Dynamic Panel Data Models with Unobserved Groupings and Factor Structure

Bachelor Thesis in
Econometrics and Operations Research

Nando Vermeer
447763

Supervisor: Wang, W.
Second assessor: Naghi, A.A.

Erasmus School of Economics
ERASMUS UNIVERSITY ROTTERDAM
July 7th, 2019

Contents

Abstract

In this paper we extend the panel data model proposed by Ando and Bai (2016) by including an autoregressive component. In order to estimate the model parameters, the original estimation method was adapted in two different ways, a direct approach and an iterative approach. Of these, the iterative approach resulted in consistent estimation of the model parameters, while the direct approach had problems with convergence. By means of Monte Carlo simulation, we empirically show that independence over time of the error terms is necessary for consistent estimation of the model parameters. However, dependence over the different dependent variables of the error terms is allowed. By a similar method, it is shown that the model is robust against overspecification of the true number of lags.

1 Introduction

The analysis of panel data has been a central subject in the field of econometrics since its inception. Because of this, many different models have been proposed to capture the relations between certain variables, the most famous being linear regression. While still frequently used in many situations, this model does not fit all data. Sometimes, the relationships between the dependent variables and the the explanatory variables are more complicated. For example, the dependent variables could also depend on unobserved factors. In order to better study which dependent variable is depended on which unobserved factors, the dependent variables can be grouped based on their dependence on these factors. Such a model is introduced in Ando and Bai (2016), where they consider grouped panel data with unobserved groups and factors.

After deriving the properties of this model, they performed a Monte Carlo simulation to show that their model was able to correctly find the number of groups, as well as estimate the model parameters very well. After this, the model described above was fitted on panel data of US mutual funds, as well as panel data of China's mainland stock market. The main purpose of this analysis seems to be to identify the different groups that the panel data consists of, as well as the factors which drive the dependent variables in each group.

In this paper we replicate some of the results given in Ando and Bai (2016), as well as extend them. The main extension that we propose consist of equipping the model of Ando and Bai (2016) with an underlying autoregressive structure. This is interesting for a variety of reasons. This extension is important in a theoretical sense, since this specific extension has not been studied before, but it is also interesting to consider the empirical applications. A practical example might be the analysis of a macroeconomic panel data set. It has been shown that macroeconomic time series often have a significant autoregressive component (Meese (1984), Bagliano (2009), An (2014)). Furthermore, it could also be argued that certain macroeconomic variables move together. Since the GDP of Germany likely has a higher correlation with the unemployment of Germany than the unemployment in Brazil, it might be wise to consider different groups which are moved by different

factors. Because of these reasons, we justify the proposed extension.

In order to investigate if this extension is practical, we first need to consider a few sub questions. One of these questions concerns the consistency of parameter estimation, and if this still holds under the proposed structure. Because of the extra model parameters, we may have to make certain adjustments to the estimation procedure, as well as the assumptions in order for the parameter estimation to be consistent. Furthermore, it might also be of interest to consider what happens if the number of lags are misspecified, and if there is evidence that, under these conditions, consistent estimation still holds.

2 Literature

In order to obtain a better grasp on the model we investigate, we consider the literature on the subject. Here we mainly consider papers which are closely related to Ando and Bai (2016), as well as papers concerning dynamic panel data models.

First of all, let us consider the paper by Bai (2009), which is closely related to the paper by Ando and Bai (2016). This paper introduces a panel data model with known regressors, as well as a unknown factor structure. In order to estimate the regression coefficients and the factor structure, Bai (2009) proposes an iterative estimation procedure, where the regression coefficients and the factor structure are estimated separately from each other. They also note that this iterative method has a higher rate of convergence compared to a joined estimation method. We use this fact in the construction of one of the estimation methods we propose for our Dynamic Model.

The model proposed by Ando and Bai (2016) can be seen as an extension to the model given in Bai (2009). While the dependent variable still contains a regressive component, as well as a factor structure, it is now also assumed that each dependent variable is a member of a certain group. Here each group has a certain number of factors, which moves each individual dependent variable of that group. The model that the paper of Bai (2009) proposes can be seen as a restricted version of this model, where the number of groups is set to 1.

Concerning the dynamic part of the extension, it might be of interest to also consider some of the literature that is written on this specific subject. Kiviet (1995) gives an overview of the standard dynamic panel data model, and analyzes the finite sample properties of different estimation procedures for the unknown elements. We use this paper for inspiration concerning the assumptions which are needed in order to obtain desirable estimation properties.

Since we need both the dynamic structure and a model similar in nature to that of Ando and Bai (2016), it would be wise to consider the papers already written on such a combination. A paper which achieves a model similar to the one that is studied in our extension is Moon and Weidner (2017). Here they analyze how the model from Bai (2009) handles the effect of regressors with interactive fixed effects, for example autoregressors. They show that estimation is asymptotically biased, and provide bias corrected versions of the Wald, LR and LM test statistics. Note that this model differs from our extension, since they consider the case where all dependent variables are

placed in a single group, instead of allowing multiple groups. This paper originally gave us the inspiration for the extension, as well as provide a general idea of how to achieve the addition of a dynamic component.

3 Methodology

3.1 Original Model

3.1.1 Model Introduction

Let us first introduce the main model of this paper, which is the same as in Ando and Bai (2016). Consider N dependent variables, named $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$, such that $\boldsymbol{y}_i$ is a $T \times 1$ vector. We assume that each of these dependent variables belongs to exactly one group. This is denoted as $g_i = j$, which means that dependent variable $\boldsymbol{y}_i$ belongs to group j, where $j \in \{1, \ldots, S\}$. Given this classification, we assume that each dependent variable $y_{i,t}$ can be expressed as

$$y_{i,t} = \boldsymbol{x}_{i,t}'\boldsymbol{\beta} + \boldsymbol{f}_{g_i,t}'\boldsymbol{\lambda}_{g_i,i} + \epsilon_{i,t}, \quad i = 1, \ldots, N, t = 1, \ldots, T \tag{1}$$

With $\boldsymbol{x}_{i,t}$ a $p \times 1$ vector of observable regressors, $\boldsymbol{f}_{g_i,t}$ the $r_j \times 1$ vector of factors for group g_i at time t, and $\boldsymbol{\lambda}_{g_i,i}$ the $r_j \times 1$ vector of factor loadings for stock i. In order to differentiate between this model and the extended model, we refer to this model as the **Orginal Model**. For a range of reasons, it might be desirable to write this model in matrix notation, which is often used in the rest of this paper. Note that the model given in (1) can also be written as follows:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{F}_{g_i}\boldsymbol{\lambda}_{g_i,i} + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, N, t = 1, \ldots, T \tag{2}$$

$$\boldsymbol{y}_i = \begin{pmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,T} \end{pmatrix}, \boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{x}_{i,1}' \\ \boldsymbol{x}_{i,2}' \\ \vdots \\ \boldsymbol{x}_{i,T}' \end{pmatrix}, \boldsymbol{F}_j = \begin{pmatrix} \boldsymbol{f}_{j,1}' \\ \boldsymbol{f}_{j,2}' \\ \vdots \\ \boldsymbol{f}_{j,T}' \end{pmatrix}, \boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{i,1} \\ \epsilon_{i,2} \\ \vdots \\ \epsilon_{i,T} \end{pmatrix}$$

3.1.2 Estimating the Model Parameters

For now, we assume that the number of groups, as well as the respective number of factors for each group are already found. How these can be estimated is described in section 3.1.3. In order to find estimates for the rest of the model parameters, Ando and Bai (2016) propose an iterative approach where the regression coefficients, the factor structure and the group membership are updated independently of each other. This algorithm can be described in the following way:

1. First initial estimates for the group membership, the factor structure and the regression coefficients are found.

2. Given the regression coefficients and the factor structure, update the group membership.

3. Given the regression coefficients and the group membership, update the factor structure.

4. Given the group membership and the factor structure, update the regression coefficients.

5. Repeat the three steps above until convergence.

Let us first consider the way we update the group membership, given the regression coefficients and the factor structure. Given both the regression coefficients and factor structure, we appoint each dependent variable to the group that minimizes the average norm of the error terms. Note that we also have to estimate the factor loadings for these dependent variables, which is done by means of least squares (Heij et al, 2004). In mathematical notation, this is expressed as follows:

$$\begin{aligned}
&\widehat{g}_i = \frac{1}{T} \arg\min_{j\in\{1,\ldots,S\}} ||\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{F}}_j\widehat{\boldsymbol{\lambda}}_{j,i}||^2 \\
&\widehat{\boldsymbol{\lambda}}_{j,i} = \boldsymbol{F}_j'(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}})/T
\end{aligned}$$

