Accurate statistical analysis in dynamic panel data models

Bun, M.J.G.

Citation for published version (APA):
Chapter 1

Introduction

1.1 Motivation

Economic behaviour is often characterised by dynamic adjustment processes, i.e. current behaviour of economic agents is determined by past actions, present circumstances and expectations about the future. In time series econometrics the usual way of modelling such behaviour empirically is by fitting dynamic linear regression models to the data. An important feature of these models is the inclusion of lagged values of the dependent variable as regressors to model the relation between current and past outcomes. Next to lagged dependent variable regressors, both contemporaneous and lagged values of explanatory variables may be included in the model. The resulting class of regression models is often referred to as Autoregressive Distributed Lag (ADL) models. Regarding the marginal effects of explanatory variables on the dependent variable, in ADL models a clear distinction can be made between short- and long-run multipliers. A specific representation of ADL models is the well-known error-correction model (Davidson et al., 1978), in which the long-run equilibrium relationship and short-run dynamics are modelled more explicitly.

In modelling dynamic economic processes by fitting ADL models to the data, panel data analysis may have advantages over either pure cross-section or time series analysis\(^1\). Under specific assumptions cross-sectional estimates may be interpreted as long-run effects (van den Doel and Kiviet, 1994; Pesaran and Smith, 1995). However, it is the repeated measurement of the cross-section which enables the researcher to estimate also the dynamics, which may be informative both for the magnitude of immediate effects and the speed of adjustment to the long-run relationship. The obvious alternative for dynamic

\(^1\)For a more general discussion of the relative merits of panel data, see Hsiao (1985).
panel data models are econometric models based on pure time series. However, often one needs rather long time series in order to have sufficient degrees of freedom, which may not be available at the desired level of aggregation. For example, when analysing household or firm behaviour often a few time periods are available only. Also, economic relationships may have changed over time, e.g. due to regime switches or other factors. This may complicate the estimation of structural parameters from long time series. As the cross-sectional dimension in a panel provides extra data to estimate the same number of unknown parameters, with panel data techniques it is possible to use fewer time observations than in a pure time series approach.

Typically, panel data are associated with a short time series and a large cross-section\(^2\). Some well-known examples are the data used in Balestra and Nerlove (1966), i.e. annual data on gas consumption for the years 1950-1962 in 36 US states, or Arellano and Bond (1991), who exploit annual data on 140 UK companies for the period 1976-1984 to estimate labour demand. Least squares estimation techniques give consistent estimators in a dynamic setting only when the number of time observations \(T\) in the data set approaches infinity. Various authors have paid attention to this issue and have proposed IV and GMM alternatives, which are consistent for finite \(T\) and an infinite number of cross-sectional observations \(N\) (Anderson and Hsiao, 1982; Arellano and Bond, 1991; Ahn and Schmidt, 1995; Blundell and Bond, 1998). GMM estimators typically use more orthogonality conditions than their simple IV counterparts and they take the covariance structure of the disturbances into account. Therefore they are asymptotically more efficient.

In an increasing number of studies panel data techniques have been exploited for analysing data sets where \(T\) and \(N\) are of similar magnitude or where \(T\) is larger than \(N\). Examples are Pesaran and Smith (1995), who use annual data on 38 UK industries for the period 1956-1984 to estimate labour demand, or Baltagi and Griffin (1997), who use annual data on gasoline consumption in 18 OECD countries for the period 1960-1990. An important question is whether the statistical analysis of this type of panel data should differ from that of the typical small \(T\), large \(N\) panel. Here, least squares based methods, which are consistent for \(T\) large, may be preferred.

A complicating issue in the large \(T\), large \(N\) panel is the appropriate modelling of unobserved heterogeneity, especially between cross-sectional units. Traditional panel data models incorporate heterogeneity across individuals and/or time by allowing for individual and/or time specific constants, whereas the parameters of interest, i.e. the reaction

\(^2\)In this thesis we will consider models for balanced panel data only. Also, we abstain from attrition or rotating panels, hence each individual time series is complete and contains repeated observations on a unique subject.
coefficients, are assumed equal across individuals and over time. Recent contributions to the panel data literature (Robertson and Symons, 1992; Pesaran and Smith, 1995; Maddala et al., 1997) question this homogeneity assumption and propose alternative models, which allow the complete parameter vector to vary across individuals. More in particular, Pesaran and Smith (1995) show that in case of heterogeneity of the slope parameters traditional panel data estimators are inconsistent in dynamic models. However, based on a comparison of forecast performance Baltagi and Griffin (1997) conclude that traditional panel data estimators are superior if slope heterogeneity is not too strong.

In this thesis we will analyse panel data with both $T$ and $N$ small or moderate. There are several reasons which justify the analysis of panels with a limited number of observations in both the time and cross-sectional dimensions. First, as mentioned above there is a growing number of panels available, e.g. cross-country data, with a limited number of cross-section units $N$. Second, heterogeneity in large panels may be mitigated by analysing subsamples with both $T$ and $N$ small or moderate. In this way, it is hoped that structural breaks in the slope vector across cross-sectional units or over time are avoided. Hence, traditional panel data models with individual and/or time specific constants may be a valid alternative when applied to panels where both $T$ and $N$ are moderately small.

When both dimensions in the panel are small first-order asymptotic theory developed either for the finite $T$, large $N$ panel or large $T$, large $N$ panel may be misleading. It is rather straightforward to examine the actual performance in finite samples of asymptotic inference techniques by simulation. In this thesis we will present outcomes from such Monte Carlo experiments. Regarding dynamic panel data models there are only few analytical finite sample results available. Also results on higher-order asymptotic properties of statistical inference techniques are scarce. In this thesis we will focus on these higher-order properties and try to improve on existing first-order asymptotic inference methods.

In Section 1.2 we will review some of the major developments regarding the estimation of dynamic regression models for combined cross-sectional and time series data. The literature on this particular class of econometric models is vast, hence we will present some selected models and inference methods only. Next, in Section 1.3 we will present empirical results, which illustrate the potential problems in producing accurate inference in these models. The examples presented will emphasise the importance of the availability of accurate inference procedures for the analysis of panels with a limited number of both time and cross-sectional observations. In Section 1.4 we will focus in more detail on the contributions of this thesis to the field and we will give an outline of the remaining
1.2 Cross-sectional time series data

1.2.1 Model

Throughout the thesis we will consider inference procedures for particular members of the following class of linear regression models

$$y_{it} = \sum_{p=1}^{P} \gamma_{pi} y_{it-p} + \beta_i' x_{it} + \eta_i + \varepsilon_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T. \quad (1.1)$$

The indices $i$ and $t$ in (1.1) refer to cross-sectional and time series observations respectively. Examples of the former are households, firms, industries or countries, while the latter may be years, quarters or even a shorter time period. In the introduction of this chapter several examples of combined cross-sectional time series data have been given.

