Large Dynamic Covariance Matrices

Robert F. Engle[1], Olivier Ledoit[2], and Michael Wolf[2]

[1]Department of Finance
New York University

[2]Department of Economics
University of Zurich

The Importance of Good Forecasts

The Importance of Good Forecasts

Good forecasts of time-varying objects can make the difference between life and death.

The Importance of Good Forecasts

Good forecasts of time-varying objects can make the difference between life and death.

Here is a weather-related example from the movie Sharknado 2:

The Importance of Good Forecasts

Good forecasts of time-varying objects can make the difference between life and death.

Here is a weather-related example from the movie Sharknado 2:

The remainder of the talk will address finance-related examples.

Outline

Outline

Problem & Aim of the Paper

Problem:

- Multivariate GARCH models are popular tools for risk management and portfolio selection
- However, the number of assets in the investment universe generally poses a challenge to such models
- In other words, many multivariate GARCH models suffer from the curse of dimensionality

Problem & Aim of the Paper

Problem:

- Multivariate GARCH models are popular tools for risk management and portfolio selection
- However, the number of assets in the investment universe generally poses a challenge to such models
- In other words, many multivariate GARCH models suffer from the curse of dimensionality

Aim of the paper:

- Robustify the DCC model of Engle (2002, JBES) against large dimensions
- Comparison to all kinds of other multivariate GARCH models is left to future research

Outline

Notation

Subscripts:

- $i = 1, \ldots, N$ indexes assets
- $t = 1, \ldots, T$ indexes time

Ingredients:

- $r_{i,t}$: observed return, stacked into $\mathbf{r}_t := (r_{1,t}, \ldots, r_{N,t})'$
- $d_{i,t}^2 := \mathsf{Var}(r_{i,t}|\mathcal{F}_{t-1})$: conditional variance
- D_t: diagonal matrix with generic entry $d_{i,t}$
- $H_t := \mathsf{Cov}(\mathbf{r}_t|\mathcal{F}_{t-1})$: conditional covariance matrix; $\mathsf{Diag}(H_t) = D_t^2$
- $s_{i,t} := r_{i,t}/d_{i,t}$: devolatilized return, stacked into $\mathbf{s}_t := (s_{1,t}, \ldots, s_{N,t})'$
- $R_t := \mathsf{Corr}(\mathbf{r}_t|\mathcal{F}_{t-1}) = \mathsf{Cov}(\mathbf{s}_t|\mathcal{F}_{t-1})$: conditional correlation matrix
- $\sigma_i^2 := \mathbb{E}(d_{i,t}^2) = \mathsf{Var}(r_{i,t})$: unconditional variance
- $C := \mathsf{Corr}(\mathbf{r}_t) = \mathsf{Cov}(\mathbf{s}_t)$: unconditional correlation matrix

Model Definition

Univariate volatilities governed by a GARCH(1,1) process:

$$d_{i,t}^2 = \omega_i + a_i r_{i,t-1}^2 + b_i d_{i,t-1}^2$$

Model Definition

Univariate volatilities governed by a GARCH(1,1) process:

$$d_{i,t}^2 = \omega_i + a_i r_{i,t-1}^2 + b_i d_{i,t-1}^2$$

DCC model of Engle (2002, JBES) with correlation targeting:

$$Q_t = (1 - \alpha - \beta)\, C + \alpha\, \mathbf{s}_{t-1}\mathbf{s}_{t-1}' + \beta\, Q_{t-1} \tag{1}$$

where Q_t is a pseudo conditional correlation matrix.

Model Definition

Univariate volatilities governed by a GARCH(1,1) process:

$$d_{i,t}^2 = \omega_i + a_i r_{i,t-1}^2 + b_i d_{i,t-1}^2$$

DCC model of Engle (2002, JBES) with correlation targeting:

$$Q_t = (1 - \alpha - \beta)\, C + \alpha\, \mathbf{s}_{t-1}\mathbf{s}_{t-1}' + \beta\, Q_{t-1} \tag{1}$$

where Q_t is a pseudo conditional correlation matrix.

