A note on the estimation of functional autoregressive models

Xavier de Luna, Suad Elezović
Department of Statistics, Umeå University

Abstract

Consider situations where a real valued function is observed over time and has a dynamic dependence structure. Linear autoregressive models, which have been proven useful to model dynamics of “pointwise” time series, can be generalized to such a functional time series situation. We call such models functional autoregressive models. Their parameters are functions of a real valued argument (as the data) and we consider a two-step estimation procedure inspired by Fan and Zhang’s (2000) proposal for functional linear models. The latter proposal is based on a first step where the ordinary least squares is used to estimate pointwise linear models for given values of the argument of the functions observed. The second step smoothes the first-step estimates, regressing the latter on the mentioned arguments. The second step does not only yield smooth estimates of the functional parameters but also provides less variable pointwise estimates at the price of a bias. We do not only contribute by presenting an autoregressive model for functional data but also by proposing a two-stage estimator where the first step takes into account the contemporaneous correlation structure through a multivariate generalized least squares estimator. Some of the properties of the resulting two-step procedure are given. Financial functional data is used as an illustration.

Keywords: Financial data; functional time series; multivariate generalized least squares; seemingly unrelated autoregression.
1 Introduction

In this note we introduce a procedure for the estimation of a functional time series model first utilized in Elezović (2008) to model volatility in Swedish electronic limit order book. For an observed time series of functions of $w \in \mathbb{R}$, $y_t(w)$, where $t \in \mathbb{Z}$ indexes time, we consider the following autoregressive model

$$y_t(w) = \theta_0(w) + \sum_{i=1}^{p} \theta_i(w)y_{t-i}(w) + \varepsilon_t(w), \quad (1)$$

where $\theta_i(\cdot)$, $i = 0, 1, \ldots, p$, are $p + 1$ unknown real valued parameter functions, $\varepsilon_t(w)$ is a stochastic process with mean zero and covariance function $\gamma(v, w) = \text{cov}\{\varepsilon_t(v), \varepsilon_t(w)\}$, $\forall t$. We call such a model a functional autoregressive model not to be confused with the functional coefficient autoregressive models (e.g. Fan and Yao, 2003), which are models for pointwise time series, where parameters are functions of covariates. On the other hand, Fan and Zhang (2000) consider a similar model to (1). However, while they work with functionals of time observed on independently sampled individuals, our functional time series model consider a time series of functionals. Fan and Zhang (2000) proposed a two step estimator. In a first step they perform a univariate estimation pointwise for given observed values of $w$. Then, in a second step, they refine these estimates by smoothing them over the values of $w$. The second step is natural when the functions $\theta_i(w)$, $i = 0, \ldots, p$, are assumed to be smooth. Fan and Zhang also argue that efficiency is gained by smoothing since the univariate estimators do not take into account the information at neighboring $w$s. In this paper, we first argue that the univariate estimators should be replaced by multivariate estimators that take into account the correlation structure of the error term $\varepsilon_t(w)$. We derive some properties of a two step estimation procedure (multivariate estimation followed by smoothing) for the functional autoregressive model (1).

Section 2 describes the model, the different estimation procedures and their properties. In Section 3, the estimation and inference is
applied to high frequency financial data. The paper is concluded in Section 4.

2 Method and theory

2.1 Model

We assume we have observations of the function $y_t(w_j)$ at $W$ distinct values, $j = 1, \ldots, W$ and at $T$ time points $t = 1, 2, \ldots, T$. For model (1) at value $w_j$, the data is generated by a classical linear autoregressive model of order $p$ (e.g., Brockwell and Davis, 1991)

$$y_t(w_j) = \theta_0(w_j) + \sum_{i=1}^{p} \theta_i(w_j)y_{t-i}(w_j) + \epsilon_t(w_j),$$

where $\epsilon_t(w_j)$ are independently distributed with mean zero and variance $\gamma(w_j, w_j)$. We assume that the process $y_t(w)$ is causal (p. 262 Brockwell and Davis, 1991) for any $w$, i.e. it has a representation $y_t(w) = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}(w), t \in \mathbb{Z}$, with $\sum_{j=0}^{\infty} |\psi_j| < \infty$. This, in particular, ensures that the multivariate process

$$Y_j = (y_1(w_j), y_2(w_j), \ldots, y_T(w_j))'$$

is stable and stationary as defined in Lütkepohl (2005).