In model (1.1) for each cross-sectional unit $i$ the dependent variable $y_{it}$ is regressed on a $K \times 1$ vector of explanatory variables $x_{it}$ with parameter vector $\beta_i$, $P$ lagged values of the dependent variable with parameter vector $\gamma_i = (\gamma_{i1}, \ldots, \gamma_{ip})'$ and a constant term. Note that $x_{it}$ may contain both current and lagged values of explanatory variables, so (1.1) essentially specifies an ADL model for $N$ distinct cross-sectional units.

We assume that all elements of $\beta_i$, $\gamma_i$ and $\eta_i$ are constant through time. Furthermore, the relationship (1.1) between $y_{it}$ and $x_{it}$ is assumed to be dynamically stable. For $P = 1$ this implies $|\gamma_{i1}| < 1$, but in higher-order dynamic models more complicated restrictions on the autoregressive coefficients are required for stability\(^3\). The explanatory variables in $x_{it}$ are assumed to be predetermined, i.e.

$$E[x_{it} \varepsilon_{js}] = 0, \quad \forall i, j, t \leq s. \quad (1.2)$$

The disturbances $\varepsilon_{it}$ are uncorrelated through time, but we allow for heteroscedasticity across individuals\(^4\) and for non-zero contemporaneous cross-correlations, i.e.

$$E[\varepsilon_{it} \varepsilon_{js}] = 0, \quad \forall i, j, t \neq s,$$

$$E[\varepsilon_{it} \varepsilon_{jt}] = \sigma_{ij}, \quad \forall i, j, t. \quad (1.3)$$

\(^3\)A necessary, but not sufficient restriction for stability in higher-order dynamic regression models is $\sum_{p=1}^{P} \gamma_{pi} < 1$.

\(^4\)We will use the terms cross-sectional unit and individual interchangeably.
1.2. Cross-sectional time series data

Stacking the observations over time we get

$$y_t = \sum_{p=1}^{P} \gamma_p y_{t-p} + X_t \beta_t + \eta_t + \epsilon_t,$$  \hspace{1cm} (1.4)

where $y_{t-p} = (y_{t-1-p}, \ldots, y_{t-T-p})'$, $X_t = (x_{t1}, \ldots, x_{tT})'$ and $\epsilon_t = (1, \ldots, 1)'$ a $T \times 1$ vector of ones. This can be written more compactly as

$$y_t = W_t \delta_t + \eta_t + \epsilon_t,$$  \hspace{1cm} (1.5)

where $\delta_t = (\gamma_t', \beta_t')'$ and $W_t = [y_{t-1}; \ldots; y_{t-p}; X_t]$. The assumptions about $\epsilon_t$ can be written as

$$E[\epsilon_t] = 0, \quad \left\{ \begin{array}{l} E[\epsilon_t \epsilon'_t] = \sigma_{ij} I_T. \end{array} \right.$$  \hspace{1cm} (1.6)

In this thesis we will consider various inference methods for system (1.5) with disturbances (1.6). In the case of micro-economic data on households, firms or industries, we may have a large cross-section and only few time observations. In the case of macro-economic data, e.g. cross-country data, the time series dimension may be large as compared with the cross-sectional dimension. Depending on the type of data, e.g. dimensions of the data set and degree of aggregation, we will employ different approaches. We will make use of distinct models, which differ regarding the restrictions imposed on (1.5) and (1.6) as we shall see.

1.2.2 Dynamic panel data models

Regarding the modelling of typical micro-economic panel data, i.e. with $T$ small and $N$ large, often particular restrictions are imposed on model (1.5) and the assumptions (1.6) in order to make inference more tractable. First, it is common to assume stochastic independence between individuals, i.e. $\sigma_{ij} = 0$ for $i \neq j$. For example, in a household panel it may be reasonable to assume that individual household decisions are not influenced by actions of the other households in the population. Second, it is common to impose homogeneity of the reaction coefficients, i.e. $\delta_t = \delta$. As the typical panel data set has only a few time observations, it is infeasible to allow for heterogeneity in the full parameter vector. Hence, modelling of the heterogeneity is limited to the constant term $\eta_t$.

Assuming $\sigma_{ij} = 0 (i \neq j)$ and $\delta_t = \delta$ and stacking the observations across cross-sectional units in (1.5) one gets

$$y = W \delta + u,$$  \hspace{1cm} (1.7)
where $\delta = (\gamma', \beta')', \eta_i = (\eta_1, ..., \eta_N)'$, $u = S \eta + \varepsilon$, $S = I_N \otimes \iota_T$, $y$ and $W$ are $NT \times 1$ and $NT \times (K + P)$ matrices of stacked observations and $\varepsilon$ is the $NT \times 1$ vector of disturbances. Model (1.7) is a higher-order dynamic panel data model and has been used widely in practice.

In panel data models the individual specific effect $\eta_i$ is assumed either fixed or random. In this section (and in Chapter 2 and 3) we will assume that $\eta_i$ is random. In case of random effects, $\eta_i$ has constant mean and finite variance $\sigma^2_{\eta_i}$, is uncorrelated with $\eta_j$ $(j \neq i)$ and not correlated with the general disturbance term $\varepsilon_{it}$.

We will discuss some well-known procedures for estimating the unknown parameter vector $\delta$ in (1.7). For ease of exposition, we will consider a simplified version of model (1.7) first. We assume that $P = 1$ or $W = [y_{-1}:X]$ and that $y_{i0}$, $i = 1, ..., N$, is observed (but not earlier observations). Regarding the assumptions (1.2) and (1.3) we assume strictly exogenous regressors $X$ and a scalar disturbance covariance matrix for $\varepsilon$, i.e. we simplify to $E [X' \varepsilon] = 0$ and $E [\varepsilon \varepsilon'] = \sigma^2_{\varepsilon} I_{NT}$. We will use several transformations of the original data, i.e. we will exploit the following $(T - 1) \times T$ tranformations

$$J_T = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}, \quad J_T^* = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & 1 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

(1.8)

and also $D_T = J_T - J_T^*$. Note that $D_T$ transforms a $T$ element vector for individual $i$ into $T - 1$ first differences, because $J_T$ skips the first observation and $J_T^*$ skips the final observation. We define also $D = I_N \otimes D_T$, $J = I_N \otimes J_T$ and $J^* = I_N \otimes J_T^*$.

LSDV estimator

It is clear from (1.7) and $W = [y_{-1}:X]$ that when the individual specific effects are random, the regressor $y_{i,t-1}$ is correlated with the composite disturbance term $u_{it} = \eta_i + \varepsilon_{it}$. Hence, standard estimators for the random effects model are inconsistent (results on the asymptotic bias can be found in Sevestre and Trognon, 1985). Therefore, one usually avoids to estimate the coefficients $\delta$ by error component techniques, because this leads to many complications (see Sevestre and Trognon, 1996). In dynamic panel data models it is more common to assume that the individual effects are fixed, or otherwise to treat genuinely random individual effects as fixed anyhow. In that way we deal in a rather straightforward manner with their likely non-zero correlation with the regressors $W$. 

