When “Structural Change” meets “Transaction Cost” in Volatility Timing

Y. Liao[a,*], **D. Bu**[b], **A. Gibberd**[c,**]

[a] *QUT Business School, Economics and Finance, The Queensland University of Technology, Australia*
[b] *Department of Applied Finance, Macquarie University, Australia*
[c] *Department of Mathematics & Statistics, Lancaster University, UK*

April 2, 2019

Abstract

We propose a new volatility timing strategy that are particularly favourable in the presence of both structural changes in covariance matrix and transaction costs. The approach relies on an assumption that the return covariance matrix changes sparsely over time, along with a novel regularized rolling window technique for tracking the significant changes. The new strategy strikes a good balance between reducing portfolio risk exposure and turnovers, earning larger out-of-sample certainty equivalent returns after transaction costs, compared with a set of commonly used alternatives. These gains are robust to economic states, large scale portfolios, investment horizons and investor's risk aversion.

*JEL classification:*C13, C51, C61, G11

Keywords: Volatility Timing Strategy; Structural Change; Transaction Cost; Sparse Time Variation; Large Scale Portfolio

---

*Principal corresponding author

**Corresponding author

Email addresses: yin.liao@qut.edu.au. (Y. Liao), di.bu@mq.edu.au (D. Bu), a.gibberd@lancaster.ac.uk (A. Gibberd)

I. Introduction

We motivate our study from a mean-variance investor perspective, who adjusts his allocation based on changes in the estimated conditional covariance matrix of returns, which are commonly referred to as "volatility timing" strategies. These strategies have gained popularity since the seminal work of Fleming et al. (2001), and recalled its importance during the recent global financial crisis (GFC)[1]. Despite the theoretical attractiveness, implementing "volatility timing" remains an empirical challenge to both academics and practitioners due to the presence of transaction costs. The purpose of "volatility timing" strategy is to overcome the challenge that changing market conditions present to traditional static asset allocation, and hence requires investors to actively adjust portfolio weights based on sample information of the volatility dynamics. However, the presence of transaction costs makes otherwise optimal rebalancing costly, outweighing the benefit of timing volatility. This presented investors with a dilemma: how should we conduct "volatility timing" in the presence of transaction cost? Answering this question is relevant not only for portfolio allocation during extreme periods of volatility like those seen in 2008, but also much more generally as reducing risk exposure (through active trading) and raising transaction cost is a salient trade-off in portfolio allocation. To address the dilemma, we propose a new "volatility timing" strategy that detects significant changes (or structural breaks) in asset return covariance matrix, and rebalances portfolio solely when a change point has been detected. We document that this new strategy simultaneously lowers portfolio risk exposure and turnovers, and largely improves out-of-sample certainty equivalent returns (CER) after transaction costs, compared with other commonly used alternatives. These results suggest that investors can benefit from change detection in volatilities in the presence of transaction cost. We then interpret the new strategy from both portfolio choice under turnover penalization and a Bayesian portfolio construction perspective.

A commonly employed approach in the literature to mitigate the impact of transaction cost in portfolio allocation is to slow down trading and allow investors to deviate from the zero-cost optimal positions. For example, the strategy discussed in Gârlranu and Pedersen (2013) and Gârlranu and Pedersen (2016) suggests invetors to trade only partially toward the

[1]Fleming et al. (2001) showed that portfolios formed using "volatility timing" outperform the unconditionally efficient static portfolios that have the same target expected return and volatility. A recent work of Moreira and Muir (2019) documented that a long-term investor who ignores variation in volatility gives up the equivalent of 2.4% of wealth per year.

desired position, e.g. trading only 15% toward the zero-cost optimal targets each day. Another strategy proposed by Kirby and Ostdiek (2012) allows the sensitivity of portfolio weights to volatility changes to be adjusted via a tuning parameter, and thus, the portfolio turnover can be controlled to a low level. Although these strategies successfully control over portfolio turnovers through reducing trading frequency, a common practice in these strategies is using a constant trading reduction rate over time, which would not be appropriate in reality. Intuitively, when the asset return volatility evolves smoothly such as what we have seen during the market calm periods, using a large reduction rate for trading can be appropriate as it can remarkably reduce transaction costs without highly increasing portfolio risk exposure. However, in more turbulent periods, a lower level of trading reduction rate would be required so that the portfolio weights can adjust faster and adapt with the significantly changed market conditions. Therefore, to better timing the volatility, we need a mechanism to locally detect significant variation in volatilities, and then, to tell investors when significant updates are needed for portfolio weights.

So far, the canonical approach to assessing time variations in asset return volatility is to use a rolling window estimation (see DeMiguel et al. 2009b,a, Kirby and Ostdiek 2012, Kourtis et al. 2012, Goto and Xu 2015, for example). This analysis involves recursively estimating sample covariance matrix of the asset returns by re-weighting new observations according to a rolling window. Subsequently, analysis can be performed directly on the covariance matrix estimate to infer the dependence structure of asset returns when new observations arise. While rolling windows are a valuable tool for investigating dynamic changes, there are two main issues associated with its use. First, the choice of window length can be a difficult parameter to tune. It is advisable to set the window length to be large enough to allow for a robust estimation but without making it too large, which can result in overlooking short-term fluctuations. Second, the rolling window faces the potential issue of variability between temporally adjacent estimates. This arises as a direct consequence of the fact that each covariance matrix across the rolling windows is estimated independently without any mechanism present to encourage temporal homogeneity. This additional variability can jeopardise the accuracy of the estimation as well as hugely increase turnovers in the context of portfolio allocation.

To address these issues, we propose a *regularized rolling window (RRW)* approach in this paper to assess the time variation of covariance matrices. Our RRW approach regularizes the standard rolling window estimation using a penalty term that assists which exploits temporal similarity between consecutive window estimates, resulting in a piecewise constant estimate. Specifically, our approach estimates the covariance matrix using traditional quasi-maximum likelihood but with an additional constraint in order to shrinking the difference between contemporaneous and lagged estimates. Without the constraint, our approach is reduced back to the standard rolling window analysis, but with the constraint, the overall size of element-wise differences between consecutive covariance matrix estimates are penalized, producing two-fold estimation effects: (i) the time variation of any element of the covariance matrix (e.g. the $(i, j)th$ entry) is set to be zero if the sample variation is below than a constraint whose strength relates to a turning parameter in the RRW optimisation problem; (ii) the time variation of any element of the covariance matrix is decreased toward zero by the magnitude of the threshold when the sample variation is above the threshold. As such, the penalty term leads the covariance matrix estimate to achieve both shrinkage and sparsity in time variations, eliminating unnecessary evolutions embedded in the standard rolling window analysis. The motivation of our RRW approach is similar to the recent advances in machine learning and statistics on change detection in time-varying regression parameters: the "fused lasso" estimator of Tibshirani et al. (2005) that operates in a linear (least-squares) regression setting and acts to shrink the insignificant changes in consecutive parameter estimates towards zero. We extend the idea to detecting changes in a matrix at element level, and employ the matrix version of "fused lasso", that is, the "graphical fused lasso" algorithm of Gibberd and Nelson (2017) to solve the estimation problem.

Next, we apply our new covariance matrix estimate to mean-variance portfolio decision to develop a new volatility timing strategy. Under our approach, the portfolios are re-balanced monthly based solely on the significant changes detected in the covariance matrix, otherwise, remain as if a buy-and-hold portfolio. We control the sensitivity of change detection in covariance matrix, and also, the portfolio turnovers via a tuning parameter in the RRW approach. The tuning parameter is hence treated as a measure of change detection aggressiveness, and allows us to keep the turnover of the proposed strategies to a level competitive with other existed strategies, e.g. $1/N$ naive diversification. To better understand the economic motivation of the new strategy, we offer alternative interpretation on the strategy from both portfolio decision with turnover penalization and Bayesian portfolio decision perspective. A recent study of Hautsch and Voigt (2019) established a link between turnover penalization and covariance shrinkage in portfolio allocation. Following this study, we show that a zero-cost mean-variance portfolio formed using our new covariance matrix estimator is equivalent to a sample mean-variance portfolio achieved through using a turnover penalization, where (i) the turnover is measured by a weighted quadratic transaction cost, and (ii) the weights on transaction costs are determined by a term measuring the difference between the standard and our regularized rolling window based covariance matrix estimates. A large difference between the standard and our regularized rolling sample estimates implies that the real covariance structure is relatively stable so that the regularization removes hugely the unnecessary evolutions generated by the standard rolling window analysis. Regularizing the weighted transaction costs, therefore, ensures that the strength of penalization on portfolio turnover is time-varying, depending on the magnitude of time variation in covariance matrix. To the best of our knowledge, the time varying turnover penalization is novel in the literature, as previous studies (e.g. DeMiguel et al. 2009b, DeMiguel and Olivares-Nadal 2018, Engle et al. 2012) usually adopted a constant tuning parameter to control the level of penalization on portfolio turnovers. Turning to the Bayesian interpretation, we know Bayesian investors often employ useful prior information about quantities of interest. We show that the covariance matrix estimate from our RRW approach can be interpreted as being the *maximum a-posteriori (MAP)* Bayesian estimate associated with Gaussian likelihood for asset returns and a Laplace prior on the time change of the inverse covariance matrix. The resulting portfolios are hence a Bayesian portfolio formed by the investors who have a prior belief on the changes of covariance matrix, and where they construct a portfolio that maximizes the posterior distribution of the change in the covariance matrix.

Lastly, we investigate the economic value of our new volatility timing strategy using a range of real data sets. We evaluate the out-of-sample empirical gains associated with investing in mean-variance portfolios using our new covariance matrix estimate, where the portfolio expected returns are measured using the sample estimate which is a constant. Therefore, the portfolio weights mainly focus on our estimation on the structure of covariance matrix. We find that the new strategy outperforms a set of commonly used mean-variance portfolio alternatives, including the standard rolling sample strategy where the covariance matrix is measured by the rolling sample estimate, the $1/N$ naive diversification strategy and the volatility timing strategies using covariance matrix forecasts from dynamic models, e.g., Exponential weighted moving average (EWMA) model, in terms of both out-of-sample portfolio risk and turnover control. More important, the new strategy earns significantly larger certainty equivalent return (CER) after transaction cost. For example, our strategy has an estimated annual CER of 10.238% after transaction cost for a data set comprising 25 portfolios formed on size and book-to-market characteristics. In comparison, the three competing strategies have estimated CER (after transaction cost) of 4.924%, 3.927% and 8.458%, respectively. We further provide additional insights into the gains generated by our new strategy, and attribute the gains to the improved covariance matrix estimation accuracy as well as the better timing ability in significant changes of the covariance matrix.

Methodologically, our RRW approach relates to a burgeoning literature on estimating covariance matrix using the shrinkage technique developed in statistics and machine learning fields. The idea behind is to shrink an unbiased estimator towards a lower variance (or more stable) target so that the shrunk estimator can strike a balance between mis-specification biases and estimation risk. The existing shrinkage estimators of covariance matrix, so far, mainly focus on addressing the static single-period estimation problem, e.g. the "sparse" estimator of Goto and Xu (2015) that shrinks the off-diagonal elements of the (inverse) covariance matrix towards zero and thus reduces the cross-sectional dimension (or the number) of assets in the portfolio selection. The estimators of Ledoit and Wolf (2003) and Chan et al. (1999) shrink the sample covariance matrix towards a more parsimonious target matrix, such as a constant correlation matrix or a covariance matrix with industry factor structure. Our RRW estimator extends the shrinkage idea to a dynamic setting, casting attention on the time change of the covariance matrix. The improved covariance matrix estimator is hence particular conducive to dynamic portfolio selection.

