The wedding of smoothed periodogram and one factor model in the frequency domain

Hilmar Böhm   Rainer von Sachs

Institut de Statistique
Université Catholique de Louvain

September 21st, 2007

An appetizer

Suppose that we have a p-variate ($p > 2$) Gaussian random variable X that is distributed

$$X \sim \mathcal{N}(\mu, \mathrm{Id}_p),$$

and want to estimate the unknown mean vector μ, for which we have a sample of size $N = 1$ only. What kind of estimator can we use?

The canonical solution

The usual parametric estimator would be to take the sample mean,

$$\hat{\mu} = X. \quad (1)$$

This estimator is least squares, maximum likelihood,...
Thus, it is the best possible choice (?)

A decision theoretic approach

Consider the risk function

$$\mathcal{R}(\hat{\mu}, \mu) = \mathsf{E} \left\|\hat{\mu} - \mu\right\|^2 , \quad (2)$$

where $\|\cdot\|$ means the Hilbert-Schmidt norm (i.e., sum of squares).

An estimator $\hat{\mu}_1$ is *dominated* by another estimator $\hat{\mu}_2$ if, for all possible μ,

$$\mathcal{R}(\hat{\mu}_1, \mu) \geq \mathcal{R}(\hat{\mu}_2, \mu).$$

Shrink it !

James and Stein (1961) have shown that, for $p > 2$, the ML-estimator is dominated by

$$\hat{\mu}_{JS} = \left(1 - \frac{(p-2)}{\|X\|^2}\right) X \tag{3}$$

Time and space warp

And now to something completely (?) different.

Time domain and spectral domain

A centered stationary time series $(X_t)_{(t \in \mathbb{Z})}$ can be described by its **autocovariance** function

$$\gamma(h) = \mathrm{E}\, X_t X'_{t+h}$$

or by its **spectrum**

$$f(\omega) = \sum_{h \in \mathbb{Z}} \gamma(h) \exp(-2\pi i \omega h), \quad \omega \in [0, 2\pi)$$

where $i = \sqrt{-1}$.
The above definitions are valid for p-variate time series ($p \geq 1$).

Periodogram

A nonparametric estimator of the spectrum is based on the **periodogram**:

$$I_T(\omega_j) = \frac{1}{2\pi T} \underbrace{\left(\sum_{t=1}^{T} X_t \exp(-i\omega t)\right)}_{\text{DFT}} \left(\sum_{t=1}^{T} X_t \exp(-i\omega t)\right)^*$$

Simulated AR 1

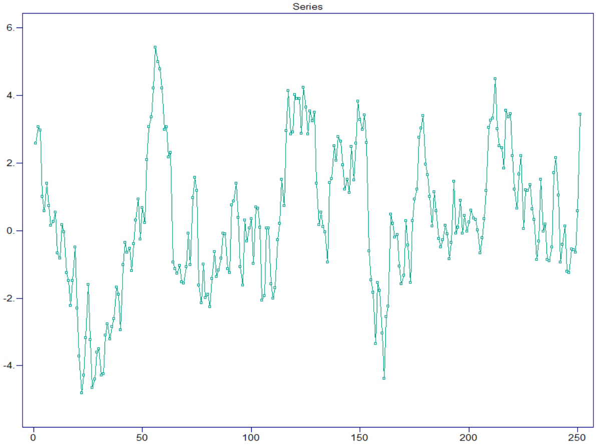
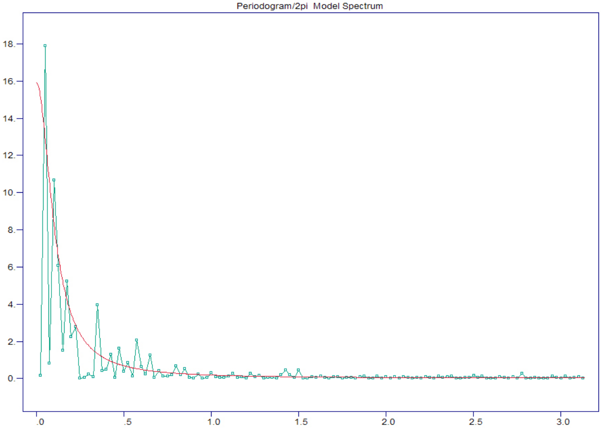


