Serial Correlation in Dynamic Panel Data Models with Weakly Exogenous Regressors and Fixed Effects

Reinhard Hujer*, Paulo J. M. Rodrigues† and Christopher Zeiss‡

J.W.Goethe-University, Frankfurt

Working Paper – March 9, 2005

Abstract

Our paper wants to present and compare two estimation methodologies for dynamic panel data models in the presence of serially correlated errors and weakly exogenous regressors. The first is the first difference GMM estimator as proposed by Arellano and Bond (1991) and the second is the transformed Maximum Likelihood Estimator as proposed by Hsiao, Pesaran, and Tahmiscioglu (2002). Thereby, we consider the fixed effects case and weakly exogenous regressors. The finite sample properties of both estimation methodologies are analysed within a simulation experiment. Furthermore, we will present an empirical example to consider the performance of both estimators with real data.

Keywords: Dynamic Panel Data, Serial Correlation, Simulation, Fixed Effects, Beveridge Curve

JEL Classification: C23, J64

*Reinhard Hujer is Professor of Statistics and Econometrics at the J.W.Goethe-University of Frankfurt/M and Research Fellow of the IZA, Bonn and ZEW, Mannheim, e-mail: hujer@wiwi.uni-frankfurt.de.

†Paulo J. M. Rodrigues is Research Assistant at the Chair of Statistics and Econometrics, e-mail: prodrigues@wiwi.uni-frankfurt.de.

‡Christopher Zeiss is Research Assistant at the Chair of Statistics and Econometrics, e-mail: zeiss@wiwi.uni-frankfurt.de.

1. Introduction

In recent years an increasing fraction of empirical macroeconomic research is done on the basis of panel data. Applications are for example empirical research on long run growth (see Mankiw, Romer, and Weil (1992), Fischer (1993) and Levine and Renelt (1992)) or the aggregate impact analysis of active labour market policy (see e.g. Calmfors and Skedinger (1995), Boeri and Burda (1996) or Hujer and Zeiss (2005)). Typically, methods for panel data are designed for microeconomic applications where the data contains information on individuals or firms and hence the data usually provides a large cross section in combination with a short time series dimension. In contrast, panel data for macroeconomic research consist of aggregated time series data for a set of countries (e.g. OECD countries) or regional districts. Therefore, regional panel data usually provide a larger time series dimension and often also a smaller cross section dimension compared to typical panel data. Furthermore, empirical research on the basis of regional data is usually guided by macroeconomic theory which often suggest a dynamic specifications for the empirical model. The performance of several estimators for dynamic panel data models in the macroeconomic context was analysed by Judson and Owen (1999). But, besides the general problem to obtain consistent estimates in the context of dynamic panel data models, a large time series dimension often comes along with specific time series issues like serial correlation or non-stationarity. In order to obtain consistent estimates, especially serial correlation is a severe problem in dynamic panel data models.

Our paper wants to present and compare two estimation methods for dynamic panel data models in the presence of serially correlated errors and weakly exogenous regressors. The first is the first difference GMM estimator as proposed by Arellano and Bond (1991) and the second is the transformed maximum likelihood estimator as proposed by Hsiao, Pesaran, and Tahmiscioglu (2002). Thereby, we will only consider the fixed effects model. This is primarily reasoned by the fact that in most applications of dynamic panel data models in the macroeconomic context, the data is available only for a few regions or the data cover the full population of regions for the country of consideration. Furthermore, the fixed effects assumption generally avoids a parametrisation of the distribution function for the regional specific unobserved heterogeneity term, which might be hard to justify from a theoretical point of view. In particular, spatial correlation as well as non-zero correlation of the residual with the regressors should not be neglected within regional data. A further issue which often arises in dynamic panel data models is the problem that the regressors can hardly be treated as strictly exogenous. In a dynamic macroeconomic framework strictly exogenous variables are mostly not available. However, in the majority of cases regressors turn out to be weakly exogenous (or predetermined), i.e. only future values of the explanatory variables are affected by the current value of the dependent variable. In this case identification of the parameters requires to account for the weak exogeneity of the regressors.

In applied research mostly the first differences GMM estimator is used for empirical analysis. But, as several simulation experiments have shown, the first differences GMM estimator suffers from a considerable finite sample bias.[1] The bias is usually associated with a highly persistent pattern of the dependent variable (see Blundell, Bond, and Windmeijer (2000)). As pointed

[1] See for example Blundell and Bond (1998), Kiviet (1995) and Hsiao, Pesaran, and Tahmiscioglu (2002).

out by Arellano and Bond (1991), the identification of the parameters with the GMM estimator hinges heavily on the lack of serial correlation. But as the utilisation of aggregated time series data often involves serial correlation, GMM often does not lead to consistent estimates. An alternative estimator for the fixed effects case was suggested by Hsiao, Pesaran, and Tahmiscioglu (2002). The transformed maximum likelihood estimator avoids similar to the GMM estimator the incidental parameter problem and thus yields consistent estimates. Although simulation experiments from Hsiao, Pesaran, and Tahmiscioglu (2002) have shown that maximum likelihood performs superior to GMM, especially in the case of a high persistent pattern of the dependent variable, serial correlation remains a severe problem that leads to inconsistent estimates.

Our analysis wants to compare the performance of both estimators in the presence of serially correlated errors. Furthermore, we will present a modification of both estimators in order to account explicitly for serial correlation and analyse the performance of the modified estimators in the finite sample. Similar work that considers serial correlation in panel data models with weakly exogenous regressors can be found in Keane and Runkle (1992). In the framework of the first differences GMM estimator, serial correlation can be counteracted by imposing a sufficient number of lags for the instruments. Although this is a very simple way to account for serial correlation, it can make the size of the sample in time dimension diminish in a drastic way (Sevestre and Trognon, 1996). Furthermore, this approach may lead to severe efficiency losses especially if the number of lags imposed is large. Within the maximum likelihood framework serial correlation can be straightforwardly incorporated by directly including moving average terms. Although this procedure can become burdensome for higher orders of moving average terms, it should provide a consistent and efficient estimator in the presence of serial correlation. The finite sample properties of both estimation methodologies are analysed within a simulation experiment. Furthermore, we will present an empirical example to consider the performance of both estimators with real data.

The structure of the paper is as follows: Section 2 presents the consequences of serial correlation for the first differences GMM and the transformed maximum likelihood estimator. Furthermore, we will discuss how to control for serial correlation in both estimation methodologies. Section 3 contains a simulation experiment to assess the consequences of serial correlation and to analyse the performance of both estimators if they account for serial correlation. Section 4 presents an empirical example that analyses the Beveridge Curve for West Germany with regional data. Finally, section 5 concludes with a short upshot.

2. Serial Correlation in Dynamic Panel Data Models

2.1. First Difference GMM Estimator

Consider the first order autoregressive model

$$y_{it} = \alpha y_{i(t-1)} + \boldsymbol{\beta}' \mathbf{x}_{it} + \mu_i + u_{it}, \tag{1}$$

where $\mathbf{x}_{it}$ $(K \times 1)$ is a set of K regressors, μ_i is an individual specific constant term and u_{it} is a residual term which varies over the cross section and time.[2] $i = (1, 2, \cdots, N)$ is an index

[2] The analysis can be easily extended to higher order autoregressive models.

over the cross section and $t = (1, 2, \cdots, T)$ denotes the time dimension. The residual term u_{it} is assumed to follow an MA(1) process

$$u_{it} = \varepsilon_{it} + \theta \varepsilon_{i(t-1)}, \tag{2}$$

where ε_{it} is i.i.d. with zero mean and variance σ_ε^2. Other forms of serial correlation are generally ruled out by the assumption that the model (1) has been transformed so that the coefficients satisfy some set of common factor restrictions. The explanatory variables are assumed to be weakly exogenous (or predetermined) forcing variables with $E(\varepsilon_{it}\mathbf{x}_{is}) \neq 0$ for $s > t$ and zero otherwise. Note, that $\mathbf{x}_{it}$ is weakly exogenous with respect to the i.i.d. disturbances ε_{it}, i.e. if θ is nonzero $\mathbf{x}_{it}$ is an endogenous variable with respect to u_{it} since $E(\mathbf{x}_{it}u_{is}) \neq 0$ for $s \geq t$. Thus, not only the identification of α but also the identification of $\boldsymbol{\beta}$ is directly affected by serial correlation.