The second contribution of our study links with the work of Fleming et al. (2001), Fleming et al. (2003) and Moreira and Muir (2017) who study volatility timing empirically in the context of a short-term mean-variance investor. We go well beyond the results in these papers by proposing a sparse rather than continuous assumption on the time variation of the covariance matrix. We then design an algorithm to pick up the sparse set of time points where the covariance matrix experiences structural changes. In the context of portfolio allocation with the presence of transaction cost, the sparse assumption on the covariance matrix time variation is particularly attractive, as the resulting piecewise constant covariance matrix estimate would largely reduce portfolio turnovers without losing capacity in detecting significant changes.

The rest of the paper is organized as follows. Section 2 provides a small simulation study to address the motivation of our new RRW method. Section 3 introduces the RRW method for change detection in conditional covariance matrix and provides discussion on several empirical implementation issues. Section 4 develops our new volatility timing strategy using the RRW covariance matrix estimator and offers two alternative interpretations on the new strategy. Section 5 describes the data and presents the empirical analysis results. Section 6 examines the robustness of our results. Section 7 concludes.

II. Motivating Examples

There is a tradition in the portfolio allocation literature of using a rolling window to obtain sample estimates for the covariance matrix with the purpose of capturing the time-varying market condition. This section provides a small motivating example showing that the sample estimate from standard rolling window approach is an unreliable estimate for two significant reasons. The first is that the standard rolling window may create extra variability in the consecutive window estimates, due to the estimator assuming independence across rolling windows without a mechanism to exploit the temporal similarity. The second is that the spurious variation introduced by rolling windows can potentially mask the existence of significant changes in true covariance matrix. How severe are these statistical issues? We address this question using the following simulation study and contrast the standard rolling window with our novel regularised rolling window.

Consider a simple bivariate Gaussian process $r_t \sim \mathcal{N}(0, \Sigma_t)$ where the covariance Σ_t can change over time. To illustrate the issue of spurious estimator time variation, we will assume the covariance matrix is constant across time with $\Sigma_{11} = 1$, $\Sigma_{22} = 0.4$, and $\Sigma_{12} = 0.1$. However, we let the covariance Σ_{12} exhibits a changepoint at time τ, and its value changes to 0.2 throughout the rest of sample periods. We simulate $T = 2000$ return data from the bivariate process and set up $\tau = 1000$. At first, we simply use a standard rolling window approach with width $M = 100$ to estimate the covariance matrix, and then compares this to our newly proposed regularised rolling window (RRW) estimator.

Table 1 represents estimation results of Σ_{12} across 300 simulation trials. μ_1 and μ_2 denote the average of the estimators across the 300 trials before and after the change point $\tau = 1000$, and σ_1 and σ_2 denote the standard deviation of the estimators. We also report the percentage of time points where the estimators exhibit changes from last period estimate at the last column of the table. First, looking at the average (μ_1 and μ_2) and standard deviation (σ_1 and σ_2), we find that while the standard rolling window estimator performs slightly better than our estimator in terms of estimation bias, the estimation standard deviation is largely reduced by our estimator. This is consistent with the objective of regularized estimation that shrinks the estimator from the unbiased estimator (the sample-based estimator) in the direction that reduces estimation errors. Second, turning to the average percentage of time points at which the estimates changes from the previous time-step, our estimator changes at much less time points than the standard rolling window counterpart, proving that with the regularization our estimator shows piece-wise constant over time, largely constraining spurious time variations exhibited by the standard rolling sample estimator. As we will see later in the paper, this property proves extremely valuable for non-zero cost dynamic portfolio selection. Third, as our estimator involves a turning parameter $(\lambda) tocontrolthestrengthoftheregularization, weseethatwhenweincreasethevalueof \lambda$ from 0.05 to 0.1, both the standard deviation and percentage of changed time points of the estimator are further reduced. To further provide a visual evidence, Figure 1 plots the time series of standard rolling sample estimator v.s our estimator (by setting $\lambda = 0.1$) for Σ_{12} from a single simulation trial. Clearly, our RRW estimator is much more stable than the standard rolling sample estimator when the true covariance is constant, and also, has a better capacity to reflect the significant changes in the true covariance.

In summary, the standard rolling window estimates vary substantially over time, even in periods where the true covariance is constant. As a result, when these estimates are used for portfolio allocation, the additional time variation would have a strong influence on the portfolio weights and turnovers, as even small time variations in the sample estimates of the covariance matrix can lead to large fluctuations in the portfolio weights (Kirby and Ostdiek 2012). In the next section we will introduce our new regularised rolling window approach to estimating the covariance matrix in more detail.

III. Econometric Methodology

A. The Regularised Rolling Window Approach

We start with the standard rolling window approach, where the covariance matrix at each time t is estimated by minimizing a loss function (or negative log-likelihood) with return

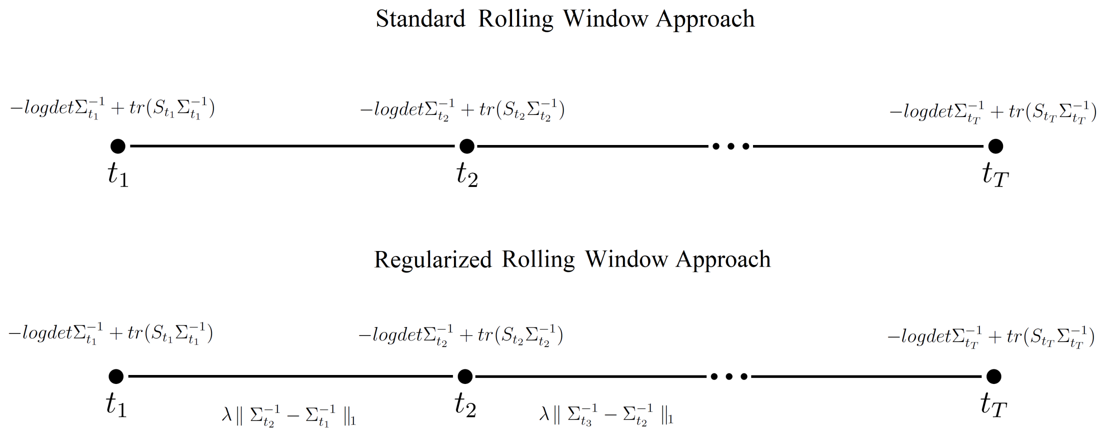

Figure 1: **A graphical comparison between the standard and regularized "rolling window" approach.** The top panel shows the estimation procedure of standard rolling window approach and the bottom panel shows the estimation procedure of the regularized rolling window approach.

observations from the time window $[t-1-h, t-1]$, where h is the window length. Specifically, we consider

$$\hat{\Sigma}_t^{-1} := \arg\min_{\Sigma_t^{-1}} \left[l(\Sigma_t^{-1}) \right] \ .$$

with the loss

$$l(\Sigma_t^{-1}) := -\log\det(\Sigma_t^{-1}) + \mathrm{trace}(\hat{S}_t \Sigma_t^{-1}) \ . \tag{1}$$

where $\hat{S}_t$ and Σ_t^{-1} denote the sample and our estimate of the covariance matrix, respectively. Thus, the time-varying covariance matrix through each time point are estimated recursively by including new observations according to the rolling window. Clearly, with the rolling window analysis, the covariance matrix is estimated independently across estimation windows without any mechanism present to encourage temporal similarity. A potential issue arises that the estimates for two adjacent time points might be largely different due to estimation errors which contradicts with the reality especially when the market is relatively stable. Additionally, the extra variability caused by the independent estimation across rolling windows can potentially mask significant changes in covariance matrix.

To address these empirical features, we propose a regularized rolling window approach in order to enforce temporal homogeneity and thus reflect significant time variations in covariance matrix. We minimize a penalized loss function which contains the negative log-likelihood as shown in Equation 1 and an additional penalty term that regularizes the difference between the contemporaneous and the lagged estimates produced by previous estimation window:

$$l(\Sigma_t^{-1}) = -\log\det(\Sigma_t^{-1}) + \mathrm{trace}(\hat{S}_t\Sigma_t^{-1}) + \lambda\|\Sigma_t^{-1} - \hat{\Sigma}_{t-1}^{-1}\|_1, \tag{2}$$

where $\hat{\Sigma}_t^{-1}$ is the lagged estimate from the previous estimation window, and the ℓ_1 norm, defined as

$$\|\Sigma_t^{-1} - \Sigma_{t-1}^{-1}\|_1 := \sum_{i,j} |\Sigma_{ij,t}^{-1} - \Sigma_{ij,t-1}^{-1}| \ ,$$

measuring the difference between the current and lagged inverse covariance matrix estimates (that is, the sum of the absolute values of edgewise differences between the two estimates.). The λ is a tuning parameter that controls the degree of regularization on the difference, becoming a soft threshold.

Clearly, the new approach nests the standard rolling window estimation as a special case when the regularization parameter (λ) is equal to zero. Figure 2 gives a graphical interpretation about the relation between our RRW approach and the standard rolling window analysis. The regularization term, $\lambda\|\Sigma_t^{-1} - \hat{\Sigma}_{t-1}^{-1}\|_1$, restricts the difference between the current inverse covariance matrix Σ_t^{-1} and the estimate of previous period $\hat{\Sigma}_{t-1}^{-1}$. For off-diagonal entries, if the consecutive difference in the inverse-covariance is below than a pre-determined soft threshold, e.g. related to the size of λ, we set the difference to be zero, when the difference is above the soft threshold we shrink the difference by the the magnitude of the soft threshold toward zero. As such, the regularization achieves both sparsity and shrinkage in time variations of the covariance matrix estimation, encouraging a temporally stable (or piecewise constant) estimator which has capacity to reflect big changes in the matrix at element level, but is less sensitive to small ones.

B. Empirical Implementation

We next discuss several empirical issues related to the implementation of the RRW approach. First, we turn to the problem of solving Equation 2. Due to the presence of the penalty terms, this loss function is convex, however, because of the ℓ_1 norm it is not continuously differentiable. In this context, the traditional optimization algorithms used for maximum likelihood estimation, or generalized method of moments, etc., cannot be adopted without further modifications. Instead, we employ the popular alternating directions method of multipliers

(ADMM) algorithm (Boyd et al. 2010) to solve the optimisation problem. The ADMM method is a form of augmented Lagrangian algorithm that is particularly well suited to addressing the highly structured nature of problems such as the one proposed here, for instance Danaher et al. (2013), Gibberd and Nelson (2017) also use this approach for fused estimation of inverse covariance matrices. We provide more detail on the estimation procedure using the ADMM algorithm in Appendix B.

The second major challenge when implementing the RRW approach relates to the selection of the regularisation parameter λ. In this paper, we employ a heuristic parameter-tuning technique inspired by the Akaike information criterion (AIC). We define the AIC for each window $t = 1, \ldots T$ as

$$\mathrm{AIC}_t(\lambda) = -\log\det(\Sigma_t^{-1}) + \mathrm{trace}(\hat{S}_t \Sigma_t^{-1}) + K_t, \tag{3}$$

where K_t is an estimate of the "degrees of freedom". Further discussion of the degrees of freedom is deferred to Appendix C. In practice, we use the first $T = 120$ months as a training period (with window width $M = 60$) and use it to search for the value of λ that minimizes the average, $(T - M)^{-1} \sum_{t=M}^{T} \mathrm{AIC}_t(\lambda)$. Following this training period, we adhere to this choice throughout the out-of-sample testing period[2].

IV. The Portfolio Problem

In this section, we use our RRW approach based covariance matrix estimates to develop a new volatility timing strategy. To better understand the motivation of the new strategy, we offer two alternative interpretations from the portfolio choice with turnover penalization and the Bayesian portfolio choice perspectives. We conclude the section by introducing several commonly used mean-variance strategies and performance metrics which we will use to evaluate RRW portfolios.

A. Volatility Timing Strategies

To develop our volatility timing strategies, we consider a risk-averse investor who allocates wealth across N risky assets plus a riskless asset, e.g. cash. The investor uses conditional mean-variance analysis to make his allocation decisions and re-balances his portfolio monthly.

[2]Goto and Xu (2015) stated that when the "rolling horizon" approach is used, re-estimating the tuning parameters period by period can certainly improve the performance, but they did not adopt this approach due to the intensive computation burden associated with it. We refer to their statement here to justify our choice.

