SCHOOL OF STATISTICS
UNIVERSITY OF THE PHILIPPINES DILIMAN

WORKING PAPER SERIES

Estimation Procedure for a Multiple Time Series Model

by

Rolan Paul K. Veron Cruz
and
Erniel B. Barrios

UPSS Working Paper No. 2012-04
June 2012

School of Statistics
Ramon Magsaysay Avenue
U.P. Diliman, Quezon City
Telefax: 928-08-81
Email: updstat@yahoo.com

Estimation Procedure for a Multiple Time Series Model

Rolan Paul K. Veron Cruz
School of Statistics, University of the Philippines Diliman
Erniel B. Barrios
School of Statistics, University of the Philippines Diliman

Abstract

Given a multiple time series sharing common autoregressive patterns, we estimate an additive model. The autoregressive component and the individual random effects are estimated by integrating maximum likelihood estimation and best linear unbiased predictions in a backfitting algorithm. The simulation study illustrated that the estimation procedure provides an alternative to the Arellano-Bond GMM estimator of the panel model when T > N and the Arellano-Bond generally diverges. The estimator has high predictive ability. In cases where T ≤ N, the backfitting estimator is at least comparable to Arellano-Bond estimator.

Keywords: multiple time series, panel data, additive models, GMM estimator, backfitting, mixed model

Math Classification Codes: 62G05, 62M10

1 Introduction

Understanding the credit risk component of a loan portfolio is one of the most critical functions of a financial institution. According to (BIS, 1999), credit risk is most simply defined as the potential that a bank borrower or counterparty will fail to meet its obligations in accordance with agreed terms. Over the last decade, financial institutions all over the world have developed sophisticated models in an attempt to understand the credit risk arising from important areas of their businesses. These models are intended to aid banks in quantifying and managing credit risk, and have contributed largely and increasingly to quantifying and managing these institutions' performance, compensation processes, customer profitability, pricing, and capital decision structures, see (Twala, 2010).

We postulate a model that characterizes loan portfolios and subsequently propose a procedure to estimate the parameters of this model. The significance of this model is two-fold. First, it enables managers and analysts to predict outstanding balances of any individual. This not only supplies them with a means to valuate their portfolio at any given time, but it also equips them with a warning signal for possible default by an individual, allowing them to take preemptive corrective action. Secondly, this model provides an understanding into the dynamics of their portfolio and could be used for strategic planning in various aspects of their businesses: for marketing, it could be used in market segmentation, analysis, and decision-making; in finance, such insights could be used in policy creation for asset-liability management; for operations and human resources, findings could be used for setting operational standards and performance incentives.

Panel data models are very popular in studying the behavior of a group of units with longitudinal data. Typically, panel data is analyzed using an analysis of covariance approach, where the covariance matrix is assumed to be non-diagonal, and estimated by generalized least squares. To characterize loan porfolio, we assume that clients are units and their longitudinal behavior constitute a time series data. These data from the clients are not treated as panel data but rather as multiple time series.

We propose an alternative analysis of multiple time series where instead of extracting the common components, the individual time series are analyzed and combined to characterize the common behavior. The proposed model assumes independence among the individuals, and takes advantage of this in postulating additive models. An innovation of this model over panel data analysis is the assumption that individuals have identical autoregressive coefficients. We also take advantage of the additivity of the model and use backfitting in estimation.

2 Strategies in Modeling Multiple Time Series Data

Often the variable of interest is not only related to its predecessors in time, but also on past values of other variables. For example, household consumption expenditures may depend on variables such as income, interest rates, and investment expenditures. If all these variables are related to the consumption expenditures, it makes sense to use their possible additional information content in forecasting consumption expenditures. In other words, denoting the related variables by $Y_{1,t}, Y_{2,t}, \ldots, Y_{K,t}$, the forecast of $Y_{1,T+h}$ at the end of period T may be of the form

$$Y_{1,T+h} = f_1\left(Y_{1,T}, Y_{2,T}, \ldots, Y_{K,T}, Y_{1,T-1}, Y_{2,T-1}, \ldots, Y_{K,T-1}, Y_{1,T-2}, Y_{2,T-2}, \ldots\right) \tag{1}$$

Similarly, a forecast for the second variable may be based on past values of all variables in the system. More generally, a forecast of the K-th variable may be expressed as

$$Y_{K,T+h} = f_K\left(Y_{1,T}, \ldots, Y_{K,T}, Y_{1,T-1}, \ldots, Y_{K,T-1}, \ldots\right) \tag{2}$$

Specifically, Equation (2) can be represented by a vector autogregressive moving average (VARMA) model allow for feedback relationships among the K time series as illustrated in the model:

$$\Phi_{\mathbf{p}}(B)\mathbf{z}_{\mathbf{t}} = \Theta_{\mathbf{q}}(B)\mathbf{a}_{\mathbf{t}} \tag{3}$$

where

$$\Phi_{\mathbf{p}}(B) = \mathbf{I} - \phi_1 B - \ldots - \phi_{\mathbf{p}} B^p \tag{4}$$

$$\Theta_{\mathbf{p}}(B) = \mathbf{I} - \boldsymbol{\theta}_1 B - \ldots - \boldsymbol{\theta}_{\mathbf{p}} B^p \tag{5}$$

are matrix polynomials in B (backshift operator), and the ϕ's and $\boldsymbol{\theta}$'s are kxk matrices, $\mathbf{z_t}$ is the vector of deviations from the mean of stationary time series, and $\{\mathbf{a_t}\}$ = $(a_{1t},\ldots,a_{kt})'$ is a sequence of random shock vectors identically and normally distributed with zero mean and covariance matrix Σ.

(Box and Tiao, 1977) proposed a canonical transformation of a k-dimensional stationary autoregressive process in characterizing structure of k related time series. They illustrated how a k-dimensional space of the observations could be decomposed into independent stationary and nonstationary subspaces. Variables in the nonstationary space represented dynamic growth, while those in the stationary and independent subspaces reflected relationships which remain stable over time.

(Tiao and Box, 1981) extended the univariate approach of (Box and Jenkins, 1970) in modeling and analysis of multiple time series. Realizing that vector ARMA models in Equation (3) contain a dauntingly large number of $\{k^2(p+q)+1/2k(k+1)\}$ parameters, thus complicating methods for model building, (Tiao and Box, 1981) sketch an iterative approach that initially fits (3) but subsequently leads to simplification of its structure. This was then compared to (Granger and Newbold, 1977), (Wallis, 1977), and (Chan and Wallis, 1978), and concluded that their methodology is far simpler to use in practice, and do not require the multitude of steps the previous approaches need to arrive at even a simple model. (Tiao and Tsay, 1983) further illustrates the approach of (Tiao and Box, 1981), and introduces a new method called the extended sample cross-correlation appraoch in identifying the order (p,q) of a vector ARMA model. The usefulness of the vector ARMA model and the proposed modeling approach is demonstrated by analyzing US Hog Data.

