Vast Volatility Matrix Estimation using High Frequency Data for Portfolio Selection

by J Fan, Y Li and K Yu

Discussion by Paolo Zaffaroni

Imperial College Business School
Imperial College London

11th September 2010

- This paper is concerned with consequences of practical implementation of *mean-variance* trading strategies, with emphasis on when dealing with many assets p
and high-frequency data is available.

- In particular, it focuses at global minimum variance portfolio (not a tangency portfolio except when means equal across assets):

$$\mathbf{w}^{gmv} = \mathrm{argmin}_{\mathbf{w}} \left(\mathbf{w}' \boldsymbol{\Sigma} \mathbf{w} \right), \text{ such that } \mathbf{w}'\mathbf{e} = \mathbf{1}, \qquad (1)$$

where $\boldsymbol{\Sigma}$ is covariance matrix of asset returns.

- gmv portfolio does not depend on mean parameterization and some evidence it works 'well' in practice (Jagannathan and Ma, 2003).

- Problem is that when replacing $\boldsymbol{\Sigma}$ by an estimate, say the sample covariance matrix,

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=0}^{n-1} (\mathbf{r}_{t-i} - \bar{\mathbf{r}})(\mathbf{r}_{t-i} - \bar{\mathbf{r}})'$$

then bad performance and properties (minimum variance!) not warranted.

- Culprits are estimation and model uncertainty (see nice survey of Brandt (2004)). Both serious, for different reasons, especially when p (number of assets) large.

- To deal with estimation uncertainty, look for a 'better' estimator. Popular choices:

  - Factor model-based estimators:

    in another paper Fan, Fan and Lv (2007) established factor model-based estimators better for $\mathbf{\Sigma}^{-1}$ (the $inverse$) when used into gvm portfolio. See also empirical work of Chan et al (1999).

  - Shrinkage-Stein estimators:

    Jagannathan and Ma (2003) show that imposing no-short selling constraint yields shrinkage covariance matrix.

  - Fan, Fan and Yu (2008), henceforth FFY, provide statistical insight on this.

THIS PAPER takes another angle and investigates the possible gain, if any, when implementing gmv portfolio (with gross constraints) using HIGH FREQUENCY DATA.

THIS PAPER builds upon FFY which can be summarized as follows.

Consider the so-called oracle approach (for $c = 1$ short-selling ruled out)

$$\mathbf{w}^{gmv*} = \operatorname{argmin}_{\mathbf{w}} (\mathbf{w}'\mathbf{\Sigma}\mathbf{w}) = \operatorname{argmin}_{\mathbf{w}}\mathbf{R}(\mathbf{w}) \quad (2)$$
$$\text{such that } \mathbf{w}'\mathbf{e} = \mathbf{1}, \ \| \mathbf{w}^{\mathbf{gmv}*} \|_{\mathbf{1}} \leq \mathbf{c}, \quad \text{some } \mathbf{c} > \mathbf{0}.$$

In practice we use the estimated portfolio weights (we are not the oracle!)

$$\hat{\mathbf{w}}^{gmv*} = \operatorname{argmin}_{\mathbf{w}} \left(\mathbf{w}'\widehat{\mathbf{\Sigma}}\mathbf{w}\right) = \operatorname{argmin}_{\mathbf{w}}\mathbf{R}_{\mathrm{n}}(\mathbf{w})$$
$$\text{such that } \mathbf{w}'\mathbf{e} = \mathbf{1}, \ \| \mathbf{w}^{\mathbf{gmv}*} \|_{\mathbf{1}} \leq \mathbf{c}, \quad \text{some } \mathbf{c} > \mathbf{0}.$$

- Main result is

$$\mid R(\mathbf{w}^{gmv*}) - R_n(\hat{\mathbf{w}}^{gmv*}) \mid \leq 3c^2 a_n,$$

where

$$a_n = \| \mathbf{\Sigma} - \widehat{\mathbf{\Sigma}} \|_\infty = O(\frac{\sqrt{p}}{n}).$$

when using sample covariance matrix.

- Key result: the decision we make based on $R_n(\hat{w})$ is basically the same as if we knew $R(w)$!

- This result complements Jagannathan and Ma (2003) argument: adding gross exposure constraint improves estimation of risk *as if* an estimate better than $\hat{\mathbf{\Sigma}}$ was used: a large a_n can be compensated by a small c.