Conditional correlation and covariance matrices then:

$$R_t = \mathsf{Diag}(Q_t)^{-1/2}\, Q_t\, \mathsf{Diag}(Q_t)^{-1/2}$$
$$H_t = D_t R_t D_t$$

Model Definition

Univariate volatilities governed by a GARCH(1,1) process:

$$d_{i,t}^2 = \omega_i + a_i r_{i,t-1}^2 + b_i d_{i,t-1}^2$$

DCC model of Engle (2002, JBES) with correlation targeting:

$$Q_t = (1 - \alpha - \beta)\, C + \alpha\, \mathbf{s}_{t-1}\mathbf{s}_{t-1}' + \beta\, Q_{t-1} \tag{1}$$

where Q_t is a pseudo conditional correlation matrix.

Conditional correlation and covariance matrices then:

$$\begin{aligned} R_t &= \mathsf{Diag}(Q_t)^{-1/2}\, Q_t\, \mathsf{Diag}(Q_t)^{-1/2} \\ H_t &= D_t R_t D_t \end{aligned}$$

Data-generating process:

$$\mathbf{r}_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, H_t)$$

Outline

Making Estimation Feasible

Estimating the model with a large number of assets is challenging.

Major difficulty:

- Inverting the conditional covariance matrix H_t for the likelihood

Making Estimation Feasible

Estimating the model with a large number of assets is challenging.

Major difficulty:

- Inverting the conditional covariance matrix H_t for the likelihood

Solution by Pakel et al. (2017, WP):

- Instead of using the full conditional covariance matrix, use a selection of two-by-two blocks
- The composite likelihood is obtained by combining the likelihoods of (contiguous) pairs of assets

Making Estimation Feasible

Estimating the model with a large number of assets is challenging.

Major difficulty:

- Inverting the conditional covariance matrix H_t for the likelihood

Solution by Pakel et al. (2017, WP):

- Instead of using the full conditional covariance matrix, use a selection of two-by-two blocks
- The composite likelihood is obtained by combining the likelihoods of (contiguous) pairs of assets

Three-stage estimation scheme:

1. Fit a GARCH(1,1) model to each asset
2. Estimate the unconditional correlation matrix C of the devolatilized returns for correlation targeting
3. Maximize the composite likelihood to estimate (α, β)

Outline

Nonlinear Shrinkage to Counter Large Dimensions

Main contribution:

- Improved estimation of the unconditional correlation matrix C, which serves as the correlation target in equation (1)

Nonlinear Shrinkage to Counter Large Dimensions

Main contribution:

- Improved estimation of the unconditional correlation matrix C, which serves as the correlation target in equation (1)

Naïve approach:

- Use the sample correlation matrix of the devolatilized returns $\widehat{\mathbf{s}_t}$
- Corresponds to the original proposal of Engle (2002, JBES)
- This approach does not work well in large dimensions, and cannot even be used when $N > T$

Nonlinear Shrinkage to Counter Large Dimensions

Main contribution:

- Improved estimation of the unconditional correlation matrix C, which serves as the correlation target in equation (1)

Naïve approach:

- Use the sample correlation matrix of the devolatilized returns $\widehat{\mathbf{s}}_t$
- Corresponds to the original proposal of Engle (2002, JBES)
- This approach does not work well in large dimensions, and cannot even be used when $N > T$

Superior approach:

- Apply nonlinear shrinkage to the devolatilized returns $\widehat{\mathbf{s}}_t$
- This approach works well in large dimensions, even when $N > T$

Nonlinear Shrinkage: Starting Point

Generic setting:

- I.i.d. data $\mathbf{y}_t \in \mathbb{R}^N$ with covariance matrix Σ
- Stacked into $T \times N$ matrix Y

Nonlinear Shrinkage: Starting Point

Generic setting:

- I.i.d. data $\mathbf{y}_t \in \mathbb{R}^N$ with covariance matrix Σ
- Stacked into $T \times N$ matrix Y

The sample covariance matrix S admits a spectral decomposition

$$S = U \Lambda U'$$

Here:

- U is an orthogonal matrix whose columns are the sample eigenvectors $(u_1, \ldots, u_N)$
- Λ is a diagonal matrix whose diagonal entries are the sample eigenvalues $(\lambda_1, \ldots, \lambda_N)$

