Macroeconomic Forecasting with Large Data Sets under Asymmetric Loss

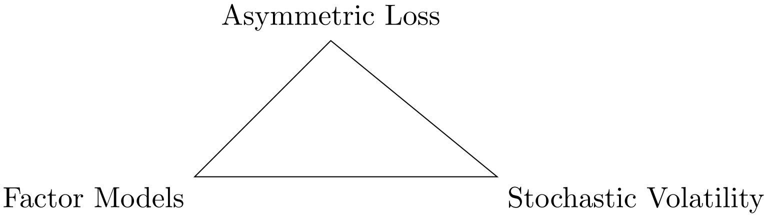
Sinem Hacioglu (joint work with Matei Demetrescu)

IWH-CIREQ Macroeconometric Workshop

3 December 2013

Outline

We seek to provide answers to possible questions arising from forecasting of factor augmented models under generalised loss functions.



Why Factor Models?

- Increased availability and use of comprehensive data sets.
- Large number of series can be summarized by the use of Principal Component Analysis (PCA).
- Stock and Watson (2002) show that using PCA extracted factors in the predictive regression does not affect the consistency of the forecast function.
- Bai (2003) showed the rate of convergence and the limiting distributions of the PCA estimated factors, factor loadings, and common components.

The Basic Model

We start with the usual linear forecasting model

$$y_{t+h} = c + \sum_{j=1}^{q} a_j y_{t-j+1} + \sum_{k=1}^{r} b_k f_{t,k} + v_{t+h}, \quad t = 1, 2, \ldots, T\,, \tag{1}$$

where the forecast error v_{t+h}. In practice, one resorts to a two-stage procedure, given that observations on N auxiliary variables $x_{t,i}$ are available, from which $f_{t,k}$ may be estimated in a first stage. Assuming linear measurement equations for the factors, we have that

$$x_{t,i} = \sum_{k=1}^{r} \lambda_{i,k} f_{t,k} + u_{t,i}. \tag{2}$$

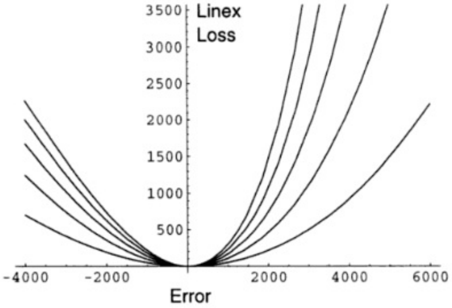
This ultimately takes us to the feasible predictive regression

$$y_{t+h} = c + \sum_{j=1}^{q} a_j y_{t-j+1} + \sum_{k=1}^{r} b_k \hat{f}_{t,k} + v_{t+h}, \tag{3}$$

Why Asymmetric Loss Functions?

- There are recent studies which focus on more general loss functions, in particular asymmetric loss functions.
  - Artis and Marcellino (2001): IMF and OECD forecasts of the deficit of G7 countries were found to be systematically biased towards over or under prediction in compared to mean-squared optimal forecasts.
  - Christodoulakis and Mamatzakis (2008, 2009) found asymmetric preferences of EU institutional forecasts.
  - Pierdzioch et al. (2011) analysed the loss function of Bank of Canada, along with Pierdzioch et al. (2013) where evidence on the use of asymmetric loss functions of various forecasters on yen/dollar exchange rate forecasts is found.
  - Elliott et al. (2005), Clements et al. (2007), Patton and Timmermann (2007), Capistrán (2008) so on.

An example for asymmetric loss function



Source: Variants in Economic Theory: Selected Works of Hal R. Varian

Why not Linex Loss function in our study as well? $\rightarrow$ Requires infinitely many moments!

The loss function we are going to use has the following structure (Elliott et al. (2005))

$$\mathcal{L}\left(y_{t+h} - y^f_{t+h}\right) = \left(\alpha + (1-2\alpha)\,\mathbb{I}\left(y_{t+h} - y^f_{t+h}\right)\right)\left|y_{t+h} - y^f_{t+h}\right|^p. \tag{4}$$

Why Stochastic Volatility?

- The volatility of macroeconomic variables is not constant in general.
- The question we ask is: does usual factor extraction capture all the relevant information under a given loss function?
- Patton and Timmermann (2007) showed that for the loss functions of the type we use, the optimal forecast has the form

$$y_{t+h}^{opt} = \mathrm{E}\left(y_{t+h} \middle| y_t, y_{t-1}, \ldots, x_{t,i} \right) + C\sqrt{\mathrm{Var}\left(y_{t+h} \middle| y_t, y_{t-1}, \ldots, x_{t,i} \right)} \tag{5}$$

for some constant C depending on the loss function and the shape of the conditional distribution.

