INSTITUT FÜR HÖHERE STUDIEN
INSTITUTE FOR ADVANCED STUDIES
Vienna

Big (Time Series) Data: Mixed Frequency, Forecasting and Graphical Time Series Models

Philipp Gersing [1*]

[1] Department of Mathematics, Vienna University of Technology, Institute for Advanced Studies, Vienna

Supervisor: Prof. Manfred Deistler

* gersing@ihs.ac.at

1 Mixed Frequency: A New Estimator and Extension to GDFMs

In many high dimensional economic time series the univariate time series occur at different sampling frequencies. The approach described in [1], [2], and related papers is to directly model the high-frequency system as a VAR or VARMA system. In [1] generic identifiability for high frequency VAR systems is established via a blocking approach which is the starting point for the construction of the new estimator.

1.1 Estimator and Asymptotics

We consider a weakly stationary stochastic process of stock variables (flow and linear aggregations are also treated) $(y_t) = (y_t^{f\prime}\ y_t^{s\prime})'$, where (y_t^f) is a $n_f \times 1$ vector observed at $t \in \mathbb{Z}$ but (y_t^s) is a $n_s \times 1$ vector observed only at a lower sampling frequency $N\mathbb{Z}$ for some $N \in \mathbb{N}$. The underlying high frequency systems is a VAR-system:

$$y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + \nu_t \quad \nu_t \sim WN(\Sigma_\nu) \tag{1}$$

where $\Sigma_\nu = q \leq n$ is a white noise process. Define $a(z) = I_n - A_1 z - ... - A_p z^p$ as the lag polynomial. Set $\nu_t = b\varepsilon_t$ where $\varepsilon_t \sim WN(I_q)$, then $\Sigma_\nu = bb'$ for b being a $n \times q$ matrix with rank q. Note that we allow $q < n$, i.e. we allow (y_t) to be a *singular VAR process*. The parameter space is

$$\Theta = \left\{(A_1, ..., A_p) | \det(a(z)) \neq 0 \quad \forall z \leq 1\right\} \times \left\{\Sigma_\nu | \Sigma_\nu = \Sigma_\nu', \Sigma_\nu \geq 0, \Sigma_\nu = q\right\}. \tag{2}$$

The following second moments of (y_t) are observed:

$$\begin{aligned} \gamma^{ff}(h) &= y_{t+h}^f y_t^{f\prime} \quad h \in \mathbb{Z} \\ \gamma^{sf}(h) &= y_{t+h}^s y_t^{f\prime} \quad h \in \mathbb{Z} \\ \gamma^{ss}(h) &= y_{t+h}^s y_t^{s\prime} \quad h \in N\mathbb{Z}. \end{aligned} \tag{3}$$

The observed outputs $(y_t^f \mid t \in \mathbb{Z})$ and $(y_t^s \mid t \in N\mathbb{Z})$ can be represented by the blocked process

$$\tilde{y}_t = \begin{pmatrix} y_t^f \\ y_t^s \\ y_{t-1}^f \\ \vdots \\ y_{t-N+1}^f \end{pmatrix} \quad t \in N\mathbb{Z}$$

$(\tilde{y}_t \mid t \in N\mathbb{Z})$ has a state space representation (see [1]). The population second moments of $(\tilde{y}_t)$ are exactly the observable population second moments of the high frequency process (y_t) described in (3). The goal is to find an estimator for the underlying high frequency VAR system based on the blocked process $(\tilde{y}_t \mid t \in N\mathbb{Z})$. In a first step a state space system for $(\tilde{y}_t \mid t \in N\mathbb{Z})$ is estimated by subspace methods (see [3]). This state space system does in general not correspond to a VAR system because VAR processes are not closed under marginalisation. In the second step, we approximate the estimates obtained in the first step by a VAR system. This is done by taking the appropriate root of the state transition matrix and by approximating it with the companion matrix corresponding to the high dimensional VAR system. The main point here is to find the statistically optimal approximation.

1.2 Mixed Frequency Generalised Dynamic Factor Models

We consider a generalised dynamic factor model for stationary square integrable double sequences $(x_{it} \mid i \in \mathbb{N}, t \in \mathbb{Z})$ as described in [4]:

$$x_{it} = \underbrace{b_i(z)u_t}_{:= \chi_{it}} + \xi_{it}. \tag{4}$$

where (u_t) is a q - dimensional orthonormal white noise process orthogonal to the idiosyncratic components ξ_{it} and $b_i(L)$ is a rational filter. We suppose that the underlying high frequency process has a generalised dynamic factor representation (4) but is observed at mixed sampling frequency. Our goal is to develop a denoising procedure, i.e. a procedure providing an estimate for the latent variables (χ_{it}) for this case. This relates to the singular VAR systems with mixed frequency.

2 Targeted Low Dimensional Subspaces

We aim to find a new dimension reduction method and structure theory for forecasting a target output variable with a (ultra) high dimensional predictor vector. We focus on three aspects here: Firstly, the method should be designed for the incorporation of *highly disaggregated/granular predictor data* (big time series data) in the regression equation. Secondly, the extraction of the low dimensional process should be *targeted* to the output variable that we want to forecast. Thirdly the procedure should capture sparse and dense structures in the data correctly (see explanation below).Suppose we have one or more high dimensional stationary $n_k \times 1$ predictor processes (input processes) $\{x_t^{(k)} \mid k = 1, ..., K\}$ and a scalar target variable (y_t) which is to be predicted, where all univariate components have zero mean and finite second moments. For a $n \times 1$ process (x_t) set $\mathbb{H}(x) := \overline{\text{sp}}\{x_{it} \mid 1 \leq i \leq n, t \in \mathbb{Z}\}$. Our goal is to construct low dimensional processes $(v_t^{(k)})$ $1 \leq k \leq K$ from $(x_t^{(k)})$ which are specifically designed for the purpose of forecasting (y_t). The low dimensional vectors $(v_t^{(k)})$ might be filtered versions or linear combinations of the $(x_t^{(k)})$'s. The final prediction equation might look as follows:

$$\begin{aligned} y_{t+h} = \alpha_0 + b_1(z)v_t^{(1)} + ... + b_K(z)v_t^{(K)} \\ + \alpha_1 y_{t-1} + ... + \alpha_p y_{t-p} + \varepsilon_{t+h} \end{aligned} \tag{5}$$

where $b_k(z)$ are filters.

3 Graphical Models for High Dimensional Time Series

Networks and graphical models for time series have obtained increasing attention recently. We are interested in the structure, estimation and testing of (conditional) Granger causality networks. Our interest is on changes of causality pattern under additive noise and time aggregation, certain kinds of non-stationarity, and again on the use of mixed frequency data in this context.

References

[1] B Anderson, M Deistler, E Felsenstein, B Funovits, and M Zamani. Multivariate AR systems and mixed frequency data: g-identifiability and estimation. *Econometric Theory*, pages 793–826, 2016.

[2] B Anderson, M Deistler, E Felsenstein, and L Koelbl. The structure of multivariate AR and ARMA systems: Regular and singular systems, the single and the mixed frequency case. *Journal of Econometrics*, pages 1–8, 2016.

[3] M Deistler, K Peternell, and W Scherrer. Consistency and relative efficiency of subspace methods. *Automatica*, pages 1865–1875, 1995.

[4] M Forni and M Lippi. The generalized dynamic factor model: representation theory. *Econometric Theory*, 2001.