Readme file for: A replication of "A Quasi-Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models" (Review of Economics and Statistics, 2012)

Riccardo (Jack) Lucchetti[1] Ioannis A. Venetis[2]

May 12, 2020

[1]Università Politecnica delle Marche, Dipartimento di Scienze Economiche e Sociali (DiSES), Ancona, Italy. Email: r.lucchetti@univpm.it

[2]University of Patras, School of Economic Sciences and Business Administration, Department of Economics, Greece. Email: ivenetis@upatras.gr

Main replication scripts

Please note that, in order to have the scripts run correctly, you will need the `DFM` package (version 0.42 at least) for gretl.[1] After starting gretl, go to `File-->Function packages-->On server...`, scroll down to find `DFM`, right click on it and install it.

1. run `tables-2-4-6.inp` to replicate Tables 2, 4 and 6 in our main paper. As it is, the script file replicates Table 2. Uncomment line 17 and set the `TABLE` number to either 4 or 6 to replicate the corresponding tables (tables 4 and 6 in the main paper). With 5000 runs, the simulation experiment can be time consuming. Use instead 100 in line 26 to just evaluate the script's operational performance. Upon completion, a txt file is created to the working directory under the name `table_2_100.txt` or `table_4_100.txt` or `table_6_100.txt` (if 100 replications are employed)

2. run `table-7.inp` to replicate table 7 in our main paper. It writes a txt file to the working directory named

   `table_7_rep_100_tau_0.5_u_0.1_a_0.9_d_0.5.txt`

   if 100 replications used). It contains the six sub-panels incorporated in Table 7 of the paper.

3. run `table-8-MSE.inp` to replicate table 8. It writes a txt file to the working directory named

   `table_8_MSE_rep_100_tau_0.5_u_0.1_a_0.9_d_0.5.txt`

   if 100 replications used. It contains the six sub-panels incorporated in Table 8 of the paper.

Supporting scripts

1. The file `simfuncs.inp` contains functions that are specific to the paper's simulation exercise (for example the data generation process). It is employed by `tables-2-4-6.inp` and `table-7.inp`.

2. The file `simfuncs2.inp` contains functions that are specific to the paper's simulation exercise for MSE computations. It is employed by the `table_#_MSE.inp` scripts.

[1] gretl is an open source software. We use gretl version **gretl 2020b**. It can be downloaded: gretl for Windows at http://gretl.sourceforge.net/win32/ or gretl on macOS at http://gretl.sourceforge.net/osx.html

3. In order to obtain table A7, run script `table-A7.inp`. Prior to running the script, uncomment lines 396-398 in `simfuncs.inp` and comment line 394.

4. In order to obtain table A8, run script `table-A8-MSE.inp`. Prior to running the script, uncomment lines 397-399 in `simfuncs2.inp` and comment line 395.

A. Some extra tables I

Tables A7 and A8 correspond to tables 7 and 8 in the paper but the experiment is extended with

(i) more elaborate autoregressive dynamics for the factors,

(ii) heterogeneous autoregressive coefficients for idiosyncratic components and

(iii) factor loadings that generate a factor structure which is more pervasive than simply relying on the zero mean Normal distribution.

Table A7: Extended experiment

	$N = 10$	$N = 25$	$N = 50$	$N = 100$
		TR_{ML}		
$T = 50$	0.48	0.52	0.55	0.58
$T = 100$	0.56	0.61	0.63	0.67
		Number of iterations		
$T = 50$	43	27	21	17
$T = 100$	29	19	15	12
		TR_{ML}/TR_{TS}		
$T = 50$	1.37	1.12	1.04	1.04
$T = 100$	1.58	1.11	1.08	1.05
		TR_{ML}/TR_{PC}		
$T = 50$	1.19	1.18	1.18	1.18
$T = 100$	1.22	1.18	1.18	1.18
		$TR_{ML}/TR_{PC,1}$		
$T = 50$	1.36	1.32	1.34	1.39
$T = 100$	1.30	1.25	1.26	1.31
		$TR_{ML}/TR_{PC,1,lags}$		
$T = 50$	1.03	1.04	1.07	1.10
$T = 100$	1.04	1.04	1.06	1.11

