Title: Estimation of Multiple Time Series with Volatility

1. Introduction

Many practical problems in modeling involve multiple time series that share common autoregressive patterns. For examp!e, prices of individual stocks in a given market such as the Philippine Stock Exchange can behave similarly to one anotherin reaction to nationwide or international phenomena such as an upgrade in the country's credit rating or another global financial crisis. Similarly, credit card payment history of a cohort of credit card clients can depend on common factors such as the price of basic commodities in the country or the country's unemployment rate. Thus, there is significant impetus to explore estimation procedures for multiple time series that can be assumed to share autoregressive patterns.

1.1 Motivation

Veron Cruz and Barrios (2014) recently published a procedure for estimating multiple time series using integrated maximum likelihood estimation and best linear unbiased predictions in a backfitting algorithm to estimate both the autoregressive component and the individual random effects. Simulations conducted in Veron Cruz and Barrios (2014) assumed that each of N time series had similar pattern specifications throughout its generation. However, such ideal circumstances are often not found in practice. Rather, observed time series are typically subject to various phenomena that can affect the validity of estimations that do not take such phenomena into consideration. One such phenomenon is volatility, which are a series of aberrant observations (Campano, 2012) caused by the sudden occurrence of events that are relevant to the time series observed. The unaccounted presence of volatility in time series has been known to significantly impact the ability of existing models to accurately describe characteristics of the series (Campano, 2012). In order to address this, Campano (2012) developed a robust method for estimating the stationary part of a time series through the integration of a block bootstrap procedure and an AR-sieve into a forward search algorithm.However, the work of Campano (2012) was only conducted on single time series. It has not yet been determined if the application of the method to the case of multiple time series can yield similarly desirable outcomes. As such, this study sought to develop a hybrid method from Campano (2012) and Veron Cruz and Barrios (2014) in order to estimate multiple time series with the presence of volatility.

1.2 Objectives

Specifically, this study seeks to accomplish the following objectives:

- Develop a hybrid algorithm combining the backfitting algorithm with the block bootstrap procedure to estimate parameters of multiple time series with volatility
- Conduct simulations of multiple time series with and without volatility and estimate their parameters using the hybrid algorithm
- Compare the predictive ability and parameter estimate bias of the hybrid algorithm with the backfitting algorithm

1.3 Significance

This study seeks to contribute to statistical knowledge by exploring the possibility of combining two existing methods in order to better address multiple time series problems that emerge in practice. Volatility is an important concern in time series estimation that has been addressed through various procedures in literature. The block bootstrap procedure offers a convenient and intuitive way of de-emphasizing volatile segments of single time series so as to provide more robust estimates. Since practical problems in time series estimation typically involve more than one time series, it is important to determine if the block bootstrap procedure can work even in the multiple time series setting, how well it fares against a similar estimation procedure that does not make use of block bootstrapping, and what limitations the hybrid procedure can have when implemented in this setting.

2. Estimation of Multiple Time Series

The origins of the algorithm developed by Veron Cruz and Barrios (2014) can be traced back to the systematic procedure for applying autoregressive integrated moving average (ARIMA) models constructed by Box and Jenkins (1970). This procedure, which Box and Jenkins (1970) only applied to single time series forecasting, used a three-stage iterative process which consisted of model identification, parameter estimation, and model evaluation. The procedure was extended by Tiao and Box (1981) to apply to multiple time series settings. As with other studies that proposed estimation procedures for multiple time series, Tiao and Box (1981) recognized the value of jointly modeling two or more time series in order to make use of information from one or more series to improve the accuracy of modeling the other series. The procedure developed by Tiao and Box (1981) was found to be much simpler to implement than other procedures that were developed before it.However, the procedure assumed that each time series corresponded to different autoregressive patterns that can be estimated as a vector of parameters (Tiao & Box, 1981).Several studies considered the particular case of the time series is a multiple time series problem measuring the same unit of quantity (Akman & Gooijer, 1996;Back & Weigend, 1997; Arellano & Bond, 1991), such as stock prices of different stocks in a single market measured using a common currency. The method developed by Akman and Gooijer (1996) assumed a set $\{Z_i(t)\}$ of *n*correlated time series of common length *N* such that each can be separately represented by a stationary and invertible autoregressive moving average (ARMA) model of order (P_i, Q_i) as follows: $\phi_i(B)Z_i(t) = \theta_i(B)\xi_i(t)$, where $\phi_i(B)$ and $\theta_i(B)$ are polynomials of orders P_i and Q_i respectively, $\xi_i(t)$ is a zero-mean white noise process and B is a backshift operator defined by $B^k Z_i(t) = Z_i(t-k)$. From this model, Akman and Gooijer (1996) extracted purely AR and MA components, and then forced these components to be mutually uncorrelated and applied univariate time series modeling and forecasting techniques. Simulations conducted showed favorable results except when the parameter is set close to the invertibility or stationarity regions (Akman & Gooijer, 1996). A major limitation of the procedure is the need to identify (P_i, Q_i) for each $Z_i(t)$ before application of the method, which a complicated process for high order mixed ARMA models (Akman & Gooijer, 1996). Back and Weigend (1997) applied independent component analysis similar to the algorithm developed by Akman and Gooijer (1996) to modeling stock prices of 28 stocks in the Tokyo stock exchange. In this practical application, it was found that infrequent but large shocks to the market have greater impacts on the stability of the model than frequent but small fluctuations, which implies the importance of properly addressing the issue of temporarily but considerable volatility that the present study is concerned with.

