Comment on “Generalised Linear Dynamic Factor Models - An Approach via Singular Autoregressions”

Brett Ninness and Adrian Wills

While this paper makes an important contribution in a systems and control context, it is also a fascinating example of bi-directional contribution between that field, and the field of econometrics.

More specifically, the genesis of the paper arises in economic and financial modeling scenarios where the dimension N of the vector of system measurements y_t^N can be large, and possibly bigger than the number of observations of this vector T. For example, as noted in [1], y_t^N may contain measurements of “*different macro variables, returns on different assets, [and] data disaggregated by sector or region*”.

Importantly, models of this sort are likely to be of increasing importance in systems and control settings too as the demands for analyzing larger scale systems rises, particularly in relation to communication, transport and energy networks.

Within economic and finance fields, this paper notes that various methods for estimating the parameters modeling such processes have been developed, with particular attention paid to exploiting structural aspects in order to restrict the dimension of the parameter space. One example is the frequency domain type approach in [1].

This paper carefully develops a new and efficient approach based on AR modeling and solving subsequent Yule–Walker equations, where the latter is not straightforward due to the “innovations” process being rank deficient.

A distinctive feature is the significant depth of underpinning theory provided to support and develop the proposed approach. This draws heavily on concepts and results developed over many years in the systems and control community - many of them by the first two authors. Hence while the problem has its original roots in econometrics and finance, it is highly relevant in a systems and control setting, and indeed as illustrated by this paper, the tools from that field can contribute substantially back into econometrics and related areas.

The authors do not arrive at an AR modeling approach casually, and the paper contains a careful analysis of alternatives, such as ARMA modeling via a state-space description

$$x_{t+1} = Fx_t + G\varepsilon_{t+1}, \qquad \widehat{y}_t = Hx_t. \tag{1}$$

Interestingly, the first named author together with colleagues has made important contributions on the topic of certain minimal (injective) parametrizations of such models now known as “*data driven local co-ordinates*” (DDLC) [2]. More specifically, a fully (and hence over) parametrization β of (1) would be of dimension $n_\beta = n(n+q+N)$, while DDLC methods provide an efficient mechanism for working with a minimal injective parametrization θ of dimension $n_\theta = n(q+N)$. This is employed, for example, in the widely used Matlab System Identification Toolbox in the context of gradient-based search for prediction error (PE) and maximum likelihood (ML) estimates.

In the case of ML estimation, this requires a Kalman filter, with floating point operation (FLOP) load $O(n^2T)$ to be run for each parameter θ_i, resulting in an overall FLOP load *per iteration* of the gradient based search of $O(n^3NT)$.

Clearly then, there is an important practical motivation for this paper in seeking efficient alternate estimation methods when N is large.

In relation to this, the authors here would like to raise the important work [3] on this topic that does not seem to enjoy the attention it possibly deserves. This involves the ML estimation case wherein the estimate $\widehat{\theta}$ of θ is taken as

$$\widehat{\theta} = \arg\min_{\theta} L(\theta), \quad L(\theta) \triangleq \log p(Y_T;\theta) \tag{2}$$

where $Y_T = \{y_1, \cdots, y_T\}$. In [3] it is noted that by Fisher's identity, the gradient of $L(\theta)$ is identical to that

$$\mathcal{Q}(\theta) = \mathbf{E}_\theta\{\log p(X_T, Y_T;\theta) \mid Y_T\} \tag{3}$$

where the derivative is taken only with respect to the density inside the conditional expectation, and $X_T = \{x_1, \cdots, x_T\}$.

The practical relevance of this is that only *one* Kalman smoother needs to be run, with FLOP load $O(n^3T)$, in order to compute all necessary gradients of $L(\theta)$ with FLOP load $O(\max(n^3, nN)T)$. In the context of this comment, this has the potential to reduce the FLOP load per iteration of gradient-based search by the substantial factor of N.

This naturally leads to the question of empirical analysis profiling the performance of the various estimation options available, including the newly developed one in this paper, which could be a valuable topic for future work.

Brett Ninness and Adrian Wills are with the School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW 2308, Australia {Adrian.Wills, Brett.Ninness}@newcastle.edu.au

References

[1] M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic-factor model:Identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554, 2000.

[2] T. Ribarits, M. Deistler, and T. McKelvey. An analysis of the parameterization by data driven local coordinates for multivariable linear systems. *Automatica*, 40:789–803, 2004.

[3] Mordechai Segal and Ehud Weinstein. A new method for evaluating the log-likelihood gradient (score) of linear dynamic systems. *IEEE Transactions on Automatic Control*, 33(8):763–766, 1988.