In order to circumvent the incidental parameter problem pointed out by Neyman and Scott (1948), the first difference GMM estimator as suggested by Arellano and Bond (1991) relies on the equations in first differences where the fixed effects are eliminated

$$\Delta y_{it} = \gamma \Delta y_{i(t-1)} + \boldsymbol{\beta}' \Delta \mathbf{x}_{it} + \Delta u_{it}. \tag{3}$$

To identify α and $\boldsymbol{\beta}$ in a dynamic model with weakly exogenous regressors, Arellano and Bond (1991) suggest the following moment conditions

$$E(y_{i(t-s)}\Delta u_{it}) = 0; \quad \text{for } t = 3, \ldots, T \text{ and } 2 \leq s \leq t-1 \tag{4}$$

$$E(\mathbf{x}_{i(t-s)}\Delta u_{it}) = 0; \quad \text{for } t = 2, \ldots, T \text{ and } 1 \leq s \leq t-1. \tag{5}$$

But if u_{it} is generated by an MA(1) process, the orthogonality conditions $E(y_{i(t-2)}\Delta u_{it}) = 0$ and $E(x_{i(t-1)}\Delta u_{it}) = 0$ are not valid. Thus, if these moment conditions were used the GMM estimator would become inconsistent. The easiest way to account for the serial correlated errors is to exclude the invalid moment conditions. As the serial correlation due to an MA(1) process implies $E(\Delta u_{it}\Delta u_{is}) = 0$ for $s \geq 3$, the set of valid moment conditions simply reduces to

$$E(y_{i(t-s)}\Delta u_{it}) = 0; \quad \text{for } t = 4, \ldots, T \text{ and } 3 \leq s \leq t-1 \tag{6}$$

$$E(\mathbf{x}_{i(t-s)}\Delta u_{it}) = 0; \quad \text{for } t = 3, \ldots, T \text{ and } 2 \leq s \leq t-1. \tag{7}$$

A major advantage of this procedure is that the extension to higher order MA processes is straightforward. The number of lags imposed in the set of instruments simply rises with the order of the MA process (see Arellano and Bond (1991)). Major drawback of this procedure is that it reduces the size of the sample in the time dimension in a drastic way for higher order MA processes (Sevestre and Trognon, 1996). To compute the GMM estimator we require in the case of a MA(1) process $T \geq 4$, in the case of an MA(2) process $T \geq 5$ and so on. Furthermore, for a higher number of lags imposed for the instruments, the explanatory power of the instruments may be very poor. For this reason, the first difference GMM estimator may suffer from severe efficiency losses when accounting for higher order of serial correlation.

To obtain the first difference GMM estimator, the moment conditions can be summarised as $m(\gamma) = E(\mathbf{W}_i \boldsymbol{\Delta}\mathbf{u}_i)$ with the sample counterpart $\frac{1}{N}\sum_{i=1}^{N} \mathbf{W}_i \boldsymbol{\Delta}\hat{\mathbf{u}}_i = 0$ where

$$\mathbf{W}_i = \begin{pmatrix} [y_{i1}, \mathbf{x}'_{i1}, \mathbf{x}'_{i2}] & & & 0 \\ & [y_{i1}, y_{i2}, \mathbf{x}'_{i1}, \mathbf{x}'_{i2}, \mathbf{x}'_{i3}] & & \\ & & \ddots & \\ 0 & & & [y_{i1}, y_{i2}, \ldots, y_{i(T-3)}, \mathbf{x}'_{i1}, \mathbf{x}'_{i2}, \ldots, \mathbf{x}'_{i(T-2)}] \end{pmatrix}$$

and the parameters α and $\boldsymbol{\beta}$ are collected into the vector $\boldsymbol{\gamma}$. The first difference GMM estimator is calculated from

$$\hat{\gamma} = \left(\mathbf{Z}'\mathbf{W}\mathbf{V}_N^{-1}\mathbf{W}'\mathbf{Z}\right)^{-1} \left(\mathbf{Z}'\mathbf{W}\mathbf{V}_N^{-1}\mathbf{W}'\Delta\mathbf{y}\right), \tag{8}$$

where $\Delta\mathbf{y}_i = (\Delta y_{i4}, \ldots, \Delta y_{iT})'$, $\mathbf{Z}_i = \left[(\Delta y_{i3}\ \Delta\mathbf{x}'_{i4}), \ldots, (\Delta y_{i(T-1)}\ \Delta\mathbf{x}'_{iT})\right]$ and to construct $\Delta\mathbf{y}$, $\mathbf{W}$ and $\Delta\mathbf{Z}$ we stack the observations over the individuals i. Following Hansen (1982) the optimal choice for the weighting matrix is

$$\widetilde{\mathbf{V}}_N = \sum_{i=1}^{N} \mathbf{W}'_i \Delta\hat{\mathbf{u}}_i \Delta\hat{\mathbf{u}}'_i \mathbf{W}_i, \tag{9}$$

where $\Delta\hat{\mathbf{u}}_i = (\Delta\hat{u}_{i4}, \ldots, \Delta\hat{u}_{iT})'$ are the estimated residuals from a consistent one step estimator. The asymptotic covariance matrix of the first difference GMM estimator is given by

$$\text{Asy.Var}(\hat{\gamma}) = (\mathbf{Z}'\mathbf{W}\widetilde{\mathbf{V}}_N^{-1}\mathbf{W}'\mathbf{Z})^{-1}. \tag{10}$$

Although the application of the first difference GMM estimator is straightforward, the consistency can suffer from a weak instruments problem. As discussed by Blundell, Bond, and Windmeijer (2000) and Binder, Hsiao, and Pesaran (2002), a weak instruments problem arises if the autoregressive parameter is near unity. In the extreme case of a unit root process, the first-differences GMM estimator breaks down completely. Furthermore a weak instruments problem arises if the variables in levels are for a major part driven by the individual effect. In this case the explanatory power of the instruments in levels for the variables in first differences is very poor. To overcome these problems Ahn and Schmidt (1995) and Blundell and Bond (1998) have proposed additional moment conditions. Unfortunately, under the fixed effects assumption these moment conditions require the identification of the individual effects and therefore do not solve the incidental parameter problem.[3]

2.2. Transformed Likelihood Estimator

In order to set up the transformed maximum likelihood estimator as suggested by Hsiao, Pesaran, and Tahmiscioglu (2002), we again consider the model given in (1) where the residual is determined by an MA(1) process. Fixed effects are eliminated by taking first difference as in equation (3). By collecting Δy_{it} and $\Delta\mathbf{x}_{it}$ into $\Delta\boldsymbol{\omega}_{it}$, we can write the joint probability distribution of $(\Delta\boldsymbol{\omega}_{i1}, \Delta\boldsymbol{\omega}_{i2}, \ldots, \Delta\boldsymbol{\omega}_{iT})$ as

$$f(\Delta\boldsymbol{\omega}_{iT}|\boldsymbol{\mathcal{I}}_{i(T-1)})f(\Delta\boldsymbol{\omega}_{i(T-1)}|\boldsymbol{\mathcal{I}}_{i(T-2)}), \ldots, f(\Delta\boldsymbol{\omega}_{i1}|\boldsymbol{\mathcal{I}}_{i0}), \tag{11}$$

[3] Note, that all these estimators were developed for the random effects model.

where $\boldsymbol{\mathcal{I}}_{it}$ contains the information up to time t, i.e. $\boldsymbol{\mathcal{I}}_{it} = (\Delta\boldsymbol{\omega}_{i1}, \Delta\boldsymbol{\omega}_{i2}, \ldots, \Delta\boldsymbol{\omega}_{it})$ for $t = 1, 2, \ldots, T-1$ and $\boldsymbol{\mathcal{I}}_{i0}$ is normalised to unity. Since

$$f(\Delta\boldsymbol{\omega}_{it}|\boldsymbol{\mathcal{I}}_{i(t-1)}) = f(\Delta\mathbf{x}_{it}|\boldsymbol{\mathcal{I}}_{i(t-1)}) f(\Delta y_{it}|\Delta\mathbf{x}_{it}, \boldsymbol{\mathcal{I}}_{i(t-1)}) \tag{12}$$

for $t = 1, \ldots, T$, the relevant part of the log-likelihood is given by

$$\sum_{i=1}^{N}\sum_{t=1}^{T} \ln f(\Delta y_{it}|\Delta\mathbf{x}_{it}, \boldsymbol{\mathcal{I}}_{i(t-1)}). \tag{13}$$

For $t \geq 2$ the elements in (13) are fully specified by equation (3). To define $\ln f(\Delta y_{i1}|\Delta\mathbf{x}_{i1}, \boldsymbol{\mathcal{I}}_{i0}) = \ln f(\Delta y_{i1}|\Delta\mathbf{x}_{i1})$, we need an assumption about the initial condition of the data generating process since Δy_{i0} is not observable. The problem that remains is to find the density for the initial observation that does not depend on incidental parameters. By backward substitution we can write

$$\Delta y_{i1} = \alpha^m \Delta y_{i(-m+1)} + \boldsymbol{\beta}' \sum_{j=0}^{m-1} \alpha^j \Delta\mathbf{x}_{i(1-j)} + \sum_{j=0}^{m-1} \alpha^j \Delta u_{i(1-j)}, \tag{14}$$

where we assume that the process has started at $-m$. Taking expectations conditional on $\Delta y_{i(-m+1)}$ and $\Delta\mathbf{x}_{i1}, \ldots, \Delta\mathbf{x}_{i(-m)}$, we obtain

$$E(\Delta y_{i1}|\Delta y_{i(-m+1)}, \Delta\mathbf{x}_{i1}, \ldots, \Delta\mathbf{x}_{i(-m)}) = \alpha^m \Delta y_{i(-m+1)} + \boldsymbol{\beta}' \sum_{j=0}^{m-1} \alpha^j \Delta\mathbf{x}_{i(1-j)}. \tag{15}$$

Since $\Delta\mathbf{x}_{i(1-j)}$ for $j = 1, 2, \ldots, m$ is not observable, this expected value is unknown. Treating the expected value as a free parameter to be estimated would result in an incidental parameter problem. In order to specify the expected value as a function of a finite number of parameters Hsiao, Pesaran, and Tahmiscioglu (2002) suggest to write them as a function of the observable $\Delta\mathbf{x}_{i1}$. Note that, since the models contains exogenous variables, the likelihood has to be approximated. See also Bhargava and Sargan (1983) for a exposition of the same method in the random effects case. The case with no explanatory variables in considered in appendix A1., note that for this case a exact likelihood function exists. Due to the weak exogeneity assumption only $\Delta\mathbf{x}_{i1}$ is a valid explanatory variable for the conditional expectation of Δy_{i1}. In what follows, Hsiao, Pesaran, and Tahmiscioglu (2002) assume either that

1. the process has been going on for a long time, i.e. $m \to \infty$ and $|\alpha| < 1$. Hence $E(\Delta y_{i1}|\Delta\mathbf{x}_{i1}) = 0$, or
2. m is finite and $E(\Delta y_{i1}|\Delta\mathbf{x}_{i1})$ is the same across individuals.