Chapter 1. Introduction
1.2. **Cross-sectional time series data**

If the individual specific constants are considered fixed, then the $N$ components of the vector $\eta = (\eta_1, ..., \eta_N)'$ constitute $N$ unknown parameters corresponding to the dummy variables $I_N \otimes \iota_T$ in (1.7). Note that their estimation will lead to a considerable loss of degrees of freedom, especially when $N$ is large. Estimation of the $N + K + 1$ coefficients of (1.7) by ordinary least squares yields estimates which are called Least Squares Dummy Variables (LSDV) or fixed effect estimates. Using standard partitioned regression results, the resulting estimator for $\delta$ can be expressed as

$$
\hat{\delta}_{LSDV} = (W'AW)^{-1}W'Ay
$$

where the $NT \times NT$ matrix $A$ is equal to

$$
A = I_{NT} - S(S'S)^{-1}S'

= I_N \otimes (I_T - \frac{1}{T}I_T'\iota_T')

= I_N \otimes A_T.
$$

Note that $A$ is the within transformation which wipes out the individual effects. For ease of exposition it is assumed that all the explanatory variables are time variant so that $W'AW$ is invertible.

The LSDV estimator (1.9) can be written also as

$$
\hat{\delta}_{LSDV} = (W^{*}W^{*})^{-1}W^{*}y^{*},
$$

where $y^{*} = Py$, $W^{*} = PW$ and $P = I_N \otimes P_T$ is the forward orthogonal deviations operator (Arellano and Bover, 1995). This transformation will prove to be useful when constructing and analysing method of moment (MM) estimators. The $(T - 1) \times T$ upper-triangular matrix $P_T$ transforms as follows, i.e.

$$
y^{*}_t = c_t \left[ y_t - \frac{1}{T-t}(y_{t+1} + ... + y_T) \right]
$$

where $c^*_t = (T - t)/(T - t + 1)$. Regarding $P_T$ one can deduce that $P_TP_T' = I_{T-1}$ and $P_T'P_T = A_T$ (Arellano and Bover, 1995). Also $P_T = (D_TD_T')^{-\frac{1}{2}}D_T$ where $D_T$ is the first difference operator as defined above. To see this, note that the columns of $D_T$ span the orthogonal complement of $\iota_T$. Hence, projection on the orthogonal complement of $\iota_T$ is equal to projection on $D_T'$ or

$$
A_T = I_T - \iota_T(\iota_T'\iota_T)^{-1}\iota_T' = D_T(D_T'\iota_T')^{-1}D_T.
$$
This implies that $A_T = P_T'P_T$ with $P_T = (D_TD_T')^{-\frac{1}{2}}D_T$ and also $P_TP_T' = I_{T-1}$. It can be shown (Arellano and Bover, 1995) that when an upper-triangular matrix for $(D_TD_T')^{-\frac{1}{2}}$ has been chosen, the transformation (1.12) follows. If applied to IID disturbances, the transformation $P_T$ preserves the orthogonality between the transformed disturbances, hence it is referred to as orthogonal deviations.

Although strict exogeneity of the regressors $X$ implies $E(X'A\varepsilon) = 0$, we have $E(W'A\varepsilon) \neq 0$, since

$$E[y_{-1}'A\varepsilon] = \sum_{i=1}^{N} \sum_{t=1}^{T} E\left[\left(y_{i,t-1} - \frac{1}{T} \sum_{s=1}^{T} y_{i,s-1}\right) \varepsilon_{it}\right]$$

$$= -N \sum_{i=1}^{T} E\left[\varepsilon_{it} \sum_{s=t+1}^{T} y_{i,s-1}\right]$$

$$= -\sigma_{\varepsilon}^2 N \frac{1}{1 - \gamma} \left(1 - \frac{1 - \gamma^T}{T(1 - \gamma)}\right). \tag{1.14}$$

Therefore, the LSDV estimator of $\delta$ is consistent for $T \to \infty$ but inconsistent for $N \to \infty$ and $T$ finite, see Nickell (1981).

IV estimation

A different transformation of (1.7) for removing the individual specific effects is first differencing. Noting that $W = [y_{-1}:X]$ we can write

$$Dy = \gamma Dy_{-1} + DX\beta + D\varepsilon, \tag{1.15}$$

where the first difference operator $D$ is as defined above. Since

$$E(y_{-1}'D'D\varepsilon) = -E(y_{-1}'\varepsilon_{-1}) = -E(\varepsilon_{-1}'\varepsilon_{-1}) = -\sigma_{\varepsilon}^2 N(T - 1), \tag{1.16}$$

ordinary least squares is inconsistent now, irrespective of how the sample size is extended. For this situation Anderson and Hsiao (1982) proposed two simple instrumental variables (IV) estimators. As instrument for $Dy_{-1}$ they suggest either to use the two-period lagged level $J^*y_{-1}$ (we will label this implementation AHL) or the two-period lagged first difference $Dy_{-2}$. The latter requires the omission of another initial observation in estimation. The instruments $J^*y_{-1}$ and $Dy_{-2}$ are uncorrelated with the disturbance term $D\varepsilon$ and correlated with $Dy_{-1}$, so they are valid. Furthermore, in case of strict exogeneity of $X$ we may use $DX$ itself as instrument for $DX$.

Considering the equation in levels (1.7), several authors (Arellano and Bover, 1995; Kiviet, 1995; Blundell and Bond, 1998) have noticed that lagged first differences, i.e.
1.2. Cross-sectional time series data

$Dy_{-1}$ and $DX$, are valid instruments (we will label this estimator IVld). Note that by constructing these instruments the first time period is lost, i.e. the levels equation (1.7) becomes

$$Jy = JW\delta + Ju,$$

where $J$ skips the first time observation of each individual time series.

In general, we can express simple IV estimators of $\delta$ in either (1.15) or (1.17) as

$$\hat{\delta}_{IV} = (Z'W)^{-1}Z'y,$$

where $Z$ is the $N(T-1) \times (K+1)$ matrix of instruments, $W$ the $N(T-1) \times (K+1)$ matrix of (transformed) regressors ($DW$ or $JW$) and the $N(T-1) \times 1$ vector $y$ the (transformed) dependent variable ($Dy$ or $Jy$).

Many more implementations of IV are possible. For example, one could use combinations of instruments in first differences and levels or exploit instruments in orthogonal deviations. Although IV estimators, unlike $\hat{\delta}_{LSDV}$, are consistent for $N \to \infty$ and finite $T$, they are asymptotically inefficient because they do not exploit all available moment conditions. Also IV estimators as in (1.18) do not take into account that the disturbance terms in (1.15) and (1.17), i.e. $De$ and $Ju$ respectively, do not have a scalar covariance matrix. A unifying approach that copes with both the correlation structure of the disturbances and the exploitation of any further instruments is Generalised Method of Moments (Hansen, 1982).