Now the updating of the factor structure is considered, given the group membership and the regression coefficients. Since the betas are kept constant at this stage of estimation, we can subtract $\boldsymbol{x}_{i,t}'\boldsymbol{\beta}$ from both sides in (2) to obtain

$$\begin{aligned}
&\boldsymbol{w}_{j,i} \equiv \boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} = \boldsymbol{F}_{g_i}\boldsymbol{\lambda}_{g_i,i} + \boldsymbol{\epsilon}_i \\
&g_i = j
\end{aligned}$$

It can be seen that $\boldsymbol{w}_{j,i}$ now has a pure factor structure, and by principal components, the factors as well as the factor loadings can be obtained. If $\boldsymbol{W}_j = [\boldsymbol{w}_{j,1}, \ldots, \boldsymbol{w}_{j,N_j}]$, then the estimated factors $\widehat{\boldsymbol{F}}_j$ can be obtained as $\sqrt{T}$ times the eigenvectors corresponding to the largest r_j eigenvalues of $\boldsymbol{W}_j\boldsymbol{W}_j'$. Furthermore, the factor loadings $\boldsymbol{\Lambda}_j$ can be estimated as $\widehat{\boldsymbol{\Lambda}}_j = \boldsymbol{W}_j'\widehat{\boldsymbol{F}}_j/T$. For a more in depth derivation, we refer to Bai and Ng, (2002).

After updating the factor structure, the regression coefficients must be updated. This is done by performing a penalized regression, where the penalty function is set equal to the SCAD penalty function described in Fan and Li, (2001). The objective function that has to be minimized can then be written as follows:

$$L_{NT}\left(\boldsymbol{\beta}, \boldsymbol{G}, \boldsymbol{F}_1, \ldots, \boldsymbol{F}_S, \boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_S\right) = \sum_{j=1}^{S}\sum_{i:g_i=j} \|\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{F}_{g_i}\boldsymbol{\lambda}_{g_i,i}\|^2 + NT\cdot\sum_{j=1}^{p} p_{\kappa,\gamma}(|\beta_j|)$$

Where we define $p_{\kappa,\gamma}(|\beta_j|)$ as

$$p_{\kappa,\gamma}\left(|\beta_j|\right) = \begin{cases} \kappa\,|\beta_j| & (|\beta_j| \le \kappa) \\ \frac{\gamma\kappa|\beta_j| - 0.5\left(\beta_j^2+\kappa^2\right)}{\gamma-1} & (\kappa < |\beta_j| \le \gamma\kappa) \\ \frac{\kappa^2\left(\gamma^2-1\right)}{2(\gamma-1)} & (\gamma\kappa < |\beta_j|) \end{cases}$$

The function L_{NT} can be minimized by making use of a property derived by Fan and Li, (2001), where we iteratively perform the following ridge regression:

$$\begin{aligned}
&\boldsymbol{\beta}^{(k)} = \left\{\mathbf{X}'\mathbf{X} + NT \cdot \Sigma_\lambda\left(\boldsymbol{\beta}^{(k-1)}\right)\right\}^{-1}\mathbf{X}'\mathbf{z} \\
&\mathbf{X} \equiv [\boldsymbol{X}_1', \ldots, \boldsymbol{X}_n']' \\
&\boldsymbol{z} \equiv [\boldsymbol{y}_1', \ldots, \boldsymbol{y}_n']' \\
&\Sigma_\lambda(\boldsymbol{\beta}) = \operatorname{diag}\left\{p_\lambda'(|\beta_1|)/|\beta_1|, \ldots, p_\lambda'(|\beta_p|)/|\beta_p|\right\}
\end{aligned}$$

Let us also consider the parameters of the penalty function. Similar to Fan and Li, (2001), as well as Ando and Bai (2016), we set $\gamma = 3.7$ to minimize the Baysian risk of the estimator.

These three steps are then repeated until convergence of the parameters is reached. While the exact definition of convergence is not discussed in Ando and Bai (2016), we interpret it as an convergence in the average norm of the error terms.

The last part that is left to discuss is the initialization of the parameters in order to kick-start the iterative process. The first aspect which is initialized is the group specification, which is done by utilization of the K-Means algorithm (MacQueen, 1967). Subsequently the regression coefficients are estimated using the algorithm specified above, and setting the factor structure equal to zero since it is unknown. After having found initial estimates of both the groups and the regression coefficients, the factor structure is estimated using the algorithm as specified above.

3.1.3 Estimating the Number of Groups and Factors

As stated in the previous section, the model parameters are estimated under the assumption that the number of groups and factors are known. However, in practice this structure has to be estimated as well. This is done using the same method as in Ando and Bai (2016), which utilizes a criterion function which has to be minimized. The criterion function of choice is a generalized C_p criterion of Mallows (1973), which is based on the expected mean squared error of the model. The C_p criterion is therefore given as

$$\begin{aligned}
C_p(S, k_1, \ldots, k_S, \kappa) =& \frac{1}{NT}\sum_{j=1}^{S}\sum_{i;\hat{g}_i=j}\left\|\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{F}}_{\hat{g}_i}\hat{\boldsymbol{\lambda}}_{\hat{g}_i,i}\right\|^2 \\
&+ \frac{1}{TN}\operatorname{tr}\left[\boldsymbol{K}_x\boldsymbol{V}_\beta\left(\hat{\boldsymbol{F}}_1, \ldots, \hat{\boldsymbol{F}}_S, \kappa\right)\right] + \sum_{j=1}^{S} k_j\hat{\sigma}^2\frac{N_j}{N}\left(\frac{T+N_j}{TN_j}\right)\log(TN_j)
\end{aligned}$$

Where the first term is the estimated mean squared error for the model, the second term is equal to the estimation bias of the first term, and the last term consists of a penalty function to discourage overidentification. The $\hat{\sigma}^2$ term in the penalty function is an estimator of $(NT)^{-1}\sum_{j=1}^{S}\sum_{i:g_i^0=j}\|\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{F}}_{g_i^0}\hat{\boldsymbol{\lambda}}_{g_i^0,i}\|^2$, used to scale the penalty. In practice this term can be replaced with $(NT)^{-1}\sum_{j=1}^{S}\sum_{i;\hat{g}_i=j}\left\|\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{F}}_{\hat{g}_i}\hat{\boldsymbol{\lambda}}_{\hat{g}_i,i}\right\|^2$, estimated under the maximum

number of regressors, groups and factors.

In order to estimate the correct model specification, a search is performed over the possible values for $S, k_1, \ldots, k_S, \kappa$, where we pick $S, k_1, \ldots, k_S$ out of finite subsets of the natural numbers, and κ out a finite subset of the nonnegative real numbers.

While mathematically the search over the parameters is without issue, there does exist a practical issue that needs to be addressed. In order to guarantee an optimal specification based on the criterion, all combinations of the number of factors and groups have to be considered. This implies that the model has to be estimated many times, which is computationally expensive. If we define $S_{\max},\ k_{\max}$ respectively as the maximum number of groups and factors, and let $S \in \{1, \ldots, S_{\max}\}$, $k_i \in \{0, \ldots, k_{\max}\}$, the number of times the model has to be estimated can be expressed as

$$|\boldsymbol{\kappa}| \cdot \sum_{x=1}^{S_{\max}} (k_{\max} + 1)^x \tag{3}$$

Where $|\boldsymbol{\kappa}|$ is defined as the number of values κ is allowed to take. While estimating the model this many times is necessary to guarantee an optimum, by one assumptions we can find a heuristic which greatly reduces the number of times the model has to be estimated. The assumption can be stated as follows:

$$\begin{aligned}
&k_i^* = \arg\min_{k_i \in \{0,\ldots,k_{\max}\}} \{C_p(S, k_1, \ldots, k_S, \kappa)\} \implies \\
&k_i^* = \arg\min_{k_i \in \{0,\ldots,k_{\max}\}} \left\{C_p\left(S, \tilde{k}_1, \ldots, \tilde{k}_S, \kappa\right)\right\} \\
&\forall \tilde{k}_1, \ldots, \tilde{k}_{i-1}, \tilde{k}_{i+1}, \ldots, \tilde{k}_S \in \{0, \ldots, k_{\max}\}, \forall i \in \{1, \ldots, S\}
\end{aligned}$$

This condition guarantees that the number of factors can be optimized independently of each other. Using the assumptions above, we can iteratively optimize each k_i for a specific value for S and κ. This reduces the number of times the model has to be estimated to the following:

$$|\boldsymbol{\kappa}| \cdot \sum_{x=1}^{S_{\max}} (k_{\max} + 1) \cdot x = |\boldsymbol{\kappa}| \frac{S_{\max}(S_{\max} + 1)}{2} \cdot (k_{\max} + 1)$$

In order to illustrate the improvement, consider the following example: if we set $S_{\max} = 4,\ k_{\max} = 8$ and $|\boldsymbol{\kappa}| = 2$, then the number of times the model has to be estimated according to (3) is equal to 14760, while it is reduced to 180 by using the method described above. This reduction is especially desirable when performing Monte Carlo simulation, where this optimization has to be performed multiple times.