Note, that assumption 2 only requires that the expected changes in the initial endowments are the same across individuals and does not require $|\alpha| < 1$. The expected value of Δy_{i1} conditional on $\Delta\mathbf{x}_{i1}$ is given by

$$E(\Delta y_{i1}|\Delta\mathbf{x}_{i1}) = b^* + E\left[\boldsymbol{\beta}' \sum_{j=0}^{m-1} \alpha^j \Delta\mathbf{x}_{i(1-j)}|\Delta\mathbf{x}_{i1}\right] + E\left[\sum_{j=0}^{m-1} \alpha^j \Delta u_{i(1-j)}|\Delta\mathbf{x}_{i1}\right], \tag{16}$$

where $b^* = 0$ under assumption 1 and $b^* = b$ under assumption 2.
Hsiao, Pesaran, and Tahmiscioglu (2002) suggested a projection technique in order to get a

computable expression for expected values in equation (16). The elements of the sum of the first expected value in equation (16) can be projected with

$$E(\Delta \mathbf{x}_{i1-j}|\Delta \mathbf{x}_{i1}) = \mathbf{g}_j + \mathbf{\Psi}_j \Delta \mathbf{x}_{i1}, \tag{17}$$

where $\mathbf{\Psi}_j$ is a $(K \times K)$ matrix and $\mathbf{g}_j$ is a $(K \times 1)$ vector. In our case $\mathbf{\Psi}_j$ and $\mathbf{g}_j$ depend not only on the autocovariances of the processes determining $\mathbf{x}_{it}$. Additionally, dependencies between the processes of the regressors determine the structure of $\mathbf{\Psi}_j$. In the extreme case of independent processes, determining $\mathbf{x}_{it}$ the off-diagonal elements of $\mathbf{\Psi}_j$ would be zero.

For the elements of the sum of the second expected value in equation (16) we write

$$E(\Delta u_{1-j}|\Delta \mathbf{x}_{1i}) = d_j + \boldsymbol{\varphi}_j \Delta \mathbf{x}_{i1}, \tag{18}$$

where d_j is a scalar and $\boldsymbol{\varphi}_j$ is a $(1 \times K)$ vector. The elements in d_j and $\boldsymbol{\varphi}_j$ can be derived from the joint distribution of $(\mathbf{x}_{i1}, \mathbf{x}_{i0}, \mathbf{x}_{i,-1}, \cdots, u_{i1}, u_{i0}, u_{i,-1}, \cdots)$ (Hsiao, Pesaran, and Tahmiscioglu, 2002).

The projection technique suggested by Hsiao, Pesaran, and Tahmiscioglu (2002) allows $\mathbf{x}_{it}$ to follow trend stationary and first difference stationary data generating processes. Important is however, that the process generating $\mathbf{x}_{it}$ does not follow different trends (stochastic or deterministic) for different i. In this case the expected value would not be a function of finite parameters, i.e. the incidental parameter problem could not be solved and inconsistent estimates would be the result for finite T.[4]

Inserting the projections into equation (16) we get

$$E(\Delta y_{i1}|\Delta \mathbf{x}_{i1}) = b^* + \boldsymbol{\beta}' \sum_{j=0}^{m-1} \alpha^j \left[\mathbf{g}_j + \mathbf{\Psi}_j \Delta \mathbf{x}_{i1}\right] + \sum_{j=0}^{m-1} \alpha^j \left[d_j + \boldsymbol{\varphi}_j \Delta \mathbf{x}_{i1}\right]. \tag{19}$$

Therefore, we can write for Δy_{i1} the following simplified expression

$$\Delta y_{i1} = \boldsymbol{\lambda}_0 + \Delta \mathbf{x}_{i1} \boldsymbol{\lambda}_1 + \xi_{i1}, \tag{20}$$

where $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$ $(K \times 1)$ are unknown coefficients. The parameter vectors $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$ are functions of $b^*, \boldsymbol{\beta}, \alpha, \mathbf{g}_j, \mathbf{\Psi_j}, d_j$ and $\boldsymbol{\varphi}_j$. The residual ξ_{i1} with $E(\xi_{i1}|\mathbf{x}_{i1}) = 0$ is defined as

$$\xi_{i1} = \Delta y_{i1} - E(\Delta y_{i1}|\Delta \mathbf{x}_{i1}). \tag{21}$$

The expression for ξ_{i1} can be obtained from (14) and (19).

$$\begin{aligned} \xi_{i1} &= \left(\alpha^m \Delta y_{i(-m+1)} - b^*\right) + \boldsymbol{\beta}' \sum_{j=0}^{m-1} \alpha^j \left\{\Delta \mathbf{x}_{i(1-j)} - [\mathbf{g}_j + \mathbf{\Psi}_j \Delta \mathbf{x}_{i1}]\right\} \\ &\quad + \sum_{j=0}^{m-1} \alpha^j \left\{\Delta u_{i(1-j)} - \left[d_j + \boldsymbol{\varphi}_j \Delta \mathbf{x}_{i1}\right]\right\} \end{aligned} \tag{22}$$

To construct the likelihood function we set up the vector of residuals $[\xi_{i1}, u_{i2}, u_{i3}, \cdots, u_{iT}]$ from

$$\Delta \mathbf{u}_i^* = \left[\Delta y_{i1} - \boldsymbol{\lambda}_0 + \Delta \mathbf{x}_{i1} \boldsymbol{\lambda}_1,\ \Delta y_{i2} - \alpha \Delta y_{i1} - \boldsymbol{\beta} \Delta x_{i2}\ , \ldots,\ \Delta y_{iT} - \alpha \Delta y_{i(T-1)} - \boldsymbol{\beta} \Delta x_{iT}\right]'. \tag{23}$$

[4] Notice that in this case the projection would depend on i, i.e. $E(\Delta \mathbf{x}_{i1-j}|\Delta \mathbf{x}_{i1}) = g_{i,j} + \mathbf{\Psi}_{i,j} \Delta \mathbf{x}_{i1}$.

As u_{it} is generated by an MA(1)-process we additionally have to consider the structure of the covariance matrix $E(\Delta \mathbf{u}_i^* \Delta \mathbf{u}_i^{*\prime}) = \sigma_\varepsilon^2 \mathbf{\Omega_i}^* = \mathbf{\Omega_i}$. Concerning the variance of the initial condition we specify $E(\xi_{i1}^2) = \sigma_\varepsilon^2 \omega$ and treat ω as a free parameter to be estimated. As noted by Hsiao, Pesaran, and Tahmiscioglu (2002) this procedure poses no problem as long as m is unknown and identical across i. For the covariances regarding the initial condition, the MA(1) process of u_{it} implies

$$\begin{aligned} E(\xi_{i1}\Delta u_{i2}) &= -\sigma_\varepsilon^2[(1-\theta)^2 - \theta\alpha], \\ E(\xi_{i1}\Delta u_{i3}) &= -\sigma_\varepsilon^2\theta \text{ and} \\ E(\xi_{i1}\Delta u_{it}) &= 0 \text{ for } t = 4, \ldots, T. \end{aligned}$$

The remaining elements of the covariance matrix $\mathbf{\Omega_i}$ are given by

$$\begin{aligned} E(\Delta u_{it}^2) &= 2(1+\theta^2-\theta)\sigma_\varepsilon^2, \\ E(\Delta u_{it}\Delta u_{i(t-1)}) &= -\sigma_\varepsilon^2(1-\theta)^2, \\ E(\Delta u_{it}\Delta u_{i(t-2)}) &= -\theta\sigma_\varepsilon^2 \text{ and} \\ E(\Delta u_{it}\Delta u_{i(t-j)}) &= 0 \text{ for } j \geq 3. \end{aligned}$$

Under the assumption $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ the log-likelihood function is given by

$$LL = -\frac{NT}{2}\ln(2\pi) - \frac{N}{2}\ln|\mathbf{\Omega_i}| - \frac{1}{2}\sum_{i=1}^{N} \Delta \mathbf{u}_i^{*\prime} \mathbf{\Omega_i}^{-1} \Delta \mathbf{u}_i^*. \tag{24}$$

Since the log-likelihood function given in (24) is highly nonlinear, maximisation is performed by iterative methods like Newton-Raphson. The likelihood function depends on a fixed number of parameters and satisfies standard regularity conditions, so the MLE is consistent and asymptotically normal distributed as $N \to \infty$ (Hsiao, Pesaran, and Tahmiscioglu, 2002). Note, that in the case of weakly exogenous regressors this holds whether T is fixed or tends to infinity. Monte-Carlo experiments by Hsiao, Pesaran, and Tahmiscioglu (2002) have shown that in the finite sample the MLE generally performs superior compared to the first difference GMM estimator. Especially if the AR-coefficient is near unity, MLE produces a considerably lower finite sample bias.