GMM estimation

The assumptions on the stochastic part of model (1.7) imply for individual $i$ a set of $m$ linear moment conditions\(^5\) embodied in the $(T - 1) \times m$ matrix $Z_{it}$. In dynamic panel data models, typically the columns of $Z_{it}$ contain lagged values of the dependent variable as instruments in addition to instruments concerning the $x$'s. For example, the GMM estimator used in the simulation study and empirical application of Arellano and Bond (1991) exploits

$$Z_{it} = \begin{bmatrix} y_{i,0} & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \Delta x'_{i,2} \\ 0 & y_{i,0} & y_{i,1} & 0 & 0 & 0 & \cdots & 0 & \Delta x'_{i,3} \\ 0 & 0 & 0 & y_{i,0} & y_{i,1} & y_{i,2} & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & y_{i,0} & \cdots & y_{i,T-2} & \Delta x'_{i,T} \end{bmatrix}, \quad (1.19)$$

\(^5\)For ease of exposition, we postpone until Chapter 2 a more detailed description of the assumptions made on the stochastic part of model (1.7).
where $\Delta$ means first difference. Note that they do not use all available overidentifying restrictions arising from the assumed strict exogeneity of the $x$'s. The number of instruments used in (1.19) is $m = \frac{1}{2} T(T - 1) + K$.

In general, the set of moment conditions for individual $i$ can be expressed as (Arellano and Honoré, 2000)

$$E[Z'_{i}B_{T}u_{i}] = E[Z'_{i}B_{T}(\eta_{Ti} + \varepsilon_{i})] = E[Z'_{i}B_{T}\varepsilon_{i}] = 0,$$  \hspace{1cm} (1.20)

where $B_{T}$ is any $(T - 1) \times T$ upper-triangular matrix with rank $(T - 1)$ and $B_{TuT} = 0$. Summing over individuals one gets

$$E[Z'_{i}B_{i}] = E \left[ \sum_{i=1}^{N} Z'_{i}B_{T}\varepsilon_{i} \right] = 0,$$  \hspace{1cm} (1.21)

where $Z_{i} = (Z'_{i1}, ..., Z'_{iN})'$ and $B = I_{N} \otimes B_{T}$. Arellano and Honoré (2000) mention two possibilities for $B$, i.e. either the first difference operator $D$ or the orthogonal deviations transformation $P$.

The GMM estimator of $\delta$ is based on the sample moments $\frac{1}{N} \sum_{i=1}^{N} Z'_{i}B_{T}\varepsilon_{i} = 0$ and is obtained as

$$\hat{\delta}_{GMM} = \arg \min_{\delta} \varepsilon'Z_{i}GZ'_{i}B_{i}\varepsilon,$$  \hspace{1cm} (1.22)

where $G$ is a weighting matrix. The optimal choice for $G$ in (1.22) is to make it proportional to the inverse of $V = E[Z'_{i}B_{i}\varepsilon'Z'_{i}B_{i}]$, which is the covariance matrix of $Z'_{i}B_{i}\varepsilon$. The estimator in (1.22) is equivalent with the generalised least squares (GLS) estimator of $\delta$ in

$$Z'_{i}B_{i}y = Z'_{i}BW\delta + Z'_{i}B_{i}\varepsilon.$$  \hspace{1cm} (1.23)

Note that the covariance matrix of the transformed disturbance term $Z'_{i}B_{i}\varepsilon$ in (1.23) is equal to $V$. Then the optimal GMM estimator equals

$$\hat{\delta}_{OPT} = (W'B'Z_{i}\hat{V}^{-1}Z'_{i}BW)^{-1}W'Z_{i}\hat{V}^{-1}Z'_{i}B_{i}y,$$  \hspace{1cm} (1.24)

which is a two-step estimator because a preliminary consistent estimate for $\hat{V}$ is needed.

An estimate of the inverse of the optimal weighting matrix is

$$\hat{V} = \frac{1}{N} \sum_{i=1}^{N} Z'_{i}B_{T}\hat{\varepsilon}_{i}B'_{T}Z_{i},$$
1.2. Cross-sectional time series data

where $\hat{\varepsilon}_t$ are the residuals of an initial consistent estimator. For example, a particular one-step GMM estimator based on $G^{-1} = \sigma^2\varepsilon_i B B' Z_i$, i.e. the theoretical covariance matrix of $Z_i' B \varepsilon$ in case of IID disturbances, can be written as

$$
\hat{\delta}_{GMM} = (W'B'Z_i (Z_i' B B' Z_i)^{-1} Z_i' B W)^{-1} W'B'Z_i (Z_i' B B' Z_i)^{-1} Z_i' B y.
$$

(1.25)

It can be shown that under the IID assumption the one-step estimator in (1.25) is asymptotically efficient.

Arellano and Honoré (2000) mention two possibilities for $B$, i.e. either the first difference operator $D$ or the orthogonal deviations transformation $P$. When $B = D$ the estimator in (1.25) specialises to the one-step estimator of Arellano and Bond (1991). When $B = P$ the resulting estimator is similar to the GMM estimator analysed in Alvarez and Arellano (1998). It can be shown that when all available linear moment conditions are used in estimation both estimators are equivalent. However, in many situations it may be wise not to use all moment conditions available as otherwise finite sample properties may deteriorate, as we shall see.

Note that the moment conditions used so far have been based on a transformation $B$ of the original level equation (1.7), which eliminates the individual specific effects, and therefore lagged levels are valid instruments under specific assumptions regarding the error components and initial conditions. Under additional assumptions regarding mean stationarity of the processes for $y_{it}$ and the elements in $x_{it}$, also first differences are valid instruments for the levels equation (1.17) (Arellano and Bover, 1995; Blundell and Bond, 1998). More in particular, the set of moment conditions for individual $i$ can be expressed as (Blundell et al., 2000)

$$
E[Z_{di}' J_T u_i] = 0.
$$

(1.26)

For example, analogue to (1.19) we may use

$$
Z_{di} = \begin{bmatrix}
\Delta y_{i,1} & 0 & 0 & \cdots & 0 & \cdots & 0 & \Delta x_{i,2} \\
0 & \Delta y_{i,1} & \Delta y_{i,2} & 0 & \cdots & 0 & \Delta x_{i,3} \\
0 & 0 & \Delta y_{i,1} & \Delta y_{i,2} & 0 & \cdots & 0 & \Delta x_{i,3} \\
& \vdots & & & & & & \\
0 & \cdots & \Delta y_{i,1} & \cdots & \Delta y_{i,T-1} & \cdots & \Delta x_{i,T} & \\
\end{bmatrix}.
$$

(1.27)

Again note that not all available overidentifying restrictions have been used in (1.27) in case of strict exogeneity of the $x$'s. The construction of a GMM estimator using the

---

6This does not hold, for example, for $Z_{di}$ as defined in (1.19).
moment conditions (1.26) is similar as before. An optimal combination of instruments in levels and first differences has been exploited by Blundell and Bond (1998). Note that by taking such a combination many moment conditions become redundant. For more details, see Blundell and Bond (1998) and Blundell et al. (2000).