Let R_{t+1}, $\mu = E[R_{t+1}]$, and $\Sigma_t = E_t[(R_{t+1} - \mu)(R_{t+1} - \mu)']$ denote an $N \times 1$ vector of risky asset returns, the expected value of R_{t+1}, and the conditional covariance matrix of R_{t+1}. For each date t, to minimize conditional volatility subject to a given expected return, the investor solves the following quadratic program:

$$\begin{aligned} w_{t+1} &:= \min_w [w^\top \Sigma_t w] \\ s.t. \quad \mu_p &= w^\top \mu + (1 - w^\top 1) r_{\text{free}}, \end{aligned} \tag{4}$$

where w is an $N \times 1$ vector of portfolio weights on the risky assets, r_{free} is the return on the riskless asset, and μ_p is the target expected return. The solution to this optimization problem is given by:

$$w_{t+1} = \frac{\mu_p \Sigma_t^{-1} \mu}{\mu^\top \Sigma_t^{-1} \mu}\,, \tag{5}$$

where w delivers the risky asset weights, and the weight on the riskless asset is $1 - w^\top 1$.

The trading strategy implicit in Equation 5 identifies the dynamically re-balanced portfolio that has minimum conditional variance for any choice of expected return, which is referred to as *minimum volatility (MV)* strategy. Similarly, we could conduct an analysis where the objective is to maximize the expected return subject to achieving a particular conditional variance. To save the space, we only report results relating to the minimum volatility strategy in the empirical section. The results relating to the maximum return strategy are available upon request. The MV strategy given via Equation 5 implies that to identify portfolio weights requires us to have one-step-ahead estimates of both the vector of conditional means, μ, and the conditional covariance matrix, Σ_t. Merton (1980) shows that a very long sample period would be needed to produce reliable coefficient estimates in a predictive regression. We assume, therefore, that our investor models expected returns as constant (and we estimate it using sample mean), and the portfolio weights in the strategy primarily focus on time variation in covariance matrix. Plugging the RRW covariance matrix estimator into equation 5, we form a new MV strategy that updates portfolio weights solely based on significant changes in Σ_t.

B. A Transaction Cost Interpretation of RRW

In Hautsch and Voigt (2019), a link between turnover penalization and covariance shrinkage in portfolio allocation is investigated. Specifically, they show that the optimization problem with quadratic transaction costs can be interpreted as a classical mean-variance problem without transaction costs, however, where the covariance matrix is regularized towards the identity matrix. Following this line of interpretation, we show that how optimal MV portfolio achieved using our RRW based covariance matrix estimate without transaction costs links with MV portfolio optimization problem with a time-varying quadratic transaction costs, where the time dependence is determined by the difference between our regularized and the sample covariance matrix.

Proposition 1. *RRW Equivalence to Transaction Cost Penalisation*

In the case of our RRW approach, the resulting covariance matrix is derived from the following optimization problem:

$$\begin{aligned} \hat{\Sigma}_{t,RRW} \quad &:= \quad \arg\max_{\Sigma_t} \left[\log\det(\Sigma_t) - \mathrm{tr}(\Sigma_{t-1}^{-1}, \Sigma_t - \hat{S}_t)\right] \ . \\ s.t. \quad &\|\Sigma_t - \hat{S}_t\|_\infty \leq \lambda \end{aligned} \tag{6}$$

where $\|X\|_\infty := \max_{ij} |X_{ij}|$ *is the dual norm of* $\|X\|_1$. *Thus, the RRW covariance matrix estimator* $\hat{\Sigma}_{t,RRW}$ *can be expressed associated with the sample covariance estimator* $\hat{S}_t$ *as*

$$\hat{\Sigma}_{t,RRW} = \hat{S}_t + \lambda\Delta_t, \tag{7}$$

where $\Delta_t := (\hat{\Sigma}_{t,RRW} - \hat{S}_t)/\lambda$. *The proof of the above is given in the Appendix A. and follows from basic duality properties of the RRW optimization problem. Plugging the RRW estimator into the standard mean-variance portfolio optimization, the portfolio allocation* w_{t+1}^* *can then be stated as*

$$\begin{aligned} w_{t+1}^* \quad &= \quad \arg\min_{w} \left[w^\top \hat{\Sigma}_{t,RRW} w - w^\top \mu\right] \\ &= \quad \arg\max_{w} \left[w^\top \mu^* - w^\top \hat{S}_t w - \lambda(w - w_t^+)^\top \Delta_t (w - w_t^+)\right] \ , \end{aligned} \tag{8}$$

where

$$\mu_t^\star = \mu - 2\Delta_t w_t^+ \ ,$$

and the weights are normalized such that $w^\top 1 = 1$.

Intuitively, our RRW estimator regularizes the sample estimator $\hat{S}_t$ through an additional matrix Δ_t, with λ serving as shrinkage parameter. Note that the regularization effect of

RRW estimator exhibits some similarity to the implications of the shrinkage approach proposed by Ledoit and Wolf (2003), but the RRW approach replaces the indentity matrix in the Ledoit and Wolf estimator with Δ_t which is time-varying. The resulting zero-cost mean-variance portfolio using our RRW estimator is thus equivalent to a portfolio formed with a time-varying penalty on transaction cost. Figure 3 provides an example of time evolution of $\lambda\Delta_t$[3] and highlights recession periods according to the NBER business cycle classification in grey. Clearly, in contrast with the constant identity matrix, the figure shows the Δ_t changes over time and peaks at the recession periods.

To further illustrate the effect of time-varying transaction cost penalty on portfolio optimization, Figure. 4 plots the surface of mean-variance optimization function over time with both our time-varying and traditional constant quadratic transaction cost penalty. The optimal solution for first asset allocation (that is, w_1) is also provided in the lower panel of the figure. The NBER defined recession periods are also highlighted in grey. We observe that the function surface under the time-varying transaction cost penalty (top left panel) exhibits much more stable than the one using constant penalty (top right panel) in most of time without losing capacity to reflect abrupt changes of market conditions during recession periods, such as the observed spikes in the function surface during 2008 crisis period. These findings are further collaborated by the time series plot of w_1 in the lower panel of the figure. Taken together, the time-varying transaction cost penalty resulting from our RRW covariance matrix estimator helps to impose a market-condition-dependent regularization on the transaction cost, assisting investors to achieve better marketing timing and better balance between portfolio risk exposure and turnovers.

C. Bayesian Portfolio Interpretation

Kyung et al. (2010) and Wang (2012) respectively give Bayesian interpretations for regularised regression via the lasso Tibshirani (1996) and the graphical lasso Friedman et al. (2008). Since these estimators are closely aligned with the RRW optimisation problem we can follow their line of reasoning to give a Bayesian interpretation for our RRW covariance matrix estimator as well as the resulting portfolios.

[3] To generate the examples in Figs. 3 and 4, we construct a two-asset portfolio using two randomly selected assets from the Fama-French 48 Industry portfolio dataset. We consider both the max and min elements of the matrix $\Delta_t to illustrate its time evolution in Figure 3$

Assumption 1. *We start with the assumption that future stock returns are independently normally distributed according the previously estimated covariance, i.e.*

$$r_{t+1} \sim \mathcal{N}(0, \Sigma_t) \ , \tag{9}$$

and $r_{t+1} \perp r_t$, i.e. there is no auto-correlation structure in returns.

Assumption 2. *In order to understand the RRW estimator, we now make a further assumption, and put ourselves in the shoes of an investor who has a prior belief that the inverse covariance may change in a sparse manner over time, i.e. changes will not be at every time-step, but occur rarely. Specifically, we will assume that the temporal variation of the inverse-covariance follows a Laplace (double exponential) distribution:*

$$p(\Theta_{t+1} - \hat{\Theta}_t|\rho) = Z^{-1} \prod_{i<j} \left\{ f_{\mathrm{DE}}(\Theta_{t;ij} - \hat{\Theta}_{t-1;ij}|\rho) \right\} \times \prod_{i=1}^{N} \left\{ f_{\mathrm{Exp}}(\Theta_{t+1;ii} - \hat{\Theta}_{t;ii}|\rho/2) \right\} \mathbf{1}_{\Theta_t \succeq 0} \ ,$$

where $f_{\mathrm{DE}}(x|\rho) = (\rho/2)\exp(-\rho|x|)$ has the form of the double exponential density, $f_{\mathrm{Exp}}(x|\rho) = \rho \exp(-\rho x)\mathbf{1}_{x>0}$ has the form of the exponential density, and Z is a normalising constant. The notation $\mathbf{1}_{\Theta \succeq 0}$ is used to denote the indicator function, in this case for the space of positive definite matrices for Θ_t.

Proposition 2. *Given that assumptions 1 and 2 hold, and we further assume $\Sigma_t = \hat{\Sigma}_t$ in Equation 9, then one can interpret the RRW estimator (minimiser of Eq. 2) as being the maximum-a-posteriori (MAP) estimate for the inverse covariance at time $t+1$. Specifically, assume that investor believes the temporal variation of inverse covariance matrix has a prior distribution as above, there exists a threshold parameter ρ such that our RRW estimator is the mode of the posterior distribution of Θ_t.*

Now it is clear that under our framework, choosing the portfolio that maximizes the posterior distribution of the change of (inverse) covariance matrix guarantees that the investor is choosing the portfolio with the highest probability of being the MV portfolio given the investors prior distribution on the temporal change of the (inverse) covariance matrix and the observed asset-return data. In other words, in our setting, the investor chooses the portfolio that maximizes the posterior probability (i.e. the posterior mode) of the change of (inverse) covariance matrix. This interpretation is a bit different from the traditional Bayesian portfolio choice literature in which the investor either chooses the portfolio that maximizes expected utility with respect to the posterior distribution of stock returns (for instance, Jorion 1986) or the portfolio that maximizes the posterior distribution of portfolio weights directly (see DeMiguel et al. 2009a, Tu and Zhou 2010). In our framework the investor has a prior belief on the change of (inverse) covariance matrix rather than on the asset-return distribution or on the portfolio weights. Consequently, while the Bayesian investor in the traditional setting chooses the portfolio that maximizes expected utility with respect to the posterior distribution of asset returns, or chooses the portfolio that maximizes portfolio weights with respect to the posterior distribution of portfolio weights, in our setting the investor chooses the portfolio that maximizes the posterior distribution of the (inverse) covariance matrix changes.

D. *Performance Evaluation Metrics*

To measure the economic value of our new approach, we compare its performance with several competing MV strategies using a series of performance evaluation metrics. Our approach builds on the standard rolling window estimation, and thus, a natural choice for the benchmark is the MV portfolio using standard rolling sample estimates of covariance matrix, which we refer to as $\mathrm{MV}_{\mathrm{sample}}$. This approach uses the rolling sample covariance matrix as a predictor for the future covariance-matrix (Chan et al. 1999, DeMiguel et al. 2009b,a, Kirby and Ostdiek 2012, Kourtis et al. 2012, Goto and Xu 2015). To further reduce the estimation error of the sample covariance matrix, we use the shrunk version of the sample estimator (Ledoit and Wolf 2003) that shrinks the sample estimate of covariance matrix towards an identity matrix. Counterparts for the RRW approach can be generated by using the shrunk sample estimator for S_t in Equation 1. Secondly, we consider the naïve $1/N$ strategy, denoted by $\mathrm{MV}_{\mathrm{equal}}$. The naïve $1/N$ strategy demonstrates favorable out-of-sample performance and has been found very hard to beat in practice, especially in the presence of high transaction costs. We use it as a benchmark in order to examine the ability of our new approach in terms of controlling transaction costs. Thirdly, we consider other volatility timing strategies, specifically, the MV portfolios using covariance forecasts from dynamic models, e.g. the *Exponentially Weighted Moving Average (EWMA)* model (Zakamulin 2015). This method of estimating the covariance matrix is popularized by the RiskMetrics group, and Zakamulin (2015) also find that the simple EWMA covariance matrix forecast performs comparably with the multivariate GARCH forecast. In this approach, the exponentially weighted covariance matrix is estimated using the following recursive form:

$$\hat{S}_t^{\text{EWMA}} = (1 - \lambda_{\text{EWMA}})\varepsilon_{t-1}\varepsilon_{t-1}^{\top} + \lambda_{\text{EWMA}}\hat{S}_{t-1}^{\text{EWMA}} \quad , \tag{10}$$

where $0 < \lambda_{\text{EWMA}} < 1$ is the decay constant, and ε_{t-1} is the return residual. We follow the recommendations of the RiskMetrics group and estimate the EWMA covariance-matrix using $\lambda_{EWMA} = 0.97$. This comparison is of particular useful for examining whether investors benefit from sparse rather than continuous time variation assumption on covariance matrix.