In practice it often occurs that several multiple time series of essentially the same quantity are available. Examples of such situations are seismic waves recorded at different geographical locations, wind-speeds measured at neighboring stations, and closely related biological time series measured on the same variable. In such cases, (Akman and Gooijer, 1996) proposed the component-extraction analysis methodology as a tool to detect common underlying sources of variation in the observed multivariate time series. They based their methodology on a theorem developed by (Granger and Morris, 1976) that states: if $X(t) : ARMA(p_1, q_1)$ and $Y(t) : ARMA(p_2, q_2)$ and $\{X(t)\},\{Y(t)\}$ be independent processes such that $Z(t) = X(t) + Y(t)$ then $Z(t) : ARMA(p,q)$ where $p \le p_1 + q_1$ and $q \le max(p_1 + q_2, p_2 + q_1)$.(Akman and Gooijer, 1996) then showed that common underlying independent processes may be extracted from correlated real-valued time series.

Instead of extracting common paterns of multiple time series, it can be modeled as panel data instead, e.g., dynamic panel models. (Arellano and Bond, 1991) developed estimators for dynamic panel models using the generalized method of moments (GMM). This relaxes the assumption that the error term and the explanatory variables are orthogonal (uncorrelated). Monte Carlo studies indicated negligible finite sample biases in the GMM estimators and substantially smaller variances than those associated with simpler estimators. The estimators are optimal for short time series i.e. large N and small T, and perform very well for unbalanced panel data.

The complicated nature of multiple time series data can be simplified by introducing additivity and subsequent estimation via backfitting algorithm. (Hastie and Tibshirani, 1991) described the additive model as a generalization of the usual linear regression model. An important feature of the linear model that has made it so popular for statistical inference is that it includes additive predictor effects. Additive models retain this important feature, i.e., they are additive in predictor effects as defined by:

$$Y = \alpha + \sum_{j=1}^{p} f_j(X_j) + \varepsilon \qquad (6)$$

where the errors ε are independent of the X_js, $E(\varepsilon)$ = 0 and $var(\varepsilon) = \sigma^2$. The f_js are arbitrary univariate functions, one for each predictor.

The backfitting algorithm is the most general method for estimating additive models that allows us to estimate each function by an arbitrary smoother. It enables one to fit an additive model using any regression-type fitting mechanisms.

The conditional expectations provide a simple intuitive motivation for the backfitting algorithm. If the additive model (6) is correct, then for any k, $E(Y - \alpha - \sum_{j \neq k} f_j(X_j) \mid X_k) = f_k(X_k)$. This immediately suggests an iterative algorithm for computing all the f_j which is given in terms of data and arbitrary scatterplot smoothers S_j. See (Hastie and Tibshirani, 1991) for details of the backfitting algorithm.

The backfitting algorithm has been used successfully in estimation procedures for spatio-temporal models. (Landagan and Barrios, 2007) developed an estimation procedure that imbeds the Cochrane-Orcutt procedures in to the backfitting algorithm, and using agricultural data, showed that their method yield superior forecasts compared to some common approaches. (Dumanjug, et.al., 2010), also using a hybrid of the backfitting algorithm, the Cochrane-Orcutt procedure, and bootstrap methods, achieved very low prediction error in the estimation of a spatio-temporal model. (Bastero and Barrios, 2011) infused the forward search algorithm and maximum likelihood estimation into the backfitting framework, and showed the capability of this method in producing robust estimates of the parameters even in the presence of structural change.

The bootstrap method can be used to adjust the estimates from the results of the iterative process in the backfitting algorithm. Suppose a random sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ from an unknown probability distribution F has been observed, and it is desired to estimate a parameter of interest $\theta = t(F)$ on the basis of $\mathbf{x}$. For this purpose, an estimate $\hat{\theta} = s(\mathbf{x})$ is calculated from $\mathbf{x}$. How accurate is $\hat{\theta}$? The *bootstrap* was introduced in 1979 as a computer based method for estimating the standard error of $\hat{\theta}$. And, according to (Efron and Tibshirani, 1993), it enjoys the advantage of being completely automatic, requires no theoretical calculations, and is available no matter how complicated the estimator $\hat{\theta} = s(\mathbf{x})$ may be.

Corresponding to a bootstrap data set $\mathbf{x}^*$ is a bootstrap replication of $\hat{\theta}$, $\hat{\theta}^* = s(\mathbf{x}^*)$. The quantity $s(\mathbf{x}^*)$ is the result of applying the same function $s(\cdot)$ to $\mathbf{x}^*$ as was applied to $\mathbf{x}$. The bootstrap estimate $se_F(\hat{\theta})$, the standard error of a statistic $\hat{\theta}$, is a plug-in estimate that uses the empirical distribution function $\hat{F}$ in place of the unknown distribution function F. Specifically, the bootstrap estimate of $se_F(\hat{\theta})$ is defined by

$$se_{\hat{F}}(\hat{\theta}^*). \tag{7}$$

In other words, the bootstrap estimate of $se_F(\hat{\theta})$ is the standard error of $\hat{\theta}$ for data sets of size n randomly sampled from $\hat{F}$. Equation (7) is called the ideal bootstrap estimate of standard error of $\hat{\theta}$. The bootstrap algorithm works by drawing many independent bootstrap samples, evaluating the corresponding bootstrap replications, and estimating the standard error of $\hat{\theta}$ by the empirical standard deviation of the replications.

According to (Efron and Tibshirani, 1993), the limit of $\hat{s}e_B$ as B goes to infinity is the ideal bootstrap estimate of $se_F(\hat{\theta})$,

$$\lim_{B\to\infty} \hat{s}e_B = se_{\hat{F}} = se_{\hat{F}}(\hat{\theta}^*) \tag{8}$$

The fact that $\hat{s}e_B$ approaches $se_{\hat{F}}$ as B goes to infinity amounts to saying that an empirical standard deviation approaches the population standard deviation as the number of replication goes large. The "population" of values $\hat{\theta}^* = s(\mathbf{x}^*)$, where $\hat{F} \to (x_1^*, x_2^*, \ldots, x_n^*) = \mathbf{x}^*$.

3 Methodology

Consider N time series $\{Y_{1,t}, Y_{2,t}, \ldots, Y_{N,t}\}$, $t = 1,2,\ldots,T$. Suppose the N units are uncorrelated but assumes similar pattern-specifications, i.e, the N time series shares a common autocorrelation parameter ϕ. The multiple time series is summarized in the following:

$$Y_{i,t} = \phi Y_{i,t-1} + \lambda_i + \varepsilon_{i,t} \tag{9}$$

$$\lambda_i : \mathrm{N}(\mu_i, \sigma^2_{\lambda_i}), \quad \varepsilon_i : \mathrm{N}(0, \sigma^2_{\varepsilon}) \tag{10}$$

Independence of the N units leads for the model above to become a dynamic panel data model. While the N time series may not be theoretically correlated, the assumption that they share common autoregressive patterns suggests a common time series component can be expected, see (Akman and Gooijer, 1996), (Tiao and Tsay, 1983), and (Box and Tiao, 1977).

To illustrate the model in Equations (9) and (10), consider a loan portfolio composed of N individual loans, with data from T periods. Assume that $Y_{i,t}$ is the outstanding balance of the ith customer of the portfolio, at some time point t. $Y_{i,t}$ can then be modeled as an autoregressive process, with an autoregressive component common to each individual to account for the credit card policies applicable to all cardholders. A random λ_i component is included to represent the customer's periodic personal transactions like payments. The individual components λ_i are assumed to be normally distributed, with each individual component characterized by its own mean and variance. This reflects variations in individual units' behavior, e.g. monthly payments among cardholders.

