in the separate chapters in this thesis we come to several general conclusions. First, the empirical studies show that allowing for dynamic adjustment is cru-
cial for an adequate description of the data. More in particular, including lagged values of both dependent and explanatory variables as regressors in the model is usually inevitable for getting plausible estimates. However, especially the inclusion of lagged dependent variable regressors complicates statistical inference considerably. In Chapters 2 and 3 we partly explained the large differences found between different estimation techniques as emphasised in Chapter 1. We have found some regularities across inference methods, which may provide the applied researcher with guidelines on deciding which technique to use in a particular situation. However, it seems that there is no method available which has superior performance over a wide range of parametrisations and dimensions.

Second, analysing dynamic models for panel data one has to take into account the particular dimensions of the panel as both finite sample and asymptotic distributions depend on these dimensions. For example, in deriving higher-order properties of the LSDV estimator the approach in Chapters 2 and 4 is from a large $NT$ and large $T$ perspective respectively. As a result the leading term in the bias approximation of the LSDV estimator is different in these chapters. More generally the simulation results of Chapter 2 and bias approximations in Chapter 3 show that dimensions play an important role with respect to the accuracy of estimators. The dependence of finite sample distributions on the dimensions of the panel has an analogue in first-order asymptotic theory. Regarding asymptotic theory we considered in this thesis only some particular cases, i.e. letting either $T$ or $N$ or both going to infinity. However, letting both dimensions go to infinity a whole spectrum of asymptotic distributions emerges depending on the relative rates of $N$ and $T$ going to infinity. The asymptotic distribution of the LSDV estimator, for example, exhibits an asymptotic bias when $N$ and $T$ grow large at the same rate, while this is absent when $T$ goes faster to infinity than $N$ (Alvarez and Arellano, 1998; Hahn and Kuersteiner, 2000).

Third, modelling heterogeneity in panel data models through individual and/or time specific effects may be inadequate. In many data sets homogeneity of slope vectors across cross-sectional units and/or time may be violated leading to serious misspecification. The approach we take in this thesis is to split the available sample into parts and then analyse them separately. An alternative approach is to incorporate a particular specification for the heterogeneity in the model directly. An example is to allow for heterogeneity in the short-run coefficients but impose homogeneity of long-run parameters (Pesaran et al., 1999). Another approach is to assume stochastically varying slope coefficients (in addition to random individual specific effects) resulting in so-called random coefficient models.