Nonlinear Shrinkage: Class of Estimators

Rotation Equivariance

- Observed $T \times N$ data matrix: Y
- W is an N-dimensional orthogonal / rotation matrix
- $\widehat{\Sigma} := \widehat{\Sigma}(Y)$ is a generic estimator of Σ
- It is rotation-equivariant if $\widehat{\Sigma}(YW) = W'\widehat{\Sigma}(Y)W$

Nonlinear Shrinkage: Class of Estimators

Rotation Equivariance

- Observed $T \times N$ data matrix: Y
- W is an N-dimensional orthogonal / rotation matrix
- $\widehat{\Sigma} := \widehat{\Sigma}(Y)$ is a generic estimator of Σ
- It is rotation-equivariant if $\widehat{\Sigma}(YW) = W'\widehat{\Sigma}(Y)W$

Without specific knowledge about Σ, rotation equivariance is a desirable property of an estimator.

Nonlinear Shrinkage: Class of Estimators

Rotation Equivariance

- Observed $T \times N$ data matrix: Y
- W is an N-dimensional orthogonal / rotation matrix
- $\widehat{\Sigma} := \widehat{\Sigma}(Y)$ is a generic estimator of Σ
- It is rotation-equivariant if $\widehat{\Sigma}(YW) = W'\widehat{\Sigma}(Y)W$

Without specific knowledge about Σ, rotation equivariance is a desirable property of an estimator.

We use the following class of rotation-equivariant estimators going back to Stein (1975, 1986):

$$\widehat{\Sigma} := UDU' \quad \text{where} \quad D := \mathsf{Diag}(d_1, \ldots, d_N) \text{ is diagonal}$$

Nonlinear Shrinkage in Action

Generic estimator in the class $\widehat{\Sigma} := UDU'$.

Keep the sample eigenvectors.

Shrink the sample eigenvalues:

- $D := \mathsf{Diag}(d(\lambda_1), \ldots, d(\lambda_N))$
- Based on nonlinear shrinkage function $d : \mathbb{R} \to \mathbb{R}$

Nonlinear Shrinkage in Action

Generic estimator in the class $\widehat{\Sigma} := UDU'$.

Keep the sample eigenvectors.

Shrink the sample eigenvalues:

- $D := \mathsf{Diag}(d(\lambda_1), \ldots, d(\lambda_N))$
- Based on nonlinear shrinkage function $d : \mathbb{R} \to \mathbb{R}$

Approach of Ledoit and Wolf (2012, AOS; 2015, JMVA):

- Use large-dimensional asymptotics where $N/T \to c > 0$
- Consistently estimate optimal limiting shrinkage function d^*
- Feasible estimator: $\widetilde{\Sigma} := U \times \mathsf{Diag}(\widetilde{d}(\lambda_1), \ldots, \widetilde{d}(\lambda_N)) \times U'$

Nonlinear Shrinkage in Action

Generic estimator in the class $\widehat{\Sigma} := UDU'$.

Keep the sample eigenvectors.

Shrink the sample eigenvalues:

- $D := \mathsf{Diag}(d(\lambda_1), \ldots, d(\lambda_N))$
- Based on nonlinear shrinkage function $d : \mathbb{R} \to \mathbb{R}$

Approach of Ledoit and Wolf (2012, AOS; 2015, JMVA):

- Use large-dimensional asymptotics where $N/T \to c > 0$
- Consistently estimate optimal limiting shrinkage function d^*
- Feasible estimator: $\widetilde{\Sigma} := U \times \mathsf{Diag}(\widetilde{d}(\lambda_1), \ldots, \widetilde{d}(\lambda_N)) \times U'$

Further Details on Nonlinear Shrinkage

Proposed Estimation of the DCC Model

Estimation of the correlation target C:

- Apply nonlinear shrinkage to the devolatilized returns $\widehat{\mathbf{s}}_t$
- The resulting estimator is not a proper correlation matrix
- Post-processing the estimator takes care of this problem, that is, convert covariance matrix into a correlation matrix

Proposed Estimation of the DCC Model

Estimation of the correlation target C:

- Apply nonlinear shrinkage to the devolatilized returns $\widehat{\mathbf{s}}_t$
- The resulting estimator is not a proper correlation matrix
- Post-processing the estimator takes care of this problem, that is, convert covariance matrix into a correlation matrix

Three-stage estimation scheme:

1. Fit a GARCH(1,1) model to each asset
2. Use nonlinear shrinkage to estimate C
3. Maximize the composite likelihood to estimate (α, β)

Simpler alternative:

- Use linear shrinkage of Ledoit and Wolf (2004, JMVA) in step 2.

