High Dimensional Forecasting

Francis J. DiTraglia

University of Pennsylvania

Econ 722

What Have We Learned So Far?

What's Different About High-Dimensional Problems?

- OLS performs very badly if the number of regressors is large relative to sample size.
- Estimation uncertainty can be a problem
- Noise accumulation in PCA

Main References

Stock & Watson (2006) – "Forecasting with Many Predictors"

Overview of high-dimesional forecasting with a review of forecast combination, factor models, and Bayesian approaches.

Ng (2013) – "Variable Selection in Predictive Regressions"

Reviews and relates a number of shrinkage & selection methods.

Overview

- What have we covered in the course so far? Try to tie everything together today, suggest extensions and open problems.
- Survey articles
- Diffusion index forecasting (Stock and Watson 2002)
- Other ways of extracting factors: Sparse PCA, ICA
- Other ways of forecasting: Stock and Watson 2012, Kim and Swanson
- Target or not? Bai and Ng (2008), Kelly & Pruitt, PLS, etc.
- Nonlinear stuff? Kernel methods, Bai and Ng (2008), that recent paper from a job candidate...
- Boosting and Bagging
- Open Questions: selection with generated predictors, how should we choose lasso and ridge parameters for dependent data?
- Inference post-selection with shrinkage estimators.

Diffusion Index Forecasting – Stock & Watson (2002a,b)

JASA paper has the theory, JBES paper has macro forecasting example.

Basic Setup

Forecast scalar time series y_{t+1} using N-dimensional collection of time series X_t where we observe periods $t = 1, \ldots, T$.

Assumption

Static representation of Dynamic Factor Model:

$$\begin{aligned} y_t &= \beta' F_t + \gamma(L) y_t + \epsilon_{t+1} \\ X_t &= \Lambda F_t + e_t \end{aligned}$$

"Direct" Multistep Ahead Forecasts

"Iterated" forecast would be linear in F_t, y_t and lags:

$$y^h_{t+h} = \alpha_h + \beta_h(L) + \gamma_h(L) + \gamma_h(L) y_t + \epsilon^h_{t+h}$$

This is really just PCR

Diffusion Index Forecasting – Stock & Watson (2002a,b)

Estimation Procedure

1. Data Pre-processing
   1.1 Transform all series to stationarity (logs or first difference)
   1.2 Center and standardize all series
   1.3 Remove outliers (ten times IQR from median)
   1.4 Optionally augment X_t with lags
2. Estimate the Factors
   - No missing observations: PCA on X_t to estimate $\widehat{F}_t$
   - Missing observations/Mixed-frequency: EM-algorithm
3. Fit the Forecasting Regression
   - Regress y_t on a constant and lags of $\widehat{F}_t$ and y_t to estimate the parameters of the "Direct" multistep forecasting regression.

Diffusion Index Forecasting – Stock & Watson (2002b)

Recall from last time that, under certain assumptions, PCA consistently estimates the space spanned by the factors. Broadly similar assumptions are at work here.

Main Theoretical Result

Moment restrictions on (ϵ, e, F) plus a "rank condition" on Λ imply that the MSE of the procedure on the previous slide converges to that of the infeasible optimal procedure, provided that $N, T \to \infty$.

Diffusion Index Forecasting – Stock & Watson (2002a)

Forecasting Experiment

- Simulated real-time forecasting of eight monthly macro variables from 1959:1 to 1998:12
- Forecasting Horizons: 6, 12, and 24 months
- "Training Period" 1959:1 through 1970:1
- Predict h-steps ahead out-of-sample, roll and re-estimate.
- BIC to select lags and # of Factors in forecasting regression
- Compare Diffusion Index Forecasts to Benchmark
  - AR only
  - Factors only
  - AR + Factors

Diffusion Index Forecasting – Stock & Watson (2002a)

Empirical Results

- Factors provide a substantial improvement over benchmark forecasts in terms of MSPE
- Six factors explain 39% of the variance in the 215 series; twelve explain 53%
- Using all 215 series tends to work better than restricting to balanced panel of 149 (PCA estimation)
- Augmenting X_t with lags isn't helpful