A Simple Method for Predicting Covariance Matrices of Financial Returns

Other titles in Foundations and Trends® in Econometrics

Factor Extraction in Dynamic Factor Models: Kalman Filter Versus Principal Components
Esther Ruiz and Pilar Poncela
ISBN:978-1-63828-096-5

Performance Analysis: Economic Foundations and Trends
Valentin Zelenyuk
ISBN: 978-1-68083-866-4

Experimetrics: A Survey
Peter G. Moffatt
ISBN: 978-1-68083-792-6

Climate Econometrics: An Overview
Jennifer L. Castle and David F. Hendry
ISBN: 978-1-68083-708-7

Foundations of Stated Preference Elicitation: Consumer Behavior and Choice-based Conjoint Analysis
Moshe Ben-Akiva, Daniel McFadden and Kenneth Train
ISBN: 978-1-68083-526-7

A Simple Method for Predicting Covariance Matrices of Financial Returns

Kasper Johansson
Stanford University
kasperjo@stanford.edu

Mehmet G. Ogut
Stanford University
giray98@stanford.edu

Markus Pelger
Stanford University
mpelger@stanford.edu

Thomas Schmelzer
Stanford University
Abu Dhabi Investment Authority
thomas.schmelzer@adia.ae

Stephen Boyd
Stanford University
boyd@stanford.edu

the essence of knowledge
Boston — Delft

Foundations and Trends® in Econometrics

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

K. Johansson *et al.*. *A Simple Method for Predicting Covariance Matrices of Financial Returns*. Foundations and Trends® in Econometrics, vol. 12, no. 4, pp. 324–407, 2023.

ISBN: 978-1-63828-309-6

Foundations and Trends® in Econometrics
Volume 12, Issue 4, 2023

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Econometrics publishes survey and tutorial articles in the following topics:

- Econometric Models
- Simultaneous Equation Models
- Estimation Frameworks
- Biased Estimation
- Computational Problems
- Microeconometrics
- Treatment Modeling
- Discrete Choice Modeling
- Models for Count Data
- Duration Models
- Limited Dependent Variables
- Panel Data
- Time Series Analysis
- Latent Variable Models
- Qualitative Response Models
- Hypothesis Testing
- Econometric Theory
- Financial Econometrics
- Measurement Error in Survey Data
- Productivity Measurement and Analysis
- Semiparametric and Nonparametric Estimation
- Bootstrap Methods
- Nonstationary Time Series
- Robust Estimation

Information for Librarians

Foundations and Trends® in Econometrics, 2023, Volume 12, 4 issues. ISSN paper version 1551-3076. ISSN online version 1551-3084. Also available as a combined paper and online subscription.

Contents

A Simple Method for Predicting Covariance Matrices of Financial Returns

Kasper Johansson[1], Mehmet G. Ogut[2], Markus Pelger[3], Thomas Schmelzer[4] and Stephen Boyd[5]

[1] *Department of Electrical Engineering, Stanford University, USA; kasperjo@stanford.edu*
[2] *Department of Electrical Engineering, Stanford University, USA; giray98@stanford.edu*
[3] *Department of Management Science and Engineering, Stanford University, USA; mpelger@stanford.edu*
[4] *Department of Electrical Engineering, Stanford University, USA, and Abu Dhabi Investment Authority, UAE; thomas.schmelzer@adia.ae*
[5] *Department of Electrical Engineering, Stanford University, USA; boyd@stanford.edu*

ABSTRACT

We consider the well-studied problem of predicting the time-varying covariance matrix of a vector of financial returns. Popular methods range from simple predictors like rolling window or exponentially weighted moving average (EWMA) to more sophisticated predictors such as generalized autoregressive conditional heteroscedastic (GARCH) type methods. Building on a specific covariance estimator suggested by Engle in 2002, we propose a relatively simple extension that requires little or no tuning or fitting, is interpretable, and produces results at least as good as MGARCH, a popular

Kasper Johansson, Mehmet G. Ogut, Markus Pelger, Thomas Schmelzer and Stephen Boyd (2023), "A Simple Method for Predicting Covariance Matrices of Financial Returns", Foundations and Trends® in Econometrics: Vol. 12, No. 4, pp 324–407. DOI: 10.1561/0800000047.