While (9) appears like a dynamic panel model, typical estimation procedures like the generalized method of moments (that relies only on the variance-covariance matrix of the error term) may fail to take advantage of the presence of a common autoregressive pattern. Furthermore, the additivity of the autoregressive term and individual effects can be used to justify a backfitting algorithm. Thus, we propose to estimate ϕ and λ_i by integrating the maximum likelihood estimation of ϕ and the best linear unbiased predictions (BLUP) of λ_i in a backfitting algorithm. The algorithm is implemented in two Phases as follows:

Phase I. Initialization

1. Ignoring the term $\phi Y_{i,t-1}$ in (9), estimate the random effects λ_i as $\hat{\lambda}_i$ using the best linear unbiased predictions (BLUP) method.

2. Compute residuals $r_{i,t}^{ar} = Y_{i,t} - \hat{\lambda}_i$.

3. Estimate $\hat{\phi}^{BS}$ using the following bootstrap procedure:

   (a) For each of the N time series $r_{i,t}^{ar}, r_{i,t-1}^{ar}, \ldots, r_{i,1}^{ar}$ $i = 1 \ldots N$, perform ARIMA modeling to obtain the estimate $\hat{\phi}_i$.

   (b) Perform bootstrap resampling on $\hat{\phi}_i$ to obtain $\hat{\phi}^{BS}$, the boostrap estimate of ϕ.

This phase now yields initial estimates $\hat{\lambda}_i$ and $\hat{\phi}^{BS}$ that are possibly biased because of misspecification error induced by ignoring certain terms in the model while deriving the estimates.

Phase II. Iteration

Given $\hat{\phi}^{BS}$ from Phase I,

1. Compute residuals $r_{i,t}^{re} = Y_{i,t} - \hat{\phi}^{BS} Y_{i,t-1}$.

2. Estimate $\hat{\lambda}_i$ using best linear unbiased prediction, i.e., the residuals are modeled as $r_{i,t}^{re} = \lambda_{i,t} + \varepsilon_{i,t}^{'}$, where $\varepsilon_{i,t}^{'} \sim \mathrm{N}(0, \sigma^2)$.

3. Compute new residuals $r_{i,t}^{ar} = Y_{i,t} - \hat{\lambda}_i$.

4. Compute for a new $\hat{\phi}^{BS}$.

   (a) For each of the N time series $r_{i,t}^{ar}, r_{i,t-1}^{ar}, \ldots, r_{i,1}^{ar}$ $i = 1 \ldots N$, perform ARIMA modeling to obtain the estimate $\hat{\phi}_i$.

   (b) Perform bootstrap resampling on $\hat{\phi}_i$ to obtain $\hat{\phi}^{BS}$, the boostrap estimate of ϕ.

5. Repeat from Step 1 until convergence.

4. Simulation Study

To evaluate the proposed estimation procedure, N time series of length T, weRE generated for various settings of N, T, and the other parameters of the postulated model in (9) and (10). The different scenarios to be simulated are given in Table 1. Each scenarios are replicated 100 times.

The values of N and T were varied to represent different conditions on data availability. Values of ϕ ranges from as low as 0.10 to as high as 0.95 to simulate the behavior of a nearly nonstationary time series. We also vary σ_ϵ^2 to assess the varying effects of measurement error.

The choices for the distribution of λ_i are the normal and Poisson distributions. In the example on portfolio of credit card holders, the symmetric normal distribution represent a propensity to pay outstanding balances regularly. On the otherhand, the skewed Poisson distribution represents a tendency of the card holders not to pay so that there is an accumulation of balance nearing towards the credit limit. To choose the parameters of the distribution, the degree of compactness of the distribution was considered while attempting to keep the distributions somewhat comparable in terms of means and variances.

Table 1: Simulation Settings

Parameters	Values
N	30 60 150
T	30 60 150
ϕ	0.1 0.5 0.75 0.95
σ_ε^2	1 5 10
λ_i Distribution	$\mathsf{N}(\mu_i, \sigma_{\lambda_i}^2)$ $Poisson(\mu_i)$

Table 1.a Parameters for $\lambda_i : \mathsf{N}(\mu_i, \sigma_{\lambda_i}^2)$

μ_i distribution	U(50,100) N(75,208)
Coefficient of Variation	5% 20% 100%

Table 1.b Parameters for $\lambda_i : Poisson(\mu_i)$

μ_i distribution	$\mathsf{U}(\mu_p, \sigma_p^2)$ $\mathsf{N}(\mu_p, \sigma_p^2)$
μ_{pois}	25 75
σ_{pois}^2	5 10 20

To generate N time series, we proceed as follows:

1. For each $i, i = 1 \ldots N$
2. Generate μ_i, from either U(50,100) or U(75,208). If λ_i is normal, compute the corresponding $\sigma_{\mu_i}^2$ based on the specified coefficient of variation.
3. Generate λ_i, from either $\mathsf{N}(\mu_i, \sigma_{\lambda_i}^2)$ or $Poisson(\mu_i)$.

4. Build a time series of length $2T$ based on (9) and (10), i.e., for each $t, t = 1 \dots T$, we first generate an ε_{it} based on its specified parameters, then compute for $Y_{i,t}$ as $Y_{i,t} = \phi Y_{i,t-1} + \lambda_i + \varepsilon_{i,t}$.

5. Discard the first half of the time series, since these values have been influenced by the choice of $Y_{i,0}$.

The result is thus N time series of length T.

$$\begin{matrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,T} \\ Y_{2,1} & Y_{2,2} & \cdots & Y_{2,T} \\ \vdots & \vdots & \ddots & \\ Y_{N,1} & Y_{N,2} & \cdots & Y_{N,T} \end{matrix} \tag{11}$$

To assess the postulated model and the estimation procedure, some measures were computed. The mean absolute prediction error (MAPE) is used to assess how closely the estimation procedure is able to capture the behavior of the time series. This was then compared against a benchmark, the MAPE obtained from using the generalized method of moments estimator of (Arellano and Bond, 1991). The mean squared error (MSE) of the time series is another measure of how well the estimated time series compares with the simulated time series. This also allow us to compute standard errors of predictions. The bias of the estimates of λ_i, ϕ, and $\sigma^2_{\lambda_i}$ will further exhibit optimal behavior of the estimation procedure.

5 Results and Discussions

From Table 1, there is a total of 1944 possible combinations of settings in the simulation study for different values of N, T, ϕ, σ_{ε}^{2}, distribution of λ_i (and the parameters). In the discussion, we first partition the data into two: (a) those settings where λ_i is normally distributed, of which there are 648 possible settings combinations; (b) and those where λ_i is Poisson distributed, of which there are 1296 possible combinations. For each of these, we then discuss the results in the following order: non-convergence of the estimation procedure, predictive ability as measured by mean absolute prediction errors (MAPE) and its mean squared errors (MSE), and the biases of the estimates.