Accounting for serial correlation within the maximum likelihood framework provides an asymptotically normal distributed and efficient estimator. Thus, if the serial correlation is driven by an MA process, MLE directly accounting for the MA terms provides a superior procedure to control for the serial correlation. However, compared to the first difference GMM estimator, computation of the MLE is associated with more computational burden. Especially for higher order MA processes, which is a straightforward extension, the construction of the covariance matrix for the MLE can become tedious. Therefore, the following simulation experiment wants to assess the gains in bias and efficiency that results from the application of MLE compared to the first difference GMM estimator.

3. Monte Carlo Evidence

To assess the finite sample performance of the GMM and the transformed maximum likelihood estimator in the presence of weakly exogenous regressors and serial correlation, we have conducted a simulation experiment. The design of the data generating process is closely related to Hsiao, Pesaran, and Tahmiscioglu (2002). The model for the dependent variable y_{it} is given by

$$y_{it} = \alpha y_{i(t-1)} + x_{it}\beta + \mu_i + u_{it}. \tag{25}$$

Along with the assumptions of the previous sections, serial correlation is specified as an MA(1) process

$$u_{it} = \varepsilon_{it} + \theta_1 \varepsilon_{i(t-1)}, \tag{26}$$

where ε_{it} is distributed as $N(0, \sigma_\varepsilon^2)$.

The weakly exogenous regressor x_{it} is generated by the process

$$x_{it} = \psi_{it} + \nu_i + \lambda \sum_{s=1}^{t-1} \frac{1}{t-s}\varepsilon_{is}, \tag{27}$$

where ψ_{it} follows an ARMA(1,1) process

$$\psi_{it} = \phi\psi_{i(t-1)} + \omega_{it} + \rho\omega_{i(t-1)} \tag{28}$$

with $\omega_{it} \sim N(0, \sigma_\omega^2)$. Weak exogeneity is incorporated into the regressor process via the term $\lambda \sum_{s=1}^{t-1} \frac{1}{t-s}\varepsilon_{is}$. By setting λ unequal to zero we ensure that $E(\varepsilon_{it}x_{ik})$ is zero for $k \leq t$ and non-zero for $k > t$. For the latter case this correlation is given by $E(\varepsilon_{it}x_{i(t+j)}) = \frac{\lambda}{j}\sigma_\varepsilon^2$ for $j = (1, 2, 3, \cdots)$. Therefore, we ensure that the correlation decreases as the distance between the observations increases. An important feature of this specification is that in the case of serial correlation the regressor is in fact endogenous since $E(u_{it}x_{it}) \neq 0$. Hence β is not identified if the estimation procedure does not account for the serially correlated error term. Furthermore, it should be noted that the specification of the x_{it} leads basically to a non-stationary process since variance and autocovariance of x_{it} changes over time due to the term $\lambda \sum_{s=1}^{t-1} \frac{1}{t-s}\varepsilon_{is}$. Considering the moments of x_{it} we find that in the limit with respect to t, i.e. for a sufficiently long history of the process the variance converges to

$$\lim_{t\to\infty} E(x_{it}^2) = \left[1 + \frac{(\rho+\phi)^2}{(1-\phi^2)}\right]\sigma_\omega^2 + \sigma_\nu^2 + \frac{(\lambda\sigma_\varepsilon\pi)^2}{6}. \tag{29}$$

In particular, the part of the variance $E(x_{it}^2)$ associated with the term $\lambda \sum_{s=1}^{t-1} \frac{1}{t-s}\varepsilon_{is}$ converges form an initial level $(\lambda\sigma_\varepsilon)^2$ to a final level $\frac{(\lambda\sigma_\varepsilon\pi)^2}{6}$ as t increases. Therefore, in the limit the variance of x_{it} can be handled as a constant given by (29). Furthermore, in the limit the process becomes covariance-stationary as the autocovariances do not depend on t. This can be seen from the limit of the first order and second order autocovariances for x_{it}, which are given by

$$\lim_{t\to\infty} E(x_{it}x_{i(t-1)}) = g(1) + \sigma_\nu^2 + \lambda^2\sigma_\varepsilon^2 \tag{30}$$

$$\lim_{t\to\infty} E(x_{it}x_{i(t-2)}) = g(2) + \sigma_\nu^2 + \frac{3}{4}\lambda^2\sigma_\varepsilon^2, \tag{31}$$

where $g(z)$ denotes the autocovariance generating function for the ARMA(1,1) process ψ_{it} (see Hamilton (1994)). Hence, as t increases the variance and the autocovariances of the x_{it}-process converge form an initial level to a final level given by (29), (30) and (31). For the variance the speed of convergence is such that for $t = 100$ the variance has reached 99.4% of (29). In our simulation experiment we generate the series form $t = -100$ and exclude the first 100

Tab. 1: Simulation design

Simulation Design	α	β	θ	ϕ	ρ	g	λ	σ^2_ε	σ^2_ω	σ^2_μ	σ^2_ν	$\sigma_{\mu\nu}$
1	0.4	0.6	0	0.5	0.5	0.01	0.2	0.85	1	1	1	0.7
2	0.4	0.6	0.5	0.5	0.5	0.01	0.2	0.85	1	1	1	0.7
3	0.4	0.6	0.9	0.5	0.5	0.01	0.2	0.85	1	1	1	0.7
4	0.4	0.6	0	0.5	0.5	0.01	0.2	1.25	1	1	1	0.7
5	0.4	0.6	0.5	0.5	0.5	0.01	0.2	1.25	1	1	1	0.7
6	0.4	0.6	0.9	0.5	0.5	0.01	0.2	1.25	1	1	1	0.7
7	0.4	0.6	0	0.9	0.5	0.01	0.2	0.85	1	1	1	0.7
8	0.4	0.6	0.5	0.9	0.5	0.01	0.2	0.85	1	1	1	0.7
9	0.4	0.6	0.9	0.9	0.5	0.01	0.2	0.85	1	1	1	0.7
10	0.8	0.2	0	0.5	0.5	0.01	0.2	0.85	1	1	1	0.7
11	0.8	0.2	0.5	0.5	0.5	0.01	0.2	0.85	1	1	1	0.7
12	0.8	0.2	0.9	0.5	0.5	0.01	0.2	0.85	1	1	1	0.7
13	0.8	0.2	0	0.5	0.5	0.01	0.2	1.25	1	1	1	0.7
14	0.8	0.2	0.5	0.5	0.5	0.01	0.2	1.25	1	1	1	0.7
15	0.8	0.2	0.9	0.5	0.5	0.01	0.2	1.25	1	1	1	0.7
16	0.8	0.2	0	0.9	0.5	0.01	0.2	0.85	1	1	1	0.7
17	0.8	0.2	0.5	0.9	0.5	0.01	0.2	0.85	1	1	1	0.7
18	0.8	0.2	0.9	0.9	0.5	0.01	0.2	0.85	1	1	1	0.7

observations to ensure that for $|\phi| < 1$, x_{it} is a covariance stationary process and thus y_{it} is also stationary if $|\alpha| < 1$. Furthermore, the exclusion of the initial observations avoids an unnecessary impact of the initial values on ψ_{it} and y_{it}.

The individual effects μ_i and ν_i are generated according to a multivariate normal distribution

$$(\mu_i \ \nu_i) \sim N\left((0 \ 0), \begin{pmatrix} \sigma^2_\mu & \sigma_{\mu\nu} \\ \sigma_{\mu\nu} & \sigma^2_\nu \end{pmatrix}\right). \tag{32}$$

By allowing the individual effects to be correlated, we make sure that the random effects estimates would be inconsistent.

