Forecasting Multiple Time Series Using One-Sided Dynamic Principal Components

Ezequiel Smucler, joint work with Daniel Peña and Victor J. Yohai

Department of Statistics, University of British Columbia

Introduction

Forecasting a large number, m, of cross-correlated time series is challenging, since the usual parametric models for multivariate time series have a number of parameters that grows as m^2, and hence estimation requires a very large number of observations.

Introduction

A possible approach is to first reduce the dimensionality of the data and then apply a low-dimensional forecasting procedure to the resulting low-dimensional time series.

In this talk, we introduce a dynamic version of principal components, called **one-sided dynamic principal components (ODPC)**, that, unlike previous proposals for principal components for time series (Brillinger 81', Peña and Yohai 16'), is useful for forecasting.

Overview

Principal Components

We have $\mathbf{z}_t \in \mathbb{R}^m$, $t = 1, \ldots, T$, time series with zero mean. Let $\mathbf{Z} \in \mathbb{R}^{T \times m}$ be the data matrix with $\mathbf{z}_t'$ as rows,

$$\mathbf{Z} = \begin{pmatrix} z_{1,1} & z_{1,2} & \cdots & z_{1,m} \\ \vdots & \vdots & \vdots & \vdots \\ z_{T,1} & z_{T,2} & \cdots & z_{T,m} \end{pmatrix}$$

We want to find a simpler, lower dimensional structure.

Principal Components

One of the classical statistical techniques for dimension reduction is principal component analysis (PCA), (Pearson 1901, Hotelling 1933).

Principal components do not assume any model for the data and aim at finding linear combinations of the data with optimal properties.

Principal Components

Let $\widehat{\mathbf{a}} \in \mathbb{R}^m$ be the unit norm eigenvector of the covariance matrix of the data, corresponding to its maximum eigenvalue.

$\widehat{\mathbf{a}}$ is called the first principal direction. The first principal component is defined by

$$\hat{f}_t = \mathbf{z}_t' \widehat{\mathbf{a}} \quad t = 1, \ldots, T.$$

It can be shown that $\widehat{\mathbf{a}}$ is the direction over which the variance of the data is maximized.

Principal Components

Suppose we want to obtain a vector $\mathbf{f} \in \mathbb{R}^T$, $\mathbf{f}' = (f_1, \ldots, f_t)$ and loadings $\mathbf{b} \in \mathbb{R}^m$, such that if we reconstruct $\mathbf{z}_t$ with $\mathbf{f}$ and $\mathbf{b}$ as

$$\mathbf{b} f_t,$$

the reconstruction mean squared error

$$\frac{1}{T} \sum_{t=1}^{T} \|\mathbf{z}_t - \mathbf{b} f_t\|^2$$

is minimal.

By the Eckart-Young Theorem, the first principal component, $\widehat{\mathbf{f}}' = (\mathbf{z}_1' \widehat{\mathbf{a}}, \ldots, \mathbf{z}_T' \widehat{\mathbf{a}})$ is an optimal solution to the reconstruction problem:

$$(\widehat{\mathbf{f}}, \widehat{\mathbf{a}}) \in \arg \min_{\mathbf{f} \in \mathbb{R}^T, \mathbf{b} \in \mathbb{R}^m} \frac{1}{T} \sum_{t=1}^{T} \|\mathbf{z}_t - \mathbf{b} f_t\|^2$$

Principal Components

PCA does not take into account the dynamic structure of the data. The reconstruction of $\mathbf{z}_t$ given by PCA only uses $\widehat{f}_t = \mathbf{z}_t' \widehat{\mathbf{a}}$.

The method we propose (ODPC), is based on adding dynamics to the reconstruction criterion for which the classical principal components are optimal.