extension of GARCH that handles multiple assets. To evaluate predictors we introduce a novel approach, evaluating the regret of the log-likelihood over a time period such as a quarter. This metric allows us to see not only how well a covariance predictor does overall, but also how quickly it reacts to changes in market conditions. Our simple predictor outperforms MGARCH in terms of regret. We also test covariance predictors on downstream applications such as portfolio optimization methods that depend on the covariance matrix. For these applications our simple covariance predictor and MGARCH perform similarly.

1

Introduction

1.1 Covariance Prediction

We consider cross-sections, *e.g.*, a vector time series of n financial returns, denoted $r_t \in \mathbf{R}^n$, $t = 1, 2, \ldots$, where $(r_t)_i$ is the return of asset i from $t-1$ to t. We focus on the case where the mean $\mathbf{E}\, r_t$ is small enough that the second moment $\mathbf{E}\, r_t r_t^T \in \mathbf{R}^{n\times n}$ is a good approximation of the covariance matrix $\mathbf{cov}(r_t) = \mathbf{E}\, r_t r_t^T - (\mathbf{E}\, r_t)(\mathbf{E}\, r_t)^T$, where $\mathbf{E}$ denotes expectation. This is the case for most daily, weekly, or monthly stock, bond, and futures returns, factor returns, and index returns. We start by focussing on the case where the number of assets n is modest, say, on the order 10–100 or so; in Section 8 we explain how to extend the method to much larger universes using ideas such as factor models.

We model the returns r_t as independent random variables with zero mean and covariance $\Sigma_t \in \mathbf{S}^n_{++}$ (the set of symmetric positive definite matrices). We focus on the problem of predicting or estimating Σ_t, based on knowledge of $r_1, \ldots, r_{t-1}$. The prediction is denoted as $\hat{\Sigma}_t \in \mathbf{S}^n_{++}$. The predicted volatilities of assets are given by

$$\hat{\sigma}_t = \mathbf{diag}(\hat{\Sigma}_t)^{1/2} \in \mathbf{R}^n,$$

where **diag** with a matrix argument is the vector of diagonal entries of the matrix, and the squareroot of a vector above is elementwise. We denote the predicted correlations as

$$\hat{R}_t = \mathbf{diag}(\hat{\sigma}_t)^{-1}\hat{\Sigma}_t\,\mathbf{diag}(\hat{\sigma}_t)^{-1},$$

where **diag** with a vector argument is the diagonal matrix with entries from the vector argument.

Covariance estimation comes up in several areas of finance, including Markowitz portfolio construction (Markowitz, 1952; Grinold and Kahn, 2000), risk management (McNeil *et al.*, 2015), and asset pricing (Sharpe, 1964). Much attention has been devoted to this problem, and a Nobel Memorial Prize in Economic Sciences was awarded for work directly related to volatility estimation (Engle, 1982).

While it is well known that the tails of financial returns are poorly modeled by a Gaussian distribution, our focus here is on the bulk of the distribution, where the Gaussian assumption is reasonable. For future use, we note that the log-likelihood of an observed return r_t, under the Gaussian distribution $r_t \sim \mathcal{N}(0, \hat{\Sigma}_t)$, is

$$l_t(\hat{\Sigma}_t) = \frac{1}{2}\left(-n\log(2\pi) - \log\det\hat{\Sigma}_t - r_t^T\hat{\Sigma}_t^{-1}r_t\right). \tag{1.1}$$

The Gaussian log-likelihood is closely related to a popular metric for evaluating covariance predictors in econometrics, called the (Gaussian) quasi-likelihood (QLIKE) (Patton, 2011; Patton and Sheppard, 2009; Laurent *et al.*, 2013). QLIKE is the negative log-likelihood, under the Gaussian assumption, up to an additive constant and a positive scale factor. Roughly speaking, we seek covariance predictors that achieve large values of log-likelihood, or small values of QLIKE, on realized returns. We will describe the evaluation of covariance predictors in detail in Section 4.