To provide comparability of the simulation experiment to our empirical example, we set $N = 150$ and $T = 20$ and use the following parameter values for the simulations. For the individual effects we set σ^2_μ and σ^2_ν to unity and impose a correlation between both effects of 0.7. Concerning the regressor we introduce weak exogeneity by setting λ to 0.2. Furthermore, we impose a trend with $g = 0.01$ and the parameters of the ARMA process are given by $(\phi, \rho) = (0.5, 0.5)$ and $(0.9, 0.5)$. Hence, we either allow for a slight or a strong persistence in the regressor process. The variance of ω is set to $\sigma^2_\omega = 1$. For the dependent variable we fix similar to Kiviet (1995) the long run multiplier $\frac{\beta}{1-\alpha}$ to unity and specify $(\alpha, \beta) = (0.4, 0.6)$ and $(0.8, 0.2)$. The variance of the residual is set to $\sigma_\varepsilon^2 = 0.85$ and 1.25, i.e. we consider either a situation with $\sigma_\varepsilon^2/\sigma_\mu^2 < 1$ or a situation with $\sigma_\varepsilon^2/\sigma_\mu^2 > 1$. As pointed out by Blundell, Bond, and Windmeijer (2000), a high variance of the individual effects μ_i relative to the variance of ε_{it} leads to a weak instruments problem because the predictive power of the instrument in levels for the variables in first differences becomes very poor. A similar problem arises if the autoregressive parameter tends to unity. Finally, we assess the consequences of serial correlation by setting θ to 0, 0.5 and 0.9. Table 1 summarises our simulation design and the simulation results from 1000 replications can be found in table 2 and 3.

For the calculation of the GMM estimates we cut the history of the series as instruments at $t-5$. The results for the simulation designs with $\alpha = 0.6$ and $\beta = 0.4$ are collected in table 2. For the specification without serially correlated errors, i.e. designs 1,4 and 7, we find a superior performance for the maximum likelihood estimator by comparing the RMSE. Note, that both estimators perform best when the process originating the forcing variable is highly persistent. Also the superior performance of the MLE is more clear cut for the autoregressive parameter than for the parameter of the weakly exogenous variable.

By introducing serial correlation through a moderate MA coefficient, i.e. designs 2,5 and 8, we find some huge improvements of the RMSE for the modified estimation methods in comparison to the methods that do not take the problem into account. This is not a surprising finding, since we compare unbiased with biased estimation methods. The interesting postulation is that the RMSE for the mis-specified MLE is higher than for the mis-specified GMM estimator. The comparison of both estimators changes in favour to the MLE when autocorrelation is taken into account. The RMSE falls dramatically in both cases, but is reduced by a factor around 8 for the MLE and a factor around 4 for the GMM estimator. The results of the simulation designs 3,6 and 9, i.e. imposing a moving average coefficient of $\theta = 0.9$ are similar to the findings for $\theta = 0.5$, although it can be stated that the estimation methods loose some precision in finite samples if serial correlation increases. The only case where this finding is contradicted is in the comparison of design 5 with design 6, where the RMSE falls from 0.0302 to 0.0275.

A loss of precision can also be observed for the designs with a higher variance of ε_{it}. For example, by comparing model 2 with model 5 we find that the RMSE for the GMM estimator increases from 0.0266 to 0.0302 for the autoregressive parameter and from 0.0309 to 0.0449 for the β parameter. The numbers for the MLE are 0.0161 to 0.0181 and 0.0144 to 0.0200, respectively. It can be stated that the RMSE for the MLE do not only increase by a lesser absolute amount, 0.002 against 0.036, but they are also robuster in a relative sense. The RMSE of the MLE only increases by 12.42% against an increase of 13.53% for the GMM estimator. In the case of a moving average parameter of 0.9 the RMSE for the MLE increases by a higher amount than the RMSE of the GMM estimator. Nevertheless, MLE remains efficient than GMM by the means of the RMSE.

For the model specifications which impose a highly persistent regressor process, we can state that the ML-estimators become more accurate. This can be seen from the comparison of design 2 with design 8, and design 3 with design 9. In the case of $\theta = 0.5$ we find for the MLE a reduction of the RMSE of 0.0038 and 0.0032 for α and β respectively. For the GMM estimator the figures are an increase of 0.0051 and a decrease of 0.0037. Similar to the case of an increasing variance, the results change by imposing an MA coefficient of 0.9. For the likelihood method we even obtain an increase in RMSE for the β parameter, whereas for the GMM method we obtain a increase for α. But again we have a better performance of the MLE in terms of RMSE.

Considering the direction of the bias in the case GMM and MLE ignores serial correlation, we find that for the autoregressive parameter both estimators tend to be upward biased. For the regressor coefficient we find a contrary direction of the bias, i.e. GMM tends to be upward biased, whereas MLE tends be downward biased.

The results for the simulation designs with parameter values of 0.8 and 0.2 for α and β are collected in table 3. The general conclusions for the designs with $\alpha = 0.4$ carry over to the designs with $\alpha = 0.8$, i.e. MLE dominates the GMM estimator by comparing the RMSE. An important finding is however that the advantage of MLE with respect to the autoregressive parameter is considerably higher if the dependent variable is highly persistent. By comparing the RMSE of the MLE and the GMM estimator for the designs with a moderate autoregressive parameter we find a ratio of 54% up to 86%, i.e. MLE-RMSE reaches between 54% and 86% of the GMM-RMSE. In contrast, the range for this ratio is 21% up to 62% in the case of a high autoregressive parameter. Furthermore, this finding is obtained for each individual design. For the estimator of the β parameter this finding cannot be confirmed, since the performance of the MLE is nearly identical compared to the designs with a moderate autoregressive parameter. Regarding the direction of the bias in the case of a high α coefficient, we again find the tendency of MLE to be upward biased while GMM tends to be downward biased. For the regressor coefficient no clear cut picture with respect to the direction of the bias can be found.

Our simulation results show that in the presence of serial correlation the first difference GMM and the transformed maximum likelihood estimator suffer from a severe finite sample bias. We find that the bias due to serial correlation is generally larger for the autoregressive parameter compared to parameters for the explanatory variables. Accounting for serial correlation we find a significant bias-reduction in terms of RMSE compared to the estimation methods that do not account for this data characteristic. If the estimators account for serial correlation, we find a superior performance of the maximum likelihood estimator relative to the GMM estimator. That is, the explicit allowance for MA terms within the MLE framework leads to a considerable gain in precision of the parameter estimates. Moreover, we find that the advantage of the MLE over the GMM estimator increases with a higher autoregressive coefficient. The inferior performance of the GMM estimator in the case of a high persistent dependent variable is carried over from dynamic specifications without serial correlated errors. So, if one is primarily interested in the autoregressive coefficient or wants to compute the lag coefficients, it is advisable to use MLE. With respect to the coefficients for the explanatory variables we generally find smaller differences between MLE and GMM.

Tab. 2: Simulation Results ($\alpha = 0.4$ and $\beta = 0.6$)

Simulation Design		Bias α	Bias β	Std α	Std β	RMSE α	RMSE β
1	GMM	-0.0103	-0.0018	0.0170	0.0178	0.0198	0.0179
	GMM MA(1)	-0.0249	0.0073	0.0309	0.0274	0.0397	0.0283
	MLE	-0.0077	-0.0114	0.0126	0.0132	0.0119	0.0141
	MLE MA(1)	-0.0113	-0.0101	0.0155	0.0136	0.0154	0.0136
2	GMM	0.0923	0.0139	0.0189	0.0239	0.0943	0.0276
	GMM MA(1)	-0.0021	0.0009	0.0266	0.0309	0.0266	0.0309
	MLE	0.1210	-0.0533	0.0131	0.0168	0.1210	0.0533
	MLE MA(1)	-0.0123	-0.0085	0.0155	0.0157	0.0161	0.0144
3	GMM	0.1046	0.0383	0.0196	0.0280	0.1064	0.0474
	GMM MA(1)	0.0033	0.0036	0.0272	0.0353	0.0274	0.0355
	MLE	0.1685	-0.0574	0.0136	0.0200	0.1685	0.0574
	MLE MA(1)	-0.0200	0.0115	0.0146	0.0147	0.0212	0.0149
4	GMM	-0.0142	-0.0028	0.0206	0.0257	0.0250	0.0258
	GMM MA(1)	-0.0388	0.0114	0.0397	0.0384	0.0555	0.0400
	MLE	-0.0060	-0.0183	0.0162	0.0189	0.0138	0.0217
	MLE MA(1)	-0.0140	-0.0153	0.0209	0.0193	0.0203	0.0199
5	GMM	0.1464	0.0406	0.0198	0.0334	0.1477	0.0525
	GMM MA(1)	0.0001	-0.0002	0.0302	0.0449	0.0302	0.0449
	MLE	0.1662	-0.0689	0.0149	0.0229	0.1662	0.0689
	MLE MA(1)	-0.0120	-0.0119	0.0188	0.0217	0.0181	0.0200
6	GMM	0.1502	0.1025	0.0193	0.0426	0.1514	0.1109
	GMM MA(1)	0.0047	0.0064	0.0272	0.0532	0.0275	0.0535
	MLE	0.2114	-0.0546	0.0147	0.0272	0.2114	0.0550
	MLE MA(1)	-0.0227	0.0275	0.0171	0.0210	0.0241	0.0292
7	GMM	-0.0069	0.0031	0.0139	0.0172	0.0155	0.0175
	GMM MA(1)	-0.0159	0.0074	0.0244	0.0222	0.0291	0.0234
	MLE	-0.0015	-0.0043	0.0117	0.0115	0.0095	0.0097
	MLE MA(1)	-0.0018	-0.0041	0.0129	0.0121	0.0105	0.0100
8	GMM	0.0624	-0.0108	0.0174	0.0245	0.0648	0.0268
	GMM MA(1)	0.0063	-0.0006	0.0311	0.0272	0.0317	0.0272
	MLE	0.0998	-0.0717	0.0121	0.0137	0.0998	0.0717
	MLE MA(1)	-0.0058	-0.0013	0.0140	0.0140	0.0123	0.0112
9	GMM	0.0701	-0.0004	0.0178	0.0312	0.0723	0.0312
	GMM MA(1)	0.0162	0.0026	0.0370	0.0328	0.0404	0.0329
	MLE	0.1435	-0.0977	0.0137	0.0163	0.1435	0.0977
	MLE MA(1)	-0.0198	0.0163	0.0138	0.0139	0.0207	0.0179