Linear Shrinkage

Easiest way to think about it:

- Convex linear combination of the sample covariance matrix and (a multiple of) the identity matrix:

$$\widehat{\Sigma} = \delta(\bar{s}^2 I) + (1 - \delta) S$$

- $\bar{s}^2$ is the average of the N sample variances s_i^2
- $\delta \in [0, 1]$ is the shrinkage intensity

Linear Shrinkage

Easiest way to think about it:

- Convex linear combination of the sample covariance matrix and (a multiple of) the identity matrix:

$$\widehat{\Sigma} = \delta(\bar{s}^2 I) + (1 - \delta)S$$

- $\bar{s}^2$ is the average of the N sample variances s_i^2
- $\delta \in [0, 1]$ is the shrinkage intensity

Alternative way to think about it:

- This estimator is also of the form UDU', but d is restricted to be a certain linear function:

$$d(\lambda_i) := \delta\overline{\lambda} + (1 - \delta)\lambda_i$$

- $\overline{\lambda}$ is the average of the N sample eigenvalues λ_i

Outline

Big Picture

Goal:

- Examine out-of-sample properties of Markowitz portfolios via backtest exercises

Big Picture

Goal:

- Examine out-of-sample properties of Markowitz portfolios via backtest exercises

Two applications:

- Global minimum variance (GMV) portfolio
- Full Markowitz portfolio with a signal

Big Picture

Goal:

- Examine out-of-sample properties of Markowitz portfolios via backtest exercises

Two applications:

- Global minimum variance (GMV) portfolio
- Full Markowitz portfolio with a signal

(Out-of-sample) Performance measures:

- Standard deviation
- Information ratio

Data & Portfolio Rules

Data:

- Download daily return data from CRSP
- Period: 01/01/1980–12/31/2015

Data & Portfolio Rules

Data:

- Download daily return data from CRSP
- Period: 01/01/1980–12/31/2015

Updating:

- 21 consecutive trading days constitute one 'month'
- Update portfolios on 'monthly' basis

Data & Portfolio Rules

Data:

- Download daily return data from CRSP
- Period: 01/01/1980–12/31/2015

Updating:

- 21 consecutive trading days constitute one 'month'
- Update portfolios on 'monthly' basis

Out-of-sample period:

- Start investing on 01/08/1986
- This results in 7560 daily returns (over 360 'months')

Data & Portfolio Rules

Portfolio sizes:

- We consider $N \in \{100, 500, 1000\}$

Data & Portfolio Rules

Portfolio sizes:

- We consider $N \in \{100, 500, 1000\}$

Portfolio constituents:

- Select new constituents at beginning of each 'month'
- Find the N largest stocks that have
  (i) a complete 1250-day return history
  (ii) a complete 21-day return future

Data & Portfolio Rules

Portfolio sizes:

- We consider $N \in \{100, 500, 1000\}$

Portfolio constituents:

- Select new constituents at beginning of each 'month'
- Find the N largest stocks that have
  (i) a complete 1250-day return history
  (ii) a complete 21-day return future

Covariance matrix estimation:

- Use previous $T = 1250$ days to estimate the covariance matrix

Global Minimum Variance Portfolio

Problem Formulation

$$\min_{w} w' H_t w$$

$$\text{subject to} \quad w'\mathbf{1} = 1$$

(where **1** is a conformable vector of ones)

Global Minimum Variance Portfolio

Problem Formulation

$$\min_w w' H_t w$$
$$\text{subject to} \quad w'\mathbf{1} = 1$$

(where **1** is a conformable vector of ones)

Analytical Solution

$$w^* = \frac{H_t^{-1}\mathbf{1}}{\mathbf{1}' H_t^{-1}\mathbf{1}}$$

Global Minimum Variance Portfolio

Problem Formulation

$$\min_{w} w' H_t w$$

$$\text{subject to} \quad w'\mathbf{1} = 1$$

(where $\mathbf{1}$ is a conformable vector of ones)