One-Sided Dynamic Principal Components

Given $k \in \mathbb{N}$, $\mathbf{A} \in \mathbb{R}^{(k+1)\times m}$ with rows $\mathbf{a}_h'$ and $\mathbf{B} \in \mathbb{R}^{(k+1)\times m}$ with rows $\mathbf{b}_h'$, we can define

$$f_t = f_t(\mathbf{A}) = \sum_{h=0}^{k} \mathbf{z}_{t-h}' \mathbf{a}_h, \quad t = k+1, \ldots, T$$

and reconstruct $\mathbf{z}_t$ as

$$\widehat{\mathbf{z}}_t(\mathbf{A}, \mathbf{B}) = \sum_{h=0}^{k} \mathbf{b}_h f_{t-h}(\mathbf{A}).$$

One-Sided Dynamic Principal Components

The reconstruction error is measured as

$$\text{MSE}(\mathbf{A}, \mathbf{B}) = \frac{1}{T - 2k} \sum_{t=2k+1}^{T} \left\| \mathbf{z}_t - \widehat{\mathbf{z}}_t(\mathbf{A}, \mathbf{B}) \right\|^2 .$$

Optimal values of $\mathbf{A}$ and $\mathbf{B}$, say $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$, are defined as those that minimize $\text{MSE}(\mathbf{A}, \mathbf{B})$.

We define $\widehat{\mathbf{f}} = \widehat{\mathbf{f}}(\widehat{\mathbf{A}})$, associated with some optimal $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$ as the first ODPC of the data. For definiteness, we take $\widehat{\mathbf{A}}$ with $\|\widehat{\mathbf{A}}\|_F = 1$.

The number of lags used, k, is a tuning parameter that needs to be chosen.

One-Sided Dynamic Principal Components

Note that $\widehat{\mathbf{A}}$ plays a role in the definition of the ODPC analogous to that played by the $\widehat{\mathbf{a}}$, the first principal direction.

$$\hat{f}_t = \hat{f}_t(\widehat{\mathbf{A}}) = \sum_{h=0}^{k} \mathbf{z}'_{t-h}\widehat{\mathbf{a}}_h, \quad t = k+1, \ldots, T$$

The second ODPC is defined as the first ODPC of the residuals.

Computing ODPC

It can be shown that minimizing $\text{MSE}(\mathbf{A}, \mathbf{B})$ only over $\mathbf{A}$ or only over $\mathbf{B}$ amounts to solving a least squares problem.

Starting from an initial guess, we alternate solving a least squares problem for $\mathbf{B}$ and solving a least squares problem for $\mathbf{A}$ until convergence.

Forecasting using ODPC

Let $\widehat{\mathbf{f}}' = (\widehat{f}_{k+1}, \ldots, \widehat{f}_T)$ be the first ODPC and $\widehat{\mathbf{B}}$ be the corresponding matrix of loadings, with entries $\widehat{b}_{h,j}$.

Given some univariate forecasting procedure, forecast $\widehat{f}_{T+1}$ and compute the 1-step ahead forecast of $\mathbf{z}_T$ as

$$\widehat{z}_{T+1,j} = \sum_{h=0}^{k} \widehat{b}_{h,j} \widehat{f}_{T+1-h} \quad j = 1, \ldots, m.$$

Higher order forecasts can be defined in a similar fashion.

Simulation

We conduct a simulated forecasting exercise, 500 replications, compute one-step ahead forecast MSE.

Data is generated according to

$$z_{t,j} = \sin(2\pi j/m) f_t + \cos(2\pi j/m) f_{t-1} + (j/m) f_{t-2} + f_{t-3} + e_{t,j},$$

where the $e_{t,j}$ are i.i.d. standard normal random variables. The factor f_t satisfies an MA(2) model.

Simulation

We assume the number of lags and factors is known, and compute the ODPC and the procedures of Stock and Watson (2002) (SW), Forni et al. (2005) (FHLR) and Forni et al. (2015) (FHLZ), based on factor models.

To compute the ODPC forecasts, we use an automatic algorithm to select an ARIMA model (Hyndman and Khandakar, 2008) for the dynamic principal component.

Simulation

T	m	ODPC	FHLR	FHLZ	SW
50	50	**1.46**	1.51	1.74	1.54
	100	**1.41**	1.45	1.64	1.53
	200	**1.44**	1.45	1.64	1.53
100	50	**1.33**	1.46	1.57	1.49
	100	**1.29**	1.37	1.51	1.44
	200	**1.26**	1.33	1.49	1.38
200	50	**1.28**	1.37	1.48	1.41
	100	**1.22**	1.31	1.40	1.37
	200	**1.23**	1.30	1.45	1.36

Table 1: Prediction MSEs of ODPC, FHLR, FHLZ and SW in a dynamic factor model for different values of T and m.