Tab. 3: Simulation Results ($\alpha = 0.8$ and $\beta = 0.2$)

Simulation Design		Bias α	Bias β	Std α	Std β	RMSE α	RMSE β
10	GMM	-0.0672	-0.0142	0.0391	0.0202	0.0778	0.0247
	GMM MA(1)	-0.0998	-0.0116	0.0628	0.0252	0.1179	0.0277
	MLE	-0.0175	-0.0200	0.0128	0.0113	0.0183	0.0204
	MLE MA(1)	-0.0236	-0.0194	0.0150	0.0113	0.0242	0.0198
11	GMM	-0.0693	0.0161	0.0473	0.0297	0.0839	0.0338
	GMM MA(1)	-0.0296	-0.0025	0.0431	0.0303	0.0523	0.0304
	MLE	0.0711	-0.0190	0.0117	0.0162	0.0711	0.0208
	MLE MA(1)	-0.0156	-0.0157	0.0156	0.0148	0.0181	0.0178
12	GMM	-0.0643	0.0159	0.0462	0.0295	0.0791	0.0335
	GMM MA(1)	-0.0272	-0.0025	0.0412	0.0301	0.0494	0.0302
	MLE	0.0900	-0.0082	0.0119	0.0197	0.0900	0.0170
	MLE MA(1)	-0.0184	0.0038	0.0154	0.0147	0.0200	0.0122
13	GMM	-0.0617	-0.0120	0.0375	0.0273	0.0721	0.0298
	GMM MA(1)	-0.0884	-0.0103	0.0555	0.0365	0.1044	0.0379
	MLE	-0.0100	-0.0252	0.0162	0.0164	0.0153	0.0260
	MLE MA(1)	-0.0159	-0.0245	0.0185	0.0164	0.0200	0.0254
14	GMM	-0.0075	0.0724	0.0358	0.0375	0.0366	0.0815
	GMM MA(1)	-0.0204	-0.0002	0.0376	0.0426	0.0428	0.0426
	MLE	0.0878	-0.0123	0.0134	0.0225	0.0878	0.0205
	MLE MA(1)	-0.0100	-0.0171	0.0170	0.0205	0.0158	0.0215
15	GMM	-0.0644	0.1248	0.0387	0.0510	0.0751	0.1348
	GMM MA(1)	-0.0110	0.0095	0.0327	0.0511	0.0345	0.0520
	MLE	0.1009	0.0208	0.0133	0.0293	0.1009	0.0289
	MLE MA(1)	-0.0159	0.0222	0.0163	0.0218	0.0186	0.0255
16	GMM	-0.0226	-0.0048	0.0200	0.0169	0.0302	0.0176
	GMM MA(1)	-0.0352	-0.0068	0.0321	0.0205	0.0476	0.0216
	MLE	-0.0055	-0.0086	0.0109	0.0083	0.0099	0.0099
	MLE MA(1)	-0.0067	-0.0083	0.0115	0.0083	0.0108	0.0097
17	GMM	-0.0346	0.0177	0.0296	0.0253	0.0455	0.0309
	GMM MA(1)	-0.0146	-0.0023	0.0317	0.0253	0.0349	0.0254
	MLE	0.0508	-0.0242	0.0112	0.0114	0.0508	0.0244
	MLE MA(1)	-0.0082	-0.0089	0.0135	0.0113	0.0127	0.0118
18	GMM	-0.0879	0.0293	0.0369	0.0382	0.0953	0.0482
	GMM MA(1)	-0.0108	0.0023	0.0311	0.0305	0.0329	0.0306
	MLE	0.0704	-0.0271	0.0116	0.0135	0.0704	0.0274
	MLE MA(1)	-0.0156	0.0029	0.0141	0.0124	0.0173	0.0100

4. Empirical Example

In this section we estimate the Beveridge curve relationship for West Germany with the first difference GMM and the transformed maximum likelihood estimator and account for possible serial correlation as presented in the previous sections. The Beveridge curve represents a steady-state relationship between the unemployment rate and the vacancy rate that can be derived from a matching function that approximates the trading frictions on the labour market due to a time and cost consuming search process. The existence of an non-instantaneous job matching involves a negative relationship between the unemployment and the vacancy rate that can be investigated with aggregate data. Beveridge curves were estimated by a number of authors, e.g. Jackman, Pissarides, and Savouri (1990) for several countries (see Petrongolo and Pissarides (2001) for an overview of the empirical studies).

To estimate the Beveridge curve for West Germany we use a regional data set consisting of 141 West German adminstration areas of the Federal Employment Services.[5] For these 141 regions we have monthly data from Jan. 2002 up to Dec. 2003 at hand. The empirical model is given by

$$\ln u_{it} = \alpha \ln u_{it-1} + \beta \ln v_{it-1} + \mu_i + \lambda_t + \xi_{it}, \tag{33}$$

where u_{it} denotes the unemployment rate and v_{it} the vacancy rate relative to the labour force. Furthermore the model includes a regional specific constant term μ_i, a time specific constant term λ_t and a possible serial correlated residual term ξ_{it} varying over i and t. For all estimations the time specific constant term is removed by the transformation $y_{it} - \frac{1}{N}\sum_{i=1}^{N} y_{it}$. The vacancies v_{it} are assumed to be endogenous with $E(\xi_{it} v_{is}) \neq 0$ for $s \geq t$. With this assumption the vacancies are determined by the current and lagged unemployment rate but not by future values. To avoid a simultaneity bias the vacancies enter equation (33) with one lag and therefore can be handled as weakly exogenous regressors. Furthermore, in the case of monthly data it is reasonable to assume that the current unemployment rate is determined by the state of the labour market at the end of the previous month. Following Jackman, Pissarides, and Savouri (1990) we specify a dynamic model in order to account for the persistence of the unemployment rate.

For the calculation of the first difference GMM estimator we utilise the following moment conditions

$$E(\Delta\xi_{it} u_{it-s}) = 0 \quad \text{for } s = 2+k, 3+k, \cdots, t-1 \tag{34}$$

$$E(\Delta\xi_{it} v_{it-s}) = 0 \quad \text{for } s = 2+k, 2+k, \cdots, t-1, \tag{35}$$

where the choice of k depends then on the assumed order of serial correlation. That is, if we assume there is no serial correlation, then we set $k = 0$ and for first or second order serial correlation k is set to 1 or 2 respectively. The relatively large time dimension leads to a huge amount of moment conditions if the whole history of the series as instruments is used. As mentioned by Arellano and Bond (1998) using too many instruments may lead to an (small sample) overfitting bias. For the calculation of the GMM estimators we therefore cut the history of the series as instruments at $t - 5$. Monte Carlo studies, see e.g. Arellano and Bond (1991) have often found

[5] Due to the immense differences between the East and West German labour market we restrict the analysis to West Germany.

Tab. 4: First Difference GMM Estimates[1]

	No Serial Correlation		MA(1) Process		MA(2) Process	
	Parameter	t-Value	Parameter	t-Value	Parameter	T-Value
u_{it-1}	0.8320	64.66	0.8612	62.53	0.8382	40.78
v_{it-1}	-0.0498	-5.12	-0.0464	-4.95	-0.0423	-3.92
Wald-Test (df)	4488.68	(2)	3974.73	(2)	1665.10	(2)
Sargan (df)	140.52	(162)	125.32	(118)	104.14	(76)
$m(1)$ (df)	-5.05	(141)	-5.61	(141)	-2.96	(141)
$m(2)$ (df)	-6.26	(141)	-6.36	(141)	-6.11	(141)
Hausman-Test (df)			4.70	(2)	4.38	(2)

Degrees of freedom for test statistics are in parenthesis.

[1] Two step estimates with corrected standard errors.

severely downward biased asymptotic standard errors of the two-step estimates in small samples. Therefore we present in the following analysis the two-step estimates with corrected standard errors as proposed by Windmeijer (2005).[6] This is because the expression for the asymptotic variance ignores the presence of the estimated one-step estimates in the weight matrix (Bond and Windmeijer, 2002). Monte Carlo results have shown that the corrected variance of the two-step estimator often provides more reliable inference with size proportions similar to those of the one-step variance, see Bond and Windmeijer (2002) and Windmeijer (2005). The calculation of the maximum likelihood estimator was done according to the presentations in section 2.2. The weak exogeneity assumption for v_{it-1} enables us to use Δv_{i0} in the linear projection for the initial condition Δu_{i1} with the additional parameter $\pi(\Delta v_{i0})$.[7] The variance of the initial condition is parameterised as $(1/\sigma_\varepsilon^2)\text{Var}(\Delta u_{i1}) = (1 - 1/T) + w^2$ to ensure a positive definite variance covariance matrix.