3.2 Dynamic Model

3.2.1 Model Introduction

Apart from considering the effects that the grouped factors and the regressors have on the dependent variables, it might also be of interest to consider the effect that previous values of

the dependent variable might have. A model that encapsulates this idea in the form of a linear relationship is called an autoregressive model (Das (1994)). An AR model of degree h regresses the current value of the dependent variable against the previous h values. In mathematical terms, this can be described as the following relationship:

$$y_{i,t} = \sum_{k=1}^{h} \phi_k y_{i,t-k} + \epsilon_{i,t}$$

If we want to equip the model specified in Ando and Bai (2016) with such a structure, it may be useful to add the structure to the right side of the equation, to obtain the following model specification:

$$y_{i,t} = \sum_{k=1}^{h} \phi_k y_{i,t-k} + \boldsymbol{x}'_{i,t}\boldsymbol{\beta} + \boldsymbol{f}'_{g_i,t}\boldsymbol{\lambda}_{g_i,i} + \epsilon_{i,t}, \quad i = 1, \dots, N,\ t = 1, \dots, T \tag{4}$$

Note that we assume that the model has homogeneous autoregressive coefficients. In other words, the effect that previous observations have on the current observation is constant over all dependent variables, and does not depend on the groupings. The model that is given in (4) will henceforth be referred to as the **Dynamic Model**. As is mentioned before, the added autoregressive coefficients have to be estimated as well as the rest of the unknown parameters. How this is done is described in the next section.

3.2.2 Estimating the Model Parameters

For the Dynamic Model described in section 3.2.1, some alterations have to be made to the estimation procedure, since the autoregressive structure also has to be estimated. Because of this, we propose two different estimation methods. How these will be tested is described in the Monte Carlo section. For now we assume that the autoregressive structure is homogeneous over the groups, however a very similar estimation procedure can be applied to a Dynamic Model with heterogeneous coefficients, for which the same arguments should hold.

The first estimation method makes use of the linear relationship between the current and previous dependent variables, which is similar to that of the regressors. Let us first introduce some notation. Similar to the standard AR(h) models, we have to remove the first h values from the dependent variable because of the autoregressive nature. We therefore define

$$\boldsymbol{y}_i^{(j,k)} = \begin{bmatrix} y_{i,j} \\ \vdots \\ y_{i,k} \end{bmatrix}$$

such that the full model can be written in matrix notation as

$$\boldsymbol{y}_i^{(h+1,N)} = \sum_{k=1}^{h} \boldsymbol{y}_i^{(h-k+1,N-k)} \phi_k + \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{F}_{g_i} \boldsymbol{\lambda}_{g_i,i} + \boldsymbol{\epsilon}_i$$

Now observe that we can use block matrix notation to write the model as follows:

$$\begin{aligned}
&\boldsymbol{y}_i^{(h+1,N)} = \widetilde{\boldsymbol{X}}_i \widetilde{\boldsymbol{\beta}} + \boldsymbol{F}_{g_i} \boldsymbol{\lambda}_{g_i,i} + \boldsymbol{\epsilon}_i \\
&\widetilde{\boldsymbol{X}}_i \equiv \begin{bmatrix} \boldsymbol{y}_i^{(h,N-1)} & \dots & \boldsymbol{y}_i^{(1,N-h)} & \boldsymbol{X}_i \end{bmatrix} \\
&\widetilde{\boldsymbol{\beta}} \equiv \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_h \\ \boldsymbol{\beta} \end{bmatrix}
\end{aligned}$$

Using this notation, we can estimate the model parameters in the same fashion as we did in section 3.1.2. However, we do note that one of the assumptions that is necessary for consistency does not hold for this model. The assumption in question is C6 from Ando and Bai (2016), which is stated as follows: $\epsilon_{i,t}$ is independent of $\boldsymbol{x}_{i,s}$, $\boldsymbol{\lambda}_{j,i}$ and $\boldsymbol{f}_{j,s}$ for all i, j, s, t. However, since $\epsilon_{i,t}$ is allowed to be dependent over t, this assumption is violated. For an example why this is the case, consider the fact that an AR(1) model can be written as a special instance of the Dynamic Model we propose, for which estimation is not consistent if we allow for time dependence.

In order to solve this problem, it is necessary to add another assumption. Specifically, we assume that every $\epsilon_{i,t}$ is generated independently over time. Note that this does not imply that there exists no correlation between the error terms at time t, which is still allowed. We will henceforth refer to this estimation method as the **Expanded Regressor** method of estimation.

Under this condition of independence over time, as well as a few technical assumptions which are given in the appendix, the estimation of the autoregressive and regressive coefficients is consistent. More formally, we have proven the theorem given below, which can be found in the appendix as well.

Theorem 1. Consistency of the autoregressive and regressive structure: Let assumptions a-e hold, and let $\kappa \to 0$ and $\min\{N, T\} \to \infty$ as $T, N \to \infty$. Then we have that $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\phi}} \xrightarrow{p} \boldsymbol{\phi}_0$.

The second method of estimation makes use of a similar iterative estimation structure as the one used by Ando and Bai (2016). Let us first assume that we already know the parameters $\phi_1^0, \dots, \phi_h^0$. Then, subtracting the autoregressive structure from both sides in (2), we obtain.

$$y_{i,t} - \sum_{k=1}^{h} \phi_k y_{i,t-k} = \boldsymbol{x}'_{i,t} \boldsymbol{\beta} + \boldsymbol{f}'_{g_i,t} \boldsymbol{\lambda}_{g_i,i} + \epsilon_{i,t}$$

Note that $y_{i,t} - \sum_{k=1}^{p} \phi_k y_{i,t-k}$ can now be seen as the Original Model from Ando and Bai (2016),

and all the remaining parameters can be estimated by the same algorithm as described in 3.1.2. However, in practice we still have to estimate the autoregressive parameters. If we assume that all other parameters are known, we can now subtract these elements from both sides in (4), to obtain the following equation:

$$y_{i,t} - (\boldsymbol{x}'_{i,t}\boldsymbol{\beta} + \boldsymbol{f}'_{g_i,t}\boldsymbol{\lambda}_{g_i,i}) = \sum_{k=1}^{h} \phi_k y_{i,t-k} + \epsilon_{i,t}$$

Here we can see that $y_{i,t} - (\boldsymbol{x}'_{i,t}\boldsymbol{\beta} + \boldsymbol{f}'_{g_i,t}\boldsymbol{\lambda}_{g_i,i})$ has a linear relationship with the lagged dependent variables. This implies that we can estimate the autoregressive parameters ϕ_k by means of least squares. Now that it is known how to update the model parameters, we propose a similar algorithm of iterative parameter updating as in Ando and Bai (2016), which is described as follows:

1. Set the autoregressive parameters $\phi_k = 0$, and initialize the groups, the regression coefficients and the factor structure as described in section 3.1.2.

2. Given the groups, the regression coefficients and the factor structure, update the autoregressive structure as described above.

3. Given the autoregressive structure, update the the groups, the regression coefficients and the factor structure as described in steps 2-4 in section 3.1.2.

4. Repeat steps 2 and 3 until convergence is reached.

Because of the extra step that has to be performed every iteration we will henceforth refer to this process as the **Expanded Iterative** method of estimation.

Both methods have their own advantages. For example, the Expanded Regressor method has the advantage that it is proven to be consistent, while the Expanded Iterative method depends on already having a good estimate of the autoregressive component. However, it was show in Sargan (1964) that for a similar model an iterative method gives consistent estimates. Furthermore, Bai (2009) found that an iterative approach had better results relating to the convergence of the model parameters than a direct estimation method. Since both methods have their own benefits relative to each other, it is unclear which one will perform better from theoretical reasoning alone.

3.3 Monte Carlo Simulation

3.3.1 Simulation of the Original Model

In order to find the empirical properties of this estimation method, we perform Monte Carlo simulation, and analyze the model parameters that are found. In this section we describe the different data generating processes used in the simulation, as well as the methods by which the results are compared.