1.2 Contributions

This monograph makes three contributions. First, we propose a new method for predicting the time-varying covariance matrix of a vector of financial returns, building on a specific covariance estimator suggested by

Engle in 2002. Our method is a relatively simple extension that requires very little tuning and is readily interpretable. It relies on solving a small convex optimization problem, which can be carried out very quickly and reliably (Boyd and Vandenberghe, 2004). Our method performs as well as much more complex methods, as measured by several metrics.

Our second contribution is to propose a new method for evaluating a covariance predictor, by considering the regret of the log-likelihood over some time period such as a quarter. This approach allows us to evaluate how quickly a covariance estimator reacts to changes in market conditions.

Our third contribution is an extensive empirical study of covariance predictors. We compare our new method to other popular predictors, including rolling window, exponentially weighted moving average (EWMA), and generalized autoregressive conditional heteroscedastic (GARCH) type methods. We find that our method performs slightly better than other predictors. However, even the simplest predictors perform well for practical problems like portfolio optimization.

Everything needed to reproduce our results, together with an open source implementation of our proposed covariance predictor, is available online at:

https://github.com/cvxgrp/cov_pred_finance.

1.3 Outline

In Section 2 we describe some common predictors, including the one that our method builds on. We introduce our proposed covariance predictor in Section 3. In Section 4 we discuss methods for validating covariance predictors that measure both overall performance and reactivity to market changes. We describe the data we use in our first empirical studies in Section 5, and give the results in Section 6.

In the next sections we discuss some extensions of and variations on our method, including realized covariance prediction (Section 7), handling large universes via factor models (Section 8), obtaining smooth covariance estimates (Section 9), and using our covariance model to generate simulated returns (Section 10).

References

Ait-Sahalia, Y. and D. Xiu. (2017). "Using principal component analysis to estimate a high dimensional factor model with high-frequency data". *Journal of Econometrics*. 201(2): 384–399.

Alexander, C. and A. Chibumba. (1997). "Multivariate orthogonal factor GARCH". *University of Sussex, Mimeo*.

Andersen, T., T. Bollerslev, P. Christoffersen, and F. Diebold. (2006). "Volatility and Correlation Forecasting". In: ed. by G. Elliott, C. Granger, and A. Timmermann. Vol. 1. *Handbook of Economic Forecasting*. Elsevier. 777–878.

Bai, J. (2003). "Inferential theory for factor models of large dimensions". *Econometrica*. 71(1): 135–171.

Bai, J. and S. Ng. (2008). "Large dimensional factor analysis". *Foundations and Trends® in Econometrics*. 3(2): 89–163.

Barratt, S. and S. Boyd. (2022). "Covariance prediction via convex optimization". *Optimization and Engineering*.

Bauwens, L., S. Laurent, and J. Rombouts. (2006). "Multivariate GARCH models: a survey". *Journal of Applied Econometrics*. 21(1): 79–109.

Bauwens, L. and E. Otranto. (2023). "Modeling realized covariance matrices: a class of Hadamard exponential models". *Journal of Financial Econometrics*. 21(4): 1376–1401.

Bauwens, L., G. Storti, and F. Violante. (2012). "Dynamic conditional correlation models for realized covariance matrices". *CORE DP*. 60: 104–108.

Bishop, C. (2006). *Pattern recognition and machine learning*. Vol. 4. No. 4. Springer.

Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity". *Journal of Econometrics*. 31(3): 307–327.

Bollerslev, T. (1990). "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model". *The Review of Economics and Statistics*. 72(3): 498–505.

Bollerslev, T., R. Engle, and J. Wooldridge. (1988). "A Capital Asset Pricing Model with Time-Varying Covariances". *Journal of Political Economy*. 96(1): 116–131.

Bollerslev, T., A. Patton, and R. Quaedvlieg. (2020). "Multivariate leverage effects and realized semicovariance GARCH models". *Journal of Econometrics*. 217(2): 411–430.

Boyd, S. and L. Vandenberghe. (2004). *Convex optimization*. Cambridge University Press.