In the following analysis both estimators are applied under three assumptions. First ξ_{it} is not serial correlated, i.e. $\xi_{it} = \varepsilon_{it}$ where $\varepsilon_{it} \sim \text{iid}(0, \sigma_\varepsilon^2)$. Second ξ_{it} is generated from a MA(1) process with $\xi_{it} = \varepsilon_{it} + \theta_1\varepsilon_{it-1}$. And finally ξ_{it} follows a MA(2) process with $\xi_{it} = \varepsilon_{it} + \theta_1\varepsilon_{it-1} + \theta_2\varepsilon_{it-2}$. The derivation of the variance covariance matrix of the MLE in the case of an MA(2)-process can be found in the Appendix. Table 3 contains the results for the GMM estimator and table 4 contains the results for the MLE.

In the case we do not account for serial correlation the GMM estimator leads to an AR coefficient of 0.83 and the MLE to a coefficient of 0.88. For the vacancies both estimators find a significant negative coefficient, i.e. a negative relationship between unemployment and vacancies in West Germany. The relatively small coefficient for the vacancy rate primarily results from the fact that our data includes only those vacancies that are registered by the Federal Employment Services (see Franz and Smolny (1994)). Comparing GMM and MLE we find that GMM leads nearly to a twice as large coefficient for the vacancy rate as MLE. Turning to the test statistic for first and second order serial correlation (m(1) and m(2)) we find that both statistics reject the null hypothesis. Arellano and Bond (1991) suggested the m(1) and the m(2) statistic to test

[6] All results for the GMM estimator are computed using the DPD98 Software for GAUSS.

[7] Since time specific effects are removed the linear projection does not include the constant.

Tab. 5: Maximum Likelihood Estimates

	No Serial Correlation		MA(1) Process		MA(2) Process	
	Parameter	t-Value	Parameter	t-Value	Parameter	t-Value
u_{it-1}	0.8800	67.12	0.7978	51.50	0.7315	40.49
v_{it-1}	-0.0247	-9.56	-0.0236	-7.74	-0.0230	-6.90
θ_1			0.5292	34.16	0.6279	30.55
θ_2					0.2245	10.54
σ_ε	0.0271	74.14	0.0235	75.30	0.0231	75.97
w	0.4446	9.89	0.7213	13.63	0.8029	12.90
$\pi(\Delta v_{i0})$	-0.0297	-3.53	-0.0375	-3.33	-0.0427	-3.22
Log-Likelihood	6681.42		7098.08		7150.45	
AIC	-13350.84		-14184.15		-14284.89	
Wald-Test θ_1, θ_2 (df)			1166.88	(1)	938.21	(2)

Degrees of freedom for test statistics are in parenthesis.

whether the estimates suffer from serial correlation. m(1) tests for $E(\xi_{it}\xi_{it-1}) = 0$ and m(2) tests for $E(\xi_{it}\xi_{it-2}) = 0$, where both statistics are asymptotically normal distributed. As the consistency of the GMM estimator hinges heavily on the lack of second-order serial correlation, the m(2) statistic suggest that the results from the GMM may be problematic.

If we account for an MA(1) process within the GMM and the MLE method we find that the estimate for the autoregressive parameter increases for the GMM to 0.86 and decreases for the MLE to 0.79. The contrary movent of the GMM and MLE estimate fits to our results from the simulation experiments for high persistent processes. Regarding the vacancy rate we find only marginal changes in the estimates for both estimation methods. In the case we account for an MA(2) process the estimate for the autoregressive parameter from MLE reduces again to 0.73. For the GMM the autoregressive coefficient is similar to the GMM estimate if we do not account for serial correlation. However, as indicated by the Wald test excluding the instruments for $t-2$ and $t-3$ obviously leads to considerable efficiency losses of the GMM estimator. Again the changes in the coefficient of the vacancy rate are marginal, although we find throughout a slight reduction of the coefficient if we account for serial correlation. The results show that accounting for serial correlation primarily leads to changes in the estimate of the autoregressive parameter, where these changes are more distinctive for the MLE. Therefore primarily statements that are made with respect to the autoregressive parameter or the lag coefficients are affected by serial correlation. In contrast, the unemployment vacancy relationship remains robust whether the estimation method accounts for serial correlation or not.

To test whether serial correlation is problematic in our model, we apply a Hausman-Test for the difference between the usual GMM estimator and the GMM estimator that accounts for serial correlation. This test was suggested by Arellano and Bond (1991) and utilises the test statistic

$$(\hat{\delta}_I - \hat{\delta})'[\text{est.avar}(\hat{\delta}_I) - \text{est.avar}(\hat{\delta})]^{-}(\hat{\delta}_I - \hat{\delta}) \sim \chi^2_r, \tag{36}$$

where $\hat{\delta}$ is the GMM estimator calculated from the full set of instruments, $\hat{\delta}_I$ is the GMM estimator calculated form the reduced set of instruments and $r = \text{rank}(\text{est.avar}(\hat{\delta}_I - \hat{\delta}))$. If there were no serial correlation, both estimators would be consistent but only $\hat{\hat{\delta}}$ would be efficient since it uses all available moment conditions. The Hausman-Test rejects the null hypothesis for the GMM estimator that accounts for first order serial correlation on the 10 percent level, whereas for the GMM estimator that accounts for second order serial correlation the null hypothesis cannot be rejected.

For the MLE one may test the relevance of the additional parameters θ_i with a likelihood ratio test or a Wald test. A Wald-Test for the MA parameters is presented in the table. Similar to the t-Value the Wald-Test shows that the MA parameters are significant. A second approach is to consider the fit of the model if we account for additional MA terms. For our analysis we use the Akaike-Information-Criteria (AIC) and find the lowest value for the model with an MA(2) process. In this case one may continue to add MA terms to the model until the AIC is minimised although this may be a burdensome procedure. Accounting for higher order serial correlation within the methodology of the GMM estimator is also not advisable, since the predictive power of the instruments can be expected to decrease as the lag of the instruments increases. Therefore, if higher order serial correlation seems to be present, it is advisable to consider a respecification of the model.

5. Conclusion

Regional panel data used for empirical macroeconomic research usually provides larger time series compared to typical panel data for individuals or firms. Therefore, time series issues like serial correlation are an important issue in order to obtain consistent estimates in dynamic panel data models with fixed effects and weakly exogenous regressors. Accounting for serially correlated errors within the first difference GMM framework (Arellano and Bond, 1991) may be problematic for two major reasons: First, the size of the sample in the time dimension diminishes in a drastic way for higher order serial correlation. Second, large efficiency losses may prohibit a reliable analysis on the basis of the results from the first differences GMM estimator in the presence of serial correlation. Alternatively, serial correlation can be accounted for in the framework of the transformed maximum likelihood estimator suggested by Hsiao, Pesaran, and Tahmiscioglu (2002). Directly allowing for serial correlation due to a MA process leads to an asymptotic normal and efficient estimator.

To evaluate the finite sample properties of the first difference GMM estimator and the transformed maximum likelihood estimator in the presence of serially correlated errors we conducted a simulation experiment. The results show that the presence of serial correlation leads to a severe finite sample bias. By accounting for serial correlation we find a superior performance of the maximum likelihood estimator relative to the GMM estimator. That is, the explicit allowance for MA terms within the MLE framework leads to a considerable gain in precision of the parameter estimates. Moreover, we find that the advantage of the MLE over the GMM increases with a higher autoregressive coefficient. With respect to the coefficients for the explanatory variables, we generally find smaller differences between MLE and GMM.

To assess the behaviour of the maximum-likelihood estimator and the GMM estimator in an empirical application, we present an empirical example where the Beveridge curve relationship for West Germany is estimated with regional panel data. The results show that the consideration of serial correlation primarily changes the estimate of the autoregressive parameter, where these changes are more distinctive for the MLE. If serial correlation cannot be eliminated by a respecification of the model, GMM and MLE provide a viable approach to obtain consistent estimates. Although maximum likelihood is computationally more demanding than GMM, we find that the efficiency gains, even in small samples, are worth the difficulties.