Similar to Ando and Bai (2016), we will construct the dependent variables according to the model described in (1). The factors $\boldsymbol{f}_{j,t}$ $(j = 1, \ldots, S)$ are vectors of independently generated $N(j, 1)$

variables. We also generate the factor loadings $\lambda_{j,i}$ independently from $N(0, j)$. Furthermore, the elements of $\boldsymbol{X}_i$ are generated by an uniform distribution between -2 and 2. Compared to Ando and Bai (2016), we set the number of columns p of x_t to be equal to 20, instead of 80. We do this to reduce the computational complexity of the problem, such that estimation remains feasible. The regression coefficient β has $(1, 2, 3)$ as leading coefficients, and remaining coefficients equal to zero. Furthermore, we set the number of groups $S = 3$, with group specific factors $r_i = 3$.

Keeping the above configuration, we construct the same three data generating processes for the error terms as were used by Ando and Bai (2016). These have the following specifications

1. For the first data generating process, we generate the N dimensional vector $\boldsymbol{\epsilon}_t$ from a multivariate normal distribution with $\boldsymbol{\mu}_\epsilon = \mathbf{0}$ and $\boldsymbol{\Sigma}_\epsilon = \boldsymbol{I}_n$. This specification will henceforth be referred to as the **homogeneous error specification**.

2. In the second data generating process we construct non-homoskedastic errors. These are specified as $\epsilon_{i,t} = 0.9e^1_{i,t} + \delta_t 0.9e^2_{i,t}$, where $\delta_t = 1$ if t is odd and zero if t is even. Furthermore, $\boldsymbol{e^1_t}$ and $\boldsymbol{e^2_t}$ are independently generated from the same multivariate normal distribution with $\boldsymbol{\mu}_\epsilon = 0$ and $\boldsymbol{\Sigma}_\epsilon = S = (s_{i,j})$, with $s_{ij} = 0.3^{|i-j|}$. We refer to this error specification as the **heterogeneous error specification**.

3. Lastly, we also consider a data generating process with serial and cross-sectional correlation. Here the error terms are constructed as $\epsilon_{i,t} = 0.2\epsilon_{i,t-1} + e_{i,t}$, with $\boldsymbol{e_t}$ generated from a multivariate normal distribution with with $\boldsymbol{\mu}_\epsilon = 0$ and $\boldsymbol{\Sigma}_\epsilon = S = (s_{i,j})$, and $s_{i,j} = 0.3^{|i-j|}$. From this point onwards this specification is referred to as the **serial error specification**.

These DGPs will be generated with varying combinations of N, T. We consider $N \in \{300, 600\}$ and $T \in \{100, 200\}$. The number of times this simulation is replicated is set to 200.

After generating the data by utilizing the above mentioned process, we will perform estimation of both the model specification, as well as the model parameters. For more information on how this is done, see subsections 3.1.2 and 3.1.3 concerning model estimation. The parameter estimates will then be analyzed by considering the average bias and standard deviation. The analysis for the correct number of groups and factors is done with respect to the number of groups that the estimation algorithm selected. This will be either an underidentification, an overidentification or a correct identification.

3.3.2 Simulation of the Dynamic Model

Apart from analyzing the empirical properties of the Original Model from Ando and Bai (2016), we also consider the properties of the Dynamic Model described in 3.2.1. Here the focus lies on the ability to not only find the correct values for the autoregressive parameters, but also the ability to correctly identify the other model parameters. Similar to the previous section, we consider multiple data generating processes, which are consistent with the assumptions which are used in Theorem 1. The data generating processes for the Dynamic Models can be described as follows:

- First of all we note that regressive and factor structure will be chosen in the same manner as in section 3.3.1. Similarly, the number of groups and group specific factors will both be set to three for all simulations in this section.

- For the first DGP, we specify the number of lags $h = 1$, with corresponding autoregressive parameter $\phi_1 = 0.5$. For the error specification we choose the homogeneous error specification.

- With the second DGP we investigate the effect of multiple lags, such that $h = 3$, with corresponding parameters $\phi_1 = 0.4, \phi_2 = -0.3, \phi_3 = 0.2$. Note that it can be shown that this process is stationary (Das, 1994). For the errors we consider the heterogeneous error specification.

During the initial exploration of the feasibility of the extension, it was found that the estimation algorithm for the Dynamic Model converges much slower compared to the Original Model (this is expanded upon in section 4.2.1). Because of this, we only consider the estimation of the model parameters under the assumption that the underlying group and factor structure is specified correctly. Furthermore, in order to keep the simulation computationally viable, we only consider the combinations $T = 100, N = 300$ and $T = 200, N = 600$. This reduces the total estimation time, while still retaining the ability to show empirical signs of consistency. Similar to the previous section, 200 replications will be performed for each data generating process, and for each replication the parameters will be estimated using both procedures described in 3.2.2. These estimation methods will then be compared to each other based on bias, standard deviation, as well as total estimation time.

3.3.3 Robustness Analysis

The simulations described above are fine for analyzing the performance of the model under preferable circumstances. However, it might also be of interest to consider less desirable scenarios, to test the robustness of the estimation algorithm.

For the Dynamic Model, multiple angles can be considered to test robustness. The first one that we consider concerns misspecification. We consider the same data generating process as was used above, only now the data is fitted on a Dynamic Model with $h = 3$. Depending on the results of the previous section, a suitable estimation algorithm will be chosen for this analysis.

Another test of robustness concerns the extra assumption that was made to show consistent estimation. The assumption of independence of the error terms over time might not always hold in a practical situation. Therefore, it is of interest to also consider data generating processes where there exists serial correlation between the error terms. Because of this, we construct a data generating process with the serial errors specified in 3.3.1. In order for the DGP to be Dynamic, we also add three lags of the dependent variables to the model, with autoregressive parameters $\phi_1 = 0.4, \phi_2 = -0.3, \phi_3 = 0.2$.

4 Results

4.1 Simulation of the Original Model

4.1.1 Estimating the Number of Groups and Factors

As stated before, we performed Monte Carlo simulation to study the models capability of finding the correct number of groups, factors and the correct parameter values. Here we used the same simulations as were used by Ando and Bai (2016), except we set the number of regressors equal to 20 instead of 80, and performed 200 replications instead of 1000. As mentioned earlier, this was done in order to make estimation computationally feasible.

First of all, we consider the models ability to find the correct specification. This is measured in terms of under-, correct and overidentification (respectively given as U, C and O) for both the groups and the factors. Note that for the factors, only the replications which found the correct number of groups are counted. The results of these simulations can be found in the Table 1.

Table 1: Monte Carlo results concerning the number of groups and factors

Homogeneous Errors		S			r_1			r_2			r_3		
T	N	U	C	O	U	C	O	U	C	O	U	C	O
100	300	0	174	26	0	174	0	0	173	1	0	172	2
200	300	1	169	30	0	169	0	0	169	0	0	169	0
100	600	1	180	19	0	180	0	0	180	0	0	180	0
200	600	0	195	5	0	195	0	0	195	0	0	194	1

Heterogeneous Errors		S			r_1			r_2			r_3		
T	N	U	C	O	U	C	O	U	C	O	U	C	O
100	300	1	167	32	0	167	0	0	165	2	0	163	4
200	300	3	160	37	0	160	0	0	160	0	0	160	0
100	600	0	180	20	0	180	0	0	180	0	0	177	3
200	600	0	190	10	0	190	0	0	190	0	0	189	1

Serial Errors		S			r_1			r_2			r_3		
T	N	U	C	O	U	C	O	U	C	O	U	C	O
100	300	2	162	36	0	162	0	0	162	1	0	160	2
200	300	2	168	30	0	168	0	0	168	0	0	168	0
100	600	0	187	13	0	187	0	0	187	0	0	187	0
200	600	0	186	14	0	186	0	0	186	0	0	186	0

As can be seen in all tables, as we increase both T and N, the chance of finding the correct group specification increases. Note that this property is in line with Theorem 4 from Ando and Bai (2016), which states that the estimation of the number of groups and factors is consistent. Another matter which is observed is the fact that if the number of groups is misspecified, then the misspecification almost always is overidentification. This is also in agreement with the results found in the Monte Carlo section of Ando and Bai (2016).

However, we do find that some of the results differ from their Monte Carlo simulation, the main one being the number of misspecifcation of the factors. We find that given the correct number of groups, the correct number of factors is found more often than in Ando and Bai (2016). This can be caused by multiple factors. For one, we reduced the number of redundant regressors from 77 to 17, which is likely to increase the accuracy of estimation. Furthermore, in order to decrease the computational complexity of the full simulation, we only considered one value for κ, instead of letting it vary over multiple values.

4.1.2 Parameter Estimation

After an estimate was made of the number of groups and factors, the model was estimated using these specification. By means of the iterative estimation algorithm described in section 3.1.2, the model parameters were found. Similar to Ando and Bai (2016), we mainly focus on the estimated regression coefficients. Of these coefficients, the averages and standard errors were found, and reported in Table 2, which can be found below.

Table 2: The descriptive statistics of the regression coefficients found using Monte Carlo simulation.