Arellano and Bond (1991) modeled multiple time series as panel data using generalized method of moments.Similar to the present study, Arellano and Bond (1991) sought to estimate a single autoregressive parameter *α*and individual time series effects η_i of the model $y_{it} = \alpha y_{i(t-1)} + \eta_i + v_{it}$ where the number of time series *N*is large and the common length of the series *T*is small. Assuming lack of serial correlation over time, Arellano and Bond (1991) posited that *y*values lagged two or more periods can be used as valid instruments for equations in first differences. As such, parameters can be estimated using $m = (T-2)(T-1)/2$ linear moment restrictions. Simulation studies conducted by Arellano and Bond (1991) showed low bias of parameter estimates (4%-6%) across different parameter values for α. However, since the procedure explicitly assumed that *N* is large and *T*is small, it tended to diverge when the time series length is greater than the number of time series in the panel. The algorithm developed by Veron Cruz and Barrios (2014) was meant as an alternative to the Arellano-Bond procedure in such situations. The procedure made use of a backfitting algorithm that had been previously found to work for spatio-temporal models (Landagan & Barrios, 2007; Dmanjug et al., 2010; Campano & Barrios, 2011; Bastero & Barrios, 2011).

The procedure developed by Veron Cruz and Barrios (2014) estimated the common autoregressive parameter ϕ and the individual time series parameters λ_i ofN time series,each of length T,in two phases. In the first phase,the objective was to obtain an initial, possibly biased estimate for ϕ. The second phase of the procedure made use of the initial estimate to estimate λ_i and from this, ϕ, and then refine these estimates in an iterative process until the desired level of convergence was achieved. From simulation studies, Veron Cruz and Barrios (2014) found that the estimator generally has strong predictive ability based on Mean Absolute Percentage Error (MAPE) and is able to provide estimates of ϕ and λ_i's with minimal bias for most of 1944 settings investigated. The estimator performs well across different values of ϕ, except when the value is near nonstationarity (Veron Cruz & Barrios, 2014). The predictive ability of the estimator improved as the length of the time series increased and was robust to the number of time series included in the analysis. Predictive ability was not affected by the distribution of the λ_i's (Normal or Poisson), but was affected by the variance of the error term when the coefficient of variation for the distribution of the λ_i's was 100%. As expected, the algorithm was able to provide reasonable estimates for cases where the Arellano-Bond estimator typically diverged (T>N), and produced results that were comparable to the Arellano-Bond estimator when T≤N.