Other estimators

The IV and GMM estimators considered so far exploit linear moment conditions only. Ahn and Schmidt (1995) consider also GMM inference based on additional non-linear moment conditions, which may lead to asymptotically more efficient estimators. Another approach closely related to IV estimation is Limited Information Maximum Likelihood (LIML). Note that here "LIML" refers to the instrumental variables interpretation, hence this estimator cannot be seen as mere maximum likelihood. The asymptotic distribution of the LIML estimator has been considered in Alvarez and Arellano (1998) for the first-order normal dynamic panel data model without any other explanatory variables. They find that in large $T$, large $N$ panels the LIML estimator is asymptotically Normally distributed with a variance equal to LSDV and to the GMM estimator (1.25). Alonso-Borrego and Arellano (1999) compare both by Monte Carlo experiments and in an empirical illustration the LIML estimator with symmetrically normalised GMM estimators and find that the differences in behaviour are small.

Assuming normality for the disturbances an obvious estimation method is Maximum Likelihood (ML). For random effect models, Anderson and Hsiao (1981, 1982) show that consistency of the ML estimator for finite $T$ and infinite $N$ depends crucially on the assumption for the initial conditions (see also Barghava and Sargan, 1983). Blundell and Smith (1991) and Blundell and Bond (1998) analyse conditional ML estimators, which are consistent for finite $T$ and large $N$. For ML in dynamic panel data models with fixed effects, see Hsiao et al. (1999a).

Numerous other estimators have been proposed for the dynamic panel data model, and therefore the techniques described so far are not exhaustive. See, for example, Baltagi (1995) and Mátýás and Sevestre (1996) for a broader overview of the existing techniques for the dynamic panel data model.

1.2.3 System of regression equations

The inference techniques discussed in the previous subsection other than LSDV have been designed for the case of finite $T$, large $N$ panel data. However, many available cross-section time series data, e.g. macro-economic applications, may have different characteristics.
First, the dimensions of the data may differ substantially from the small \( T \), large \( N \) panel. Often the number of cross-sectional units, i.e. industries or regions, is finite and the time span is several decades. Second, in case of regional data, e.g. cross-country data, the assumption of independence between cross-sectional units may not be valid as regional economic structures may be interrelated. Third, it is questionable whether the parameters of economic relationships are stable across cross-sectional units. In other words, there may be heterogeneity in the slope parameters in addition to the variability in the constant term\(^7\).

For these reasons it is doubtful whether inference techniques developed for the small \( T \), large \( N \) panel are appropriate in this case. Instead, inference methods, which are consistent for large \( T \), may be more suitable. We will discuss these techniques making various assumptions about the unknown parameter vector \( \delta_i \) in (1.5), i.e. whether it is equal across cross-sectional units or not. In the latter case we will distinguish two cases similar to the random and fixed effects in panel data models, i.e. randomly or deterministically varying coefficient models. For ease of exposition, we continue to assume strict exogeneity of \( X_i \), but allow for higher-order dynamic models \( (P \geq 1) \) with a non-scalar disturbance covariance matrix as in (1.6).

**Panel data model**

Model (1.5) and (1.6) can be seen as a system of \( N \) regression equations, in which the models for different cross-sectional units are interrelated via their disturbance terms. As already mentioned above when \( \delta_i = \delta \) and \( \sigma_{ij} = 0 \) \((i \neq j)\) the higher-order dynamic panel data model (1.7) with fixed individual effects results. In contrast with the previous section, we now consider inference techniques for \( \delta \) which are consistent for \( T \) large and \( N \) finite. Hence, the LSDV estimator (1.9) is a candidate as \( \operatorname{plim}_{T \to \infty} \frac{1}{T} W' A \epsilon = 0 \). Note that we assume that the covariance matrix of the disturbance term \( \epsilon \) is non-scalar, i.e. we assume cross-sectional heteroscedasticity and non-zero contemporaneous cross-correlations. The ordinary LSDV estimator in (1.9) does not take this particular covariance structure into account. An obvious alternative is the feasible generalised LSDV estimator of \( \delta \), denoted by \( \hat{\delta}_{\text{FGLSDV}} \), which can be expressed as

\[
\hat{\delta}_{\text{FGLSDV}} = (W' A \Omega^{-1} A W)^{-1} W' A \Omega^{-1} A y, \tag{1.28}
\]

\(^7\)Strictly spoken this may hold also for the finite \( T \), large \( N \) case, but there it is often infeasible to test for slope heterogeneity.
where $\Omega = \Sigma \otimes I_T$ with $\Sigma$ a $N \times N$ matrix with typical element $\sigma_{ij}$. The matrix $\Omega$ is the covariance matrix of $\varepsilon$ and is consistently estimated using the LSDV residuals, i.e.

$$\hat{\Omega} = \hat{\Sigma} \otimes I_T,$$

$$\hat{\sigma}_{ij} = \frac{(y_i - W_i\hat{\delta}_{LSDV})'(y_j - W_j\hat{\delta}_{LSDV})}{T}.$$  

(1.29)

(1.30)

Under standard regularity conditions it can be shown that when $T \to \infty$ the FGLSDV estimator is asymptotically efficient.

**Deterministically varying coefficients**

When $\delta_i$ varies across individuals we cannot use panel data techniques, but have to estimate model (1.5) separately for each cross-sectional unit. Note that model (1.5) together with the assumptions in (1.6) forms a generalised regression model and is a special case of the Seemingly Unrelated Regression (SUR) model (Zellner, 1962), since all the $\delta_i$'s have the same dimension here. Hence, ordinary least squares estimation of each equation will deliver consistent estimates for $T \to \infty$ and $N$ finite. In addition, feasible generalised least squares will be asymptotically efficient whenever a preliminary consistent estimator for the disturbance covariance matrix has been used.

**Randomly varying coefficients**

There are several intermediate ways between the fixed effects panel data model and SUR. For example, Pesaran et al. (1999) allow for heterogeneity in the short-run coefficients but impose homogeneity of long-run parameters. Another approach is assuming that the coefficients $\delta_i$ and $\eta_i$ in (1.5) are varying randomly across cross-sectional units. In Section 1.2.2 we already discussed a special case of such a model, viz. the random effects panel data model where $\eta_i$ is random but $\delta_i$ is constant across individuals.

Defining $\alpha_i = (\delta_i', \eta_i)'$ we specify

$$\alpha_i = \bar{\alpha} + \mu_i,$$

(1.31)

for $i = 1, ..., N$ with

$$\begin{cases} 
E[\mu_i] = 0, \\
E[\mu_i\mu_i'] = \Theta. 
\end{cases}$$

(1.32)

Furthermore, the elements of $\alpha_i$ are assumed to be uncorrelated with the disturbance term $\varepsilon_i$ and the regressors in $W_i$. Equations (1.5), (1.31) and (1.32) constitute the higher-order dynamic random coefficient model. The parameters of interest in random coefficient
1.3. Empirical findings

models are the first and second moments of the distribution of $\alpha_i$, i.e. inference is made about the mean parameter vector $\bar{\alpha}$ and its covariance matrix $\Theta$.