5.1 λ_i Follows a Normal Distribution

In this section, we examine which combination of settings resulted in non-convergence of the backfitting estimation procedure and compare these with the results from the Arellano-Bond GMM estimator. We look at three factors in particular: T, N, and ϕ, as these affect convergence the most. Since each possible setting is replicated 100 times, we consider a setting as convergent if of the 100 replicates, at least 90 replicates converged. We consider a setting non-converging otherwise.

Table 2 shows the percentages of scenarios where the estimation procedure resulted in non-converging solutions for the different values of T, N, and ϕ. We find that in the backfitting estimation procedure, near non-stationary time series, or those with $\phi = 0.95$, may result in non-convergence, with the number of occurrences increasing as N increases. The Arellano Bond GMM estimator may also result in non-convergence when ϕ=0.95, but will definitely result in non-convergence for cases where $T > N$. The proposed estimation procedure is superior to Arellano Bond in the sense that is has higher likelihood of convergence specially when $T > N$. Non-converging cases in the backfitting estimation procedure happens only in the nearly non-stationary cases.

Table 2: Percent of Converging and Non-Converging Settings for Different Values of N, T, and Φ

(a) Percent of Non Converging Settings for the Backfitting Estimator

		T		
N	Φ	30	60	150
30	0.95	22.22	100.00	61.11
60	0.95	38.89	100.00	100.00
150	0.95	50.00	100.00	100.00

(b) Percent of Non Converging Settings for the Arellano Bond GMM Estimator

		T		
N	Φ	30	60	150
30	0.10	-	100.00	100.00
	0.50	-	100.00	100.00
	0.75	-	100.00	100.00
	0.95	16.67	100.00	100.00
60	0.10	-	-	100.00
	0.50	-	-	100.00
	0.75	-	-	100.00
	0.95	27.78	100.00	100.00
150	0.95	44.44	100.00	100.00

(c) Percent of Converging Settings for the Backfitting Estimator

		T		
N	Φ	30	60	150
30	0.10	100.00	100.00	100.00
	0.50	100.00	100.00	100.00
	0.75	100.00	100.00	100.00
	0.95	77.78	-	38.98
60	0.10	100.00	100.00	100.00
	0.50	100.00	100.00	100.00
	0.75	100.00	100.00	100.00

(d) Percent of Converging Settings for the Arellano Bond GMM Estimator

		T		
N	Φ	30	60	150
30	0.10	100.00	-	-
	0.50	100.00	-	-
	0.75	100.00	-	-
	0.95	83.33	-	-
60	0.10	-	100.00	-
	0.50	-	100.00	-
	0.75	-	100.00	-

	0.95	61.11	-	-		0.95	27.78	-	-
150	0.10	100.00	100.00	100.00	150	0.10	100.00	100.00	100.00
	0.50	100.00	100.00	100.00		0.50	100.00	100.00	100.00
	0.75	100.00	100.00	100.00		0.75	100.00	100.00	100.00
	0.95	50.00	-	-		0.95	55.56	-	-

The remainder of our analysis will focus on the results which converged. Table 2 (c and d) show the percentages of settings which achieved convergence for different values of N, T, and ϕ, for the backfitting estimation procedure and the Arellano Bond estimator, respectively. These tables show that convergence at near non-stationarity i.e. $\phi = 0.95$ is still possible when T is small. However, as N increases, if the time-series is nearly non-stationary, the likelihood of convergence decreases.

Mean absolute percentage error summarizes how close the estimated time series is with the simulated values, i.e., it measures predictive ability of the fitted model. On average, the estimator performs better for shorter T, and larger N. The backfitting estimator is also superior to the Arellano Bond when $T > N$, since Arellano Bond did not converge in these cases. For the other cases, the backfitting estimator is either comparable to or slightly better (e.g. when $T = 60$) than the Arellano Bond, see Table 3 for details.

Table 3: Mean Absolute Percentage Error for Different Values of N, T, and Φ

(a) Backfitting Estimator

		T			
N	Φ	30	60	150	(overall)
30	0.10	5.50	6.72	6.61	6.28
	0.50	3.92	4.60	10.33	6.28
	0.75	3.64	2.74	6.01	4.13
	0.95	4.31	-	1.59	3.40
	(overall)	4.35	4.69	6.95	5.32
60	0.10	7.45	8.05	8.21	7.90
	0.50	5.27	4.88	6.18	5.44

	0.75	3.82	2.89	3.25	3.32
	0.95	4.32	-	-	4.31
	(overall)	5.31	5.27	5.88	5.48
150	0.10	5.87	6.40	9.14	7.14
	0.50	3.96	4.27	5.94	4.72
	0.75	2.40	2.84	3.32	2.85
	0.95	4.25	-	-	4.25
	(overall)	4.10	4.50	6.13	4.87
(overall)	(overall)	4.59	4.82	6.35	5.22

(b) Arellano Bond

		T			
N	Φ	30	60	150	(overall)
30	0.10	7.08	-	-	7.08
	0.50	4.12	-	-	4.12
	0.75	3.33	-	-	3.33
	0.95	2.06	-	-	2.06
	(overall)	4.24	-	-	4.24
60	0.10	8.57	10.20	-	9.38
	0.50	5.80	5.14	-	5.47
	0.75	3.76	2.81	-	3.28
	0.95	2.03	-	-	2.03
	(overall)	5.26	6.05	-	5.61
150	0.10	6.96	7.45	8.67	7.69
	0.50	4.05	4.61	5.93	4.86
	0.75	2.31	2.78	3.36	2.82
	0.95	1.99	-	-	1.99
	(overall)	4.06	4.95	5.99	4.94
(overall)	(overall)	4.52	5.50	5.99	5.03

Tables 4, 5, 6, and 7 show MAPE results for various settings of ϕ, σ_{ε}^{2}, μ_i's distribution, and λ_i's distribution's coefficient of variation.

Table 4: Mean Absolute Percent Error for Different Values of ϕ

ϕ	Backfitting	Arellano Bond
0.10	7.11	8.16
0.50	5.48	4.94
0.75	3.43	3.06
0.95	3.83	2.02

Table 5: Mean Absolute Percent Error for Different Values of σ_ε^2

σ_ε^2	Backfitting	Arellano Bond
1	2.60	2.62
5	5.68	5.31
10	7.57	7.42

Table 6: Mean Absolute Percent Error for Different Distributions of μ_i

Distribution of μ_i	Backfitting	Arellano Bond
Normal	5.85	5.41
Uniform	4.60	4.72

Table 7: Mean Absolute Percent Error for Different Coefficients of Variation

Coeff of Var	Backfitting	Arellano Bond
0.05	1.74	1.65
0.20	1.75	1.70
1.00	12.73	12.60

From Table 4, MAPE improves as ϕ increases for both the backfitting estimator and Arellano Bond up until near non-stationarity. The backfitting estimator is slightly better at very low values of ϕ, but at $\phi = 0.95$, the MAPE deteriorates, while the Arellano Bond estimator continues to improve (for as long as it converge).

Table 5 shows the effects of σ_ε^2. As σ_ε^2 increases, MAPE deteriorates for both estimators. The backfitting estimator is comparable to Arellano Bond for all values of σ_ε^2.

The effect of μ_i's distribution is shown in Table 6. Both estimators perform slightly better when the distribution is uniform.

Both estimators perform well when the distribution of λ_i is compact, i.e., the coefficients of variation are low. Both estimators deteriorate as the coefficient of variation increases, see Table 7 for details.