Notes: Comparing alternative dynamic factor extraction methods when $s = 1$. Data generation process input set at: $s = 1$, $p = 1$, $\tau = 0.5$, $u = 0.1$, $q = 2$. Choice $\alpha = 3.0$ is equivalent to drawing the autoregressive coefficient matrix for the $VAR(1)$ of the dynamic factors as in Barigozzi and Luciani (2019, p.22). Choice $d = 4.0$ means the autoregressive parameters for the idiosyncratic component d_i were drawn from a uniform distribution on $[-0.6, 0.6]$ while idiosyncratic component errors were drawn independently from the t_4 distribution. Finally, the factor loadings are such that $[\Lambda_0]_{i,j} \overset{i.i.d}{\sim} U[0.5, 1.5]$ and $[\Lambda_1]_{i,j} \overset{i.i.d}{\sim} U[-0.4, 0.4]$

Table A8: Extended experiment. Relative MSE statistics

	$N = 10$	$N = 25$	$N = 50$	$N = 100$
MSE stat for ML				
$T = 50$	0.76	0.74	0.72	0.72
$T = 100$	0.75	0.73	0.72	0.72
Number of iterations				
$T = 50$	44	27	21	17
$T = 100$	29	19	15	12
relative MSE(ML/TS)				
$T = 50$	0.92	0.97	1.00	1.00
$T = 100$	0.91	0.98	0.98	1.00
relative MSE(ML/PC)				
$T = 50$	0.99	0.99	0.99	0.99
$T = 100$	0.99	1.00	1.00	1.00
relative MSE(ML/PC,1)				
$T = 50$	1.06	1.04	1.02	1.01
$T = 100$	1.04	1.02	1.01	1.00
relative MSE(ML/PC,1,lags)				
$T = 50$	1.06	1.04	1.02	1.02
$T = 100$	1.04	1.02	1.01	1.01

Notes: Comparing alternative dynamic factor extraction methods using relative MSE calculations when $s = 1$. Data generation process input set at: $s = 1$, $p = 1$, $\tau = 0.5$, $u = 0.1$, $q = 2$. Choice $\alpha = 3.0$ is equivalent to drawing the autoregressive coefficient matrix for the $VAR(1)$ of the dynamic factors as in Barigozzi and Luciani (2019). Choice $d = 4.0$ means the autoregressive parameters for the idiosyncratic component d_i were drawn from a uniform distribution on $[-0.6, 0.6]$ while idiosyncratic component errors were drawn independently from the t_4 distribution. Finally, the factor loadings are such that $[\Lambda_0]_{i,j} \overset{i.i.d}{\sim} U[0.5, 1.5]$ and $[\Lambda_1]_{i,j} \overset{i.i.d}{\sim} U[-0.4, 0.4]$

B. Some extra tables II

Number of Monte Carlo instances with "very low" trace statics across all estimation methods.

Table B1: Number of cases where estimation "failed"

	$N = 5$	$N = 10$	$N = 25$	$N = 50$	$N = 100$
			Table 2		
$T = 50$	57	7	0	1	1
$T = 100$	1	0	0	0	0
			Table 4		
$T = 50$	-	0	0	0	0
$T = 100$	-	0	0	0	0
			Table 6		
$T = 50$	-	0	0	0	0
$T = 100$	-	0	0	0	0
			Table 7		
$T = 50$	-	2	0	0	0
$T = 100$	-	0	0	0	0

Notes: Each panel entry denotes the number of times we obtained a trace statistic $TR_{.}$ for all three methods that was less than 0.05 and it was dropped from our simulations. For example, for the results we report in table 2 case $N = 5, T = 50$, we run 5057 replications and dropped the 57 instances with all three $TR_{ML}, TR_{TS}, TR_{PC} < 0.05$ when computing the final average relative trace statistics.

Bibliography

Barigozzi M, Luciani M. 2019. Quasi maximum likelihood estimation and inference of large approximate dynamic factor models via the em algorithm. *arXiv preprint arXiv:1910.03821* .
URL https://arxiv.org/pdf/1910.03821