3. Estimation of Time Series with Volatility

As explained by Campano (2012), volatility results to heteroskedasticity, which is known to make the use of standard regression analysis tools problematic. Volatility can be generally represented by conditionality upon an exogenous variable, such as the variance of the dependent variable depending on the variance of one or more of its predictor variables (Campano, 2012). In such a representation, small changes in the variance of the predictor variable may be magnified in the variance of the dependent variable (Campano, 2012).Volatility in time series estimation has traditionally been dealt with using autoregressive conditional heteroskedastic (ARCH) and generalized autoregressive conditional heteroskedastic (GARCH) models(Bollerslev, 1986; Nelson, 1991; Enders, 2004; Brooks, 2014). ARCH models assume a functional relationship between the variance of the error terms and the actual sizes of the lag time periods' error terms (Enders, 2004). When an autoregressive moving average model is assumed for the variance of error terms, the model becomes a generalized ARCH or GARCH model (Enders, 2004). As explained by Nelson (1991),for each time point *t* in a series, ARCH models assign a scalar prediction error ξ_t, a vector of parameters *b,* avector of predetermined variables x_t, and the variance of ξ_t given information at time *t*as σ_t^2. ARCH/GARCH procedures model σ_t^2 as linear in lagged values of $\xi_t^2 = \sigma_t^2 z_t^2$ where $z_t \sim N(0,1)$ for all *t*.While these models have been found to provide good in-sample predictions, practical application of these models for forecasting has remained problematic (Politis, 2007). Politis (2007) pointed out that ARCH/GARCH models assumption of $z_t \sim N(0,1)$ often does not hold because the models imply marginal distributions for the x_t's that have heavier tails than normal. While this has been remedied by using *t*-distributions instead of normal, Politis (2007) claimed that this remedy is made arbitrarily and without sufficient justification and that out-sample predictions remain unsatisfactory because parametric modeling is simply a bad fit for forecasting such "real-world" phenomena as stock market prices (p. 360).Thus, Politis (2007) considered that the use of non-parametric modeling may be more applicable where the time series to be modeled is known to experience volatility. Building on this premise, Campano (2012) developed a robust method for estimating time series with volatility. Campano (2012) considered that the poor performance of ARCH/GARCH models is because such models assume the presence of volatility across the time series, such that when the actual volatility experienced is only temporary and the series enters a period where the volatility is no longer present, the accuracy of predictions using ARCH/GARCH models for values during that tranquil period suffers considerably (Campano, 2012). As such, Campano (2012) considered

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

Where $\lambda_i \sim N\left(\mu_i, \sigma_{\lambda_i}^2\right)$ and $\varepsilon_i \sim N\left(0, \sigma_{\varepsilon_i}^2\right)$

For all t except for some period $\{u\}$ during which the time series behaves as $u_t = h_t v$, where $v \sim N(0,1)$ and $\{h_t^2\}$ is a sequence of volatilitieswhich can be described as some function of past realizations.

The procedure is divided into three phases as follows:

For the first phase, divide each time series into 10 blocks of equal length.

For the second phase, for each of the blocks, apply the following 8 steps to obtain estimates of ϕ and λ_i's.

1.) Ignoring $\phi Y_{i,t-1}$, estimate λ_i as $\hat{\lambda}_i$ using best linear unbiased prediction (BLUP) method

2.) Compute residuals by subtracting each $Y_{i,t}$ by the corresponding $\hat{\lambda}_i$.

3.) Estimate $\hat{\phi}_i$'s by performing ARIMA on each of the residual time series obtained from the previous step to come up with N estimates for ϕ and then using bootstrap resampling to obtain a bootstrap estimate $\hat{\phi}^{BS}$ of ϕ.

4.) Computed residuals by $r_{i,t}^{re} = Y_{i,t} - \hat{\phi}^{BS} Y_{i,t-1}$.

5.) Model the residuals as $r_{i,t}^{re} = \lambda_{i,t} + \varepsilon_{i,t}$ where $\varepsilon_{i,t} \sim N(0, \sigma^2)$ to estimate the λ_i's.

6.) Use the $\hat{\lambda}_i$'s to compute for new residuals as $r_{i,t}^{ar} = Y_{i,t} - \hat{\lambda}_i$.

7.) Refine the $\hat{\phi}^{BS}$ by conducting ARIMA again on each of the $r_{i,t}^{ar}$'s and then using bootstrap resampling of the individual estimates to come up with the new $\hat{\phi}^{BS}$.

8.) Using the new $\hat{\phi}^{BS}$, iterate the process from step 4 until convergence.

For the third phase, given N sets of estimates for ϕ and the λ_i's, conduct bootstrap resampling on the N sets of estimates in order to arrive at robust final estimates of ϕ and the λ_i's.