Another measure of the predictive power of the estimator is the mean-squared-error (MSE). Tables 8, 9, 10, 11, 12 show the effects of the various settings on the MSE. MSE is largest when N is small and decreases when N increases. It is also largest when T is short, and improves as T increases. As ϕ increases, so does MSE. As ϕ approaches non-stationarity, MSE explodes. As σ_ε^2 increases, MSE improves. Although a uniform distribution for μ_i has a slightly better MSE than a normal distribution, their values are comparable. As the coefficient of variation increases, MSE declines.

Table 8: Mean Square Error for Different Values of N, T, and Φ

(a) Backfitting Estimator

		T			
N	Φ	30	60	150	(overall)
30	0.10	5.46	5.43	5.41	5.43
	0.50	7.58	7.41	7.21	7.40
	0.75	16.42	15.15	13.50	15.02
	0.95	5881.53	-	277.09	4013.39
	(overall)	1218.70	9.33	39.51	468.77
60	0.10	5.42	5.40	5.40	5.40
	0.50	7.19	7.15	7.13	7.16
	0.75	13.76	13.27	12.73	1.26
	0.95	5568.05	-	-	5568.05
	(overall)	949.59	8.61	8.42	362.10
150	0.10	5.39	5.38	5.39	5.39
	0.50	6.93	7.02	7.08	7.01
	0.75	11.89	12.18	12.17	12.08
	0.95	5109.63	-	-	5109.63
	(overall)	736.86	8.19	8.21	276.66
(overall)	(overall)	974.58	8.71	19.57	371.42

(b) Arellano Bond

N	Φ	T 30	60	150	(overall)
30	0.10	7.17	-	-	7.17
	0.50	7.37	-	-	7.37
	0.75	9.91	-	-	9.91
	0.95	1280.17	-	-	1280.17
	(overall)	270.04	-	-	270.04
60	0.10	7.29	7.47	-	7.38
	0.50	7.51	7.87	-	7.69
	0.75	10.05	11.20	-	10.63
	0.95	1216.10	-	-	1216.10
	(overall)	212.68	8.85	-	-
150	0.10	7.36	7.54	7.66	7.52
	0.50	7.60	7.97	8.21	7.93
	0.75	10.17	11.37	12.16	11.24
	0.95	1120.92	-	-	1120.92
	(overall)	167.31	8.96	9.34	67.42
(overall)	(overall)	218.00	-	-	-

Table 9: Mean Squared Error for Different Values of ϕ

ϕ	Backfitting	Arellano Bond
0.10	5.41	7.41
0.50	7.19	7.76
0.75	13.45	10.81
0.95	4671.13	1217.28

Table 10: Mean Squared Error for Different Values of σ_ε^2

σ_ε^2	Backfitting	Arellano Bond
1	513.62	162.87
5	361.44	120.12
10	229.45	84.75

Table 11: Mean Squared Error for Different Distributions of μ_i

Distribution of μ_i	Backfitting	Arellano Bond
Normal	382.56	126.38
Uniform	360.31	120.51

Table 12: Mean Squared Error for Different Coefficients of Variation

Coefficient of Variation	Backfitting	Arellano Bond
0.05	484.01	158.35
0.20	420.93	139.46

1.00	196.30	66.54

The effects of N and T on the bias of $\hat{\lambda}_i$, $\hat{\phi}$, and $\hat{\sigma}_\varepsilon^2$ are summarized in Table 13. As the number of individual time series (N) increases, the bias of the estimates for the three parameters slightly decline. However, as the time series grows in length, improved estimates for λ_i and ϕ could be expected. On the otherhand, bias of $\hat{\sigma}_\varepsilon^2$ is smallest when T is of moderate length, e.g., $T = 60$.

Table 13: Percent Biases of $\hat{\lambda}_i$, $\hat{\phi}$, and $\hat{\sigma}_\varepsilon^2$ for Different Values of N and T

$\hat{\lambda}_i$				
N	T=30	T=60	T=150	Overall
30	36.02	11.98	9.16	19.97
60	34.11	11.49	4.78	17.89
150	32.60	11.79	4.66	17.20
Overall	34.29	11.75	6.32	18.39
$\hat{\phi}$				
N	T=30	T=60	T=150	Overall
30	22.10	14.14	6.23	14.46
60	22.94	13.29	5.87	14.60
150	22.96	12.81	5.21	14.15
Overall	22.65	13.42	5.79	14.41
$\hat{\sigma}_\varepsilon^2$				
N	T=30	T=60	T=150	Overall
30	69408.08	107.50	2206.69	26558.27
60	54939.33	80.34	66.48	20687.77
150	39759.22	61.68	57.18	14685.67
Overall	55079.78	83.17	836.01	20778.74

As shown in Table 14, smaller ϕ values result in smaller biases for $\hat{\lambda}_i$ and $\hat{\sigma}_\varepsilon^2$. However, it also lead to larger bias for $\hat{\phi}$. On the other hand, Table 15 shows smaller biases for $\hat{\lambda}_i$ and $\hat{\sigma}_\varepsilon^2$, and larger biases for $\hat{\phi}$ when σ_ε^2 increases. The distribution of μ_i has negligible effect on the biases of $\hat{\lambda}_i$ and $\hat{\phi}$, but results in smaller bias for $\hat{\sigma}_\varepsilon^2$ when it is of uniform distribution, see Table 16. Finally, in Table 17, as the coefficient of variation increases, the bias of $\hat{\lambda}_i$ and $\hat{\sigma}_\varepsilon^2$ decreases, while the bias of $\hat{\phi}$ increases slightly.

Table 14: Percent Bias of $\hat{\lambda}_i$, $\hat{\phi}$, and $\hat{\sigma}_\varepsilon^2$ for Different Values of ϕ

ϕ	$\hat{\lambda}_i$	$\hat{\phi}$	$\hat{\sigma}_\varepsilon^2$
0.10	3.55	27.39	3.37
0.50	10.33	9.83	41.35
0.75	26.21	8.56	206.94
0.95	78.02	4.32	266088.49

Table 15: Percent Bias of $\hat{\lambda}_i$, $\hat{\phi}$, and $\hat{\sigma}_\varepsilon^2$ for Different Values of σ_ε^2

σ_ε^2	$\hat{\lambda}_i$	$\hat{\phi}$	$\hat{\sigma}_\varepsilon^2$
1	19.82	14.01	51261.93
5	18.55	14.47	7129.11
10	16.70	14.76	2194.89

Table 16: Percent Bias of $\hat{\lambda}_i$, $\hat{\phi}$, and $\hat{\sigma}_\varepsilon^2$ for Different Distributions of μ_i

Distribution of μ_i	$\hat{\lambda}_i$	$\hat{\phi}$	$\hat{\sigma}_\varepsilon^2$
Normal	18.38	14.43	22666.91
Uniform	18.41	14.38	18897.72

Table 17: Percent Bias of $\hat{\lambda}_i$, $\hat{\phi}$, and $\hat{\sigma}_\varepsilon^2$ for Different Coefficients of Variation

Coeff of Variation	$\hat{\lambda}_i$	$\hat{\phi}$	$\hat{\sigma}_\varepsilon^2$
0.05	20.08	14.14	21675.32
0.20	19.11	14.21	21909.69
1.00	15.79	14.90	18594.78

5.2 λ_i Follows a Poisson Distribution

Similar to the results for the Normal distribution, the backfitting estimator may diverge for nearly non-stationary time-series, with increasing likelihood of divergence as N increases or when T increases. Also, as with the Normal distribution results, Arellano Bond may diverge for nearly non-stationary time series and will definitely diverge when $T > N$, see.Table for details.