Depending on the nature of the elements of $W_i$ we can use estimation techniques using pooled data or not. If no lagged dependent variable regressors are contained in $W_i$, pooling the data and performing ordinary least squares (OLS) will give at least a consistent estimator for $\bar{\alpha}$. Moreover, asymptotically efficient inference on $\bar{\alpha}$ can be obtained with the techniques discussed in e.g. Swamy (1971). When the elements of $W_i$ do contain lagged dependent variable regressors, however, it can be shown that pooled OLS is inconsistent. Moreover, Pesaran and Smith (1995) show that panel data estimators like LSDV are inconsistent also. They propose the mean group (MG) estimator\footnote{Regarding the random coefficient model alternative so-called shrinkage estimators have been proposed by Maddala et al. (1997).}, which is essentially the average of the individual OLS estimators. Using partitioned regression results the mean group estimator for $\delta$ can be written as

$$\delta_{MG} = \frac{1}{N} \sum_{i=1}^{N} \delta_i,$$

(1.33)

with

$$\delta_i = (W'_i A_T W_i)^{-1} W'_i A_T y_i.$$

(1.34)

Assuming cross-sectional independence, i.e. $\sigma_{ij} = 0$ for $i \neq j$ in (1.6), it can be shown that (1.33) is a consistent estimator of $\bar{\delta}$ when both $T$ and $N$ go to infinity. Moreover, in Hsiao et al. (1999b) it is shown that $\delta_{MG}$ has an approximate Normal distribution and that this holds for various rates of $T$ and $N$ growing large.

1.3 Empirical findings

In this section we present some preliminary estimation results to point out the potential problems related to the estimation of dynamic models for panel data. We understand that it is not possible to make judgements about the quality of estimators simply from empirical outcomes. However, the wide range of estimates indicates the mutual differences and potential inaccuracies of some of the estimation methods employed. By the findings in the remainder of the thesis we will be able to explain some of the estimation results and differences in this section. Furthermore, we will gain insight into the quality of the various estimation methods employed here.
We will make use of two sources of data, which differ in their dimensions and in level of aggregation. The first data set, which is a typical small $T$, large $N$ panel, has been used also in Bun and El Makhloufi (2001). Recent growth theory emphasises the importance of knowledge spillovers between firms for economic development. Close proximity of firms in a specific area facilitates the circulation or transmission of ideas and innovations between firms. We analyse whether these knowledge spillovers, which are also called dynamic externalities, come from increased specialisation of similar firms within a region or from increased diversity of firms of different industries. Also the effects of local competition will be quantified. Using annual panel data on six major Moroccan urban areas and 18 industrial sectors for the period 1985-1995 an attempt will be made to distinguish which type of externality is predominant for local economic activity in Morocco.

Table 1.1 presents estimation results for some selected estimation techniques, i.e. LSDV, two simple IV estimators (AHL and IV id as described in Section 1.2.2) and two one-step GMM estimators (GMM$_{df}$ and GMM$_{lev}$). GMM$_{df}$ has been based on the equation in first differences (1.15) and use the instrument matrix (1.19), while GMM$_{lev}$ exploits instruments in first differences (1.27) for the levels equation (1.17). The dependent variable is real value added (va) and explanatory variables are regional manufacturing production (trp), real unit wage costs (wcap) and three indicators$^9$ measuring specialisation (sp), diversity (dv) and competition (cp) of firms within a region. A general dynamic specification with fixed effects has been estimated including as many lagged explanatory variables as seem required, including the one-period lagged value of the dependent variable.

The estimation results in Table 1.1 vary widely across estimation techniques. At first sight, the inconsistency of LSDV for finite $T$ seems evident from the relatively low value for the autoregressive parameter. However, the semi-consistent$^{10}$ GMM$_{df}$ estimator shows a similar picture as LSDV, which seems counterintuitive. Regarding estimated standard errors, those of the simple IV estimators are often larger than for GMM and LSDV. In the Chapters 2 and 3 we will give explanations for these features by focusing on both finite sample bias of coefficient estimators and asymptotic variance comparisons. Also we shall try to deal with possible slope heterogeneity by analysing panels with both dimensions relatively small, which may avoid structural breaks in the slope parameters. We will split the available data into parts corresponding to industrial sectors, which reduces the value of $N$ considerably, and examine whether dynamic externalities are present in each sector.

The second data set is from Vlaar and Schuberth (1998), which is an updated version

$^9$More details about the construction of these indicators can be found in Section 2.6.2.

$^{10}$With semi-consistent we mean consistent for finite $T$ and $N \to \infty$. 
1.3. **Empirical findings**

Table 1.1: Estimation results of dynamic externalities

<table>
<thead>
<tr>
<th></th>
<th>$LSDV$</th>
<th>$AHL$</th>
<th>$IVId$</th>
<th>$GMM_{du}$</th>
<th>$GMM_{iev}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln va_{t-1}$</td>
<td>0.38</td>
<td>0.86</td>
<td>0.68</td>
<td>0.36</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.25)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>$\ln trp_{it}$</td>
<td>0.45</td>
<td>0.41</td>
<td>0.60</td>
<td>0.52</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.19)</td>
<td>(0.26)</td>
<td>(0.14)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>$\ln trp_{i,t-1}$</td>
<td>-0.19</td>
<td>-0.50</td>
<td>-0.43</td>
<td>-0.22</td>
<td>-0.75</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.21)</td>
<td>(0.28)</td>
<td>(0.13)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>$\ln wcap_{it}$</td>
<td>0.73</td>
<td>0.61</td>
<td>0.77</td>
<td>0.68</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.04)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>$\ln wcap_{i,t-1}$</td>
<td>-0.22</td>
<td>-0.58</td>
<td>-0.38</td>
<td>-0.20</td>
<td>-0.53</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.20)</td>
<td>(0.07)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>$sp_{it}$</td>
<td>0.45</td>
<td>0.35</td>
<td>0.32</td>
<td>0.40</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>$sp_{i,t-1}$</td>
<td>-0.17</td>
<td>-0.34</td>
<td>-0.35</td>
<td>-0.15</td>
<td>-0.34</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.12)</td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>$dv_{it}$</td>
<td>-0.08</td>
<td>0.06</td>
<td>-0.17</td>
<td>-0.03</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.14)</td>
<td>(0.09)</td>
<td>(0.11)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>$dv_{i,t-1}$</td>
<td>-0.12</td>
<td>-0.19</td>
<td>-0.21</td>
<td>-0.18</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.13)</td>
<td>(0.19)</td>
<td>(0.11)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>$cp_{it}$</td>
<td>-0.19</td>
<td>-0.18</td>
<td>-0.24</td>
<td>-0.18</td>
<td>-0.21</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.04)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>$cp_{i,t-1}$</td>
<td>0.07</td>
<td>0.16</td>
<td>0.11</td>
<td>0.07</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.05)</td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
</tbody>
</table>

Dimensions are $T=11$, $N=95$. Figures in parentheses are standard errors.