Homogeneous Errors								
T	N		β_1	β_2	β_3	β_4	β_5	β_6
100	300	Mean	1.0000	2.0008	2.9999	-0.0001	0.0000	-0.0001
		Std.Dev.	0.0057	0.0052	0.0055	0.0008	0.0010	0.0008
200	300	Mean	1.0001	2.0000	3.0002	0.0001	0.0001	0.0000
		Std.Dev.	0.0038	0.0038	0.0036	0.0010	0.0009	0.0010
100	600	Mean	0.9998	2.0000	2.9995	-0.0001	0.0000	0.0000
		Std.Dev.	0.0035	0.0035	0.0041	0.0010	0.0009	0.0008
200	600	Mean	0.9999	1.9998	3.0000	0.0000	0.0000	0.0000
		Std.Dev.	0.0027	0.0025	0.0028	0.0011	0.0010	0.0008

Heterogeneous Errors								
T	N		β_1	β_2	β_3	β_4	β_5	β_6
100	300	Mean	1.0002	2.0004	3.0011	0.0001	-0.0001	-0.0001
		Std.Dev.	0.0084	0.0081	0.0091	0.0009	0.0008	0.0008
200	300	Mean	1.0004	1.9996	3.0001	0.0000	0.0001	0.0001
		Std.Dev.	0.0056	0.0057	0.0053	0.0010	0.0010	0.0008
100	600	Mean	1.0003	1.9997	3.0000	0.0000	0.0000	0.0000
		Std.Dev.	0.0052	0.0049	0.0052	0.0011	0.0008	0.0010
200	600	Mean	0.9998	2.0000	2.9999	0.0000	0.0000	-0.0001
		Std.Dev.	0.0038	0.0040	0.0041	0.0009	0.0010	0.0009

Serial Errors								
T	N		β_1	β_2	β_3	β_4	β_5	β_6
100	300	Mean	1.0000	2.0005	2.9999	0.0000	-0.0001	0.0000
		Std.Dev.	0.0070	0.0070	0.0074	0.0008	0.0011	0.0008
200	300	Mean	1.0003	1.9996	3.0001	-0.0001	0.0001	0.0002
		Std.Dev.	0.0050	0.0049	0.0053	0.0010	0.0009	0.0009
100	600	Mean	1.0004	1.9994	3.0006	0.0000	0.0000	-0.0001
		Std.Dev.	0.0051	0.0048	0.0051	0.0008	0.0010	0.0010
200	600	Mean	0.9996	1.9999	2.9999	0.0000	0.0000	-0.0001
		Std.Dev.	0.0037	0.0036	0.0037	0.0009	0.0009	0.0008

Once again, we note the results are satisfactory. The estimates of the first three coefficients are very close to their true values of 1, 2 and 3. Furthermore, the coefficients of the redundant regression coefficients are also close to zero. While the average of all parameters are already close to their true value, we can still see evidence of consistency in the decreasing standard deviations of the estimations. Moreover, we see that the standard deviation of the first three betas is larger in magnitude than the standard deviations of the redundant regression coefficients, even though the regressors come from the same distribution. This can be explained by the fact that the regression coefficients were estimated utilizing the SCAD penalty function. This indirectly sets the betas which would be close to zero in a normal linear regression even closer to zero, thus decreasing the standard deviation of these terms.

If we compare our results to those found by Ando and Bai (2016), we find similar results as in 4.1.1, where the estimates are more accurate for our simulation. This can once again be explained by a non varying κ, as well as a smaller number of redundant regressors.

4.2 Simulation of the Dynamic Model

4.2.1 Speed of Convergence

Before the Monte Carlo simulation was performed, there was an investigative period to study the feasibility and basic properties of the Dynamic Model. One of the more interesting properties of

this model is that the number of iterations that need to be performed until convergence is reached is significantly larger compared to the Original Model. In order to better visualize this, a plot was made of the value of the autoregressive parameters for both estimation methods over the iterations. We compare this with a plot of the regression coefficients of the Original Model over the iterations. This can be seen in Figure 1, given below.

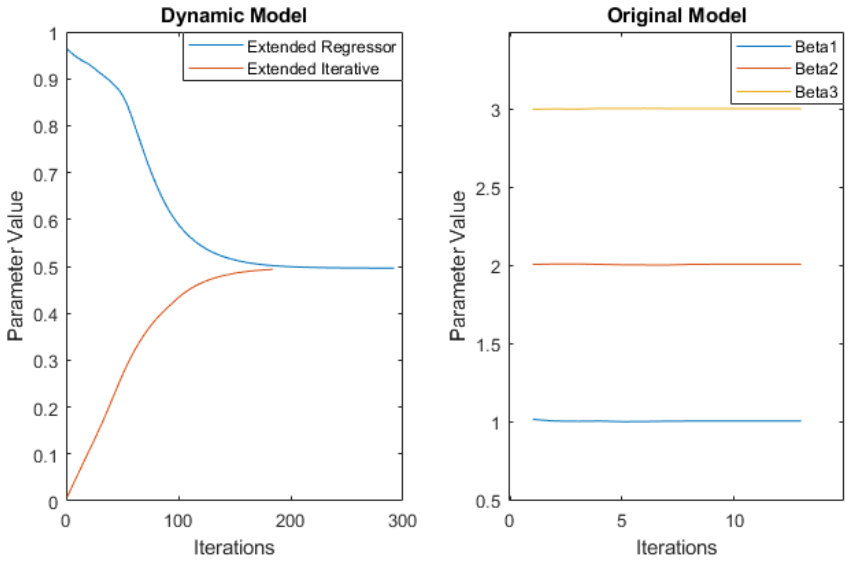

Figure 1: A comparison of the estimation of the autoregressive components of the Dynamic Model and the regression coefficients of the Original Model.

Note that these estimations were taken from one replication of respectively an Dynamic and Original Model with homogeneous errors. While one could argue that this might be insufficient, analysis of the other replications shows a similar pattern. The first thing that can be noticed in these figures is that the addition of an autoregressive component leads to slower convergence. However, we do see that both estimation methods are able to find the true value of the autoregressive parameter. What is also interesting is that the Extended Regressor approach initially overestimates the parameter value, while the Extended Iterative approach initially underestimates the value. It might be of interest to investigate if this symmetry can be used to obtain faster convergence, however due to space constraints we are unable to further investigate this hypothesis.

4.2.2 Simulation using Homogeneous Errors

As section 3.3.2 describes in detail, Monte Carlo simulation was performed using a Dynamic Model with homogeneous errors as data generating process. The averages and standard deviations of the autoregressive parameters and regression coefficients, as well as the average amount of iterations needed for convergence, are given in Table 3 for different values of N and T.

Table 3: The descriptive statistics of the Monte Carlo simulation with a dynamic DGP and homogeneous error terms.

Regressor Method								
T	N		ϕ_1	β_1	β_2	β_3	β_4	β_5
100	300	Mean	0.5314	0.9937	1.9952	2.9958	0.0080	0.0042
		Std.Dev.	0.1136	0.0878	0.1085	0.0656	0.0920	0.1125
200	600	Mean	0.5222	0.9952	2.0004	2.9973	-0.0040	-0.0021
		Std.Dev.	0.0963	0.0542	0.0380	0.0426	0.0375	0.0253

Iterative Method								
T	N		ϕ_1	β_1	β_2	β_3	β_4	β_5
100	300	Mean	0.4978	0.9997	2.0001	2.9995	-0.0006	-0.0003
		Std.Dev.	0.0032	0.0062	0.0061	0.0066	0.0050	0.0045
200	600	Mean	0.4990	1.0002	1.9999	2.9998	0.0002	0.0003
		Std.Dev.	0.0020	0.0028	0.0028	0.0028	0.0023	0.0023

Mean Iterations		Regressor	Iterative
$T = 100$	$N = 300$	217.0	202.9
$T = 200$	$N = 600$	197.6	189.1

If one considers these test statistics, then the conclusion which estimation method is better is clear. We see that the Extended Iterative method has the smallest bias, standard deviation, and on average the smallest amount of iterations until convergence is reached. We can also see an indication of consistent estimation in the test statistics of the Extended Iterative method, since both the bias and the standard deviation decrease as we increase the sample size.

4.2.3 Simulation using Heterogeneous Errors

Once again, we perform Monte Carlo simulation using a Dynamic Model. However, as described in the Methodology, we consider a autoregressive process with multiple lags, as well as heterogeneous error terms. Note that we again used both estimation methods. The results of these simulations can be found in Table 4.

Table 4: The descriptive statistics of the Monte Carlo simulation with a dynamic DGP and heterogeneous error terms.