Table 18: Mean Absolute Percentage Error for Different Values of N, T, and Φ

(a) Backfitting Estimator

N	Φ	T 30	T 60	T 150	(overall)
30	0.10	4.46	4.44	4.43	4.44
	0.50	2.75	2.78	2.79	2.77
	0.75	1.81	1.82	1.83	1.82
	0.95	4.26	1.58	1.04	2.93
	(overall)	3.20	2.97	2.79	3.00
60	0.10	4.42	4.46	4.43	4.44
	0.50	2.74	2.76	2.77	2.76
	0.75	1.75	1.79	1.79	1.78
	0.95	4.26	-	-	4.26
	(overall)	3.20	3.00	3.00	3.01
150	0.10	4.42	4.43	4.43	4.42
	0.50	2.72	2.75	2.77	2.75
	0.75	1.72	1.76	1.79	1.76
	0.95	4.26	-	-	4.26
	(overall)	3.16	2.98	3.00	3.05
(overall)	(overall)	3.20	2.98	2.92	3.05

(b) Arellano Bond

N	Φ	T 30	T 60	T 150	(overall)
30	0.10	5.14	-	-	5.14
	0.50	2.83	-	-	2.83
	0.75	1.63	-	-	1.63
	0.95	1.99	-	-	1.99
	(overall)	2.98	-	-	2.98
60	0.10	5.14	5.25	-	5.19
	0.50	2.70	2.93	-	2.90
	0.75	1.64	1.74	-	1.69
	0.95	1.99	-	-	1.99
	(overall)	2.99	3.31	-	3.13
150	0.10	5.16	5.23	5.27	5.22
	0.50	2.88	2.95	2.99	2.94
	0.75	1.66	1.75	1.81	1.74
	0.95	1.99	-	-	1.99
	(overall)	3.02	3.31	3.36	3.22
(overall)	(overall)	3.00	3.31	3.36	3.15

The backfitting estimator is robust to N and T. It is also comparable to or sometimes slightly better than Arellano Bond (e.g., $T = 60$), except when $T > N$ where Arellano Bond diverges, see Tables 19.

MAPE also improves with increasing ϕ for both estimators (Table 20), while it declines with increasing σ_ε^2 (Table 21). Just like in the normal distribution case, the backfitting estimator is robust to the distribution of μ_i, see Table 22.

Table 19: Mean Absolute Percentage Error for Different Values of N, T, and Φ

(a) Backfitting Estimator

N	Φ	T: 30	60	150	(overall)
30	0.10	4.46	4.44	4.43	4.44
	0.50	2.75	2.78	2.79	2.77
	0.75	1.81	1.82	1.83	1.82
	0.95	4.26	1.58	1.04	2.93
	(overall)	3.20	2.97	2.79	3.00
60	0.10	4.42	4.46	4.43	4.44
	0.50	2.74	2.76	2.77	2.76
	0.75	1.75	1.79	1.79	1.78
	0.95	4.26	-	-	4.26
	(overall)	3.20	3.00	3.00	3.01
150	0.10	4.42	4.43	4.43	4.42
	0.50	2.72	2.75	2.77	2.75
	0.75	1.72	1.76	1.79	1.76
	0.95	4.26	-	-	4.26
	(overall)	3.16	2.98	3.00	3.05
(overall)	(overall)	3.20	2.98	2.92	3.05

(b) Arellano Bond

N	Φ	T: 30	60	150	(overall)
30	0.10	5.14	-	-	5.14
	0.50	2.83	-	-	2.83
	0.75	1.63	-	-	1.63
	0.95	1.99	-	-	1.99
	(overall)	2.98	-	-	2.98
60	0.10	5.14	5.25	-	5.19

	0.50	2.70	2.93	-	2.90
	0.75	1.64	1.74	-	1.69
	0.95	1.99	-	-	1.99
	(overall)	2.99	3.31	-	3.13
150	0.10	5.16	5.23	5.27	5.22
	0.50	2.88	2.95	2.99	2.94
	0.75	1.66	1.75	1.81	1.74
	0.95	1.99	-	-	1.99
	(overall)	3.02	3.31	3.36	3.22
(overall)	(overall)	3.00	3.31	3.36	3.15

Table 20: Mean Absolute Percent Error for Different Values of ϕ

ϕ	Backfitting	Arellano Bond
0.10	4.44	5.20
0.50	2.76	2.91
0.75	1.79	1.71
0.95	3.62	1.99

Table 21: Mean Absolute Percent Error for Different Values of σ_ε^2

σ_ε^2	Backfitting	Arellano Bond
1	1.66	1.58
5	3.14	3.31
10	4.41	4.66

Table 22: Mean Absolute Percent Error for Different Distributions of μ_i

Distribution of μ_i	Backfitting	Arellano Bond
Normal	3.06	3.15
Uniform	3.03	3.14

Table 23: Mean Square Error for Different Values of N, T, and Φ

(a) Backfitting Estimator

		T			
N	Φ	30	60	150	(overall)
30	0.10	5.40	5.40	5.39	5.40
	0.50	7.07	7.12	7.12	7.10
	0.75	12.90	12.78	12.52	12.73
	0.95	3851.86	92.49	137.67	2255.76
	(overall)	707.26	11.44	23.18	266.30
60	0.10	5.39	5.38	5.38	5.38
	0.50	6.94	7.03	7.07	7.01
	0.75	11.86	12.15	12.19	12.07

	0.95	3838.78	-	-	3838.68
	(overall)	704.56	8.19	8.21	272.34
150	0.10	5.37	5.38	5.38	5.38
	0.50	6.84	6.97	7.05	6.95
	0.75	11.18	11.65	11.97	11.60
	0.95	4314.06	-	-	4314.06
	(overall)	708.82	8.00	8.13	270.09
(overall)	(overall)	706.86	9.23	13.59	269.52

Table 24: Mean Squared Error for Different Values of ϕ

ϕ	Backfitting
0.10	5.39
0.50	7.02
0.75	12.13
0.95	3189.29

Table 25: Mean Squared Error for Different Values of σ_ε^2

σ_ε^2	Backfitting
1	269.56
5	264.61
10	274.41

Table 26: Mean Squared Error for Different Distributions of μ_i

μ_i	Backfitting
Normal	269.93
Uniform	269.10

Biases for $\hat{\lambda}_i$ and $\hat{\phi}$ improve with increasing N and increasing T. However, bias of $\hat{\sigma}_\varepsilon^2$ is smallest when T is of moderate length, e.g., $T = 60$, see Table 27. As shown in Table 28, larger ϕ's result in larger biases for $\hat{\lambda}_i$ and $\hat{\sigma}_\varepsilon^2$. On the other hand, Table 29 shows smaller biases for $\hat{\lambda}_i$ and $\hat{\sigma}_\varepsilon^2$, and larger biases for $\hat{\phi}$ when $\hat{\sigma}_\varepsilon^2$ increases. The backfitting estimator is robust to the distribution of μ_i, see Table 30 for details.