Next, we evaluate portfolio out-of-sample performance from several perspective. First, we test out-of-sample performance in terms of risk exposure and Sharpe ratio. The portfolio risk exposure is measured by standard deviation of out-of-sample portfolio returns, and the Sharpe ratio is calculated as

$$\hat{SR} = \hat{\mu}/\hat{\sigma} \quad , \tag{11}$$

where $\hat{\mu}$ and $\hat{\sigma}$ are mean and standard deviation of portfolio return over the out-of-sample testing period. Secondly, we examine portfolio turnovers measured by

$$\text{Turnover} = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}(|\hat{w}_{i,t+1} - \hat{w}_{i,t^+}|), \tag{12}$$

where $\hat{w}_{i,t+1}$ and $\hat{w}_{i,t}$ are the desired portfolio weights in asset i at time t and $t+1$, after rebalancing, and $\hat{w}_{i,t^+}$ is the portfolio weight before rebalancing at $t+1$. The turnover quantity defined can be interpreted as the average percentage of wealth traded in each period. Lastly, we assess whether our new strategy has economic gain. We calculate the annualized *certainty equivalent excess return (CER)* of the portfolio after subtracting the *transaction costs (TCOST)*, i.e.

$$\text{CER}_{\text{Tcost}} := \hat{\mu} - \frac{\gamma}{2}\hat{\sigma}^2 - \text{TCOST}, \tag{13}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are the (annualized) mean and variance of out-of-sample portfolio excess returns. Alternatively, the $\text{CER}_{\text{Tcost}}$ can be interpreted as the return increase compared to the risk-free rate that an investor is willing to trade for a risky portfolio after accounting for transaction costs and variance.

V. Empirical Analysis

A. Data

We employ three data sets: two from Ken French's Web site for portfolio investing analysis, and one from the Center for Research in Security Prices (CRSP) database for individual asset investing analysis. The Ken French data sets contain the returns on 25 value-weighted portfolios of stocks sorted by size and book-to-market (that is, 25FF), and the 48 industry value-weighted portfolios (that is, 48Ind). For close-to-close returns, we use data from 1967 to 2017 downloaded from Ken French's Web site. The second data set consists of 50 individual stock returns from the CRSP database (that is, CRSP50), containing close-to-close returns on all stocks that were part of the S&P 500 index at some point between 1992 and 2017.

Table 2 summarizes the sample information for each data set, and also provides information regarding the question whether the sparse assumption on time variation of covariance matrix is supported by the real data. Looking at the column 8 and 9 of the table, where we report the optimal value of λ we choose for each data set (as selected via AIC), and the average percentage of non-changed off-diagonal elements between consecutive covariance matrix estimates throughout the whole sample periods. The latter measures the degree of sparsity in time variation of the covariance matrix. We find that the degree of sparsity ranges from 35.43% to 23.55% across these data sets, meaning that a significant fraction of elements in covariance matrix did not significantly change between consecutive periods. Hence, the sparse time variation assumption appears reasonable in practice.

B. Portfolio Investing: 25FF and 48Ind

We now turn to evaluating the out-of-sample portfolio performance. We start with portfolio investing by focusing on the two Fama-French portfolio data sets: 25FF and 48IND, a move towards analysis on the individual stock level is given in the next section. In each month t, we construct the MV portfolios using stock returns from past $M = 60$ months (5 years)[4]. We hold such portfolios for 1 month and calculate the portfolio returns for out-of-sample month $t + 1$. We continue this process by adding the return for the next period in the data set and dropping the earliest return from the estimation window.

[4]The choice of the rolling estimation window size, $M = 60$, follows the standard practice in the literature. To save space, we report the results only for $M = 60$. We have also conducted an analysis using a longer estimation window of $M = 120$ and found the results are generally robust. The results are available upon request.

Our primary interest is in the ability of the proposed MV strategy in reducing the out-of-sample portfolio risk. We first construct the time series of out-of-sample returns for the four portfolios: $\mathrm{MV}_{\mathrm{RRW}}$, $\mathrm{MV}_{\mathrm{sample}}$, $\mathrm{MV}_{\mathrm{equal}}$ and $\mathrm{MV}_{\mathrm{EWMA}}$. We then compare out-of-sample return variance to see whether $\mathrm{MV}_{\mathrm{RRW}}$ achieves out-of-sample risk reduction. Last, we test the significance of any difference between $\mathrm{MV}_{\mathrm{RRW}}$ and other alternatives using the stationary bootstrap of Politis and Romano (1994). Panel A of Table 3 reports the monthly out-of-sample risk for each MV portfolio strategy, and Panel B gives the difference test results. From the table, we observe: i). compared to the standard sample MV portfolio, $\mathrm{MV}_{\mathrm{sample}}$, our portfolio ($\mathrm{MV}_{\mathrm{RRW}}$) significantly reduces the portfolio out-of-sample risk, suggesting that the regularization increases the ability of the rolling sample estimates in change detection of the covariance matrix, and thereby improves the portfolio performance by better controlling portfolio risk exposure. For example, the portfolio risk decreases from 18.757($\%^2$) to 13.707($\%^2$) for 25FF, and from 44.991($\%^2$) to 13.157($\%^2$) for 48IND. ii). compared with the naïve $1/N$ strategy ($\mathrm{MV}_{\mathrm{equal}}$), our portfolio ($\mathrm{MV}_{\mathrm{RRW}}$) still achieves lower risk, implying that "volatility timing" favors portfolio out-of-sample risk reduction; iii). compared with the MV portfolio formed using covariance matrix forecasts from the EWMA model, $\mathrm{MV}_{\mathrm{EWMA}}$, our portfolio once again offers smaller risk, e.g. 13.707 v.s. 14.422 in 25FF, and 13.157 v.s. 18.043 in 48IND. This supports the assumption that sparse rather than continuous changes in the covariance matrix does not weaken, but rather strengthens the portfolio out-of-sample risk reduction.

In principle, if the out-of-sample mean return remains the same, the out-of-sample risk reduction would lead to the increase of the out-of-sample Sharpe ratio. Our experiments here validate this statement in the context of the RRW portfolio. Panel A of Table 4 reports Sharpe ratios for the six portfolios, and Panel B of the table gives the different test results. The portfolios $\mathrm{MV}_{\mathrm{RRW}}$ still outperform all the alternatives by retaining the highest Sharpe ratios across almost all the data sets. The Sharpe ratio of portfolio $\mathrm{MV}_{\mathrm{RRW}}$ ranges from 0.219 to 0.371, followed by the portfolio $\mathrm{MV}_{\mathrm{EWMA}}$.

We now turn to investigating the ability of our RRW strategy in controlling portfolio turnover. We calculate the monthly portfolio turnovers as stated in Equation 12 for the six portfolios and report these results in Panel A of Table 5. It is not surprising we observe that the equally weighted $1/N$ portfolio provides the lowest turnover for all the data sets. The naïve diversification requires only a very small amount of trades to maintain the equal weights. On the contrary, the sample portfolio MV_{sample} and the other volatility timing strategy, MV_{EWMA}, always suffer large turnovers, because it requires active trading to adapt with the changing covariance matrix in order to achieve the best risk diversification. Our portfolio (MV_{RRW}) significantly reduces the portfolio turnovers. The favorable performance in portfolio turnover control verifies that the RRW approach offers a more stable estimate for the covariance matrix that significantly reduces the portfolio turnovers and thus the associated transaction costs.

Finally, we assess the economic gain of the RRW strategy by looking at certainty equivalent returns (CER) after transaction cost. Following standard practice in most of the literature (DeMiguel et al. 2009b,a, Goto and Xu 2015), we set the risk aversion coefficient $\gamma = 5$, and the transaction costs are measured by the annualized asset turnover multiplied by 50 basis points per trade. Table 6 reports the CER of the six portfolios after subtracting transaction costs. We observe that while the sample portfolio offers negative CER after transaction cost, our portfolio realizes positive economic gains in all the data sets. Moreover, the gains compare favorably with those of other alternatives. The improvement on economic gain over other alternatives can be attributed to i) the RRW covariance matrix estimator leads to much reduced portfolio turnover, and thus reduces transaction costs; and ii) the RRW estimator substantially reduces the out-of-sample portfolio risk.

C. Individual Asset Investing: CRSP50

In this section we conduct a portfolio analysis operating based on individual assets. We randomly select 100 samples from S&P500 stocks and each sample contains 50 stocks. We then compare our RRW strategy with other alternatives in terms of out-of-sample portfolio performance among the 100 samples. The 100 random samples are useful not only for examining the average performance of our strategy in individual stock investing, but also for investigating the relation between the level of regularization on covariance matrix to achieve temporal stability and the magnitude of out-of-sample portfolio performance improvement.

The average performance across the 100 stock samples is presented in the last line of Table 3 to Table 6. Once again, these results confirm that the individual asset investor also benefits from using RRW to identify changes in the covariance matrix. The advantage of the new strategy reported in portfolio investing is still preserved here.

Then, we report summary statistics of portfolio performance across the 100 samples with an expectation that samples need higher value of regularization parameters should exhibit greater benefit when using our strategy. In Table 7 Panels A to D, we present monthly mean, standard deviation, minimum and maximum values of the estimation results for i) the out-of-sample portfolio risk, ii) the out-of-sample Sharpe ratio, iii) the portfolio turnover, and iv) the CER after transaction cost for each of the four portfolios: MV_{RRW}, $\text{MV}_{\text{sample}}$, MV_{equal}, and MV_{EWMA}.

Panel A of Table 7 shows that the portfolio (MV_{RRW} achieves lower out-of-sample portfolio risk than other alternatives in all 100 runs, which offers compelling evidence for the ability of MV_{RRW} to achieve significant reduction in out-of-sample portfolio risk. We turn to Sharpe ratio in Panel B and observe that MV_{RRW} achieves higher Sharpe ratio in all 100 runs. In terms of portfolio turnover and economic gains, we find that the MV_{RRW} provides lower turnover and higher CER after transaction cost in the majority of runs than any of the other alternative portfolio strategies. Finally, in Panel E of Table 7, we report the relation between the level of regulation on covariance matrix and the improved magnitude of out-of-sample economic gain (measured by the difference of CER after transaction cost achieved by our strategy and the standard rolling sample estimate). We show that when we regress the CER difference on the value of RRW regularization tuning parameter, the slope coefficient is 1.213 (t-statistic = 3.445) which is significantly positive. This suggests that our RRW estimator provides larger economic gains for a panel of sample that has a relatively stable covariance matrix structure and calls for a higher level of temporal stability regularization.