- THIS PAPER (theoretical part)

  Model for p-dimensional vector of (log) prices:

$$dX_t = \mu_t dt + S_t^{\frac{1}{2}} dW_t$$

  where one observes

$$X_{t_i}^o = X_{t_i} + \epsilon_i^x.$$

  To evaluate

$$R_{t,\tau}(w) = w' \Sigma_{t,\tau} w$$

  one needs

$$\Sigma_{t,\tau} = \int_t^{t+\tau} E_t S_u du,$$

  which one approximates as

$$\frac{1}{\tau} \Sigma_{t,\tau} \approx \frac{1}{h} \int_{t-h}^{t} S_u du.$$

  Assuming we know path of S_t, approximation works if τ, h are both small or large. But not so when τ small and h large due to stochastic volatility. This is where high-frequency kicks in since we can then afford to make h small yet to have lots of observations.

  Nonparametric estimation of (integrated) volatility now a large field. This paper focuses on the two-scale estimator TSCV $\widehat{\langle X, Y \rangle}_1$ (multivariate case of TSRV of Zhang et al. (2005)) and establish concentration inequality

$$P(\tilde{n}^{\frac{1}{6}} \mid \widehat{\langle X, Y \rangle}_1 - \int_0^1 \sigma_t^{(X)} \sigma_t^{(Y)} \rho_t^{(X,Y)} dt \mid > x) \leq 8 exp(-Cx^2), \;\; x = O(\tilde{n}).$$

  When TSCV used to estimate the full covariance matrix, depending on whether data are synchronized two-by-two ($pairwise$)

or all together (all), yield

$$a_n^{pairwise} = \| \mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}^{pairwise} \|_\infty = O(\frac{\sqrt{p}}{\tilde{n}_{pair}^{\frac{1}{6}}}),$$

and

$$a_n^{all} = \| \mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}^{all} \|_\infty = O(\frac{\sqrt{p}}{\tilde{n}_{all}^{\frac{1}{6}}}),$$

where typically

$$\tilde{n}_{all} < \tilde{n}_{pair} \text{ (not too distant though)}$$

Paper contains rather involved simulation (harder to follow than theoretical part!) and empirical sections which corroborate use of high-frequency data for portfolio optimization (with better estimates, less need of gross constraints) especially the $pairwise$ approach.

My thoughts:

– The estimation error bound for TSCV is less stringent (not surprisingly) than for the sample covariance estimator. Given that in realistic applications $\tilde{n} \approx 400$ then

$$a_n^{all} \text{ and } a_n^{pairwise} \text{ are } O(\frac{\sqrt{p}}{400^{\frac{1}{6}}}) = O(\frac{\sqrt{p}}{2.7})$$

Instead, when considering low-frequency approaches (with parametric rate) the bound is

$$O(\frac{\sqrt{p}}{n^{\frac{1}{2}}}),$$

and although is general $n < \tilde{n}$ then the equation

$$n^{\frac{1}{2}} = 400^{\frac{1}{6}}$$

yields

$$n \approx 7.3.$$

This means that the bound is very loose in the nonparametric case, comparable to the case when one fit a parametric model with 8 observations.

This also suggests that the value of the gross constraint c is less important. Indeed, the risk profiles for the nonparametric approaches are nearly flat in c.

– It would be nice if the empirical applications consider, as low-frequency alternative, better choices that the sample covariance matrix. Fan et al (2008) showed that factor models do a much better job when constructing portfolios. There is also interest in parametric alternatives, say DCC of Engle or tDCC of Pesaran, which are known to perform usually well.

- This paper is concerned with vast covariance matrices. However, both the simulations and the application are made with $p = 50$ and $p = 30$ assets respectively. Why not considering really large matrices?
- The paper keeps pointing out the significant differences between the low-frequency and the high-frequency. But graphically they do not appear so markedly different unless we are given some statistical guidance on this (confidence band of risk)?
- Although the paper is concerned with gmv portfolios, it would be helpful to look at the Sharpe of these strategies. Some evidence that gmv does not too bad, especially out of sample.
- One can be more ambitious and look at other strategies. This would involve modelling the conditional mean which, as we know, is all but easy. But one could explore the rich data set and consider non- and semi-parametric estimator of the conditional mean.

IN CONCLUSION

- Very interesting paper, looking forward to see the completed version.
- As usual (not the first time I discuss his papers) with Jianqing's work, I always learn a great deal: I strongly recommend it!