Analytical Solution

$$w^* = \frac{H_t^{-1}\mathbf{1}}{\mathbf{1}' H_t^{-1}\mathbf{1}}$$

Feasible Solution

$$\widehat{w} := \frac{\widehat{H}_t^{-1}\mathbf{1}}{\mathbf{1}' \widehat{H}_t^{-1}\mathbf{1}}$$

Global Mininum Variance Portfolio

Competing portfolios:

- **1/*N*:** as a simple benchmark
- **DCC-S:** based on the sample correlation matrix
- **DCC-L:** based on linear shrinkage
- **DCC-NL:** based on nonlinear shrinkage
- **RM-2006:** RiskMetrics 2006

Global Mininum Variance Portfolio

Competing portfolios:

- **1/N:** as a simple benchmark
- **DCC-S:** based on the sample correlation matrix
- **DCC-L:** based on linear shrinkage
- **DCC-NL:** based on nonlinear shrinkage
- **RM-2006:** RiskMetrics 2006

Performance measures:

- Standard deviation (primary)
- Information ratio (secondary)

Global Mininum Variance Portfolio

Competing portfolios:

- **1/*N*:** as a simple benchmark
- **DCC-S:** based on the sample correlation matrix
- **DCC-L:** based on linear shrinkage
- **DCC-NL:** based on nonlinear shrinkage
- **RM-2006:** RiskMetrics 2006

Performance measures:

- Standard deviation (primary)
- Information ratio (secondary)

Assessing statistical significance:

- Test for significant difference between DCC-S and DCC-NL uses Ledoit and Wolf (2011, WM)

Global Mininum Variance Portfolio

Annualized standard deviations:

N	$1/N$	DCC-S	DCC-L	DCC-NL	RM-2006
100	21.56	13.36	13.33	13.17***	14.69
500	19.53	10.57	10.40	9.64***	12.60
1000	19.04	10.59	9.14	8.02***	14.86

Remarks:

- In each row, the best number appears in blue
- Stars indicate significant outperformance (DCC-NL vs. DCC-S)

Global Mininum Variance Portfolio

Annualized information ratios:

N	$1/N$	DCC-S	DCC-L	DCC-NL	RM-2006
100	0.56	0.74	0.74	0.76	0.57
500	0.69	1.32	1.33	1.39	0.89
1000	0.75	1.11	1.33	1.52	0.77

Remarks:

- In each row, the best number appears in blue

Markowitz Portfolio with Signal

Problem Formulation

$$\min_{w} w' H_t w$$

$$\text{subject to} \quad w' m_t = b \quad \text{and}$$

$$w' \mathbf{1} = 1$$

(where m_t is a signal and b is a target expected return)

Markowitz Portfolio with Signal

Problem Formulation

$$\min_w w' H_t w$$
$$\text{subject to} \quad w' m_t = b \quad \text{and}$$
$$w' \mathbf{1} = 1$$

(where m_t is a signal and b is a target expected return)

Analytical Solution

$$w^* = c_1 H_t^{-1} \mathbf{1} + c_2 H_t^{-1} m$$
$$\text{where} \quad c_1 := \frac{C - bB}{AC - B^2} \quad \text{and} \quad c_2 := \frac{bA - B}{AC - B^2}$$
$$\text{with} \quad A := \mathbf{1}' H_t^{-1} \mathbf{1} \quad B := \mathbf{1}' H_t^{-1} b \quad \text{and} \quad C := m' H_t^{-1} m$$

Markowitz Portfolio with Signal

Problem Formulation

$$\min_w w' H_t w$$
$$\text{subject to} \quad w' m_t = b \quad \text{and}$$
$$w' \mathbf{1} = 1$$

(where m_t is a signal and b is a target expected return)

Analytical Solution

$$w^* = c_1 H_t^{-1} \mathbf{1} + c_2 H_t^{-1} m$$
$$\text{where} \quad c_1 := \frac{C - bB}{AC - B^2} \quad \text{and} \quad c_2 := \frac{bA - B}{AC - B^2}$$
$$\text{with} \quad A := \mathbf{1}' H_t^{-1} \mathbf{1} \quad B := \mathbf{1}' H_t^{-1} b \quad \text{and} \quad C := m' H_t^{-1} m$$

Feasible Solution $\widehat{w}$ replaces H_t with an estimator $\widehat{H}_t$.