D. *Decomposing the Performance Gain: Estimation Accuracy and the Ability to Time Significant Changes*

While the above findings support the existence of additional economic gains from using our RRW estimator in MV strategies, in this section, we attempt to answer the question: where are the gains generated from? We explain the advantage of our RRW estimator from two perspectives: *estimation accuracy* and the ability o *timing significant changes.*

Firstly, we attribute the better portfolio performance to the improved covariance matrix forecasts. We examine forecasting accuracy of our RRW approach based covariance matrix estimate, compared with the shrunk rolling sample estimate and the forecasts from EWMA model using the following log predictive likelihood:

$$l_t(\Sigma^{-1}) = \ln(\det(\hat{\Sigma}_{t-1}^{-1})) - \tilde{r}_t^{\top}\hat{\Sigma}_{t-1}^{-1}\tilde{r}_t, \tag{14}$$

where T is the total number of out-of-sample testing periods. $\widetilde{r}_t$ denotes the demeaned return vector at time t, and Σ_{t-1} is the covariance matrix estimate for time t but made at time $t-1$. We average $l_t(\Sigma^{-1})$ across the whole out-of-sample testing period, that is, $L(\Sigma^{-1}) = (1/T)\sum_{t=1}^{T} l_t(\Sigma^{-1})$. We calculate out-of-sample log predictive likelihood for each covariance matrix estimator, and then test the significance of the difference between ones from our estimator and the other alternatives. Table 8 reports the testing results. Column 2 of the table shows that our RRW estimate has a significantly higher predictive likelihood than does the shrunk rolling sample estimates in all the data sets, proving that imposing temporal similarity regularization in rolling window approach reduces the covariance matrix predictive errors. Column 3 shows that the RRW estimate outperforms the EWMA based estimate, suggesting that allowing for piecewise constancy is conducive to increase the predictive accuracy for out-of-sample covariance matrix. Overall, these results confirm that the RRW approach improves forecasting accuracy of the covariance matrix, leading to the better out-of-sample portfolio performance.

Next, we demonstrate the advantage of our RRW approach in timing significant changes of the covariance matrix. The top panel of Figure 2 plots the temporal variation of our RRW (the solid line) and the shrunk sample (the dotted line) covariance matrix estimators, measured by the trace of the estimated covariance structures ($\sum_{ii}\hat{\Sigma}_{ii,t}$). We highlight in grey the recession periods according to the NBER business cycle classification. Clearly, the RRW estimator is more stable than the sample counterpart during the calm period, such as the period between 2002-2007, but without loosing capacity to reflect large changes during the recession period, such as the GFC period from 2008-2010.

To further illustrate this point, we split the whole sample into "good" and "bad" economic periods according to the NBER business cycle classification. Then, we compare out-of-sample portfolio turnover and risk exposure of our RRW strategy with those of $1/N$ (MV_{equal}) and the shrunk rolling sample strategy ($\text{MV}_{\text{sample}}$). The results are reported in Table 9. If our strategy has better capacity in timing significant changes, we expect that it allows more aggressive updates in portfolio weights and thus larger increase in portfolio turnovers compared with the $1/N$ strategy during "bad" periods. On the contrary, during "good" periods, it should have more conservative response to the change of covariance matrix, leading much less portfolio turnovers compared with the shrunk rolling sample strategy. Looking at the portfolio turnovers reported in the table, we observe that during "good" periods, the portfolio turnover of MV_{RRW}

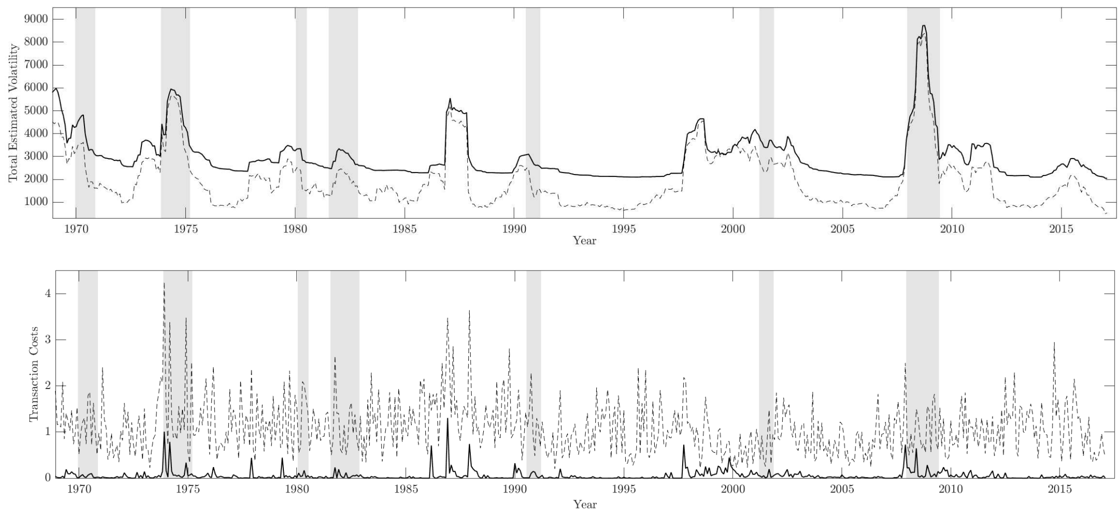

Figure 2: Top: Estimated Total Market Volatility ($\sum_{i=1}^{N} \hat{\Sigma}_{ii}$) for both the standard rolling window (dashed), and regularised rolling window (solid). Bottom: Transaction costs as measured via portfolio weights $\|w_t - w_{t-1}\|_1$. Note: In this case, we set $M = 12$ and $\lambda = 40$ to highlight the changes in variance which can occur in periods of recession.

is quite close to that of the $1/N$ portfolio, but the both are much less than that of $\text{MV}_{\text{sample}}$. For example, for the data of $25FF$, the portfolio turnover of MV_{RRW} is only 0.019 that is close to 0.017 of the $1/N$ portfolio, but much less than 0.047 of the $\text{MV}_{\text{sample}}$. On the other hand, during the bad periods, the portfolio turnover of MV_{RRW} hugely increases to 1.033, but the turnover of the $1/N$ portfolio only slightly increases to 0.018. Comparatively, the sample portfolio $\text{MV}_{\text{sample}}$ has a constantly higher portfolio turnovers, which is also increasing during the bad period. The bottom panel of Figure 2 plots the portfolio turnovers using our RRW (the solid line) and the sample (the dotted line) strategies against the business cycles. Our RRW strategy always offers less portfolio turnovers, but reasonably increases turnovers when the market is in distress.

To conclude the section we examine the portfolio risk and whether the better ability of MV_{RRW} to time changes results in a better control on portfolio risk exposure. We notice that MV_{RRW} always achieves the lowest risk, and the $1/N$ strategy outperforms the sample strategy in "good" periods by achieving lower risk, and vice versa in "bad" periods. In summary, we observe that our RRW covariance estimator has a great ability to highlight significant changes, helping the resulting portfolio to strike a better balance between portfolio turnovers and risk exposure.

VI. Robustness Checks

A. Large scale portfolio

The large scale portfolio allocations remains a challenge for econometricians and practitioners. In this section, we examine whether the economic gain achieved by our RRW approach is robust to large scale portfolios. We employ two large data sets from Ken French's Web site for portfolio investing analysis: 100 value-weighted portfolios of stocks sorted by size and book-to-market (that is, 100FF), and the combination of 100 value-weighted portfolios of stocks sorted by size and book-to-market and the 48 industry value-weighted portfolios (that is, 100FF+48Ind). We also randomly select 100 stocks from S&P stocks to construct the third large data set for individual stock investing analysis. Table 10 reports the annualized CER after transaction cost for our RRW portfolio and other competitors. We can see that the superiority of our RRW portfolio is preserved in these large scale portfolio allocations. The MV_{RRW} achieves the highest CER after transaction cost in all the three large data sets.

B. The Estimation Window Length

To some extent, one may argue that the RRW estimator uses longer samples (from the previous estimation window) than other competing methods used in empirical evaluations of the out-of-sample portfolio performance. This makes it difficult to evaluate the performance gain from the RRW approach over other methods. Does the performance gain come from the particular regularization, or does it comes from the use of a longer sample? We examine a simple way to address this question, that is, we adopt the standard rolling window approach in other competing methods with a fixed longer estimation window, e.g. $M = 120$ months. Table 11 reports the results and two patterns are observed: i) all the competing methods perform better with longer estimation window. This is not surprising as the longer sample of observations provides more historical information which helps achieving more robust estimation; ii) the RRW estimator still outperforms all the other competitors, confirming that through exploiting similarity between two consecutive estimation windows using the temporal similarity regularization, the temporally stable estimator largely reduces "spurious" time variations caused by estimation errors based on the standard rolling window approach. The resulting portfolio, therefore, exhibits more stable out-of-sample performance.

C. *Weekly Return Data*

We use monthly stock returns in the benchmark analysis; here, we evaluate the performance of the different portfolios regarding weekly return data for the five data sets to see whether the results are robust to the return data frequency. We report the CER after transaction cost in Table 12 for the cases with transaction costs of 0 and 50 basis points. We find that our results are generally robust to the use of weekly data. For instance, we find that even with weekly data, our doubly regularized portfolios with $\gamma = 5$ generally outperform the alternatives. When we compare the performance of the portfolios for monthly and weekly return data, we find that the portfolios perform slightly better with monthly than with weekly data. We believe the reason is that the benefit of more frequently adjusting the hedge trades is offset by the higher transaction costs.

D. *Robustness to Investor's Risk Aversion*

Lastly, in our main body of results for CER after transaction cost, we choose that the investor has a risk aversion of $\gamma = 5$. To investigate whether the portfolio performances are robust to this choice, we report the CER after transaction of different portfolios in Table 13 using other values for γ, for example, $\gamma = 2$ and $\gamma = 10$. Generally, as we expected, the CER after transaction costs becomes larger when $\gamma = 2$ and becomes smaller when $\gamma = 10$ in all the portfolio strategies, as the former gives less weight and the latter gives more weight to the out-of-sample portfolio risk. However, the ranking of portfolio performance is not changed at all compared with the main results. Our RRW portfolio still outperform the alternatives.

VII. Conclusion

In this paper, we propose a regularized rolling window approach to estimate the time-varying covariance matrix, which imposes a temporal variation constraint on the standard rolling window based sample estimates. This new method is simple and interpretable, whilst also yielding superior out-of-sample forecasts for the covariance matrix and capable of detecting significant changes in covariance matrix. We demonstrate that in the presence of both structural changes and transaction cost, the resulting portfolio achievs simultanenously low risk exposure and turnover, earning significant economic gains compared to a set of commonly used alternatives. These results support our initial motivation of this study: in the presence of transaction cost, investors can benefit from volatility significant change detection.

Appendix A: Theory and Methodology

A. Proof of Proposition 1

Proof. Let $\hat{S}_t$ be the empirical covariance at t and forecast until $t+1$. Let Θ_{t-1} be the estimate of the precision matrix at the previous timestep. The RRW problem we solve is

$$\hat{\Theta} = \arg\max_{U \succ 0} \left\{ f(U, \hat{S}) \right\} .$$

where

$$f(U, \hat{S}) := \log\det(U) - \mathrm{tr}(\hat{S}_t U) - \lambda \|U - \Theta_{t-1}\|_1 .$$

Using similar arguments to the regular graphical laso case (Banerjee and Ghaoui 2008), we can formulate an equivalent (due to convexity) dual problem. Noting that the ℓ_1 norm can be expressed as

$$\|U\|_1 = \max_{\|V\|_\infty \le 1} \mathrm{tr}(UV) ,$$

we can then consider the ℓ_1 norm applied to the difference of a matrix. The dual problem is thus constructed as

$$\max_U \{f(U)\} \equiv \max_{U \succ 0} \min_{\|V\|_\infty \le \lambda} \left\{ \log\det U - \mathrm{tr}(U, \hat{S}_t + V) + \mathrm{tr}(\Theta_{t-1} V) \right\} .$$

Swapping the max and min, performing the optimisation over U gives us

$$\max_U f(U) = \min_{\|V\|_\infty \le \lambda} \{ -\log\det(S + V) + \mathrm{tr}(\Theta_{t-1} V) - p \} . \tag{15}$$

Writing $W := \hat{S}_t + V$ we obtain the desired result. In general we can replace $\|\cdot\|_\infty$ with $\|\cdot\|_{\mathcal{D}}$ in Eq. 15. The final equivalence in terms of transaction cost is trivial due to the definition of Δ_t and $\hat{\Sigma}_{t;\lambda}$. A simillar analysis of transaction costs is performed for the case with the penalty $\|w - w_t^+\|_2^2$ is given in Hautsch and Voigt (2019), our result is a generalisation of this with the penalty based on $\|x\|_L^2 = x^\top L x$, the Hautsch and Voigt (2019) results relates to adding a diagonal to the empirical covariance.