Table 27: Percent Biases of $\hat{\lambda}_i$, $\hat{\phi}$, and $\hat{\sigma}_\varepsilon^2$ for Different Values of N and T

[Percentage Bias for $\hat{\lambda}_i$][Percentage Bias for $\hat{\lambda}_i$]

	$\hat{\lambda}_i$			
N	T=30	T=60	T=150	Overall
30	34.40	15.61	8.67	20.07
60	34.43	11.17	4.56	17.94
150	33.23	11.15	4.44	17.30
Overall	34.02	12.68	6.00	18.47
	$\hat{\phi}$			
N	T=30	T=60	T=150	Overall
30	22.87	13.91	6.33	14.61
60	22.50	13.17	5.88	14.45
150	22.72	12.80	5.24	14.14
Overall	22.70	13.30	5.84	14.40
	$\hat{\sigma}_\varepsilon^2$			
N	T=30	T=60	T=150	Overall
30	31729.20	103.88	1045.28	11823.53
60	31596.57	61.14	57.53	12021.73
150	31022.69	53.15	53.94	11633.31
Overall	31452.74	73.10	412.91	11826.69

Table 28: Percent Bias of $\hat{\lambda}_i$, $\hat{\phi}$, and $\hat{\sigma}_\varepsilon^2$ for Different Values of ϕ

ϕ	$\hat{\lambda}_i$	$\hat{\phi}$	$\hat{\sigma}_\varepsilon^2$
0.10	3.14	27.50	2.59
0.50	9.83	9.81	34.35
0.75	25.74	8.58	148.34
0.95	80.67	4.45	143269.40

Table 29: Percent Bias of $\hat{\lambda}_i$, $\hat{\phi}$, and $\hat{\sigma}_\varepsilon^2$ for Different Values of σ_ε^2

σ_ε^2	$\hat{\lambda}_i$	$\hat{\phi}$	$\hat{\sigma}_\varepsilon^2$
1	20.11	13.99	26855.94
5	17.41	14.52	5192.60
10	17.80	14.73	2644.58

Table 30: Percent Bias of $\hat{\lambda}_i$, $\hat{\phi}$, and $\hat{\sigma}_\varepsilon^2$ for Different Distributions of μ_i

Distribution of μ_i	$\hat{\lambda}_i$	$\hat{\phi}$	$\hat{\sigma}_\varepsilon^2$
Normal	18.51	14.40	11837.41
Uniform	18.43	14.41	11816.00

5.3 Non-Converging Arellano Bond Cases

Of the 648 simulated scenarios under the assumption that λ_i is normal, 286 or 44.14% of these resulted in non-convergence of the Arellano Bond estimator. Of these 286, the backfitting estimator converged in 169 or 59.10% of the cases. The MAPE results of these 169 cases are shown in Table 31. Results are mostly from scenarios where $T > N$. Overall, the backfitting estimator had a MAPE of 5.89%. On the otherhand, of the 1296 scenarios where λ_i is poisson, 578 or 44.60% of the cases resulted in non-convergence of the Arellano Bond. Of these, 334 or 57.80% converged for the backfitting estimator. Again, most results are from $T > N$ scenarios. Overall, the backfitting estimator had a MAPE of 2.91%, see Table 31 for details.

Table 31: MAPE of Backfitting Estimator for Non-converging Cases in Arellano Bond GMM Estimator

Normal										
		T=60			T=150			All		
	ϕ	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
N=30	0.1	1.02	6.72	31.74	1.01	6.61	27.01	1.01	6.67	31.74
	0.5	0.70	4.60	22.47	0.67	10.33	128.69	0.67	7.46	128.69
	0.75	0.56	2.74	9.48	0.50	6.01	67.85	0.50	4.38	67.85
	0.95	*	*	*	0.69	1.59	4.12	0.69	1.59	4.12
	All	0.56	4.69	31.74	0.50	6.95	128.69	0.50	5.89	128.69
N=60	0.1	**	**	**	1.01	8.21	44.63	1.01	8.21	44.63
	0.5	**	**	**	0.65	6.18	26.32	0.65	6.18	26.32
	0.75	**	**	**	0.46	3.25	14.84	0.46	3.25	14.84
	All	**	**	**	0.46	5.88	44.63	0.46	5.88	44.63
All	All	0.56	4.69	31.74	0.46	6.45	128.69	0.46	5.89	128.69
Poisson										
		T=60			T=150			All		
	ϕ	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
N=30	0.1	0.99	4.44	11.02	0.98	4.43	10.67	0.98	4.44	11.02
	0.5	0.68	2.78	6.50	0.65	2.79	6.50	0.65	2.78	6.50
	0.75	0.55	1.82	4.01	0.48	1.83	4.13	0.48	1.82	4.13
	0.95	1.34	1.58	1.66	0.77	1.04	1.84	0.77	1.16	1.84
	(all)	0.55	2.97	11.02	0.48	2.79	10.67	0.48	2.87	11.02
N=60	0.1	**	**	**	0.98	4.43	10.79	0.98	4.43	10.79
	0.5	**	**	**	0.63	2.77	6.44	0.63	2.77	6.44
	0.75	**	**	**	0.45	1.81	4.12	0.45	1.81	4.12

	All	**	**	**	0.45	3.00	10.79	0.45	3.00	10.79
All	All	0.55	2.97	11.02	0.45	2.89	10.79	0.45	2.91	11.02
* Backfitting Estimator did not converge for these scenarios. ** Arellano Bond converged for these scenarios.										

Overall bias for $\hat{\phi}$ is 8.64% (Table 32) for the normal scenarios. This can be as low as 2.01% when T is long, N is small, ϕ is large, i.e., nearly non-stationary. However, this may be as high as 28.93% when ϕ is very small. Similar results are observed for Poisson distributed scenarios.

Table 32: Bias of Backfitting Estimator for $\hat{\phi}$ for Non-converging Arellano Bond GMM Estimator

Normal										
		T=60			T=150			All		
	ϕ	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
N=30	0.1	23.54	26.32	28.93	12.34	13.71	15.94	12.34	20.01	28.93
	0.5	7.72	8.68	9.60	3.23	3.63	3.95	3.23	6.16	9.60
	0.75	7.12	7.42	7.90	2.72	2.93	3.08	2.72	5.18	7.90
	0.95	*	*	*	2.01	2.17	2.35	2.01	2.17	2.35
	All	7.12	14.14	28.93	2.01	6.23	15.94	2.01	9.95	28.93
N=60	0.1	**	**	**	9.63	11.27	12.76	9.63	11.27	12.76
	0.5	**	**	**	3.04	3.38	3.75	3.04	3.38	3.75
	0.75	**	**	**	2.79	2.95	3.09	2.79	2.95	3.09
	All	**	**	**	2.79	5.87	12.76	2.79	5.87	12.76
All	All	7.12	14.14	28.93	2.01	6.06	15.94	2.01	8.64	28.93
Poisson										
		T=60			T=150			All		
	ϕ	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
N=30	0.1	23.08	26.53	30.60	12.11	14.09	16.24	12.11	20.31	30.60
	0.5	7.90	8.53	9.45	3.01	3.61	4.18	3.01	6.07	9.45
	0.75	7.00	7.43	7.89	2.49	2.94	3.25	2.49	5.18	7.89
	0.95	6.63	6.97	7.16	1.81	2.06	2.27	1.81	3.16	7.16
	(all)	6.63	13.91	30.60	1.81	6.33	16.24	1.81	9.96	30.60
N=60	0.1	**	**	**	9.48	11.22	12.60	9.48	11.22	12.60
	0.5	**	**	**	3.15	3.46	3.81	3.15	3.46	3.81
	0.75	**	**	**	2.76	2.95	3.12	2.76	2.95	3.12
	All	**	**	**	2.76	5.88	12.60	2.76	5.88	12.60
All	All	6.63	13.91	30.60	1.81	6.12	16.24	1.81	8.67	30.60
* Backfitting Estimator did not converge for these scenarios. ** Arellano Bond converged for these scenarios.										