2.2 Univariate estimation

Several estimators of the vector $\theta_j = (\theta_0(w_j), \theta_1(w_j), \ldots, \theta_p(w_j))'$ are available, e.g., Yule-Walker, ordinary least-squares (OLS), Burger algorithm and Gaussian maximum likelihood. All these estimators have been shown to be asymptotically equivalent (Brockwell and Davis, 1991). In particular, for these estimators, denoted here $\hat{\theta}_r(w_j)$, we have

$$\sqrt{T} \left( \hat{\theta}_r^j - \theta_j \right) \xrightarrow{D} N(0, G_j),$$

where $\xrightarrow{D}$ stands for convergence in distribution as $T \to \infty$, and

$$G_j = \gamma(w_j, w_j)\Gamma_{pj},$$

3
with \( \mathbf{\Gamma}_{pj} \) being the covariance matrix of the time series \( y_t(w_j) \), \( \mathbf{\Gamma}_{pj} = [\text{Cov}\{y_s(w_j), y_t(w_j)\}]_{s,t=1}^p. \)

2.3 SUR representation and GLS

The naive univariate estimators can be improved because they do not take into account the fact that the residuals are correlated. A multivariate generalized least squares estimator (GLS) is here more efficient.

Model (1) may be interpreted as a **seemingly unrelated (auto) regression** (SUR) model (Zellner, 1962), a set of \( W \) linear autoregressive equations

\[
y_j = \mathbf{X}_j \theta_j + \mathbf{e}_j, \quad j = 1, 2, \ldots, W, \tag{4}\]

where

\[
\mathbf{e}_j = (\varepsilon_1(w_j), \varepsilon_2(w_j), \ldots, \varepsilon_T(w_j))', \quad \mathbf{\theta}_j = (\theta_0(w_j), \theta_1(w_j), \ldots, \theta_p(w_j))',
\]

and \( \mathbf{X}_j \) is a matrix of regressors defined for each \( j \) as follows

\[
\mathbf{X}_j = \begin{bmatrix}
1 & y_0(w_j) & y_{-1}(w_j) & \cdots & y_{-p+1}(w_j) \\
1 & y_1(w_j) & y_0(w_j) & \cdots & y_{-p+2}(w_j) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & y_{T-1}(w_j) & y_{T-2}(w_j) & \cdots & y_{T-p}(w_j)
\end{bmatrix}. \tag{5}\]

Then a SUR model may be specified by stacking all the equations from (4) into a single autoregression model (Judge et al., 1985, p. 466)

\[
\begin{bmatrix}
\mathbf{y}_1 \\
\mathbf{y}_2 \\
\vdots \\
\mathbf{y}_W
\end{bmatrix} = \begin{bmatrix}
\mathbf{X}_1 & 0 & \cdots & 0 \\
0 & \mathbf{X}_2 & 0 & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \mathbf{X}_W
\end{bmatrix}\begin{bmatrix}
\mathbf{\theta}_1 \\
\mathbf{\theta}_2 \\
\vdots \\
\mathbf{\theta}_W
\end{bmatrix} + \begin{bmatrix}
\mathbf{e}_1 \\
\mathbf{e}_2 \\
\vdots \\
\mathbf{e}_W
\end{bmatrix}, \tag{6}\]

which can be expressed in a compact form as

\[
\mathbf{y} = \mathbf{X}\mathbf{\theta} + \mathbf{e}, \tag{7}\]
where \( y, X, \theta \) and \( e \) are of dimensions \((WT \times 1), (WT \times W(p + 1)), (W(p + 1) \times 1) \) and \((WT \times 1)\), respectively. We have from the assumptions made earlier that \( E[e_j] = 0 \) and \( E[e_j e_k'] = \gamma(w_j, w_k)I_T \), with \( I_T \) the identity matrix of dimension \( T \times T \). The joint covariance matrix of \( e \) is then

\[
E[ee'] = \Omega = \Sigma \otimes I_T,
\]

where \( \otimes \) is the Kronecker product operator (e.g. Lütkepohl, 2005, p. 660), and

\[
\Sigma = \begin{bmatrix}
\gamma(w_1, w_1) & \gamma(w_1, w_2) & \cdots & \gamma(w_1, w_W) \\
\gamma(w_2, w_1) & \gamma(w_2, w_2) & \cdots & \gamma(w_2, w_W) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma(w_W, w_1) & \gamma(w_W, w_2) & \cdots & \gamma(w_W, w_W)
\end{bmatrix}.
\]