Markowitz Portfolio with Momentum Signal

For simplicity and reproducibility, we use momentum as the signal.

Markowitz Portfolio with Momentum Signal

For simplicity and reproducibility, we use momentum as the signal.

Competing portfolios:

- **EW-TQ:** equal-weighted portfolio of top-quintiles stocks
  $\Longrightarrow$ yields target expected return b for other portfolios
- **DCC-S:** based on the sample correlation matrix
- **DCC-L:** based on linear shrinkage
- **DCC-NL:** based on nonlinear shrinkage
- **RM-2006:** RiskMetrics 2006

Markowitz Portfolio with Momentum Signal

For simplicity and reproducibility, we use momentum as the signal.

Competing portfolios:

- **EW-TQ:** equal-weighted portfolio of top-quintiles stocks
  $\Longrightarrow$ yields target expected return b for other portfolios
- **DCC-S:** based on the sample correlation matrix
- **DCC-L:** based on linear shrinkage
- **DCC-NL:** based on nonlinear shrinkage
- **RM-2006:** RiskMetrics 2006

Performance measure:

- Standard deviation (secondary)
- Information ratio (primary)

Markowitz Portfolio with Momentum Signal

For simplicity and reproducibility, we use momentum as the signal.

Competing portfolios:

- **EW-TQ:** equal-weighted portfolio of top-quintiles stocks $\Longrightarrow$ yields target expected return b for other portfolios
- **DCC-S:** based on the sample correlation matrix
- **DCC-L:** based on linear shrinkage
- **DCC-NL:** based on nonlinear shrinkage
- **RM-2006:** RiskMetrics 2006

Performance measure:

- Standard deviation (secondary)
- Information ratio (primary)

Assessing statistical significance:

- Test for significant difference between DCC-S and DCC-NL uses Ledoit and Wolf (2008, JEF)

Markowitz Portfolio with Momentum Signal

Annualized standard deviations:

N	EW-TQ	DCC-S	DCC-L	DCC-NL	RM-2006
100	28.43	17.05	17.03	16.90***	18.87
500	24.42	12.36	12.16	11.31***	16.14
1000	22.89	13.07	10.76	9.20***	29.29

Remarks:

- In each row, the best number appears in blue
- Stars indicate significant outperformance (DCC-NL vs. DCC-S)

Markowitz Portfolio with Momentum Signal

Annualized information ratios:

N	EW-TQ	DCC-S	DCC-L	DCC-NL	RM-2006
100	0.60	0.93	0.93	0.93	0.85
500	0.70	1.34	1.37	1.48***	1.02
1000	0.76	0.98	1.30	1.62***	0.53

Remarks:

- In each row, the best number appears in blue
- Stars indicate significant outperformance (DCC-NL vs. DCC-S)

Outline

Conclusion

Multivariate GARCH models are popular tools for risk management and portfolio selection, but are often challenged in large dimensions.

Conclusion

Multivariate GARCH models are popular tools for risk management and portfolio selection, but are often challenged in large dimensions.

Two keys for making DCC model robust against large dimensions:

1. Composite likelihood makes estimation feasible
2. Nonlinear shrinkage estimation of the correlation targeting matrix ensures good performance

Conclusion

Multivariate GARCH models are popular tools for risk management and portfolio selection, but are often challenged in large dimensions.

Two keys for making DCC model robust against large dimensions:

1. Composite likelihood makes estimation feasible
2. Nonlinear shrinkage estimation of the correlation targeting matrix ensures good performance

Resulting DCC-NL model:

- Outperforms the basic DCC-S model by a wide margin
- Should become the new DCC standard

Conclusion

Multivariate GARCH models are popular tools for risk management and portfolio selection, but are often challenged in large dimensions.