Percent biases for $\hat{\lambda}_i$ are fairly reasonable for both normal and poisson settings. Table 33 shows an overall bias of 8.66%. This improves to as much as 1.09% when ϕ is very small and T is large. Similar results were observed from poisson settings.

Table 33: Bias of Backfitting Estimator for $\hat{\lambda}_i$ for Non-converging Arellano Bond GMM Estimator

Normal										
		T=60			T=150			All		
	ϕ	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
N=30	0.1	2.77	3.41	5.99	1.40	1.92	6.27	1.40	2.67	6.27
	0.5	7.79	9.78	20.70	3.22	3.86	5.22	3.22	6.82	20.70
	0.75	21.37	22.75	28.54	8.18	9.15	11.61	8.18	15.95	28.54
	0.95	*	*	*	37.95	41.45	45.12	37.95	41.45	45.12
	All	2.77	11.98	28.54	1.40	9.16	45.12	1.40	10.49	45.12
N=60	0.1	**	**	**	1.09	1.62	3.62	1.09	1.62	3.62
	0.5	**	**	**	3.06	3.61	4.71	3.06	3.61	4.71
	0.75	**	**	**	8.41	9.11	10.68	8.41	9.11	10.68
	All	**	**	**	1.09	4.78	10.68	1.09	4.78	10.68
All	All	2.77	11.98	28.54	1.09	7.10	45.12	1.09	8.66	45.12
Poisson										
		T=60			T=150			All		
	ϕ	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
N=30	0.1	2.61	3.05	3.69	1.36	1.64	2.00	1.36	2.34	3.69
	0.5	7.89	8.53	9.40	3.02	3.65	4.18	3.02	6.09	9.40
	0.75	21.04	22.31	23.87	7.50	8.82	9.84	7.50	15.57	23.87
	0.95	125.50	131.97	135.33	34.60	39.26	42.92	34.60	59.86	135.33
	(all)	2.61	15.61	135.33	1.36	8.67	42.92	1.36	11.99	135.33
N=60	0.1	**	**	**	1.07	1.33	1.62	1.07	1.33	1.62
	0.5	**	**	**	3.13	3.48	3.86	3.13	3.48	3.86
	0.75	**	**	**	8.30	8.86	9.37	8.30	8.86	9.37
	All	**	**	**	1.07	4.56	9.37	1.07	4.56	9.37
All	All	6.63	13.91	30.60	1.81	6.12	16.24	1.81	8.67	30.60

* Backfitting Estimator did not converge for these scenarios.
** Arellano Bond converged for these scenarios.

5.4 Converging Arellano Bond Cases

We also compared the backfitting and Arellano Bond GMM estimators for cases where both converged. Among the better scenarios (lower MAPE), the two estimators are comparable, with the Arellano Bond being slightly better For the relatively poor scenarios (higher MAPE), the two estimators are still comparable, with the backfitting estimator doing slightly better than the Arellano Bond GMM estimator. Similar results were noted for Poisson scenarios.

6. Conclusions

In a multiple time series model, the backfitting estimator provides an alternative to the Arellano Bond GMM Estimator when $T > N$, where Arellano Bond generally diverges in these cases. In these cases, the backfitting estimator yield high predictive ability when individual random effects is normally distributed, and even higher when it is poisson. It generally performs better with increasing value of ϕ, but only until near non-stationarity. The backfitting estimator is robust to N, and improves as T increases.

For cases where $T \leq N$, the backfitting estimator is comparable to Arellano Bond GMM estimator. For scenarios where Arellano Bond performs very well, the backfitting estimator is closely behind, differing only by a very small MAPE value. When Arellano Bond GMM performs poorly, the backfitting estimator provides very good predictive performance and minimal bias for certain parameters.

References:

Akman, I. and Gooijer, J. , (1996). Component extraction analysis of multivariate time series. Computational Statistics and Data Analysis, 21:487-499.
Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. The Review of Economic Studies, 58(2):277-297.
Bastero, R. and Barrios, E. (2011). Robust estimation of a spatiotemporal model with structural change. Communications in Statistics- Simulation and Computation, 40:448-468.
BIS (1999). Principles for the management of credit risk. Basel Committee on Banking Supervision.
Box, G. and Jenkins, G. (1970). Time Series Analysis – Forecasting and Control. San Francisco: Holden Day.
Box, G. and Tiao, G. (1977). A canonical analysis of multiple time series. Biometrika, 64(2):355-365.
Campano, W. and Barrios, E. (2011). Robust estimation of a time series model with structural change. Journal of Statistical Computation and Simulation, 81(7):909-927.
Chan, W. and Wallis, K. (1978). Multiple time series modelling: Another look at the mink-muskrat interaction. Applied Statistics, 27:168-175.
Dumanjug, C., Barrios, E., and Lansangan, J. (2010). Bootstrap procedures in a spatial-temporal model. Journal of Statistical Computation and Simulation, 80(7):809-822.
Efron, B. and Tibshirani, R. (1993). An Introduction to the Booststrap. New York: Chapman and Hall.
Granger, C. and Morris, M. (1976). Time series modeling and interpretation. Journal of the Royal Statistical Society, 139:246-257.
Granger, C. and Newbold, P. (1977). Forecasting Economic Time Series. New York: Academic Press.
Guarte, J. M. and Barrios, E. B. (2006). Estimation under purposive sampling. Communications in Statistics|Simulation and Computation, 35(2):277-284.
Hastie, T. and Tibshirani, R. (1991). Generalized Additive Models. New York: Chapman and Hall.
Landagan, O. and Barrios, E. (2007). An estimation procedure for a spatial-temporal model. Statistics and Probability Letters, 77:401-406.
Tiao, G. and Box, G. (1981). Modeling multiple time series with applications. Journal of the American Statistical Association, 76(376):802-816.
Tiao, G. and Tsay, R. (1983). Multiple time series modeling and extended sample cross-correlation. Journal of Business and Economic Statistics, 1(1):43-56.
Twala, B. (2010). Multiple classifier application to credit risk assessment. Expert Systems with Applications, 37(10):3326-3336.
Wallis, K. (1977). Multiple time series analysis and the final form of econometric models. Econometrika, 45(45):1481-1497.