Regressor Method									
T	N		ϕ_1	ϕ_2	ϕ_3	β_1	β_2	β_3	β_4
100	300	Mean	0.4224	-0.2898	0.2235	0.9995	2.0063	2.9874	0.0021
		Std.Dev.	0.0713	0.0341	0.0766	0.0676	0.0551	0.0748	0.0703
200	600	Mean	0.4222	-0.2907	0.2229	1.0036	2.0011	3.0009	0.0028
		Std.Dev.	0.0713	0.0313	0.0742	0.0512	0.0404	0.0276	0.0549

Iterative Method									
T	N		ϕ_1	ϕ_2	ϕ_3	β_1	β_2	β_3	β_4
100	300	Mean	0.3980	-0.3012	0.1978	0.9995	2.0006	2.9992	0.0002
		Std.Dev.	0.0038	0.0029	0.0040	0.0094	0.0091	0.0085	0.0073
200	600	Mean	0.3989	-0.3006	0.1988	0.9997	1.9997	2.9998	0.0001
		Std.Dev.	0.0026	0.0020	0.0028	0.0048	0.0051	0.0041	0.0039

Mean Iterations		Regressor	Iterative
$T = 100$	$N = 300$	250.3	186.9
$T = 200$	$N = 600$	244.6	198.3

As can be seen in the tables above, we find similar results as were found when the simulation is performed using homogeneous errors. Here the Extended Iterative Method once again performs the best, which can be seen in lower biases and lower standard deviations for all specifications of T and N. Furthermore, the Extended Iterative approach takes less iterations on average to reach convergence compared to the Extended Regressor method. Lastly, it can also be observed that the average bias and standard deviation decreases as we increase T and N for the Extended Iterative approach. This is an indicator that the parameter estimation remains consistent, even under heterogeneity over the dependent variables.

4.2.4 Cause of Large Differences

Although the conclusion that was made based on the statistics is fine on its own, one might wonder why the difference between these estimates is relatively large, and what might be the cause of this relatively large bias and standard deviation when the Extended Regressor method is used. In order to properly investigate this, a histogram was made of the estimates of ϕ_1 for both the first and second method ($N = 600, T = 200$, homogeneous errors), which can be seen in Figure 2.

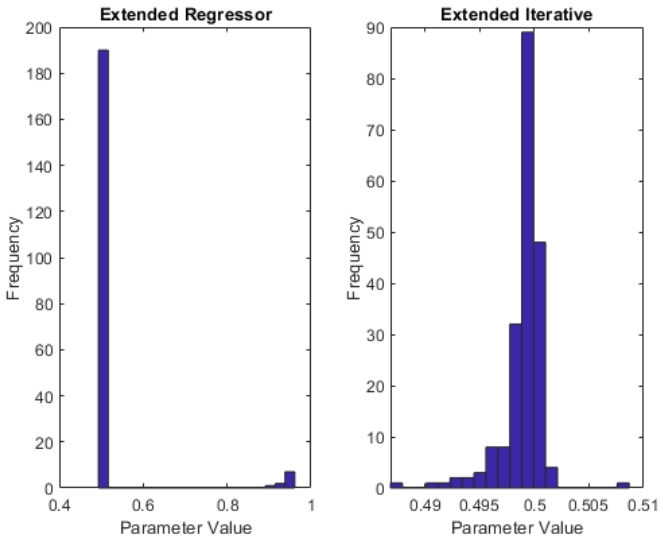

Figure 2: A histogram of the estimated autoregressive coefficients for both methods

As we can clearly see, the reason for the relative large bias and standard deviation is the fact that the Extended Regressor method does not always converge. After further investigation we hypothesize that the probable cause of this occurrence relates to the large initial estimate of the autoregressive component, since this may cause the incorrect estimation of the other model parameters, leading to non converging parameter estimation. However, we did find that increasing T and N reduced the fraction of replications where the parameters did not converge, which is in line with Theorem 1, stating that estimation is consistent as $T, N \to \infty$. However, since we are working with finite samples, we will mainly consider the Extended Iterative approach in the analysis of robustness of the Dynamic Model, which has better results relating to convergence.

4.3 Robustness Analysis

4.3.1 Overspecified Autoregressive Structure

As was mentioned earlier, it might be interesting to see what happens to the estimates when we overspecify the number of lags in our Dynamic Model, since in practice this number is unknown. This estimation was performed using the Extended Iterative method as described in the Methodology section. The results that were obtained can be found in Table 5, which is given below.

Table 5: The descriptive statistics of the Monte Carlo simulation with a dynamic DGP with misspecified autoregressive structure.

Iterative Method									
T	N		ϕ_1	ϕ_2	ϕ_3	β_1	β_2	β_3	β_4
100	300	Mean	0.4979	-0.0013	-0.0019	1.0002	2.0006	2.9997	0.0000
		Std.Dev.	0.0036	0.0034	0.0037	0.0085	0.0074	0.0076	0.0053
200	600	Mean	0.4990	-0.0004	-0.0010	1.0000	1.9998	2.9997	0.0003
		Std.Dev.	0.0026	0.0017	0.0025	0.0032	0.0033	0.0039	0.0028

In the table above it can be seen that even if the autoregressive structure is overspecified, the estimation remains consistent. This is because increasing T and N decreases the bias and standard deviation of the estimates coefficients. Because of this, we conclude that the model is robust concerning overspecification of the autoregressive structure. However, we do note that the quality of the estimated parameters will likely decrease greatly when the autoregressive structure is overspecified to a larger extend. This is because of the fact that having a large number of redundant regressors can cause spurious regressions. In order to solve this issue the autoregressive structure could be estimated by means of penalized regression similar to the SCAD method, but because of space constraints we are unable to further investigate this.

4.3.2 Estimation under Time Dependence

In order to prove that the Dynamic Model can be estimated consistently, we used the assumption of time independence of the error terms. If we allow for time dependence, it can be shown that the model suffers from endogenous regressors. As was described in the Methodology, we performed estimation under time dependence, in order to see how much the estimated parameters are affected. The results of this estimation can be found in Table 6, given below.

Table 6: The descriptive statistics of the Monte Carlo simulation with a dynamic DGP using a serial error specification.

Iterative Method									
T	N		ϕ_1	ϕ_2	ϕ_3	β_1	β_2	β_3	β_4
100	300	Mean	0.4105	-0.3030	0.2011	0.9991	2.0002	3.0008	0.0000
		Std.Dev.	0.0040	0.0028	0.0030	0.0091	0.0105	0.0115	0.0071
200	600	Mean	0.4108	-0.3022	0.2017	0.9995	2.0004	3.0005	0.0007
		Std.Dev.	0.0021	0.0016	0.0019	0.0048	0.0050	0.0047	0.0043

In the table above we see the effect that endogenity has on the parameter estimation. While the estimated parameter are still relatively close to their true value, we see that if we increase the

amount of observations in terms of both T and N, the bias does not decrease as much as in the other simulations. In some cases, the bias even increases. Even though the standard deviation does decrease as the the amount of observations increases, the relatively constant bias implies that the estimation process is likely inconsistent.

5 Conclusion

5.1 Summary of Results

So, to summarize, we investigated the effect that an autoregressive component has on the model proposed by Ando and Bai (2016). In order to do this, we proposed the straightforward extension of adding lagged dependent variables to the model equation. For the model to have consistent parameter estimation, two candidate estimation algorithms are proposed. We found that estimating the model parameters had the best results when the autoregressive structure was estimated separate from the other regressive parameters, since estimating them together often lead to estimates which did not converge to their true values. Furthermore, we showed theoretically and empirically that consistent model estimation holds for data generating processes with homogeneous error terms, as well as error terms which are correlated over the individuals. However, in proving these properties we used the assumption that the error terms were independent over time. This gave the suggestion that estimation might no longer be consistent if we relax this assumption, which we empirically showed to be true. To conclude, we found that the model of Ando and Bai (2016) is able to incorporate an autoregressive structure quite well, with only slight adaptations to the estimation procedure and assumptions.

5.2 Future Research

There are multiple directions which future research could explore. One which would complement our research relates to relaxing the assumption of independence over time. For example, if we know that the error terms follow an autoregressive specification with unknown parameters, consistent estimation might be possible by means of a method similar to system generalized least squares (Amemiya (1985)). Another extension one could further investigate is the addition of a lagged factor structure, and how well this would combine with our proposed Dynamic Model. Additionally, one could also explore how well the model is suited for forecasting. If we wish to perform k-step ahead forecast, a specific structure needs to be imposed on the regressors, as well as the factors. The most evident structure for these variable would be autoregressive structures, such that forecasting becomes feasible.