Figure: $X_t = .9X_{t-1} + e_t, \quad (e_t)_{t \in \mathbb{Z}} \sim IID(0,1)$

Periodogram of AR 1

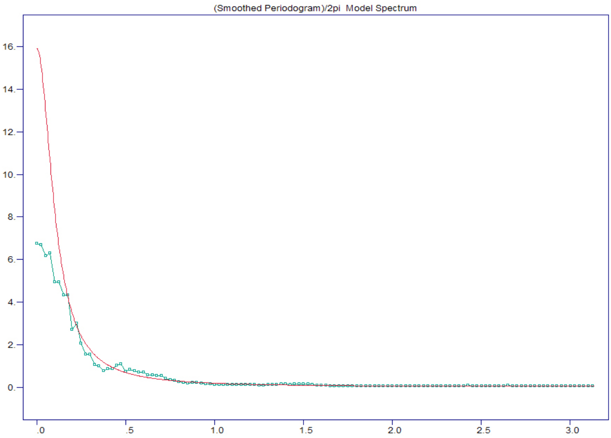


Smoothing the periodogram

The periodogram is not consistent. To construct a consistent estimator, one averages (or kernel smoothes) over neighboring Fourier frequencies

$$\hat{f}_T^0(\omega) = \frac{1}{m_T} \sum_{k=-(m_T-1)/2}^{(m_T-1)/2} I_T(\omega + \omega_k)$$

Averaged periodogram of AR 1



Properties of smoothed periodogram

The smoothed periodogram is consistent:

$$\lim_{T\to\infty} \mathrm{E} \left\| \hat{f}_T^0(\omega) - f(\omega) \right\|^2 = 0$$

and asymptotically unbiased:

$$\lim_{T\to\infty} \mathrm{E}\, \hat{f}_T^0(\omega) = f(\omega)$$

Summarizing: **It is a powerful nonparametric estimator of the spectrum.**

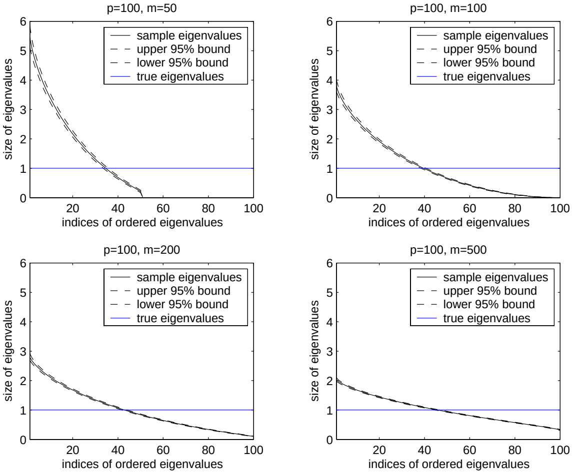
$$\hat{f}_T^0(\omega) = \frac{1}{m_T} \sum_{k=-(m_T-1)/2}^{(m_T-1)/2} I_T(\omega + \omega_k) =$$

Multivariate spectral estimation

- If (X_t) is multivariate, the periodogram is a singular matrix
- Smoothing span m_T grows less fast than T as local features have to be captured
- Problem: m_T is effective sample size of the estimator for the spectral matrix, so ratio between dimension p and sample size may be bad.
- Result: smoothed periodogram has a bad condition number and a high risk

$$\mathrm{E}\left\|\hat{f}_T^0(\omega) - f(\omega)\right\|^2$$

Graphical illustration



Possible solutions

- curse of dimensionality: sample eigensystem is the worst possible choice
- $\hat{f}_T^0(\omega)$ has thus high variance
- Solution: 'shrink' it
- shrinkage target?

An 'old' solution

In the absence of background information on the structure of the data: shrink to the identity matrix:

$$\hat{f}_T^*(\omega) = r_T(\omega)\hat{\mu}_T(\omega)\,\mathrm{Id} + (1 - r_T(\omega))\hat{f}_T^0(\omega)$$

This is done, under general asymptotics, in
Böhm, H. and von Sachs, R. (2007): *Shrinkage estimation in the frequency domain of multivariate time series*, DP0706.