Three Questions

- Can we estimate the predictive regression with extracted factors under a given loss function instead of using Ordinary Least Squares estimation(OLS) for evaluating the forecasts?
- Are the PCA-extracted factors still forecast relevant under the given loss function?
- Does usual factor extraction actually capture all relevant information under the given loss function?

First Question - Predictive regression under asymmetric loss

Let y_t be the series for which an h-step ahead forecast is required. Given the available information set $\mathcal{F}_t = \{f_{t,k}, y_t, y_{t-1}, \ldots\}$, the optimal forecast is given by

$$y_{t+h}^{opt} = \arg\min_{y^*} \mathrm{E}\left(\left|\mathcal{L}\left(y_{t+h} - y^*\right)\right| \left|\mathcal{F}_t\right.\right), \tag{6}$$

where $\mathcal{L}(\cdot)$ is the relevant loss function quantifying the cost of discrepancies between a given forecast and the future realization of y_{t+h}. The estimated optimal forecast is

$$\tilde{y}_{t+h}^{opt} = \tilde{c} + \sum_{j=1}^{q} \tilde{a}_j y_{t-j+1} + \sum_{k=1}^{r} \tilde{b}_k \hat{f}_{t,k}. \tag{7}$$

which is obtained by minimizing the observed loss

$$\tilde{c}, \tilde{a}_j, \tilde{b}_k = \arg\min_{c^*, a_j^*, b_k^*} \frac{1}{T} \sum_{t=q}^{T-h} \mathcal{L}\left(y_{t+h} - \tilde{y}_{t+h}^{\star}\right) \tag{8}$$

First Question - Predictive regression under asymmetric loss

Proposition

Let the auxiliary variables $x_{t,i}$ obey Assumptions A-E in Bai (2003). Furthermore, assume that the factors $f_{t,k}$ and the forecast errors v_{t+h} are strictly stationary and ergodic, and that $\mathrm{E}\left(\mathcal{L}'\left(v_{t+h}\right)\middle| y_t, y_{t-1}, \ldots, f_{t,k}\right) = 0$. Finally, let all series have finite moments of order p. It then holds for the estimated optimal forecast from (7) that

$$\tilde{y}_{t+h}^{opt} \xrightarrow{p} y_{t+h}^{opt} \tag{9}$$

as $N, T \to \infty$ such that $T/N \to 0$.

Second Question - Forecast relevant factors

- Factor extraction by PCA relies on the eigenvalue decomposition of the covariance matrix of the data set. Principal components are chosen by starting with the ones corresponding to the highest eigenvalue.

- What if these PCA extracted factors are *not* forecast relevant? To this extend, we employ Least Absolute Shrinkage and Selection Operator (LASSO) to make sure that we use forecast relevant factors.

Third Question - Extracting additional relevant information

To exploit the insight of capturing additional information from the volatility of macroeconomic variables we assume a stochastic volatility model

$$v_{t+h} = e_t \, e^{\frac{1}{2}\left(g_t + \sum_{l=1}^{s} \xi_l h_{t,l}\right)}$$

where g_t is an unforecastable component (stochastic or constant), while $h_{t,l}$ could be forecasted using the information from the auxiliary series. Should the conditional variance of the idiosyncratic components in the factor model depend in a similar manner on $h_{t,l}$

$$u_{t,i} = e_{t,i} \, e^{\frac{1}{2}\left(g_{t,i} + \sum_{l=1}^{s} h_{t,l}\xi_{l,i}\right)},$$

By following Nelson (1991) in using the exponential, we have

$$\log u_{t,i}^2 = \log e_{t,i}^2 + g_{t,i} + \sum_{l=1}^{s} \xi_{l,i} h_{t,l},$$

We can now extract $h_{t,l}$ from $\log u_{t,i}^2$ using PCA, leading to $\hat{h}_{t,l}$.

Forecasting US Personal Income and Industrial Production under Asymmetric Loss

We have the following basic forecasting model

$$\tilde{y}^{opt}_{t+h} = \tilde{c} + \sum_{j=1}^{q} \tilde{a}_j y_{t-j+1} + \sum_{k=1}^{r} \tilde{b}_k \hat{f}_{t,k} + \sum_{l=1}^{s} \tilde{\xi}_{l,i} \hat{h}_{t,l}, \tag{10}$$

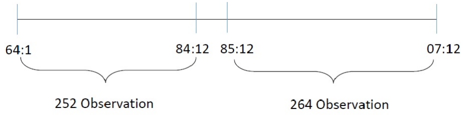
where the parameter estimates are obtained like before by minimising the observed loss

$$\mathcal{L}\left(y_{t+h} - y^{f}_{t+h}\right) = \left(\alpha + (1-2\alpha)\,\mathbb{I}\left(y_{t+h} - y^{f}_{t+h}\right)\right)\,\left|y_{t+h} - y^{f}_{t+h}\right|^2. \tag{11}$$

by setting $p = 2$.