Two keys for making DCC model robust against large dimensions:

1. Composite likelihood makes estimation feasible
2. Nonlinear shrinkage estimation of the correlation targeting matrix ensures good performance

Resulting DCC-NL model:

- Outperforms the basic DCC-S model by a wide margin
- Should become the new DCC standard

Remark:

- Nonlinear shrinkage can also help in robustifying other multivariate GARCH models against large dimensions
- A short description for the scalar BEKK model is in the paper

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15:850–859.

Ledoit, O. and Wolf, M. (2011). Robust performance hypothesis testing with the variance. *Wilmott Magazine*, September:86–89.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, 40(2):1024–1060.

Ledoit, O. and Wolf, M. (2015). Spectrum estimation: a unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis*, 139(2):360–384.

Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Review of Financial Studies*. doi: 10.1093/rfs/hhx052.

Pakel, C., Shephard, N., Sheppard, K., and Engle, R. F. (2017). Fitting vast dimensional time-varying covariance models. Working Paper FIN-08-009, NYU. Available at `https://ssrn.com/abstract=1354497`.

Stein, C. (1975). Estimation of a covariance matrix. Rietz lecture, 39th Annual Meeting IMS. Atlanta, Georgia.

Stein, C. (1986). Lectures on the theory of estimation of many parameters. *Journal of Mathematical Sciences*, 34(1):1373–1403.

Asymptotic Framework

Let $N := N(T)$ and assume $N/T \to c > 0$, as $T \to \infty$.

The following set of assumptions is maintained throughout.

- A1 The population covariance matrix Σ_T is a nonrandom N-dimensional positive definite matrix.
- A2 Let X_T be an $T \times N$ matrix of real i.i.d. random variables with zero mean, unit variance, and finite twelfth moment. One observes $Y_T := X_T \Sigma_T^{1/2}$.
- A3 Let $((\tau_{T,1}, \ldots, \tau_{T,N}); (v_{T,1}, \ldots, v_{T,N}))$ denote the eigenvalues and eigenvectors of Σ_T. The e.d.f. of the population eigenvalues, denoted by H_T, converges weakly to some limiting e.d.f. H.
- A4 $\mathsf{Supp}(H)$, the support of H, is the union of a finite number of closed intervals, bounded away from zero and infinity. Furthermore, there exists a compact interval in $(0, +\infty)$ that contains $\mathsf{Supp}(H_T)$ for all T large enough.

Ukranian Foundation

The Stieltjes transform of a nondecreasing function G is:

$$\forall z \in \mathbb{C}^+ \qquad m_G(z) := \int_{-\infty}^{+\infty} \frac{1}{\lambda - z} dG(\lambda)$$

(It has an explicit inversion formula too.)

Denote the e.d.f. of the sample eigenvalues by F_T.

Marčenko and Pastur (1967) showed that F_T converges a.s. to some nonrandom limit F at all points of continuity of F.

They also discovered how m_F relates to H and c:

$$\forall z \in \mathbb{C}^+ \qquad m_F(z) = \int_{-\infty}^{+\infty} \frac{1}{\tau\left[1 - c - c\,z\,m_F(z)\right] - z} dH(\tau) \tag{2}$$

This is the celebrated Marčenko-Pastur (MP) equation.

Transatlantic Additions

Moral: knowing H and c, one can 'solve' for F.

The particular expression (2) of the MP equation is due to Silverstein (1995).

Silverstein and Choi (1995) showed that

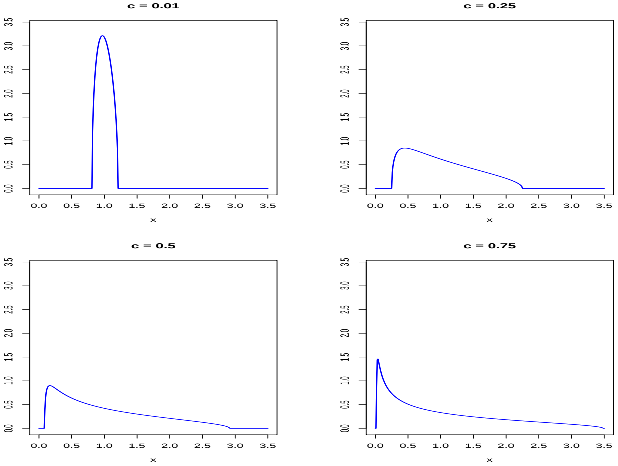
$$\forall \lambda \in \mathbb{R} \qquad \lim_{z \in \mathbb{C}^+ \to \lambda} m_F(z) =: \breve{m}_F(\lambda) \text{ exists}$$

The quantity $\breve{m}_F(\lambda)$ will be of crucial importance.