New solution

- Popular in many fields of applications, like economics, psychology: **factor models**
- Crucial problem: how many factors to choose?
- Solution: Make it too restrictive and marry it to $\hat{f}^0_T(\omega)$ by convex linear combination

One factor model

'market' time series:

$$\dot{X}_{0t}, t = 1, \ldots, T$$

with spectrum $\dot{f}_0(\omega)$

$$\text{Model:} \quad \dot{X}_{it} = \beta_i \dot{X}_{0t} + \epsilon_{it} \quad i = 1, \ldots, p \tag{4}$$

Lemma

$$\dot{d}_i(\omega) = \beta_i \dot{d}_0(\omega) + \dot{d}_i^{\epsilon}(\omega) \tag{5}$$

where $\dot{d}_i^{\epsilon}(\omega)$ is the DFT of the idiosyncratic components. Furthermore,

$$\dot{d}_i^{\epsilon}(\omega) \sim \mathcal{N}^C\left(0, (\sigma_i^{\epsilon})^2\right) \tag{6}$$

Assumptions and methods

Weights and idiosyncratic variances in the above model can be estimated by simple linear regression. We come to a parametric estimator

$$\hat{f}_T^1(\omega) = bb'\hat{f}_0^0(\omega) + D \tag{7}$$

where

$$D = \text{diag}\left(\widehat{(\sigma_1^\epsilon)^2} \ldots \widehat{(\sigma_p^\epsilon)^2}\right)$$

The shrinkage target

- This is a very simple, old fashioned model
- It has high bias, but is very easy to estimate (O $(1/T)$!!)
- We even assume that the model be misspecified : For all frequencies $\omega \in [0, 2\pi]$,

$$f_T^1(\omega) \neq f_T^0(\omega)! \tag{8}$$

$\hat{f}_T^1(\omega) = bb'\hat{f}_0^0(\omega) + D =$ 

Optimal shrinkage intensity

We search for a linear combination

$$\hat{f}^{+}(\omega) = \zeta_T(\omega)\hat{f}_T^1(\omega) + (1 - \zeta_T(\omega))\hat{f}_T^0(\omega)$$

where $\zeta_T(\omega)$ is a data driven estimator of an optimal, oracle shrinkage intensity $\zeta_T^*(\omega)$ that is the solution of the minimization problem

$$\mathrm{E}\left\|\hat{f}^{+}(\omega) - f_T^0(\omega)\right\|^2 = \min! \qquad (9)$$

Three steps to a solution

We will proceed in three steps:

- Derive the optimal shrinkage intensity $\zeta_T^*(\omega)$, which depends on background knowledge
- Identify its asymptotic behaviour
- Derive a data driven estimator $\zeta_T(\omega)$ of the optimal shrinkage intensity

1: Optimal shrinkage intensity

Simple differential calculus yields the optimal shrinkage intensity:

$$\zeta_T^*(\omega) = \frac{\sum_{i,j=1}^{p} \left(\operatorname{Var} \hat{f}_{ij}^0(\omega) - 2\Re \operatorname{Cov}\left(\hat{f}_{ij}^1(\omega), \hat{f}_{ij}^0(\omega) \right) \right)}{\sum_{i,j=1}^{p} \left(\operatorname{Var}(\hat{f}_{ij}^1(\omega) - \hat{f}_{ij}^0(\omega)) + \left| f_{ij}^1(\omega) - f_{ij}^0(\omega) \right|^2 \right)} \quad (10)$$

2: its asymptotic behaviour

$$\zeta_T^*(\omega) = \frac{1}{m_T}\frac{\pi(\omega) - 2\Re(\rho(\omega))}{\gamma(\omega)} + O\left(\frac{1}{T}\right)$$

$$\pi(\omega) = \sum_{i,j=1}^{p} \mathrm{AsyVar}\left(\sqrt{m_T}\hat{f}_{ij}^{0}(\omega)\right) \tag{11}$$

$$\rho(\omega) = \sum_{i,j=1}^{p} \mathrm{AsyCov}\left(\sqrt{m_T}\hat{f}_{ij}^{1}(\omega), \sqrt{m_T}\hat{f}_{ij}^{0}(\omega)\right) \tag{12}$$

$$\gamma(\omega) = \sum_{i,j=1}^{p} \left| f_{ij}^{1}(\omega) - f_{ij}^{0}(\omega)\right|^2 \tag{13}$$

3: and a data driven estimator

Just plug them in:

$$p(\omega) = \sum_{i,j=1}^{p_T} \left(\frac{1}{m_T} \sum_{k=-(m_T-1)/2}^{(m_T-1)/2} |I_{ij}(\tilde{\omega}_k) - \hat{f}^0_{ij}(\omega)|^2 \right)$$

$$r(\omega) = \sum_{i,j=1}^{p_T} \left(b_i b_j \hat{f}^0_{0i}(\omega) \hat{f}^0_{j0}(\omega) \right)$$

$$g(\omega) = \sum_{i,j=1}^{p_T} \left| \hat{f}^1_{ij}(\omega) - \hat{f}^0_{ij}(\omega) \right|^2$$

here it comes...