Boyd, S. and L. Vandenberghe. (2023). "Convex Optimization Additional Exercises". URL: https://github.com/cvxgrp/cvxbook_additional_exercises.

Braga, M. (2015). *Risk-based approaches to asset allocation: concepts and practical applications*. Springer.

Brooks, C., S. Burke, and G. Persand. (2003). "Multivariate GARCH models: software choice and estimation issues". *Journal of Applied Econometrics*. 18(6): 725–734.

Choueifaty, Y. and Y. Coignard. (2008). "Toward maximum diversification". *The Journal of Portfolio Management*. 35(1): 40–51.

Cochrane, J. (2009). *Asset pricing: Revised edition*. Princeton university press.

De Nard, G., O. Engle R.and Ledoit, and M. Wolf. (2022). "Large dynamic covariance matrices: Enhancements based on intraday data". *Journal of Banking & Finance*. 138: 106426.

De Nard, G., O. Ledoit, and M. Wolf. (2021). "Factor models for portfolio selection in large dimensions: The good, the better and the ugly". *Journal of Financial Econometrics*. 19(2): 236–257.

deBrito, D., M. Medeiros, and R. Ribeiro. (2018). "Forecasting Large Realized Covariance Matrices: The Benefits of Factor Models and Shrinkage". *Available at SSRN 3163668*.

Dempster, A., N. Laird, and D. Rubin. (1977). "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the royal statistical society: series B (methodological)*. 39(1): 1–22.

Engle, R. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". *Econometrica*. 50(4): 987–1007. (Accessed on 03/05/2023).

Engle, R. (2002). "Dynamic Conditional Correlation". *Journal of Business & Economic Statistics*. 20(3): 339–350.

Engle, R. and K. Kroner. (1995). "Multivariate Simultaneous Generalized ARCH". *Econometric Theory*. 11(1): 122–150.

Engle, R., O. Ledoit, and M. Wolf. (2019). "Large dynamic covariance matrices". *Journal of Business & Economic Statistics*. 37(2): 363–375.

Engle, R. and K. Sheppard. (2001). "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH".

Fama, E. and K. French. (1992). "The cross-section of expected stock returns". *the Journal of Finance*. 47(2): 427–465.

Fama, E. and K. French. (1993). "Common risk factors in the returns on stocks and bonds". *Journal of financial economics*. 33(1): 3–56.

Fan, J., A. Furger, and D. Xiu. (2016). "Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data". *Journal of Business & Economic Statistics*. 34(4): 489–503.

French, K. (2023). "Kenneth French Data Library". URL: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html%5C#Research.

Ghalanos, A. (2019). "rmgarch: Multivariate GARCH models". *R package version*: 1–3.

Golub, G. and C. Van Loan. (2013). *Matrix computations*. JHU press.

Grinold, R. and R. Kahn. (2000). "Active portfolio management".

Hansen, P., Z. Huang, and H. Shek. (2012). "Realized GARCH: a joint model for returns and realized measures of volatility". *Journal of Applied Econometrics*. 27(6): 877–906.

Hansen, P., A. Lunde, and J. Nason. (2011). "The model confidence set". *Econometrica*. 79(2): 453–497.

Hastie, T., R. Tibshirani, and J. Friedman. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer.

Hautsch, N., L. Kyj, and P. Malec. (2015). "Do high-frequency data improve high-dimensional portfolio allocations?" *Journal of Applied Econometrics*. 30(2): 263–290.

Hazan, E. (2016). "Introduction to online convex optimization". *Foundations and Trends® in Optimization*. 2(3-4): 157–325.

Hazan, E., A. Agarwal, and S. Kale. (2007). "Logarithmic regret algorithms for online convex optimization". *Machine Learning*. 69(2-3): 169–192.

Jordan, M. and R. Jacobs. (1994). "Hierarchical mixtures of experts and the EM algorithm". *Neural computation*. 6(2): 181–214.

Laurent, S., J. Rombouts, and F. Violante. (2013). "On loss functions and ranking forecasting performances of multivariate volatility models". *Journal of Econometrics*. 173(1): 1–10.

Ledoit, O. and M. Wolf. (2003). "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection". *Journal of empirical finance*. 10(5): 603–621.