A Appendix

A1. The Covariance Matrix for an ARMA(1,1) Process

Consider the model given in equation (1) with no exogenous variables. This model will be of limited practical value, but is interesting due to the fact that a exact likelihood function exists. After taking first differences we arrive at

$$\Delta y_{it} = \alpha \Delta y_{i(t-1)} + \Delta u_{it}, \tag{A.1}$$

where we assume that u_{it} is generated by an MA(1) process

$$u_{it} = \varepsilon_{it} + \theta \varepsilon_{i(t-1)}. \tag{A.2}$$

With the assumption 1, i.e. the process has been going on for a long time, i.e. $m \to \infty$ and $|\alpha| < 1$ the the mean is given by $E(\Delta y_{i1}) = 0$ and the covariance matrix $\sigma_\varepsilon^2 \mathbf{\Omega}^* = \mathbf{\Omega}$ is given by

$$\begin{aligned}
E(\Delta y_{i1}^2) &= \sigma_\varepsilon^2 \left\{ 1 + [(\theta - 1) + \alpha]^2 + \frac{\phi^2}{1-\alpha^2} \right\} && \text{(A.3)}\\
E(\Delta y_{i1} \Delta u_{i2}) &= -\sigma_\varepsilon^2 [(1-\theta)^2 + \theta\alpha] && \text{(A.4)}\\
E(\Delta y_{i1} \Delta u_{i3}) &= -\sigma_\varepsilon^2 \theta && \text{(A.5)}\\
E(\Delta y_{i1} \Delta u_{it}) &= 0 \;\; \text{for } t = 4, \ldots, T. && \text{(A.6)}
\end{aligned}$$

Under assumption 2, i.e. m is finite and $E(\Delta y_{i1}) = b$ is the same across individuals the variance is given by

$$E(v_{i1}^2) = \sigma_\varepsilon^2 \left[1 + [(\theta - 1) - \alpha]^2 + \alpha^{2(m-2)} (\alpha(\theta - 1) - \theta)^2 - \alpha^{2(m-1)} \theta^2 + \phi^2 \frac{1 - \alpha^{2(m-3)}}{1 - \alpha^2} \right], \tag{A.7}$$

with $v_{i1} = \Delta y_{i1} - b$, $\phi = \alpha^2 + \alpha(\theta - 1) - \theta$ and the covariances correspond to (A.4-A.6). The remaining entries of the covariance matrix are given in 2.2.. These moments follow directly from the assumption about the covariance structure for u_{it} and $u_{i(t-s)}$. Note, that even though $E(v_{i1}^2)$ is depending on α, we can treat it as a free parameter (see Hsiao, Pesaran, and Tahmiscioglu (2002)). In addition, note that the difference between the covariance matrix for a model with and without regressors consists basically in the term for the residual regarding the initial condition.

A2. The Covariance Matrix for the MLE with a MA(2) Residual

In order to demonstrate how the model can be generalized to higher order MA processes we consider the case in which the residual is generated by

$$u_{it} = \varepsilon_{it} + \theta_1 \varepsilon_{i(t-1)} + \theta_2 \varepsilon_{i(t-2)}. \tag{A.8}$$

Under assumption 1 the following entries for the matrix $\mathbf{\Omega}$ result

$$\begin{aligned}
E(\Delta y_{i1}^2) &= \sigma_\varepsilon^2 \left\{ 1 + [(\theta_1 - 1) + \alpha]^2 + [(\theta_2 - \theta_1) + \alpha(\theta_1 - 1) + \alpha^2]^2 + \frac{\phi^2}{1-\alpha^2} \right\} && \text{(A.9)}\\
E(\Delta y_{i1} \Delta u_{i2}) &= -\sigma_\varepsilon^2 \left[1 + (\theta_2 - \theta_1)^2 + (1 - \alpha)(\theta_2 - \theta_1) + \alpha\theta_2(\theta_1 + \alpha - 1) - \theta_1 \right] && \text{(A.10)}\\
E(\Delta y_{i1} \Delta u_{i3}) &= -\sigma_\varepsilon^2 [\theta_1 + \theta_2(\theta_1 + \alpha - 2)] && \text{(A.11)}\\
E(\Delta y_{i1} \Delta u_{i4}) &= -\sigma_\varepsilon^2 \theta_1 && \text{(A.12)}\\
E(\Delta y_{i1} \Delta u_{ik}) &= 0 \;\; k \geq 5, && \text{(A.13)}
\end{aligned}$$

with $\phi = \alpha^3 + \alpha^2(\theta_1 - 1) + \alpha(\theta_2 - \theta_1) - \theta_2$. The entry for $E(\Delta y_{i1}^2)$ can easily be modified to hold under assumption 2. The remaining covariance terms are

$$\begin{aligned}
E[(\Delta u_{it})^2] &= \sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2 - \theta_1 - \theta_1\theta_2)2 && \text{(A.14)}\\
E[\Delta u_{it}\Delta u_{i(t-1)}] &= -\sigma_\varepsilon^2[1 + (\theta_1 - \theta_2)^2 + \theta_2 - 2\theta_1] && \text{(A.15)}\\
E[\Delta u_{it}\Delta u_{i(t-2)}] &= \sigma_\varepsilon^2(2\theta_2 - \theta_2\theta_1 - \theta_1) && \text{(A.16)}\\
E[\Delta u_{it}\Delta u_{i(t-3)}] &= -\sigma_\varepsilon^2\theta_2 && \text{(A.17)}\\
E[\Delta u_{it}\Delta u_{i(t-k)}] &= 0 \text{ for } k \geq 4, && \text{(A.18)}
\end{aligned}$$

where the variance of the error for the initial condition is again treated as a free parameter and all other conclusion given in A1. remain valid.

References

AHN, S., AND P. SCHMIDT (1995): "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5–27.

ARELLANO, M., AND S. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

——— (1998): "Dynamic Panel Data Estimation using DPD98 for Gauss," http://www.ifs.org.uk/staff/steve b.shtml.

BHARGAVA, A., AND J. SARGAN (1983): "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods," *Econometrica*, 51,No.6, 1635–1660.

BINDER, M., C. HSIAO, AND M. PESARAN (2002): "Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration," Working Paper.

BLUNDELL, R., AND S. BOND (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115–143.

BLUNDELL, R., S. BOND, AND F. WINDMEIJER (2000): "Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimators," Working Paper 00/12, The Institute for Fiscal Studies.

BOERI, T., AND M. BURDA (1996): "Active Labor Market Policies, Job Matching and the Czech Miracle," *European Economic Review*, 40, 805–817.

BOND, S., AND F. WINDMEIJER (2002): "Finite Sample Inference for GMM Estimators in Linear Panel Data Models," Cemmap Working Paper cwp04/02, The Institute for Fiscal Studies.

CALMFORS, L., AND P. SKEDINGER (1995): "Does Active Labour Market Policy Increase Employment? Theoretical Considerations and Some Empirical Evidence from Sweden," *Oxford Review of Economic Policy*, 11 (1), 91–109.

FISCHER, S. (1993): "The Role of Macroeconomic Factors in Growth," *Journal of Monetary Economics*, pp. 485–512.

FRANZ, W., AND W. SMOLNY (1994): "The Measurement and Interpretaion of Vacancy Data and the Dynamics of the Beveridge Curve: The German Case," in *Measurement and Analysis of Job Vacancies*, ed. by J. Muysken. Aldershot.

HAMILTON, J. D. (1994): *Time Series Analysis.* Princeton University Press, Princeton, New Jersey.

HANSEN, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hsiao, C., M. Pesaran, and A. Tahmiscioglu (2002): “Maximum Likelihood Estimation of fixed Effects Dynamic Panel Data Models Covering Short Time Periods,” *Journal of Econometrics*, 109, 107–150.

Hujer, R., and C. Zeiss (2005): “Macroeconomic Impacts of Job Creation Schemes on the Matching Process in West Germany,” Working Paper, J.W.Goethe-University, Frankfurt.

Jackman, R., C. Pissarides, and S. Savouri (1990): “Labour Market Policies and Unemployment in the OECD,” *Economic Policy*, 5, 450–490.

Judson, A., and Owen (1999): “Estimating Dynamic Panel Data Models: A Practical Guide for Macroeconomists,” *Economics Letters*, 65, 9–15.

Keane, M., and D. Runkle (1992): “On the Estimation of Panel Data Models with Serial Correlation when Instruments are not Strictly Exogenous,” *Journal of Business and Economic Statistics*, 10, No. 1.

Kiviet, J. (1995): “On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models,” *Journal of Econometrics*, 68, 53–78.

Levine, R., and D. Renelt (1992): “A Sensitivity Analysis of Cross-Country Growth Regressions,” *American Economic Review*, pp. 942–963.

Mankiw, G. N., D. Romer, and D. N. Weil (1992): “A Contribution to the Empirics of Economic Growth,” *Quarterly Journal of Economics*, pp. 407–438.

Neyman, J., and E. L. Scott (1948): “Consistent Estimates Based on Partially Consistent Observations,” *Econometrica*, 16, 1–32.

Petrongolo, B., and C. Pissarides (2001): “Looking into the Black Box: A Survey of the Matching Function,” *Journal of Economic Literature*, XXXIX (June), 309–431.

Sevestre, P., and A. Trognon (1996): “Dynamic Linear Models,” in *The Econometrics of Panel Data. A Handbook of the Theory with Applications*, ed. by L. Mátyás, and P. Sevestre, pp. 120–144. Kluwer Academic Publishers, Dordrecht, 2nd edn.

Windmeijer, F. (2005): “A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators,” *Journal of Econometrics*, 126, 25–51.