□

B. An Alternating Directions Method of Multipliers Algorithm for RRW (RRW-ADMM)

Based on the algorithms in Danaher et al. (2013), Gibberd and Nelson (2017), we propose to solve the RRW estimator using a version of the Alternating Direction Method of Multipliers (ADMM) algorithm. The basic intuituion behind this approach is to utilise linear seperability in the objective function, in our case between the likelihood term and the regulariser, in order to decompose the optimisation problem into a series of simplified problems. The basic form of the algorithm can be explained in relation to solving the generic problem $\min_{\{u,v\}}\{f(u)+r(v)\}$ (for convex f, r) subject to the constraint that $u=v$. A standard approach is to instead solve the Lagrangian dual problem

$$\max_{p}\{g(p) := \min_{u,v}\{f(u) + r(b) + \langle p, u-v\rangle\} \ , \tag{16}$$

where p represents a set of Lagrange multipliers. Principally, we see that the Lagrangian penalty, the inner product $\langle p, u-v\rangle$, penalises divergence from the equality requirement $u=v$. As the functions f, r are both convex, strong duality holds, and the maxima $g(p^*)$ for the dual problem is equivalent to the value of the explicitly constrained problem $f(u^*)+r(v^*)$. Without further modification, the minimisation task $\min_{u,v}\{\cdot\}$ in (16) needs to be performed jointly, i.e. we should not expect to be able to sequentially minimise with respect to u and then v. However, with slight modification, through an additional *augmentation* term, one can break down this joint minimisation problem into a sequential one. Specifically, the augmented Lagrangian

$$\mathcal{L}(u,v,p) = f(u) + r(b) + \langle p, u-v\rangle + \frac{\gamma}{2}\|u-v\|_F^2 \ , \tag{17}$$

gives rise to the ADMM algorithm which performs dual ascent through the following iterations:

1. Minimise $u_{k+1} = \arg\min_u \mathcal{L}(u, v_k, p_k)$
2. Minimise $v_{k+1} = \arg\min_v \mathcal{L}(u_{k+1}, v, p_k)$
3. Tightening the Lagrangian constraint, via dual ascent $p_{k+1} = p_k + (u_{k+1} - v_{k+1})$

For further information on the ADMM formulation and convex optimisation the reader is directed to the work of Boyd et al. (2010), and the excellent book by Boyd and Vandenberghe (2004). In the following we detail the specifics required to adapt this algorithm to the RRW estimator. The augmented Lagrangian (specific version of Eq. 17) for the RRW estimator is given as:

$$
\begin{aligned}
\mathcal{L}(\{U,V,P\}) = & -\log\det(U) + \mathrm{tr}(U\hat{S}) + \lambda\|V - \hat{\Theta}_{t-1}\|_1 \\
& \ldots + \frac{\gamma}{2}\left(\|U - V + P\|_F^2 - \|P\|_F^2\right) \ . \qquad (18)
\end{aligned}
$$

The ADMM algorithm proceeds to minimise (18) subject to increasing the constraint imposed by the Lagrange multipliers P.

Step 1: Maintaining Positive Semi-definite Solutions

Initially, we minimise with respect to the primal variable U, and therefore need to solve the equation

$$U_{k+1} := \arg\min_{U \succeq 0}\left\{-\log\det(U) + \mathrm{tr}(U\hat{S}) + \frac{\gamma}{2}\|U - \Gamma\|_F^2\right\} \ ,$$

where $\Gamma := V_k - P_k$. We can interpret this minimisation as pulling the estimate for U_{k+1} towards that given by the difference between the auxilary variable and the dual $(V_k - P_k)$ and note that the prior knowledge imposed by the smoothing (and potentially sparse) regulariser is encoded within V_k.

The gradient of the above gives $-(U)^{-1} + \hat{S} + \gamma U - \gamma\Gamma) = 0$, and therefore

$$(U)^{-1} - \gamma U = \hat{S} - \gamma\Gamma$$

We construct a solution for U by equating the eigen-vectors of the left and right hand-sides. The eigen-values of each side, respectively $\{u_h\}_{h=1}^p$ and $\{s_h\}_{h=1}^p$ obbey the quadratic $u_h^{-1} - \gamma u_h = s_h$. Obtaining the eigenvectors $\{v_h \in \mathbb{R}^p\} := \mathrm{eigenvec}(\hat{S} - 2\gamma\Gamma)$ and corresponding eigenvalues $\{s_h\}$ we can update for u_h by solving the quadratic $u_h = -(2\gamma)^{-1}\{s_h \pm (s_h^2 + 4\gamma)^{1/2}\}$, for $h = 1, \ldots, p$. A positive-semi-definite update for U_{k+1} can now be constructed according to

$$U_{k+1} = \begin{pmatrix} v_1 \\ \\ v_p \end{pmatrix} \begin{pmatrix} u_1 & & \\ & \ddots & \\ & & u_p \end{pmatrix} \begin{pmatrix} v_1 \\ \\ v_p \end{pmatrix}^\top \ .$$

Step 2: Enforcing Prior Knowledge

The second step in the algorithm requires us to incorporate the influence of the regularisers. At this point, there are a variety of updates that may be required dependent on the form of the regulariser, however, in this paper we only have a single smoothing term corresponding to the ℓ_1 penalty on the differences. Minimising $\mathcal{L}(\{U,V,P\})$ with respect to V we obtain the problem

$$V_{k+1} = \arg\min_{V} \left\{ \frac{1}{2}\|\Gamma - V\|_F^2 + \frac{\lambda}{\gamma}\|V - \hat{\Theta}_{t-1}\|_1 \right\} \ , \tag{19}$$

where we utilise the substitution $\Gamma \equiv U_{k+1} + P_k$. Simply writing $V' = V - \hat{\Theta}_{t-1}$, we obtain the equivalent problem $V'^* = \arg\min_{V'}\{(1/2)\|\Gamma' - V'\|_F^2 + (\lambda/\gamma)\|V'\|_1\}$, where $\Gamma' \equiv U_{k+1} + P_k - \hat{\Theta}_{t-1}$ which has a closed form soft-threshold solution

$$V'^* = \mathrm{soft}(\Gamma'; \lambda/\gamma) \equiv \begin{cases} 0 & \text{if } |\Gamma'_{i,j}| \leq \lambda/\gamma \\ \Gamma'_{i,j} - (\lambda/\gamma)\mathrm{sign}(\Gamma'_{i,j}) & \text{if } |\Gamma'_{i,j}| > \lambda/\gamma \end{cases} .$$

The solution for V_{k+1} can then be recovered simply by adding the previous estimate, i.e. $V_{k+1} = V'^* + \hat{\Theta}_{t-1}$. As an aside, we note that this step of the algorithm may easily be altered to enable both sparse and smooth estimation, for instance, as discussed in Section **??**. To simplify the application of such combined priors, in our implementation we utilise the SLEP package (Liu et al. 2009) and *Fused Lasso Signal Approximator* to solve the update, as in Eq. 19.

Step 3: Dual Update and Convergence

The final step in the algorithm is to update the dual variables. With the Lagrangian in the simplified form of (18) this is as simple as updating according to the difference between the primal and auxilary variables, specifically, we set $P_{k+1} = P_k + (U_{k+1} - V_{k+1})$. Repeating the above steps is guranteed to lead to minimisation of the RRW objective due to its convex nature, for further details on such arguments the reader is directed towards Boyd et al. (2010).

One may note that the above algorithm introduces an additional tuning parameter $\gamma > 0$. While this does not affect the eventual solution, i.e. minima found by the algorithm, it can drastically affect the time it takes the algorithm to get there. In practice, and in our experiments, we found reasonable convergence speed with $\gamma = 1$, however, it is still an open problem how to optimally tune this parameter. Finally, it is worth noting that the RRW-ADMM algorithm can be signficantly sped up if it is initiated with a warm start. That is, if we set the initial $\{U_0, V_0, P_0\}$ to be close to the actual solution, we should expect the algorithm to converge much faster. In our work, we harness this property to speed up the parameter search over λ.

C. Estimating Degrees of Freedom and Tuning Regularisers

Given that there is no consensus on how to implement AIC/BIC type tuning for high-dimensional problems (where we may have $M < N$), we study a variety of methods to select tuning parameters and estimate degrees of freedom. As such, the AIC criteria we employ should be taken purely as a heuristic. Consider the set

$$\mathcal{D}_t = \{(i,j) \mid \hat{\Theta}_{t;ij} \neq \hat{\Theta}_{t-1;ij} \; ; \; i \neq j\} \quad ,$$

which represents the support of the difference of inverse-covariance matrices. In the graphical lasso setting, where we apply a sparsity penalty to the inverse covariance matrix entries it is common place to assume the degrees of freedom may be estmated according to the number of non-zero entries in the inverse covariance. Analogously, in our case, as we are fusing against a fixed target, i.e. the previous precision estimate, we instead propose to measure the degrees of freedom in terms of the differencing set, i.e. $\hat{K}_{\text{smooth}} = |\mathcal{D}_t|$. Interestingly, this can be linked, via the GMV construction, to counting variation in the change in portfolio weights. For instance, a large $\hat{K}_{\text{smooth}}$ implies we need to change many of our positions to reach a new portfolio, in this way, one can directly see how our approach should limit portfolio turnover.

Although, our experiments focus on using AIC to select λ we also note there are other ways one might tune these parameters for completeness. In particular, while AIC attempts to construct an in-sample estimate of out-of-sample performance, we may also consider assessing out-of-sample performance directly, equivalently to how one may evaluate the task of covariance prediction. In this case, we can simply assess the negative log-likelihood of the next data-points given our current estimate

$$l_{\text{test}}(\hat{\Theta}_t; x_{t+1}) = -\log\det(\hat{\Theta}_t) + \operatorname{tr}(x_{t+1}x_{t+1}^{\top}\hat{\Theta}_t) \; .$$

Averaging the out-of-sample likelihoods over a training period lets us assess the performance of the RRW estimator in covariance prediction.

References

Banerjee, O., Ghaoui, L. E., 2008. Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data. Journal of Machine Learning Research 9, 485–516.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., 2010. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends. Mach. Learn. 3, 1–122.

Boyd, S., Vandenberghe, L., 2004. Convex Optimization.

Chan, L., Karceski, J., Lakonishok, J., 1999. On portfolio optimization: Forecasting covariances and choosing the risk model. Review of Financial Studies 12, 937–974.

Danaher, P., Wang, P., Witten, D. M., 2013. The joint graphical lasso for inverse covariance estimation across multiple classes. Journal of the Royal Statistical Society: Series B (Statistical Methodology).

DeMiguel, V., Garlappi, L., Uppal, R., 2009a. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. Management Science 55 (5), 798–812.

DeMiguel, V., Garlappi, L., Uppal, R., 2009b. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? Review of Financial Studies 22, 1915–1953.

DeMiguel, V., Olivares-Nadal, A. V., 2018. Technical note-a robust perspective on transaction costs in portfolio optimization. Operations Research, forthcoming.

Engle, R. F., Robert, F., Jekrey, R., 2012. Measuring and modeling execution cost and risk. Journal of Portfolio Management 38 (2), 14–28.

Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility timing. Journal of Finance 56 (1), 329–352.

Fleming, J., Kirby, C., Ostdiek, B., 2003. The economic value of volatility timing using realized volatility. Journal of Financial Economics 67 (3), 473 – 509.

Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9, 432–441.

Gârlranu, N., Pedersen, L., 2013. Dynamic trading with predictable returns and transaction costs. The Journal of Finance LXVIII (6), 2309–2340.