References

Amemiya, T. (1985). Advanced econometrics. *Harvard university press.*

An, L., Jin, X., Ren, X. (2014). Are the macroeconomic effects of oil price shock symmetric?: A Factor-Augmented Vector Autoregressive approach. *Energy Economics, 45*, 217-228.

Ando, T., Bai, J. (2016). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics, 31*(1), 163-191.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica, 77*(4), 1229-1279.

Bagliano, F. C., Morana, C. (2009). International macroeconomic dynamics: A factor vector autoregressive approach. *Economic Modelling, 26*(2), 432-444.

Das, S. (1994). Time series analysis. *Princeton University Press, Princeton, NJ.*

Fan, J., Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association, 96*(456), 1348-1360.

Heij, C., de Boer, P., Franses, P. H., Kloek, T., van Dijk, H. K. (2004). Econometric methods with applications in business and economics. *Oxford University Press.*

Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of econometrics, 68*(1), 53-78.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, 1*(14), 281-297.

Mallows, C. L. (1973). Some comments on Cp. *Technometrics, 15*(4), 661-675.

Meese, R., Geweke, J. (1984). A comparison of autoregressive univariate forecasting procedures for macroeconomic time series. *Journal of Business & Economic Statistics, 2*(3), 191-200.

Moon, H. R., Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory, 33*(1), 158-195.

Sargan, J. D. (1964). Wages and prices in the United Kingdom: a study in econometric methodology. *Econometric analysis for national economic planning*, 16, 25-54.

Bai, J., Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica, 70*(1), 191-221.

Appendix

Assumptions

In this section the assumptions which are needed for the theorem are given. Note that most of these assumptions are equivalent to the assumptions given in Ando and Bai (2016), while some slightly vary. In order to differentiate between our assumptions and the assumptions from Ando and Bai (2016), our assumptions are given in lowercase letters, while the ones from Ando and Bai are given in uppercase. If alterations were made to the assumptions, the assumption is marked with a small star.

List of Assumptions

Assumption a: Group-specific pervasive factors.
The group-specific pervasive factors satisfy $E\left\|f_{j,t}\right\|^4 < \infty, \quad j = 1, \ldots, S$. Furthermore, we have that $T^{-1}\sum_{t=1}^{T} f_{j,t}f_{j,t}' \to \Sigma_{F_j}$ as $T \to \infty$, where Σ_{F_j} is an $r_j \times r_j$ strictly positive definite matrix.

Assumption b: Factor Loadings.
b1: The factor loading matrix for the group-specific pervasive factors $\Lambda_j = [\lambda_{j,1}, \ldots, \lambda_{j,N_j}$ satisfies $E\left\|\lambda_{j,t}\right\|^4 < \infty$ and $\left\|N_j^{-1}\Lambda_j'\Lambda_j - \Sigma_{\Lambda_j}\right\| \to 0$ as $N_j \to \infty$, where Σ_{Λ_j} is an $r_j \times r_j$ positive definite matrix, $j = 1, \ldots, S$. Furthermore, $\left\|\lambda_{j,i}\right\| > 0$.
b2: For each i and j , $f_{j,t}'\lambda_{j,i}$ is strongly mixing processes with mixing coefficients that satisfy $r(t) \leq \exp\left(-a_1 t^{b_1}\right)$ and with tail probability $P\left(\left|f_{j,t}'\lambda_{j,i}\right| > z\right) \leq \exp\left\{1 - (z/b_2)^{a_2}\right\}$, where a_1, a_2, b_1, b_2 are positive constants.

Assumption c: Error terms.
c1: $E\left[\varepsilon_{it}\right] = 0$ for all i and t
c2: $E\left[\varepsilon_{it}\varepsilon_{js}\right] = \tau_{ij,ts}$ with $\left|\tau_{ij,ts}\right| \leq \left|\tau_{ij}\right|$ for some τ_{ij} for all (t,s), and $N^{-1}\sum_{i,j=1}^{N}\left|\tau_{ij}\right| < C$; and $\left|\tau_{ij,ts}\right| \leq \left|\eta_{ts}\right|$ for some η_{ts} for all (i,j), and $T^{-1}\sum_{t,s=1}^{T}\left|\eta_{ts}\right| < C$. In addition, $(TN)^{-1}\sum_{i,j,t,s=1}\left|\tau_{ij,ts}\right| < C$
c3: For every (s,t), $E\left[\left|N^{-1/2}\sum_{i=1}^{N}\left(\varepsilon_{is}\varepsilon_{it} - E\left[\varepsilon_{is}\varepsilon_{it}\right]\right)\right|^4\right] < C$.
c4: $T^{-2}N^{-1}\sum_{t,s,u,v}\sum_{i,j}\left|\mathrm{cov}\left(\varepsilon_{is}\varepsilon_{it}, \varepsilon_{js}\varepsilon_{jt}\right)\right| < C$ and $T^{-1}N^{-2}\sum_{t,s}\sum_{i,j,k,l}\left|\mathrm{cov}\left(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{lt}\right)\right| < C$
c5: For each i, ϵ_{it} is strongly mixing processes with mixing coefficients that satisfy $r(t) \leq \exp\left(-a_1 t^{b_1}\right)$ and with tail probability $P\left(\left|\epsilon_{it}\right| > z\right) \leq \exp\left\{1 - (z/b_2)^{a_2}\right\}$, where a_1, a_2, b_1, b_2 are positive constants.
c6: ε_{it} is independent of $x_{js}, \lambda_{j,i}$ and $f_{j,s}$ for all i, j, t, s
c7*: ε_{it_1} is independent of ε_{it_2} for all $t_1 \neq t_2$

Assumption e: Observable predictors.
d1*: First define $\widetilde{\boldsymbol{X}}_i \equiv \left[\boldsymbol{y}_i^{(h,N-1)} \quad \ldots \quad \boldsymbol{y}_i^{(1,N-h)} \quad \boldsymbol{X}_i\right]$. Then define $D_j = \frac{1}{NT}\sum_{i;g_i=j}\widetilde{\boldsymbol{X}}_i' M_{F_j}\widetilde{\boldsymbol{X}}_i, E_j = \mathrm{diag}\left\{E_{j1}, \ldots, \quad E_{jS}\right\}, L_j = \left(L_{j1}', \ldots, L_{jS}'\right)'$ where $E_{jk},$ and L_{jk} are $E_{jk} = \frac{1}{N}\sum_{i:g_i=j,g_i^0=k}\left(\lambda_{k,i}^0\lambda_{k,i}^0\right)\otimes$

$I_T, L_{jk} = \sum_{i;g_i=j,g_i^0=k} \frac{1}{NT}\lambda_{k,i}^0 \otimes M_{Fj}\widetilde{\boldsymbol{X}}_i$ with g_i^0 denoting the true membership and $\lambda_{k,i}^0$ the true factor loadings. Let $A = \{F_j : F_j'F_j/T = I, j = 1, \ldots, S\}$. The smallest eigenvalue of the matrix $\sum_{j=1}^{S}\left(D_j - L_j'E_j^- L_j\right)$ is greater than a positive constant c for all $(F_1, \ldots, F_S) \in A$ and for all groupings with a positive fraction of membership for each group, where E_j^- is a generalized inverse of E_j.

d2*: The vector of regressors x_{it}, as well as the dependent variables y_{it} respectively satisfy $\max_{1\leq i\leq N} T^{-1}\left\|X_i\right\|^2 = O_p\left(N^\alpha\right)$ and $\max_{1\leq i\leq N} T^{-1}\left\|\boldsymbol{y}_i^{(h-k,N-1-k)}\right\|^2 = O_p\left(N^\alpha\right)$ for all $k \in \{0, \ldots, h-1\}$, with $\alpha < 1/8$, and $N/T^2 \to 0$.

d3*: All dependent variables y_{it} have finite first and second moments.

Assumption e: Number of units in each group.

All units are divided into a finite number of groups S, each of them containing N_j units such that $0 < \underline{a} < N_j/N < \overline{a} < 1$, which implies that the number of units in the jth group increases as the total number of units N grows.

Proof of Consistency

Theorem Statement

Consider the following Dynamic Model:

$$y_{i,t} = \sum_{k=1}^{h}\phi_k y_{i,t-k} + x_t'\beta + f_{g_i,t}'\lambda_{g_i,i} + \epsilon_{i,t}, \quad i = 1, \ldots, N, t = 1, \ldots, T$$

Theorem 1. Consistency of the autoregressive and regressive structure: Let assumptions a-e given above hold, and let $\kappa \to 0$ and $\min\{N, T\} \to \infty$ as $T, N \to \infty$. Then we have that $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\phi}} \xrightarrow{p} \boldsymbol{\phi}_0$.