of the data used in Fase and Winder (1998). Both studies analyse money demand in the European Union (EU) using aggregated time series. The data, which can be classified as a large $T$, small $N$ panel, are quarterly observations for the period 1970:1-1996:IV on several monetary aggregates, income, prices and interest rates for 14 EU countries. In analysing EU wide money demand most empirical studies aggregate the data across countries and use time series techniques on the aggregated data. However, considering the EU countries as a cross-section one can possibly use panel data techniques. We illustrate the different approaches by estimating a dynamic specification for the demand for real narrow money ($M1/P$). Next to the one-period lagged dependent variable regressor the explanatory variables are current and one-period lagged values of real income ($gnp$), short- ($rs$) and long-term ($rl$) interest rates and the inflation rate ($ir$).
Table 1.2 shows estimation results\(^{11}\) for the period analysed in Vlaar and Schuberth, i.e. 1979-1996 and a more recent period, i.e. 1991-1996. The entries show LSDV and MG estimates assuming equal and randomly varying slope vectors across countries respectively. The third estimator is OLS using aggregated time series (AGGR). One time observation is lost due to construction of lagged regressors, hence the estimation periods are 1979:II-1996:IV and 1991:II-1996:IV.

Regarding the period 1979-1996 the pattern of the estimation results suggests the theoretical findings of Pesaran and Smith (1995). They show that panel data estimators (e.g. the LSDV estimator) are inconsistent in the first-order dynamic panel data model with heterogenous slope coefficients and a first-order autoregressive process for the exogenous regressor. More in particular, they prove that the LSDV estimator of the coefficient of the lagged dependent variable regressor is biased towards unity, while the coefficient of the exogenous regressor tends to zero. Inspecting the LSDV estimates this picture is reasserted, i.e. the coefficient of the lagged dependent variable regressor is close to unity and the sum of the coefficients of each explanatory variable tends to zero. On the contrary, the estimates of the MG estimator are more plausible. Regarding the least squares estimator for aggregated time series (AGGR), Pesaran and Smith show that this estimator is inconsistent as well. Table 1.2 shows that the dynamics implied by the AGGR estimates are rather different from the MG estimates.

In the case of estimating EU money demand one may criticise the use of long time series for several reasons. First, some major occurrences in the period 1970-1996, e.g. oil crises and the reunification of West and East Germany, may have caused structural breaks over time in the money demand function. Hence, using long time series and assuming parameter constancy over time may lead to serious misspecification affecting all three approaches. Second, the process of convergence to one common monetary policy has started only recently for the complete group of EU countries. Hence, it is doubtful whether one can estimate one common money demand relationship for all EU countries using data from a period characterised by heterogeneous monetary policies across countries. For this reason one may avoid using the AGGR and MG estimators as their accuracy critically depends on a relatively large number of time observations. Although the MG estimator accounts for heterogeneity across countries, it nevertheless may not be accurate in this case because it needs at least a moderately large cross-section too.

In contrast, panel data models assuming constant slope parameters across cross-sectional units and over a relatively short time span may offer an alternative approach. In this case the cross-sectional dimension in the data implies extra observations to estimate

\(^{11}\)We exclude Greece from the analysis as no data on long-term interest rates are available.
1.3. Empirical findings

Table 1.2: Estimation results of the demand for $M1$

<table>
<thead>
<tr>
<th></th>
<th>LSDV</th>
<th>MG</th>
<th>AGGR</th>
<th>LSDV</th>
<th>MG</th>
<th>AGGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln(M1/P)_{i,t-1}$</td>
<td>0.95 (0.01)</td>
<td>0.81 (0.04)</td>
<td>0.74 (0.08)</td>
<td>0.79 (0.05)</td>
<td>0.60 (0.07)</td>
<td>0.68 (0.18)</td>
</tr>
<tr>
<td>$\ln gnp_{it}$</td>
<td>0.11 (0.02)</td>
<td>0.14 (0.10)</td>
<td>0.11 (0.09)</td>
<td>0.21 (0.04)</td>
<td>0.10 (0.17)</td>
<td>-0.47 (0.14)</td>
</tr>
<tr>
<td>$\ln gnp_{i,t-1}$</td>
<td>-0.06 (0.02)</td>
<td>-0.04 (0.06)</td>
<td>0.14 (0.09)</td>
<td>0.01 (0.04)</td>
<td>0.11 (0.15)</td>
<td>0.57 (0.14)</td>
</tr>
<tr>
<td>$rs_{it}$</td>
<td>-0.35 (0.12)</td>
<td>-0.32 (0.10)</td>
<td>0.08 (0.21)</td>
<td>-0.53 (0.20)</td>
<td>-0.54 (0.38)</td>
<td>0.07 (0.30)</td>
</tr>
<tr>
<td>$rs_{i,t-1}$</td>
<td>0.18 (0.11)</td>
<td>0.04 (0.09)</td>
<td>-0.20 (0.21)</td>
<td>0.29 (0.20)</td>
<td>0.18 (0.28)</td>
<td>-0.57 (0.29)</td>
</tr>
<tr>
<td>$rl_{it}$</td>
<td>-0.36 (0.20)</td>
<td>-0.47 (0.16)</td>
<td>-0.70 (0.29)</td>
<td>-0.58 (0.34)</td>
<td>-0.94 (0.52)</td>
<td>-0.71 (0.40)</td>
</tr>
<tr>
<td>$rl_{i,t-1}$</td>
<td>0.27 (0.20)</td>
<td>0.23 (0.13)</td>
<td>0.29 (0.31)</td>
<td>0.31 (0.37)</td>
<td>0.09 (0.31)</td>
<td>0.01 (0.47)</td>
</tr>
<tr>
<td>$ir_{it}$</td>
<td>-0.40 (0.09)</td>
<td>-0.44 (0.07)</td>
<td>-0.41 (0.23)</td>
<td>-0.38 (0.16)</td>
<td>-0.34 (0.23)</td>
<td>-0.65 (0.35)</td>
</tr>
<tr>
<td>$ir_{i,t-1}$</td>
<td>0.42 (0.10)</td>
<td>0.35 (0.07)</td>
<td>0.47 (0.24)</td>
<td>0.34 (0.15)</td>
<td>0.32 (0.27)</td>
<td>1.19 (0.40)</td>
</tr>
</tbody>
</table>

In the left panel $N=13$ and $T=71$, in the right panel $N=13$ and $T=23$. Figures in parentheses are standard errors.

The right hand panel of Table 1.2 gives the estimation results for the period 1991:II-1996:IV. For this short time period the MG and AGGR estimators suffer from a lack of degrees of freedom problem, while the LSDV estimators seem to give satisfactory results. The pattern of the estimates does not indicate signs of aggregation bias and the estimated standard errors show large efficiency gains. Of course, in finite samples the LSDV estimator will be biased in dynamic models like the AGGR and MG estimators are, and
so are the estimated standard errors. Also one may want to test formally on parameter homogeneity, i.e. common slopes across countries. But is the number of time observations \((T = 23)\) sufficient to have confidence in techniques which are only asymptotically valid? These and other topics will be analysed in the final two chapters of the thesis.