Data

- The data set (also known as Stock and Watson data set) consists of 131 monthly macroeconomic aggregates over the time 1964:1-2007:12 and it is now used to forecast personal income and industrial production.
- We evaluate the pseudo out–of–sample forecasts using factors recursively extracted from the auxiliary data.



- Stock and Watson initially categorised the data set by assigning each variable to one of eight main groups. Moreover, Bai and Ng (2002) information criteria show that eight factors survive the test.

Cases

- We aim to observe the averaged losses occurring while evaluating the forecasts of US personal income (PI) and industrial production (IP)
- We have 4 cases:
  - Case 1: exactly 8 factors.
  - Case 2: exactly 9 factors (including the volatility factor)
  - Case 3: selection among 8 factors.
  - Case 4: selection among 9 factors.
- We also evaluate the averaged losses both under OLS and Asymmetric Loss both for PI and IP.

Results

Alpha	Cases	OLS-PI	OLS-IP	Asym Loss-PI	Asym Loss-IP
0.1	1	0.2629	0.2081	0.2001	0.0879
	2	0.2522	0.1898	0.2011	0.0887
	3	0.2589	0.2052	0.1930	0.0871
	4	0.2516	0.1887	0.1943	0.0882
0.3	1	0.2414	0.1813	0.2223	0.1395
	2	0.2360	0.1712	0.2215	0.1396
	3	0.2380	0.1790	0.2187	0.1381
	4	0.2348	0.1702	0.2194	0.1389
0.5	1	0.2199	0.1545	0.2199	0.1545
	2	0.2198	0.1527	0.2198	0.1527
	3	0.2170	0.1527	0.2170	0.1527
	4	0.2180	0.1516	0.2180	0.1516
0.7	1	0.1984	0.1277	0.1998	0.1414
	2	0.2036	0.1341	0.2009	0.1379
	3	0.1960	0.1265	0.1970	0.1397
	4	0.2012	0.1331	0.1990	0.1366
0.9	1	0.1769	0.1009	0.1541	0.0897
	2	0.1874	0.1156	0.1576	0.0867
	3	0.1750	0.1003	0.1503	0.0881
	4	0.1844	0.1145	0.1535	0.0845

Diebold-Mariano Test

Alpha	Cases	DM Test–PI	DM Test–IP
0.1	1	-3.42	-4.12
	2	-3.21	-3.44
	3	-3.68	-4.15
	4	-3.51	-3.48
0.3	1	-3.15	-3.14
	2	-2.52	-2.36
	3	-3.48	-3.12
	4	-2.86	-2.35
0.5	1	-1.38	0.36
	2	-0.78	-0.57
	3	0.77	0.67
	4	-0.94	-0.03
0.7	1	0.26	1.21
	2	-0.46	0.32
	3	0.24	1.21
	4	-0.49	0.31
0.9	1	-1.54	-0.57
	2	-1.88	-1.37
	3	-1.89	-0.64
	4	-2.27	-1.47

Conclusions

- There were 3 questions which are related to
  - Feasible predictive regressions by using extracted factors under asymmetric loss
  - Forecasting with relevant factors
  - Extracting additional relevant information

- We used recursive forecasting method to observe forecast losses of PI and IP by making use of large number of predictor series.

- There is a clear evidence that evaluating the forecasts under the asymmetric loss leads smaller losses in compared to OLS case.

Appendix

Let y_t be the series for which an h-step ahead forecast is required. Given the available information set $\mathcal{F}_t = \{f_{t,k}, y_t, y_{t-1}, \ldots\}$, the optimal forecast is given by

$$y_{t+h}^{opt} = \arg\min_{y^*} \mathrm{E}\left(\left|\mathcal{L}\left(y_{t+h} - y^*\right)\right| |\mathcal{F}_t\right), \tag{12}$$

where $\mathcal{L}(\cdot)$ is the relevant loss function quantifying the cost of discrepancies between a given forecast and the future realization of y_{t+h}. According to Granger (1999), loss functions should be uniquely minimized at the origin, and be quasi-convex. We shall work with a specific class of loss functions, introduced by Elliott (2005); a forecast y_{t+h}^f is thus evaluated by means of

$$\mathcal{L}\left(y_{t+h} - y_{t+h}^f\right) = \left(\alpha + (1-2\alpha)\,\mathbb{I}\left(y_{t+h} - y_{t+h}^f\right)\right)\left|y_{t+h} - y_{t+h}^f\right|^p. \tag{13}$$