An empirical example

We conduct a forecasting exercise similar to the one in McCracken and Ng (2015), using the FRED-MD dataset.

We have monthly observations on 94 important indicators of the US economy ranging from January 1960 to February 2014, that is, 650 periods. Among the 94 series included are

- Initial jobless claims (CLAIMSx).
- S&P 500 index (S&P 500).
- Industrial production index (INDPRO).

An empirical example

The panel is transformed to approximate stationarity by using appropriate transformations. In particular CLAIMSx, S&P 500 and INDPRO are transformed by first difference of the logarithm.

We compare the performances of ODPC, FHLR, FHLZ, and SW when forecasting these last three series.

If W_T stands for one of the three series at time T, the target is

$$\log\left(\frac{W_{T+h}}{W_T}\right).$$

An empirical example

We compute the ODPC, FHLR and SW for one component and up to three lags and FHLZ using one dynamic factor.

We take a two year forecast horizon ($h = 24$) and use a rolling seven year window from 2005:03 to 2012:02, compute the h-steps ahead forecasts and compare the predicted values of the target variable with the actual values.

The performance of each procedure is measured by its average mean squared error relative to the average mean squared error of the one-dimensional ARIMA.

An empirical example

	CLAIMSx	S&P 500	INDPRO
ODPC 1	**0.508**	0.851	**0.754**
ODPC 2	0.522	**0.846**	0.784
ODPC 3	0.531	0.863	0.815
FHLR 1	1.012	1.019	0.871
FHLR 2	0.770	0.979	0.789
FHLR 3	0.796	0.989	0.780
FHLZ 1	1.003	0.999	0.880
SW 1	1.179	1.066	1.070
SW 2	1.213	1.081	1.095
SW 3	1.245	1.099	1.119

Table 2: MSE forecasting errors for different number of lags, relative to the MSE of the one dimensional ARIMA forecast.

Theory for ODPC

Assume $\mathbf{z}_t$ is strictly stationary and ergodic, with finite second moments and zero mean.

Let

$$MSE_0(\mathbf{A}, \mathbf{B}) = \mathbb{E}MSE(\mathbf{A}, \mathbf{B}) = \mathbb{E}\|\mathbf{z}_t - \widehat{\mathbf{z}}_t(\mathbf{A}, \mathbf{B})\|^2$$

and

$$\mathcal{I} = \{(\mathbf{A}^*, \mathbf{B}^*) : MSE_0(\mathbf{A}^*, \mathbf{B}^*) \text{ is minimal}\}$$

be the population version of the problem that defines ODPC.

Theory for ODPC

Theorem

Under appropriate assumptions

- $d((\widehat{\mathbf{A}}, \widehat{\mathbf{B}}), \mathcal{I}) \xrightarrow{a.s.} 0$ *as* $T \to \infty$.
- *If the data follows a dynamic factor model with* $\mathbf{z}_t = \boldsymbol{\chi}_t + \mathbf{e}_t$, *for all* $(\mathbf{A}^*, \mathbf{B}^*) \in \mathcal{I}$

$$\frac{1}{m}\mathbb{E}\left\|\boldsymbol{\chi}_t - \widehat{\mathbf{z}}_t(\mathbf{A}^*, \mathbf{B}^*)\right\|^2 \to 0 \quad \text{as } m \to \infty.$$

Conclusions

In the spirit of principal component analysis, ODPC is based on finding linear combinations of the observations with optimal reconstruction properties; being based only on lags of the data, it is useful for forecasting large sets of time series.

In practice, the number of components and lags used has to be chosen. One possibility is to choose them by looking to minimize the cross-validated forecasting error. This is part of our current research.

Conclusions

The paper this talk is based on is available on arXiv.

The `odpc` `R` package to compute ODPCs is available on CRAN.

Thank you