We have arrived at a **consistent** estimator

$$\hat{f}^{+}(\omega) = \frac{1}{m_T} \frac{p(\omega) - 2\Re\left(r(\omega)\right)}{g(\omega)} \hat{f}_T^1(\omega) + \left(1 - \frac{1}{m_T} \frac{p(\omega) - 2\Re\left(r(\omega)\right)}{g(\omega)}\right) \hat{f}_T^0(\omega \tag{14}$$

to which we refer as to the DDMSE, which means...

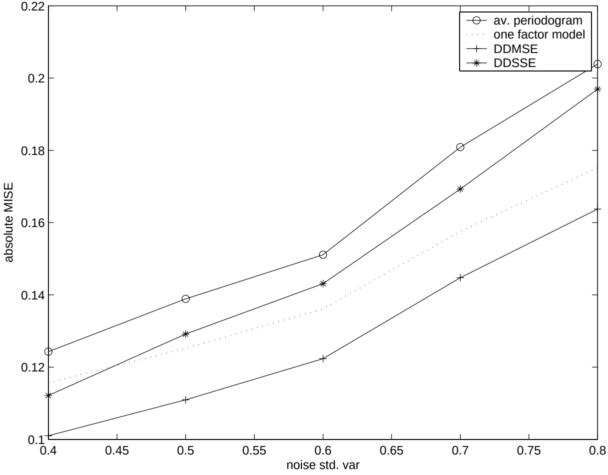
DRAGON DONKEY MIXTURE shrinkage estimator

$$\hat{f}^{+}(\omega) = \zeta_T(\omega)\hat{f}_T^{1}(\omega) + (1 - \zeta_T(\omega))\hat{f}_T^{0}(\omega) =$$

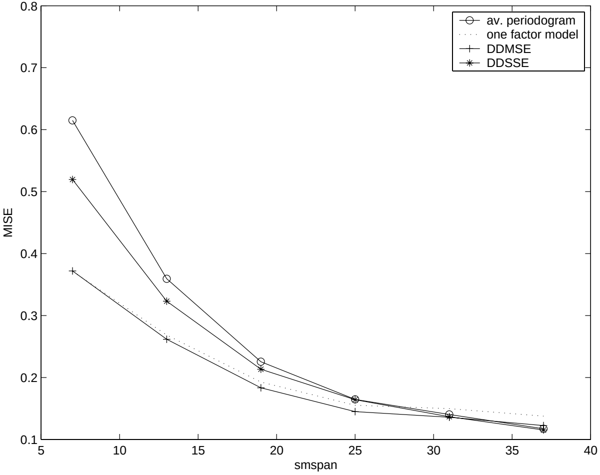
Data driven market shrinkage estimator

$$\hat{f}^{+}(\omega) = \zeta_T(\omega)\hat{f}_T^1(\omega) + (1 - \zeta_T(\omega))\hat{f}_T^0(\omega) =$$

MC for different true models



MC for different smoothing spans



Conclusions (1)

- Shrinkage is a powerful tool in multivariate spectral analysis
- In the absence of knowledge about the data, it is best to shrink to identity
- If we have some knowledge about the data, we can further improve by incorporating this knowledge in the shrinkage target

Conclusions (2)

There are many interpretations of ’shrinkage to market’ in multivariate spectral analysis

- A regularization method
- Finding the optimal tradeoff between squared bias and variance
- An innovative solution to the problem of model choice in spectral factor analysis
- A refinement of factor analysis: *stretch* to a nonparametric estimator

Literature

- Böhm, H. and von Sachs, R. (2007): *Shrinkage estimation in the frequency domain of multivariate time series*, DP0706.
- Böhm, H. (2008): *Shrinkage methods for multivariate spectral analysis*. PhD thesis (in preparation)

This is joint work with Rainer von Sachs, supported by IAP network and FSR.

The end

And they lived happily ever after...