5. Simulation

In order to test the estimation procedure, a series of simulations were conducted. Different settings for the parameters of the multiple time series were considered. Drawing from the outcomes of Veron Cruz and Barrios (2014), μ_i's were selected from $U(50,100)$ and σ^2_{λ}'s were computed from pre-set coefficients of variation. This enabled the random generation of $\hat{\lambda}_i$'s as $N(\mu_i, \sigma^2_{\lambda})$. Following this, N time series of length 2T were built by generating ε_i with preset parameters, setting ϕ and then computing for $Y_{i,t} = \phi Y_{i,t-1} + \lambda_i + \varepsilon_{i,t}$. The first half of the generated multiple time series was discarded to remove the influence of the choice of $Y_{i,0}$'s and to produce the desired N time series, each of length T. In order to induce volatility, for each time series, a random start *m*of the volatile period was determined. The random start was selected based on a pre-set range andvolatility was generated for a pre-set duration described as a percentage p of the time series length.Under this duration,i.e. from $t = m$ to $t = m + pT$,elements of the series were replaced by $Y_{i,t} = h_t v_t$, where v_t was generated as $N(0,1)$ and $h_t = \sqrt{0.1 + 0.5Z^2_{t-1}}$, where Z_{t-1} is the lag value of the original $Y_{i,t}$. The choice of h_t and v_t as well as the decision to keep these the same for all series was based on results from Campano (2012) that indicated that changing these did not affect the validity of estimates.

Overall, the following settings for the simulations were included:

N	T	ϕ	σ^2_ε	$CV_{\mu l}$	Duration of volatility	Volatility occurrence
60	100	0.5	1	5%	10%	Beginning
150	400	0.95	5	20%	20%	Middle
	1000		10	100%		End
						Differing
						None

Parameters from N to CV were selected in order to cover factors that had previously been found by Veron Cruz and Barrios (2014) to affect either predictive ability or bias of parameter estimates. Likewise, duration of volatility was varied in order to address the effect of this on bias of parameter estimates as found by Campano (2012) for single time series. Five possible settings were considered for volatility occurrence. For the "Beginning" setting, volatility was set to begin randomly anywhere in the first 10% of data points of each time series. For the "Middle" setting, this was shifted to the middle 10% of the series and for the "End" setting, volatility randomly started from the last 20% to the last 10% of each series for 10% duration and the last 30% to the last 20% for the 20% duration. Under the "Differing" setting, the volatile period for each time series began at some random point in the first 75% of the series.Finally, the "None" setting contained no volatile period.A total of 972 combinations of settings were tested. Each combination was replicated 100 times. Two methods were used to estimate the parameters of the multiple time series across each of these settings: the original procedure used in Veron Cruz and Barrios (2014), and the new procedure that integrated the block cutting technique used in Campano (2012).Results from each of these methods were evaluated using MAPE of predictions for both the entire time series and for only the non-volatile part of the series, and mean absolute bias of parameter estimates from pre-determined settings.

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.

Back, A., and Weigend, A. (1997). A First Application of Independent Component Analysis to Extracting Structure from Stock Returns. *International Journal of Neural Systems,* 8(4): 473 – 484.

Bastero, R., and Barrios, E. (2011). Robust estimation of a spatiotemporal model with structural change. *Communications in Statistics - Simulation and Computation*, 40:448–468.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity*Journal of Econometrics,* 31(3): 307–327.

Box, G., and Jenkins, G. (1970). *Time Series Analysis – Forecasting and Control.* San Francisco, CA: Holden Day.

Brooks, C. (2014). *Introductory Econometrics for Finance*. Cambridge: Cambridge University.

Campano, W. (2012). Robust Methods in Time Series Models with Volatility. *The Philippine Statistician,* 61(2): 83 – 101.

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.

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.

Enders, W. (2004). *Applied Econometrics Time Series.* NJ: John-Wiley & Sons.

Landagan, O., and Barrios, E. (2007). An estimation procedure for a spatial-temporal model. *Statistics and Probability Letters,* 77, 401–406.

Nelson, D. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59(2): 347–370.

Politis, D. (2007). Model-Free Versus Model-Based Volatility Prediction. *Journal ofFinancial Econometrics,* 5(3):358-389.

Tiao, G., and Box, G. (1981). Modeling multiple time series with applications. *Journal of the AmericanStatistical Association,* 76(376):802–816.

Veron Cruz, R., and Barrios, E. (2014). Estimation Procedure for a Multiple Time Series Model.*Communications in Statistics - Simulation and Computation,* 43(10), 2415 – 2431.