Ledoit, O. and M. Wolf. (2004). "Honey, I Shrunk the Sample Covariance Matrix". *The Journal of Portfolio Management*. 30(4): 110–119.

Ledoit, O. and M. Wolf. (2008). "Robust performance hypothesis testing with the Sharpe ratio". *Journal of Empirical Finance*. 15(5): 850–859.

Lettau, M. and M. Pelger. (2020a). "Estimating latent asset-pricing factors". *Journal of Econometrics*. 218(1): 1–31.

Lettau, M. and M. Pelger. (2020b). "Factors that fit the time series and cross-section of stock returns". *The Review of Financial Studies*. 33(5): 2274–2325.

Longerstaey, J. and M. Spencer. (1996). *Riskmetrics: Technical Document*. JP Morgan and Reuters.

Markowitz, H. (1952). "Portfolio Selection". *The Journal of Finance*. 7(1): 77–91.

McNeil, A., R. Frey, and P. Embrechts. (2015). *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press.

Menchero, J., D. Orr, and J. Wang. (2011). *The Barra US equity model (USE4), methodology notes*. MSCI Barra.

Mincer, J. and V. Zarnowitz. (1969). "The evaluation of economic forecasts". In: *Economic forecasts and expectations: Analysis of forecasting behavior and performance*. NBER. 3–46.

Mokhtari, A., S. Shahrampour, A. Jadbabaie, and A. Ribeiro. (2016). "Online optimization in dynamic environments: Improved regret rates for strongly convex problems". In: *2016 IEEE 55th Conference on Decision and Control (CDC)*. 7195–7201.

Oh, D. and A. Patton. (2016). "High-dimensional copula-based distributions with mixed frequency data". *Journal of Econometrics*. 193(2): 349–366.

Patton, A. (2011). "Volatility forecast comparison using imperfect volatility proxies". *Journal of Econometrics*. 160(1): 246–256.

Patton, A. and K. Sheppard. (2009). "Evaluating volatility and correlation forecasts". In: *Handbook of financial time series*. Springer. 801–838.

Pelger, M. (2020). "Understanding systematic risk: A high-frequency approach". *The Journal of Finance*. 75(4): 2179–2220.

Pelger, M. (2023). *Markus Pelger's Data and Code*. URL: https://mpelger.people.stanford.edu/data-and-code.

Pelger, M. and R. Xiong. (2022a). "Interpretable sparse proximate factors for large dimensions". *Journal of Business & Economic Statistics*. 40(4): 1642–1664.

Pelger, M. and R. Xiong. (2022b). "State-Varying Factor Models of Large Dimensions". *Journal of Business & Economic Statistics*. 40(3): 1315–1333.

Qian, E. (2011). "Risk Parity and Diversification". *The Journal of Investing*. 20(1): 119–127. DOI: 10.3905/joi.2011.20.1.119.

Rubin, D. and D. Thayer. (1982). "EM algorithms for ML factor analysis". *Psychometrika*. 47: 69–76.

Sharpe, W. (1964). "Capital asset prices: A theory of market equillibrium under conditions of risk". *The Journal of Finance*. 19(3): 425–442.

Silvennoinen, A. and T. Teräsvirta. (2009). "Multivariate GARCH models". In: *Handbook of financial time series*. Springer. 201–229.

Theil, H. (1961). *Economic forecasts and policy*. North-Holland Rotterdam.

Vassallo, D., G. Buccheri, and F. Corsi. (2021). "A DCC-type approach for realized covariance modeling with score-driven dynamics". *International Journal of Forecasting*. 37(2): 569–586.

Vrontos, I., P. Dellaportas, and D. Politis. (2003). "A full-factor multivariate GARCH model". *The Econometrics Journal*. 6(2): 312–334.

Weide, R. van der. (2002). "GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model". *Journal of Applied Econometrics*. 17(5): 549–564.

"Wharton Research Data Services". (2023). URL: https://wrds-web.wharton.upenn.edu/wrds/.

Zinkevich, M. (2003). "Online Convex Programming and Generalized Infinitesimal Gradient Ascent".