Proof

In order to prove the statement above, we show that the Dynamic Model can be written as the model from Ando and Bai (2016). If we show that the assumptions required by Ando and Bai (2016) still hold, then we can utilize their proof of consistency to show that the theorem holds. In order to show that addition of the autoregressive term preserves the assumptions, we first rewrite

the model equation as is done in 3.2.1.2, such that we find the following equation:

$$\begin{aligned}
&\boldsymbol{y}_i^{(h+1,N)} = \widetilde{\boldsymbol{X}}_i\widetilde{\boldsymbol{\beta}} + \boldsymbol{F}_{g_i}\lambda_{g_i,i} + \boldsymbol{\epsilon}_i \\
&\widetilde{\boldsymbol{X}}_i \equiv \begin{bmatrix} \boldsymbol{y}_i^{(h,N-1)} & \dots & \boldsymbol{y}_i^{(1,N-h)} & \boldsymbol{X}_i \end{bmatrix} \\
&\widetilde{\boldsymbol{\beta}} \equiv \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_h \\ \boldsymbol{\beta} \end{bmatrix}
\end{aligned}$$

We find that the model is now structured as the Original Model described by Ando and Bai (2016). In order to show that the assumptions A-E are not violated, we check them one by one. Since we did not alter the factor structure or the assumptions related to it, we see that by assumptions a and b assumptions A and B respectively hold. Since we also did not alter the DGP of the error terms, assumptions C1-C5 still continue to hold by c1-c5.
However, we do have that assumption C6 is violated, even if we assume independence over time (c7). To see why this is the case, set $t+1=s$. Clearly, for Dynamic Models with lag order larger than zero we have the following:

$$E(\widetilde{\boldsymbol{x}}_{i,s+1}\epsilon_{i,s}) = E\left\{ \begin{bmatrix} y_{i,s} \\ \vdots \\ y_{i,s+1-h} \\ \boldsymbol{x}_{i,s+1} \end{bmatrix} \epsilon_{i,s} \right\} = \begin{bmatrix} \tau_{ii,ss} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = E(\widetilde{\boldsymbol{x}}_{i,s+1})E(\epsilon_{i,s})$$

So independence is violated, and therefore assumption C6 does not hold in general. However, after examinating the proof from Ando and Bai (2016), one can conclude that C6 is only used to show that $\frac{1}{NT}\sum X_i'\epsilon_i = O_p((NT)^{-1/2})$. Because of this, if it can be shown that this still holds, then the proof remains valid. Since the set of stochastic variables which are $O_p(f(N,T))$ form a group which is closed under addition, it is sufficient to show that for all k that $\frac{1}{NT}\sum_i(\boldsymbol{y}_i^{(h+1-k,N-k)})'\boldsymbol{\epsilon}^{(h+1,N)} = O_p((NT)^{-1/2})$. Expanding the i-th term of this sum yields

$$(\boldsymbol{y}_i^{(h+1-k,N-k)})'\boldsymbol{\epsilon}^{(h+1,N)} = \sum_{t=h+1-k}^{N-k} y_{i,t}\epsilon_{i,t+k}$$

Note that $y_{i,t}$ can be written as an infinite sum of previously observed regressors, factors, factor loadings and error terms. Since we assumed that ϵ_{t+k} is independent of these terms, we have that $y_{i,t}$ and $\epsilon_{i,t+k}$ are independent. Thus it can be shown that $\mathrm{E}(y_{i,t}\epsilon_{i,t+k}) = 0$. Therefore, if we average over the groups, as well as time, and apply the central limit theorem (which is justified because of the finite first and second moment) to conclude that $\frac{1}{NT}\sum_i(\boldsymbol{y}_i^{(h+1-k,N-k)})'\boldsymbol{\epsilon}^{(h+1,N)} = O_p((NT)^{-1/2})$. Now we continue to check the rest of the assumptions. In order for D1 to be valid, the assumption had to be altered to also include the lagged dependent variables, resulting in the altered assumption

d1. As was mentioned in Ando and Bai (2016), this assumption is analogous to the full rank condition used in linear regression.

Assumption D2 concerns the stochastic boundness of regressors. Note that we assume d2, that is, $\max_{1\leq i\leq N} T^{-1} \|\boldsymbol{X}_i\|^2 = O_p(N^\alpha)$ and $\max_{1\leq i\leq N} T^{-1} \left\|\boldsymbol{y}_i^{(h-k,N-1-k)}\right\|^2 = O_p(N^\alpha)$ for all $k \in \{0,\ldots,h-1\}$, $\alpha < 1/8$. Note that for D2 to hold for the Dynamic model, we require that $\max_{1\leq i\leq N} T^{-1} \left\|\widetilde{\boldsymbol{X}}_i\right\|^2 = O_p(N^\alpha)$. Using properties of the Frobenius Norm, we find the following inequality:

$$\begin{aligned}
&\max_{1\leq i\leq N} T^{-1} \left\|\widetilde{\boldsymbol{X}}_i\right\|^2 = \\
&\max_{1\leq i\leq N} T^{-1} \left\{ \|\boldsymbol{X}_i\|^2 + \sum_{k=1}^{h} \left\|\boldsymbol{y}_i^{(h-k,N-1-k)}\right\|^2 \right\} \leq \\
&\max_{1\leq i\leq N} T^{-1} \|\boldsymbol{X}_i\|^2 + \sum_{k=1}^{h} \max_{1\leq i\leq N} \left\|\boldsymbol{y}_i^{(h-k,N-1-k)}\right\|^2 = O_p(N^\alpha)
\end{aligned}$$

Therefore, $\max_{1\leq i\leq N} T^{-1} \left\|\widetilde{\boldsymbol{X}}_i\right\|^2 = O_p(N^\alpha)$, and assumption D2 is also valid. Since assumption E is equivalent to our assumption e, all the assumptions that Ando and Bai (2016) require hold. Because these assumptions continue to hold if we consider a Dynamic Model, we can apply the proof of Ando and Bai (2016). By applying the proof related to Theorem 1 from Ando and Bai (2016), we therefore conclude that $\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\| = o_p(1)$ and $\left\|\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0\right\| = o_p(1)$, which implies that as $T, N \to \infty$, $\widehat{\boldsymbol{\beta}} \to \boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\phi}} \to \boldsymbol{\phi}_0$, which is what we wanted to show.

Code Dictionary

- Original Model
  - biasCalculator.m: Finds the estimation bias for the C_p criterion.
  - DGP.m: Generates the panel data used in the Monte Carlo section.
  - FactorModel.m: Estimates the model parameters related to the model from Ando and Bai (2016) for a given number of factors and groups.
  - factorTransform.m: Quality of life function, helps to work with the factor structure in the rest of the programs.
  - findErrorTerms.m: Finds the error terms, given the model parameters and panel data.
  - findFactors.m: Estimates the factor structure.
  - FullReplication.m: The program that found all the Monte Carlo results in section 4.1.1 and 4.1.2.
  - GenErr.m: Generates the error terms used by DGP.m.
  - groupEval.m: processes the data from FullReplication.m and turns it into metadata.
  - modelSelectorv2.m: The program which selects the number of factors and number of groups based on the C_p criterion.
  - MonteCarlo.m: Performs Monte Carlo simulation for the Orginal model with a certain specified error DGP.
  - perfectedSCAD.m: Performs penalized regression using the SCAD penalty.
  - SCAD.m: Evaluates the SCAD penalty function.
  - SCADGrad.m: Evaluates the derivative of the SCAD penality function.
  - updateGroups.m: Updates the groups.

- Dynamic Model
  - alphaByRegression.m: Finds the autoregressive components using the Extended Iterative approach.
  - DGP.m: Generates the Dynamic panel data used in the Monte Carlo section.
  - FactorModel.m: Estimates the model parameters for a Dynamic specification using the Extended Iterative approach.
  - FactorModelAltMethod.m: Estimates the model parameters for a Dynamic specification using the Extended Regressor approach.
  - factorTransform.m: Quality of life function, helps to work with the factor structure in the rest of the programs.
  - findErrorTerms.m: Finds the error terms, given the model parameters and panel data.
  - findFactorsAltMethod.m: Estimates the factor structure for a Dynamic model.
  - GenErr.m: Generates the error terms used by DGP.m.
  - MonteCarloDynamic.m: Performs Monte Carlo simulation for the Dynamic model with a certain specified error DGP.
  - perfectedSCAD.m: Performs penalized regression using the SCAD penalty.
  - perfectedSCADWithAlpha.m: Performs penalized regression (with the autoregressive coefficients included) using the SCAD penalty.
  - robustMonte.m: performs the Monte Carlo simulation for using the specifications described in 3.3.3.
  - SCAD.m: Evaluates the SCAD penalty function.
  - SCADGrad.m: Evaluates the derivative of the SCAD penality function.
  - updateGroups.m: Updates the groups.