Illustration

H is a point mass at one (as for identity covariance matrix).

Plot density of F for various values of c:



Optimization Problem

(Standardized) Frobenius norm:

$$||A|| := \sqrt{\mathrm{Tr}(AA')/r} \quad \text{for any matrix } A \text{ of dimension } r \times m$$

Optimization Problem

(Standardized) Frobenius norm:

$$||A|| := \sqrt{\mathrm{Tr}(AA')/r} \quad \text{for any matrix } A \text{ of dimension } r \times m$$

Loss function:

$$\mathcal{L}(U_T D_T U_T, \Sigma_T) := ||U_T D_T U_T - \Sigma_T||^2$$

Optimization Problem

(Standardized) Frobenius norm:

$$\|A\| := \sqrt{\mathrm{Tr}(AA')/r} \quad \text{for any matrix } A \text{ of dimension } r \times m$$

Loss function:

$$\mathcal{L}(U_T D_T U_T, \Sigma_T) := \|U_T D_T U_T - \Sigma_T\|^2$$

Line of attack:

- It turns out that there is nonstochastic limit of the loss function, which involves the shrinkage function d
- We minimize the limiting expression with respect to d

Nonlinear Shrinkage Estimator

We illustrate the methodology for the case $c \leq 1$.

Nonlinear Shrinkage Estimator

We illustrate the methodology for the case $c \leq 1$.

Optimal limiting shrinkage function

$$d^*(\lambda) := \frac{\lambda}{\left|1 - c - c\,\lambda\,\breve{m}_F(\lambda)\right|^2}$$

Nonlinear Shrinkage Estimator

We illustrate the methodology for the case $c \leq 1$.

Optimal limiting shrinkage function

$$d^*(\lambda) := \frac{\lambda}{\left|1 - c - c\,\lambda\,\breve{m}_F(\lambda)\right|^2}$$

A feasible estimator is obtained by:

- Replacing c with N/T
- Consistently estimating $\breve{m}_F$, which is achieved by consistently estimating H and plugging it into the MP equation together with N/T

Resulting estimator: $\widetilde{\Sigma}_T := U_T \times \mathsf{Diag}(\widetilde{d}_T(\lambda_{T,1}), \ldots, \widetilde{d}_T(\lambda_{T,N})) \times U_T'$

Nonlinear Shrinkage Estimator

We illustrate the methodology for the case $c \leq 1$.

Optimal limiting shrinkage function

$$d^*(\lambda) := \frac{\lambda}{\left|1 - c - c\,\lambda\,\breve{m}_F(\lambda)\right|^2}$$

A feasible estimator is obtained by:

- Replacing c with N/T
- Consistently estimating $\breve{m}_F$, which is achieved by consistently estimating H and plugging it into the MP equation together with N/T

Resulting estimator: $\widetilde{\Sigma}_T := U_T \times \mathsf{Diag}(\widetilde{d}_T(\lambda_{T,1}), \ldots, \widetilde{d}_T(\lambda_{T,N})) \times U_T'$

The methodology can be extended to the case $c > 1$.

Nonlinear Shrinkage Estimator

We illustrate the methodology for the case $c \leq 1$.

Optimal limiting shrinkage function

$$d^*(\lambda) := \frac{\lambda}{\left|1 - c - c\,\lambda\,\breve{m}_F(\lambda)\right|^2}$$

A feasible estimator is obtained by:

- Replacing c with N/T
- Consistently estimating $\breve{m}_F$, which is achieved by consistently estimating H and plugging it into the MP equation together with N/T

Resulting estimator: $\widetilde{\Sigma}_T := U_T \times \mathsf{Diag}(\widetilde{d}_T(\lambda_{T,1}), \ldots, \widetilde{d}_T(\lambda_{T,N})) \times U_T'$

The methodology can be extended to the case $c > 1$.

Back to Main Talk