The multivariate generalized least squares estimator (GLS) is

\[
\hat{\theta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.
\]

This estimator is more efficient than the univariate estimator of the previous section in situations where the matrix \( \Omega \) is not diagonal, otherwise both estimators are equivalent.

Under the assumptions made earlier and if the error terms have a joint normal distribution then

\[
\sqrt{T} \left( \hat{\theta} - \theta \right) \xrightarrow{\text{d}} N(0, V),
\]

where \( V = \text{plim}_{T \to \infty} T \left( X'\Omega^{-1}X \right)^{-1} \), see Lütkepohl (2005, Sec. 5.2) for a proof of similar results.
Note that we can write

\[ V = \text{plim}_{T \to \infty} T \left[ X' (\Sigma^{-1} \otimes I_T) X \right]^{-1} \]  

\[ = \text{plim}_{T \to \infty} T \begin{bmatrix} \sigma^{11}X_1 X_1 & \sigma^{12}X_1 X_2 & \cdots & \sigma^{1W}X_1 X_W \\ \sigma^{21}X_2 X_1 & \sigma^{22}X_2 X_2 & \cdots & \sigma^{2W}X_2 X_W \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{W1}X_W X_1 & \sigma^{W2}X_W X_2 & \cdots & \sigma^{WW}X_W X_W \end{bmatrix}^{-1} \]

\[ = \begin{bmatrix} \sigma^{11} \Gamma_{11} & \sigma^{12} \Gamma_{12} & \cdots & \sigma^{1W} \Gamma_{1W} \\ \sigma^{21} \Gamma_{21} & \sigma^{22} \Gamma_{22} & \cdots & \sigma^{2W} \Gamma_{2W} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{W1} \Gamma_{W1} & \sigma^{W2} \Gamma_{W2} & \cdots & \sigma^{WW} \Gamma_{WW} \end{bmatrix}^{-1}, \]

where \( \sigma^{ij} \) is the element on the \( i \)th row and \( j \)th column in \( \Sigma^{-1} \).

In typical situations \( \Sigma \) is not known and must be estimated based on univariate least squares residuals \( \hat{e}_j = y_j - X_j \hat{\theta}_j \), yielding

\[ \hat{\gamma}(w_j, w_k) = \frac{\hat{e}_j' \hat{e}_k}{T}, \quad j, k = 1, 2, \ldots, W. \]

The asymptotic properties of the estimator (10) are not modified when the estimated covariance matrix is plugged in (Lütkepohl, 2005, Sec. 5.2). Furthermore, a consistent estimator of \( V \) is obtained using \( \hat{\Gamma}_{jk} = \frac{1}{T} X_j' X_k \), \( j, k = 1, 2, \ldots, W \).

### 2.4 Improving estimation by smoothing

In situations where we expect the parameter functions \( \theta_i(w), i = 0, \ldots, p \) to be smooth we may try to improve on the naive or GLS estimators by using a smoother (Kernel, Local polynomial regression, Splines, etc.), as advocated by Fan and Zhang (2000) for functional linear models. We consider here linear smoother yielding the new estimator

\[ \hat{\theta}_i^s(w) = \sum_{j=1}^W k(w, w_j) \hat{\theta}_i(w_j), \]  

(13)
where \(k(w, w_j)\) are weights defined by the smoothing technique used and \(\hat{\theta}_i(w_j)\) is then the multivariate GLS or univariate OLS estimator. Furthermore, assuming that \(\theta_i(w)\) is \((p+1)\)-times continuously differentiable, then

\[
\hat{\theta}_i^{(q)}(w) = \sum_{j=1}^{W} k_q(w, w_j)\hat{\theta}_i(w_j)
\]

is an estimate of the \(q\)th derivative of \(\theta_i(w)\) for \(0 \leq q < p + 1\). By smoothing we introduce a bias compared with the non-smooth estimate, with the aim to improve on the variance.

