Discussion: Efficient estimation and forecasting in dynamic factor models with structural instability

Geert Mesters

Tinbergen Institute
VU University Amsterdam
The Netherlands Institute for the Study of Crime and Law Enforcement

www.geertmesters.nl

June, 2014

Overview of the paper

- Dynamic factor models where all model parameters are considered time-varying
- Mixture between observation driven and parameter driven approaches
- Variances are made TV using observation driven approach
- Loadings and VAR coefficients are made TV using parameter driven approach
- Two algorithms to approximate posterior mode: 1. filtering based, 2. simulation based
- Two applications: 1. macroeconomic forecasting, 2. yield-spread forecasting
- For the discussion I will focus on algorithm 1 and application 1

Equations

$$
\begin{aligned}
x_t &= \Lambda_t f_t + \epsilon_t & \epsilon_t &\sim N(0, V_t) \\
\lambda_t &= \lambda_{t-1} + \nu_t & \nu_t &\sim N(0, R_t) \\
f_t &= B_t f_{t-1} + \eta_t & \eta_t &\sim N(0, Q_t) \\
\beta_t &= \beta_{t-1} + \upsilon_t & \upsilon_t &\sim N(0, W_t)
\end{aligned}
$$

$$
\begin{aligned}
V_t &= \delta_1 V_{t-1} + (1-\delta_1)\mathrm{diag}(\hat{\epsilon}_t \hat{\epsilon}_t') \\
R_t &= (\mu_1^{-1} - 1)\mathrm{Cov}(\lambda_{t-1} | x_1, \ldots, x_{t-1}) \\
Q_t &= \delta_2 Q_{t-1} + (1-\delta_2)\hat{\eta}_t \hat{\eta}_t' \\
W_t &= (\mu_2^{-1} - 1)\mathrm{Cov}(\beta_{t-1} | x_1, \ldots, x_{t-1})
\end{aligned}
$$

Algorithm 1

1. Compute $\hat{f}^{\text{PCA}}$
2. Compute $\hat{\lambda}_t = \mathrm{E}(\lambda_t | X; \hat{f}^{\text{PCA}})$, for $t = 1, \ldots, T$
3. Compute $\hat{\beta}_t = \mathrm{E}(\beta_t | \hat{f}^{\text{PCA}})$, for $t = 1, \ldots, T$
4. Compute $\hat{f}_t = \mathrm{E}(f_t | X; \hat{\lambda}, \hat{\beta})$, for $t = 1, \ldots, T$

- Step (2); also gives V_t and R_t
- Step (3); also gives Q_t and W_t

General comments and questions

1. Ambitious paper!!!
2. Consistency of step (1) in algorithm 1? Bates, Plagborg-Moller, Stock & Watson (2013) give rates for $\lambda_{i,t}$? In addition B_t will also require some restrictions.
3. Imposing some structure on the loading matrix? Testing for parameter instability?
4. If forecasting and computational speed are the goals; why not entirely observation driven?
5. Reasoning the particular observation driven structure?

EWMA vs GAS update

- EWMA update per element

$$V_{i,t} = \delta_1 V_{i,t-1} + (1-\delta_1)\hat{\epsilon}^2_{i,t}$$

- GAS update per element

$$V_{i,t} = \delta_1 V_{i,t-1} + (1-\delta_1) S_t(-\frac{1}{2}F^{-1}_{i,t} + \frac{1}{2}\hat{\epsilon}^2_{i,t}F^{-2}_{i,t})$$

- where $\hat{\epsilon}_{i,t} = x_{i,t} - \hat{\lambda}_{t|t-1}\tilde{f}_t$ and $\tilde{f}_t$ is the current estimate for f_t
- and S_t is a scaling term, $F_{i,t} = \tilde{f}_t' \hat{P}_{i,t}\tilde{f}_t + V_{i,t-1}$
- Main difference is that GAS update also depends on predictive variance loadings
- Possible room for improvement see Blasques, Koopman & Lucas (2014)

Some further questions?

1. Are the variances in step (4) treated as known? if so why? Re-estimating V_t and Q_t is possible? When doing forecasting this will make a difference.
2. In general: how are the forecasts constructed?
3. Is it possible to estimate model parameters (μ's and δ's) in steps (ii) and (iii) using MLE? Similar as in Eickmeier, Lemke & Marcellino (2011). Not much work and saves grid-searches?
4. How do you initialize V_0 and Q_0 in general?

Table 2: Relative MSE for forecasting German GDP growth

Panel A: EWMA

		parameter specification					Relative MSE at horizon			
		r	δ_1	δ_2	μ_1	μ_2	1	2	3	4
1	TVP-DFM	2	0.83	0.83	1.00	1.00	0.77	0.95	0.98	1.08
2	TVP-DFM	3	0.83	0.83	1.00	1.00	0.77	0.95	0.98	1.09
3	TVP-DFM	2	0.87	0.83	1.00	1.00	0.77	0.95	0.98	1.08
4	TVP-DFM	3	0.87	0.83	1.00	1.00	0.77	0.95	0.98	1.09
5	TVP-DFM	2	0.99	0.83	1.00	1.00	0.76	0.96	1.00	1.10
6	TVP-DFM	3	0.99	0.83	1.00	1.00	0.77	0.96	1.00	1.12
7	TVP-DFM	2	0.83	0.99	1.00	1.00	0.82	1.04	1.09	1.10
8	TVP-DFM	3	0.83	0.99	1.00	1.00	0.83	1.06	1.10	1.20
9	TVP-DFM	2	0.99	0.99	1.00	1.00	0.82	1.07	1.12	1.24
10	TVP-DFM	3	0.99	0.99	1.00	1.00	0.83	1.08	1.13	1.24
11	TVP-DFM	2	0.99	0.99	0.98	0.98	0.82	1.07	1.12	1.24
12	TVP-DFM	3	0.99	0.99	0.98	0.98	0.83	1.08	1.13	1.24
13	PC	1	-	-	-	-	0.86	1.02	1.02	1.08
14	PC	2	-	-	-	-	0.87	1.01	1.01	1.09
15	PC	3	-	-	-	-	0.91	1.02	1.01	1.09
16	PC	4	-	-	-	-	0.94	1.06	1.08	1.15
17	AR	-	-	-	-	-	1.00	1.00	1.00	1.00

Comments

1. The PCA estimator is based on homoscedastic error-variances $V_t = I_n\sigma$. In the macro illustration V_t is initialized with $V_0 = I_N$. When $\delta_1 < 1$ *two* things change: (1) the variances become heteroskedastic and (2) the variances become time-varying. The improvement in forecasting is entirely attributed to the time-variation in the paper. A comparison with standard MLE would give more insight into the role for heterogeneous variances; see also Bai & Li (2012).
2. In the simulation study there is also a comparison w.r.t. two-step estimator of Doz, Giannone & Reichlin (2011), which shows that 4-steps improves two-step when time-variation in the loadings is large. Given that the loadings and factor coefficients are not time-varying in this application a comparison with two-step would be insightful.

Final remarks

- I enjoyed reading the paper.
- Thank you!