Appendix

We start with the usual linear forecasting model

$$y_{t+h} = c + \sum_{j=1}^{q} a_j y_{t-j+1} + \sum_{k=1}^{r} b_k f_{t,k} + v_{t+h}, \quad t = 1, 2, \ldots, T\,, \quad (14)$$

where the forecast error v_{t+h} cannot be predicted under $\mathcal{L}$. This does not imply, however, that v_{t+h} could not be forecasted under another loss function. The lack of predictability of v_{t+h} under $\mathcal{L}$ implies that the so-called generalised forecast error $\mathcal{L}'(v_{t+h})$ is uncorrelated with the predictors y_{t-j+1} and $f_{t,k}$; see Granger (1999) and Patton and Timmermann (2007). The optimal forecast is thus given by

$$y_{t+h}^{opt} = c + \sum_{j=1}^{q} a_j y_{t-j+1} + \sum_{k=1}^{r} b_k f_{t,k}.$$

In practice, one resorts to a two-stage procedure, given that observations on N auxiliary variables $x_{t,i}$ are available, from which $f_{t,k}$ may be estimated in a first stage. Assuming linear measurement equations for the factors, we have that

$$x_{t,i} = \sum_{k=1}^{r} \lambda_{i,k} f_{t,k} + u_{t,i}. \tag{15}$$

Particularly with orthogonality of the common and idiosyncratic components $f_{t,k}$ and $u_{t,i}$, we end up in an approximate factor model which ultimately takes us to the feasible predictive regression

$$y_{t+h} = c + \sum_{j=1}^{q} a_j y_{t-j+1} + \sum_{k=1}^{r} b_k \hat{f}_{t,k} + v_{t+h}, \tag{16}$$

to be estimated under the relevant loss in a second stage, i.e.

$$\tilde{c}, \tilde{a}_j, \tilde{b}_k = \arg\min_{c^*, a_j^*, b_k^*} \frac{1}{T} \sum_{t=q}^{T-h} \mathcal{L}\left(y_{t+h} - c^* - \sum_{j=1}^{q} a_j^* y_{t-j+1} - \sum_{k=1}^{r} b_k^* \hat{f}_{t,k} \right), \tag{17}$$

The forecast is obtained as

$$\tilde{y}_{t+h}^{opt} = \tilde{c} + \sum_{j=1}^{q} \tilde{a}_j y_{t-j+1} + \sum_{k=1}^{r} \tilde{b}_k \hat{f}_{t,k}. \tag{18}$$

We now prove by the propositon earlier that the feasible forecast from (7) consistently estimates (as $T, N \to \infty$) the unfeasible optimal forecast under the relevant loss $\mathcal{L}$.

Assume a stochastic volatility model

$$v_{t+h} = e_t \, e^{\frac{1}{2}\left(g_t + \sum_{l=1}^{s} \xi_l h_{t,l}\right)}$$

where g_t is an unforecastable component (stochastic or constant), while $h_{t,l}$ could be forecasted using the information form the auxiliary series. We follow Nelson (1991) in using the exponential, since it allows us to avoid positivity restrictions on the components g_t and $h_{t,l}$.

Should the conditional variance of the idiosyncratic components in the factor model depend in a similar manner on $h_{t,l}$,

$$u_{t,i} = e_{t,i}\, e^{\frac{1}{2}\left(g_{t,i} + \sum_{l=1}^{s} h_{t,l}\xi_{l,i}\right)},$$

where $g_{t,i}$ are individual volatility components specific for $x_{t,i}$. As usually, e_t and $e_{t,i}$ are standardised variables, mutually independent and independent of $h_{t,l}$, g_t and $g_{t,i}$. Then,

$$\log u_{t,i}^2 = \log e_{t,i}^2 + g_{t,i} + \sum_{l=1}^{s} \xi_{l,i} h_{t,l},$$

which is nothing else than a factor model for the log squares of $u_{t,i}$ with $h_{t,l}$ the common components and $\log e_{t,i}^2 + g_{t,i}$ the idiosyncratic ones. We can now extract $h_{t,l}$ from $\log u_{t,i}^2$ using PCA, leading to $\hat{h}_{t,l}$.