Gârlranu, N., Pedersen, L., 2016. Dynamic portfolio choice with frictions. Journal of Economic Theory 165, 487–516.

Gibberd, A. J., Nelson, J. D. B., 2017. Regularized Estimation of Piecewise Constant Gaussian Graphical Models: The Group-Fused Graphical Lasso. Journal of Computational and Graphical Statistics 26 (3), 623–634.

Goto, S., Xu, Y., 2015. Improving mean variance optimization through sparse hedging restrictions. Journal of Financial and Quantitative Analysis 50, 1415–1441.

Hautsch, N., Voigt, S., 2019. Large-scale portfolio allocation under transaction costs and model uncertainty. Journal of Econometrics.

Jorion, P., 1986. Bayes-stein estimation for portfolio analysis. Journal of Financial and Quantative Analysis 21, 279–292.

Kirby, C., Ostdiek, B., 2012. Its all in the timing: Simple active portfolio strategies that outperform naïve diversification. Journal of Financial and Quantitative Analysis 47, 437–467.

Kourtis, A., Dotsis, G., Markellos, R., 2012. Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix. The Journal of Banking and Finance 36, 2522–2613.

Kyung, M., J. Gill, M. G., Casalla, G., 2010. Penalized regression, standard error, and bayesian lasso. Bayesian Analysis 5, 369–412.

Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of Empirical Finance 10, 603–621.

Liu, J., Ji, S., Ye, J., 2009. SLEP: Sparse Learning with Efficient Projections. Arizona State University. URL `http://www.public.asu.edu/~jye02/Software/SLEP`

Merton, R., 1980. On estimating the expected return on the market: An exploratory investigation. Journal of Financial Economics 8, 323–361.

Moreira, A., Muir, T., 2017. Volatility-managed portfolios. The Journal of Finance 72 (4), 1611–1644.

Moreira, A., Muir, T., 2019. Should long-term investors time volatility? Journal of Financial Economics 131 (3), 507–527.

Politis, D., Romano, J., 1994. The stationary bootstrap. Journal of the American Statistical Association, 1303–1313.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K., 2005. Sparsity and smoothness via the fused lasso. Journal of Royal Statistics Society B 67 (1), 91–108.

Tu, J., Zhou, G., 2010. Incorporating economic objectives into bayesian priors: Portfolio choice under parameter uncertainty. Journal of Financial and Quantitative Analysis 45 (4), 959–986.

Wang, H., dec 2012. Bayesian Graphical Lasso Models and Efficient Posterior Computation. Bayesian Analysis 7 (4), 867–886.

Zakamulin, V., 2015. A test of covariance-matrix forecasting methods. The Journal of Portfolio Management 41 (3), 97–108.

Table 1: **Motivating Example.** Mean and standard-deviation of the covariance estimates from both the first and second periods (i.e. we take means and standard deviations over the stationary periods). The last column gives the average percentage of time-points at which the estimates change from the previous time-step. Statistics are taken over 300 trials. In this case, the true values are $\Sigma_{12} = 0.1$ and $\Sigma_{12} = 0.2$ for the first and second periods respectively, $T = 2000, \tau = 1000$.

Method	μ_1	μ_2	σ_1	σ_2	$\%_{\text{change}}$
Empirical	0.100	0.195	0.046	0.050	100
RRW ($\lambda = 0.05$)	0.100	0.187	0.036	0.042	25
RRW ($\lambda = 0.1$)	0.102	0.166	0.029	0.036	1

Table 2: **Data Description.** This table lists the data sets used in our empirical analysis. Column 2 provides the abbreviation used to refer to the testing portfolios. Column 3 gives more detailed descriptions about the data sets. Column 4 reports the number of stocks in each data set, and Column 5 reports the length of sample period. Column 6 and 7 present the training and testing period in out-of-sample analysis. Column 8 and 9 give the optimal value we used for the regularization parameter (λ) in RRW approach and the average percentage of non-changed off-diagonal elements between consecutive covariance matrix estimates throughout the whole sample period.

Data Set	Abbreviation	Description	N	T	Training Period	Testing Period	λ	Percentage of unchanged matrix off-diagonal elements
1	25FF	25 size and BM portfolios	25	553	Jan. 1967-Dec. 1972	Jan.1973-Dec.2017	0.5	23.55%
2	48IND	48 industry portfolios	48	553	Jan. 1967-Dec. 1972	Jan.1973-Dec.2017	0.5	30.23%
3	Individuals	100 stocks from SP500	100	553	Jan. 1992-Dec. 1997	Jan.1998-Dec.2017	0.4	25.21%

Table 3: **Out-of-Sample Portfolio Risk (monthly).** For each data set, this table reports the monthly out-of-sample return variances for the following four portfolios: portfolio using our RRW estimator (denoted as MV_{RRW}); portfolio using rolling window based shrunk sample estimate of the covariance matrix (denoted as MV_{sample}); equally weighted portfolio; and portfolio using EWMA based estimates of covariance matrix (denoted as MV_{EWMA}). Panel A tabulates the point estimates of the out-of-sample monthly return variances measured in $\%^2$, and Panel B tabulates the mean differences in out-of-sample return variances. We test the null of no differences using the two-sided bootstrap intervals. *, **, and *** indicate significant differences at the 10%, 5% and 1% levels, respectively.

Data Set	MV_{RRW}	MV_{equal}	MV_{sample}	MV_{EWMA}
Panel A: Return Variance ($\%^2$)				
25FF	13.707	25.589	27.629	14.422
48IND	13.157	24.725	27.833	18.043
Individuals	15.008	22.453	17.674	17.112
Panel B: Difference Test				
25FF		-5.050***	-13.922***	-0.715
48IND		-31.834***	-14.676***	-4.886***
Individuals		-4.223***	-11.257***	-2.104***

Table 4: **Out-of-Sample Sharpe Ratio (monthly).** For each data set, this table reports the monthly out-of-sample sharp ratio for the following four portfolios: portfolio using our RRW estimator (denoted as MV_{RRW}); portfolio using rolling window based shrunk sample estimate of the covariance matrix (denoted as MV_{sample}); equally weighted portfolio; and portfolio using EWMA based estimates of the covariance matrix (denoted as MV_{EWMA}). Panel A tabulates the monthly Sharpe ratio in the out-of-sample period. Panel B reports the difference in sharp ratio between the best portfolio and other alternatives. Using the portfolio (MV_{RRW}) as a benchmark, we test the null of no difference between the benchmark and other competitors indirectly by constructing the two-sided bootstrap intervals. *, **, and *** indicate significant difference at the 10%, 5% and 1% levels, respectively.

Data Set	MV_{RRW}	MV_{equal}	MV_{sample}	MV_{EWMA}
Panel A: Monthly SRs				
25FF	0.371	0.221	0.209	0.374
48IND	0.251	0.109	0.178	0.214
Individuals	0.267	0.152	0.169	0.173
Panel B: Differences in SRs				
25FF		0.149***	0.162***	0.004
48IND		0.142***	0.073*	0.038*
Individuals		0.115***	0.098*	0.094*

Table 5: **Out-of-sample Portfolio Turnover.** For each data set, this table reports the portfolio turnover for the following four portfolios: portfolio using our RRW estimator (denoted as MV_{RRW}); portfolio using rolling window based shrunk sample estimate of the covariance matrix (denoted as MV_{sample}); equally weighted portfolio; and portfolio using EWMA based estimates of the covariance matrix (denoted as MV_{EWMA}). Calculation of the portfolio turnover follows the equation 12.

Data set	MV_{RRW}	MV_{equal}	MV_{sample}	MV_{EWMA}
25FF	0.020	0.0172	0.291	0.320
48IND	0.025	0.029	0.346	0.461
Individuals	0.053	0.041	0.213	0.345

Table 6: **Out-of-sample CER after Transaction Cost (annual %).** For each data set, this table reports the values of the CER after transaction cost for the following four portfolios: portfolio using our RRW estimator (denoted as MV_{RRW}); portfolio using rolling window based shrunk sample estimate of the covariance matrix (denoted as MV_{sample}); equally weighted portfolio; and portfolio using EWA based estimates of the covariance matrix (denoted as MV_{EWMA}). The transaction cost of each is calculated as 50 basis points times monthly turnover times 12 (to annualize).

Data Set	MV_{RRW}	MV_{equal}	MV_{sample}	MV_{EWMA}
25FF	10.238	4.924	3.927	8.458
48IND	5.146	3.687	3.754	4.167
Individuals	4.834	3.593	3.693	3.732

Table 7: **Detailed Descriptions on Portfolio Performance for 50 individual stocks.** This table reports summary statistics of out-of-sample portfolio risk, sharpe ratio, turnovers and CER after transaction cost for the following four portfolio strategies: portfolio using our RRW estimator (denoted as $\mathrm{MV}_{\mathrm{RRW}}$); portfolio using rolling window based shrunk sample estimate of the covariance matrix (denoted as $\mathrm{MV}_{\mathrm{sample}}$); equally weighted portfolio; and portfolio using EWMA based estimates of covariance matrix (denoted as $\mathrm{MV}_{\mathrm{EWMA}}$), based on 100 random samples from the S&P500 stocks (each consists of 50 individual stocks).

	$\mathrm{MV}_{\mathrm{RRW}}$	$\mathrm{MV}_{\mathrm{equal}}$	$\mathrm{MV}_{\mathrm{sample}}$	$\mathrm{MV}_{\mathrm{EWMA}}$
Panel A: out-of-sample portfolio risk(monthly)				
mean	15.008	19.218	26.265	18.093
std	2.010	2.33	2.67	7.05
max	15.226	25.224	31.789	21.516
min	6.128	6.327	10.224	7.276
Panel B: portfolio Sharpe ratio				
mean	0.267	0.151	0.169	0.215
std	0.028	0.036	0.039	0.064
max	0.301	0.286	0.251	0.265
min	0.062	0.055	0.048	0.047
Panel C: portfolio turnover				
mean	0.053	0.226	0.213	0.177
std	0.003	0.013	0.014	0.352
max	0.082	0.337	0.421	0.312
min	0.022	0.035	0.052	0.251
Panel D: CER after transaction cost				
mean	3.834	3.525	3.693	3.184
std	0.035	0.049	0.055	0.156
max	4.987	3.982	4.164	3.967
min	0.998	0.598	0.616	0.879
Panel E: the relation between regularization and the improved economic gains				
Intercept	-0.135			
regularization	1.213(3.445)			

Table 8: **Out-of-sample covariance matrix prediction.** This table reports differences in log predictive likelihood (see equation 14) between different covariance matrix estimators. Column 2-3, denoted as $L_{\mathrm{RRW}} - L_{\mathrm{sample}}$ and $L_{\mathrm{RRW}} - L_{\mathrm{EWMA}}$, give the difference between our RRW estimator and the shrunk sample estimator of Ledoit and Wolf (2003), and the EWMA model implied estimate, respectively. We test the null hypothesis of no difference between our regularized estimator and other competitors using Politis and Romano (1994) stationary bootstrap to construct two-sided bootstrap intervals. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Data Set	$L_{\mathrm{RRW}} - L_{\mathrm{sample}}$	$L_{\mathrm{RRW}} - L_{\mathrm{EWMA}}$
25FF	0.987**	3.105***
48IND	1.053**	4.121***
Individuals	1.211***	3.137***

Table 9: **Portfolio performance during "good" and "bad" periods.** This table reports the portfolio turnover and CER after transaction cost for the following four portfolios: portfolio using our RRW estimator (denoted as MV_{RRW}); portfolio using rolling window based shrunk sample estimate of the covariance matrix (denoted as MV_{sample}); equally weighted portfolio; and portfolio using EWMA based estimates of covariance matrix (denoted as MV_{EWMA}). The "good" and "bad" periods are based on NBER business cycle classifications.