It is straightforward to see that

\[
\sqrt{T} \left( \hat{\theta}_i^{(q)}(w) - b_i(w) \right) \xrightarrow{D} N(0, v_i(w)),
\]

where

\[
b_i(w) = \sum_{j=1}^{W} k_q(w, w_j)\theta_i(w_j)
\]

and

\[
v_i(w) = \sum_{j=1}^{W} \sum_{l=1}^{W} k_q(w, w_j)k_q(w, w_l)\text{asCov}(\hat{\theta}_i(w_j), \hat{\theta}_i(w_l)),
\]

where for the GLS estimator \(\text{asCov}(\hat{\theta}_i(w_j), \hat{\theta}_i(w_l))\) are elements of the matrix \(V\) given above. This asymptotic result can be used to construct confidence intervals if we ignore the bias.

### 3 Application to financial data

In Elezović (2008) data from the electronic Swedish limit order book was used to compute a volatility measure (realized quadratic variation) of the spread between ask and bid prices as a function of the number of shares. This measure is typically computed as the sum
of squared inter-daily high-frequency financial returns, (see, e.g., Andersen and Bollerslev, 1998; Barndorff-Nielsen and Shephard, 2002a). Here, the RQV estimates are computed from the spreads of the inter-daily functional financial price returns of the Ericsson B shares between January 3, 2006 and August 25, 2006, thereby yielding 167 daily realized quadratic variation functions of \( w \) (a percentage number of shares with respect to those available in the order book); see Elezović (2008) for details. The data is illustrated in Figure 1, where we have chosen to compute the realized quadratic variations at 30 randomly chosen relative quantities, \( w \), within \((0\%, 100\%)\).

Model (1) with \( p = 3 \) is fitted using OLS, GLS and a smooth version of both. The smoother used is Loess (implementation \texttt{loess} in the statistical software R) (R Development Core Team, 2009), with bandwidth \( h = 0.5 \) chosen by eye-balling.

Figure 2 shows the fitted coefficient curves with \( \pm 2 \) pointwise standard errors. These pointwise standard errors are obtained from the estimated asymptotic covariance matrix \( \hat{\Sigma} \); see Section 2.3. Figure 3 depicts the widths of the 95\% pointwise confidence bands from OLS, GLS and the Loess fit.

We see from Figure 3 that the estimated asymptotic variance of the GLS estimates are clearly smaller than for the OLS estimates, while there does not seem to be gain in efficiency when smoothing the GLS estimates. Hence, while smoothing the OLS estimates has been shown to improve on efficiency in Fan and Zhang (2000) for linear models and in Elezović (2008) in a time series setting, it is not clear whether smoothing GLS would yield efficiency gains, due to the fact that GLS already takes into account the dependence structure in the residuals which the univariate OLS ignores. On the other hand, one may still be willing to smooth the GLS estimates to obtain smooth estimates of the functional parameters of the models. Further research is needed to study these issues since this is merely an illustrative example. Note, finally, that varying the values for \( h \) did not change the width of the confidence bands significantly.
4 Conclusion

In this paper we have presented a new estimator for functional autoregressive models. The estimator may also be used for the functional linear models of Fan and Zhang (2000). Our estimator is based on a SUR representation for which it is natural to use a multivariate GLS estimator instead of the univariate OLS estimator proposed by Fan and Zhang. Both the OLS and the GLS estimator may be smoothed.
in a second step in cases where the parameter functions in the model are assumed smooth. We present some asymptotic properties of the GLS estimator and its smooth version when a linear smoother is used.

We conjecture that similar asymptotic ($W \to \infty$) results to those obtained by Fan and Zhang (2000) will hold for (13) in the context of this paper. A related issue is to study methods to choose the smoothing parameter defining the weights of the smoother in an optimal way.
Figure 3: Width of confidence intervals for the fitted coefficient curves with OLS, GLS, OLS-Loess and GLS-Loess (h=0.5).

Acknowledgements

We are grateful to Maria Karlsson for comments that improved the presentation of the paper.
References