### 1.4 Outline

From the two empirical examples introduced above we conclude that the analysis of panel data where both the time series and cross-sectional dimensions are limited is important and cumbersome. We will examine and try to enhance the accuracy of inference techniques for some empirically relevant models with both \(T\) and \(N\) small. We will examine in isolation the effects on estimators and test procedures of several departures from the standard panel data model with individual specific effects, strictly exogenous regressors and a scalar disturbance covariance matrix. In developing more accurate inference procedures in the subsequent chapters we will make use of two well-known mathematical and statistical concepts, i.e. asymptotic expansion techniques and bootstrap procedures. The former is an analytical tool, while the latter is, when applied, a computer intensive empirical method. Both approaches will be used to develop higher-order asymptotic approximations of finite sample characteristics of inference techniques for dynamic panel data models.

The inclusion of lagged dependent variable regressors in the model complicates estimation of the unknown parameters when \(T\) is small. The purpose of Chapters 2 and 3 is to produce further insights into the finite sample properties of various existing inference techniques for dynamic panel data models. The workhorse model in these chapters will be the first-order stable dynamic panel data model with a scalar disturbance covariance matrix. In Chapter 2 apart from the lagged dependent variable regressor all other explanatory variables will be assumed strongly exogenous, whereas in Chapter 3 these may be weakly exogenous.

In Chapter 2 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. It is difficult to decide from which perspective, i.e. large \(T\) and/or large \(N\), we should analyse panel data when both dimensions are in fact small. Recent simulation results of Judson and Owen (1999) indicate that especially in samples with smaller values for \(N\) the semi-consistent IV and GMM techniques often perform rather poorly as far as the bias and efficiency of coefficient estimators is concerned. We will
1.4. Outline

compare these techniques with various possible implementations of the bias corrected LSDV estimator (LSDVc) as proposed by Kiviet (1995, 1999). The construction of an appropriate estimator for the standard error of the LSDVc estimator proves to be not trivial for the small T, large N case. We examine whether it is possible to exploit such an estimator in tests on coefficient values.

In Chapter 3 we continue to focus on bias correction of LSDV and other estimators in empirically more relevant models. We will 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 the presence of instantaneous feedback mechanisms, i.e. endogenous regressors, IV and GMM estimators seem to have a natural advantage over least squares based procedures. 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 panel data estimators. It is shown that, although less biased, the efficiency losses of simple IV estimators (using as many instruments as regressors) are substantial as compared with LSDV and GMM. Regarding GMM we examine the effects on finite sample bias of both the number and type of instruments used in estimation. Although the bias corrected LSDV estimator of Kiviet (1995, 1999) has been developed for strongly exogenous explanatory variables, the impact of weakly exogenous regressors on its finite sample properties will be examined by simulation.

We will use the results of Chapters 2 and 3 to analyse the determinants driving local economic activity in Moroccan cities exploiting the data described before. In contrast with Bun and El Makhloufi (2001) we will analyse sectors in isolation to detect any differences. As both dimensions of the subsamples used in estimation are very small, the results of Chapters 2 and 3 may explain the patterns of the estimation results and provide guidelines for the appropriate estimation methods to use.

In the final two chapters the focus is on panel data models with the number of cross-section units N relatively small in comparison to T. The simulations of Chapters 2 and 3 show that in the first-order dynamic panel data model with a scalar covariance matrix the bias of least squares based techniques is relatively small compared to instrumental variables based methods when T is larger than N. Based on a mean squared error criterion, least squares methods are to be preferred in this case.

In Chapter 4 we continue to focus on bias correction of the LSDV estimator, but now in higher-order dynamic panel data models with a non-scalar disturbance covariance structure. Cross-sectional heteroscedasticity and contemporaneous interdependencies between cross-sectional units may be an important empirical feature in macro-economic panel data. For example, with cross-country data domestic economic performance is often re-
lated to foreign economic developments, so disturbance terms of different cross-sectional units may be correlated. As the covariance matrix is non-scalar now, also a generalised LSDV estimator will be analysed. Using similar asymptotic expansion techniques as in Kiviet (1995, 1999) bias approximations will be derived for this estimator. Attention will be given to estimation of both impact and long-run multipliers. Also bias approximation formulae for restricted estimators will be developed along the lines of Kiviet and Phillips (1994). In the type of model analysed in this study, the estimation of the variances of the coefficient estimators by conventional asymptotic expressions can be dramatically inaccurate (Freedman and Peters, 1984). Hence, we make use of bootstrap procedures to estimate standard errors.

In Chapter 5 we focus on a fundamental assumption of panel data models, i.e. homogeneity of parameters across cross-sectional units. This is a controversial issue especially in large $T$, large $N$ panels and in the previous section we paid some attention to it already. Of course, panel data models allow for individual specific effects, but the assumption of equal slope vectors is a rather strong one. Whenever $T$ is large enough compared to $N$ it is feasible to test for heterogeneity in the full parameter vector within the framework of a system of regression equations. We will analyse both standard and generalised $F$ test statistics. The former is similar to the ordinary $F$ statistic for testing a set of linear restrictions in single equation models, while the latter has been proposed by Zellner (1962) in the context of the SUR model. The robustness of the standard $F$ test will be analysed both in dynamic regression models and in models with non-spherical disturbances. In the former case, the standard $F$ statistic can still be applied, but one has to rely on approximate distribution theory. In a simulation study we will analyse to what extent these asymptotic approximations are accurate in finite samples. In the latter case, use of the standard $F$ statistic is not appropriate. Using techniques developed in Kiviet (1980, 1991) lower and upper bounds for the critical values of the $F$ test will be calculated. As an alternative, a generalised version of the $F$ statistic can be constructed (Roy, 1957; Zellner, 1962), for which approximate distribution theory exists. By simulation the actual size of this test and a bootstrap variant will be examined for some relevant cases.

With the theoretical and simulation findings of Chapters 4 and 5 the money demand data from Vlaar and Schuberth (1998) will be revisited. The possibility of exploiting panel data techniques to analyse these cross-country data will be explored in Chapter 4. To avoid the detrimental effects of structural breaks as much as possible only the data after the German reunification will be used. It turns out that the first-order dynamic model with a scalar covariance matrix is not general enough to capture all the dynamic features and interdependencies in the data. Hence, the bias approximations developed in Chapter
4 for generalised higher-order dynamic regression models will be used in the empirical study on money demand. Apart from constancy of the parameters through time, panel data techniques require to some extent also slope homogeneity across countries. The results of Chapter 5 will be applied to the money demand data to examine the degree of heterogeneity in the slope vector across countries.

Finally, Chapter 6 summarises the major findings of this thesis. In this chapter we draw some general conclusions regarding the accuracy of statistical analysis in dynamic panel data models. Chapter 6 is followed by the bibliography and a summary in Dutch.