Date set	MV_{RRW}	MV_{sample}	MV_{equal}
Panel A: Good period			
Portfolio turnover			
25FF	0.019	0.047	0.017
48IND	0.029	0.053	0.023
Individuals	0.041	0.064	0.032
Portfolio risk			
25FF	12.543	17.112	14.121
48IND	12.897	20.206	15.654
Individuals	12.256	17.564	14.589
Panel B: Bad period			
Portfolio turnover			
25FF	1.033	2.076	0.0184
48IND	2.039	4.105	0.032
Individuals	2.567	5.453	0.058
Portfolio risk			
25FF	15.128	20.564	27.096
48IND	14.896	20.656	26.098
Individuals	16.275	21.290	25.027

Table 10: **Robustness check using large scale portfolios.** For each data set, this table reports the values of the CER after transaction cost for the following four portfolios: portfolio using our RRW estimator (denoted as MV_{RRW}); portfolio using rolling window based shrunk sample estimate of the covariance matrix (denoted as MV_{sample}); equally weighted portfolio; and portfolio using EWMA based estimates of covariance matrix (denoted as MV_{EWMA}).

Data Set	MV_{RRW}	MV_{equal}	MV_{sample}	MV_{EWMA}
100FF	4.352	-3.381	-6.093	-25.667
100FF+48IND	5.032	-3.347	-7.298	-14.490
Individuals (100 S&P stocks)	3.834	3.593	3.693	1.532

Table 11: **Robustness check using longer estimation window for competing methods.** For each data set, this table reports the values of the CER after transaction cost for the following four portfolios: portfolio using our RRW estimator (denoted as MV_{RRW}); portfolio using rolling window based shrunk sample estimate of the covariance matrix (denoted as MV_{sample}); equally weighted portfolio; and portfolio using EWA based estimates of covariance matrix (denoted as MV_{EWMA}). The RRW estimator is formed with rolling window of $M = 60$, and other covariance matrix estimators are formed with rolling window of $M = 120$. The transaction cost of each is calculated as 50 basis points times monthly turnover times 12 (to annualize).

Data Set	MV_{RRW}	MV_{equal}	MV_{sample}	MV_{EWMA}
25FF	10.238	7.519	6.942	10.114
48IND	4.146	3.962	3.990	1.408
Individuals	3.834	4.887	4.175	2.717

Table 12: **Robustness test using weekly returns.** For each data set, this table reports the values of the CER after transaction cost for the following four portfolios: portfolio using our RRW estimator (denoted as $\mathrm{MV}_{\mathrm{RRW}}$); portfolio using rolling window based shrunk sample estimate of the covariance matrix (denoted as $\mathrm{MV}_{\mathrm{sample}}$); equally weighted portfolio; and portfolio using EWMA based estimates of covariance matrix (denoted as $\mathrm{MV}_{\mathrm{EWMA}}$). The transaction cost of each is calculated as 50 basis points times weekly turnover times 52 (to annualize).

Data Set	$\mathrm{MV}_{\mathrm{RRW}}$	$\mathrm{MV}_{\mathrm{equal}}$	$\mathrm{MV}_{\mathrm{sample}}$	$\mathrm{MV}_{\mathrm{EWMA}}$
25FF	9.765	3.996	2.978	8.123
48IND	3.882	3.235	3.087	0.765
Individuals	2.365	2.106	1.987	1.032

Table 13: **Robustness test using different investor risk aversion.** For each data set, this table reports the CER after transaction using different risk aversion parameters for the following four portfolios: portfolio using our RRW estimator (denoted as $\mathrm{MV}_{\mathrm{RRW}}$); portfolio using rolling window based shrunk sample estimate of the covariance matrix (denoted as $\mathrm{MV}_{\mathrm{sample}}$); equally weighted portfolio; and portfolio using EWMA based estimates of covariance matrix (denoted as $\mathrm{MV}_{\mathrm{EWMA}}$). The transaction cost of each is calculated as 20 basis points in Panel A, and 100 basis points in Panel B, times monthly turnover times 12 (to annualize).

Data Set	$\mathrm{MV}_{\mathrm{RRW}}$	$\mathrm{MV}_{\mathrm{equal}}$	$\mathrm{MV}_{\mathrm{sample}}$	$\mathrm{MV}_{\mathrm{EWMA}}$
Panel A: $\gamma = 2$				
25FF	13.129	7.114	5.149	12.156
48IND	6.896	5.709	4.132	2.328
Individuals	4.142	4.129	4.106	2.654
Panel A: $\gamma = 10$				
25FF	9.453	4.128	3.124	9.356
48IND	3.227	3.220	2.187	0.987
Individuals	3.125	2.967	2.774	1.438

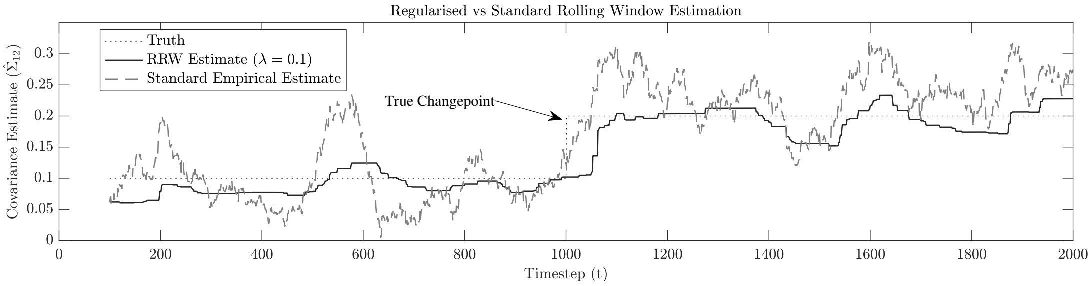

Figure 1: Estimates from a single trial of a synthetic experiment for assessing covariance estimation, in this case we have $\Sigma_{12} = 0.1$ for $t \leq \tau = 1000$ and $\Sigma_{12} = 0.2$ for $t > \tau = 1000$.

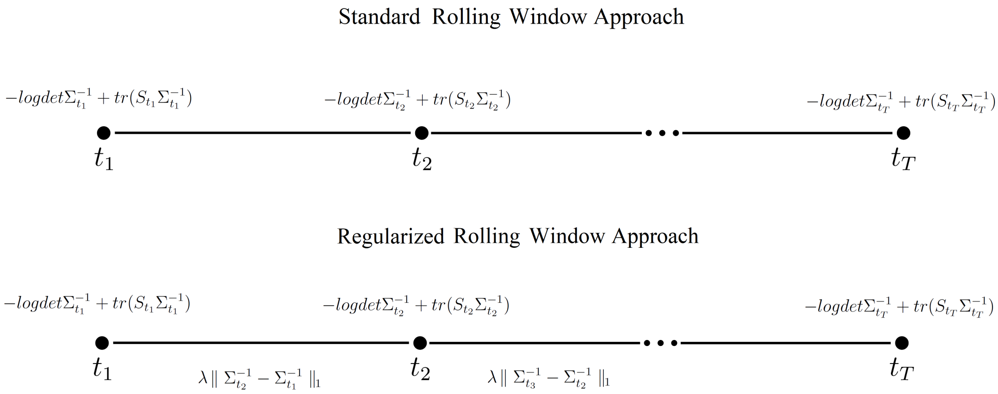

Figure 2: **A graphical comparison between the standard and regularized "rolling window" approach.** The top panel shows the estimation procedure of standard rolling window approach and the bottom panel shows the estimation procedure of the regularized rolling window approach.

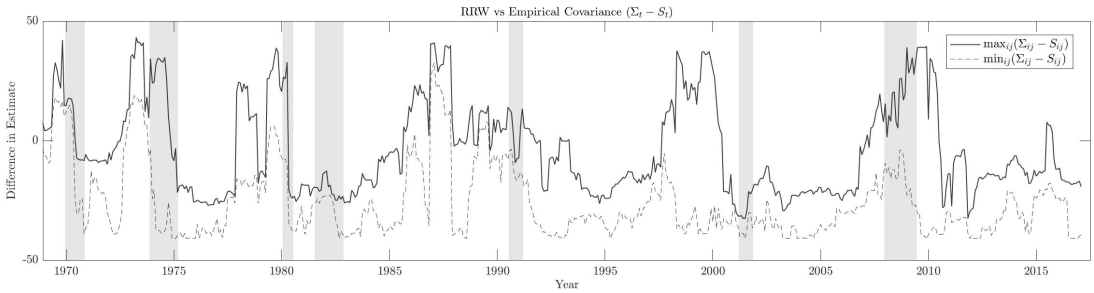

Figure 3: Plot of how $\max_{ij}(\Sigma_{RRW,t} - S_t)$ and $\min_{ij}(\Sigma_{RRW,t} - S_t)$ vary as a function of time. According to Eq. 6 the RRW estimator should always maintain $\max_{ij}(|[\hat{\Sigma}_{RRW,t} - \hat{S}_t]_{ij}|) \leq \lambda$. Note how the difference tends to increase before or around recession period, these are periods where jumps in the portfolio position (and estimated covariance) are likely to occur, see Figs. 4 for comparison.

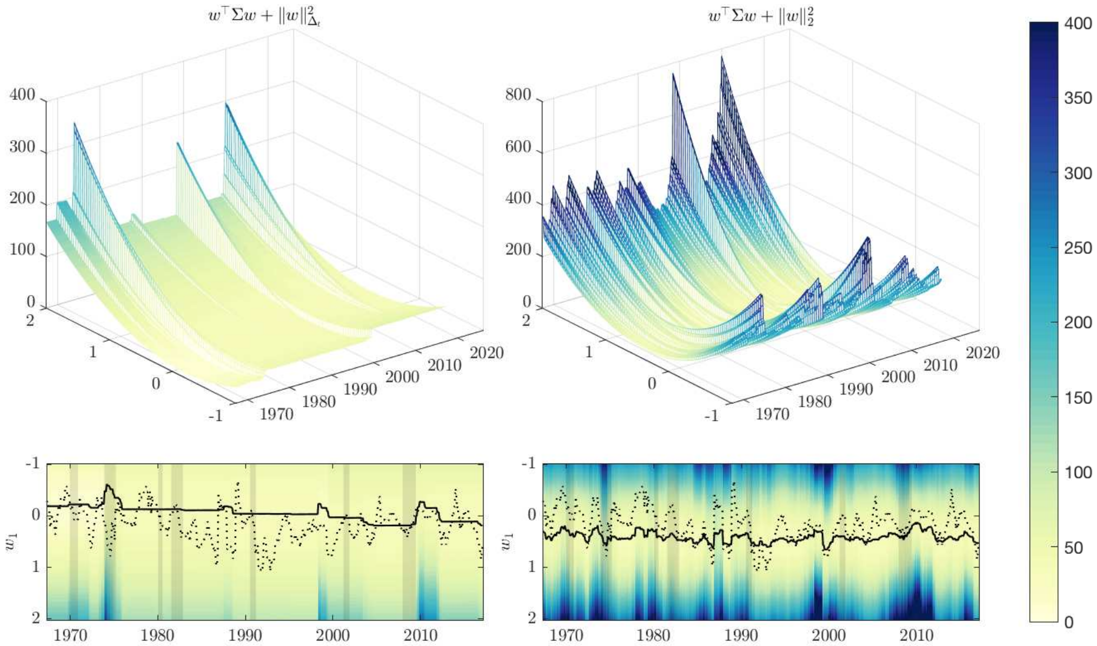

Figure 4: Surface plots of the portfolio optimization objective function over business cycles. The optimum portfolio lies at the minimiser of these objective functions and the corresponding portfolio allocation solution for first asset (that is, w_1) is indicated in the lower panels by the solid line v.s w_1 produced by minimizing $w'\Sigma w$ without transaction cost (the dashed line). We set estimation window length $M = 12$ and $\lambda = 40$ to allow comparison with Figs. 3. The grey overlaid bands (in the lower panes) denote recession periods. Note: we here use the notation $\|w\|^2_{\Delta_t} := (w - w_t^+)^\top \Delta_t (w - w_t^+)$.