economies

MDPI

Article

Maximum Trimmed Likelihood Estimation for Discrete Multivariate Vasicek Processes

Thomas M. Fullerton, Jr. [1,*], Michael Pokojovy [2], Andrews T. Anum [3] and Ebenezer Nkum [4]

[1] Department of Economics and Finance, The University of Texas at El Paso, El Paso, TX 79968, USA
[2] Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529, USA; mpokojovy@odu.edu
[3] Department of Mathematical Sciences, The University of Memphis, Memphis, TN 38152, USA; atanum@memphis.edu
[4] Cigna Healthcare, Nashville, TN 37228, USA; enkum@miners.utep.edu
* Correspondence: tomf@utep.edu

Abstract: The multivariate Vasicek model is commonly used to capture mean-reverting dynamics typical for short rates, asset price stochastic log-volatilities, etc. Reparametrizing the discretized problem as a VAR(1) model, the parameters are oftentimes estimated using the multivariate least squares (MLS) method, which can be susceptible to outliers. To account for potential model violations, a maximum trimmed likelihood estimation (MTLE) approach is utilized to derive a system of nonlinear estimating equations, and an iterative procedure is developed to solve the latter. In addition to robustness, our new technique allows for reliable recovery of the long-term mean, unlike existing methodologies. A set of simulation studies across multiple dimensions, sample sizes and robustness configurations are performed. MTLE outcomes are compared to those of multivariate least trimmed squares (MLTS), MLE and MLS. Empirical results suggest that MTLE not only maintains good relative efficiency for uncontaminated data but significantly improves overall estimation quality in the presence of data irregularities. Additionally, real data examples containing daily log-volatilities of six common assets (commodities and currencies) and US/Euro short rates are also analyzed. The results indicate that MTLE provides an attractive instrument for interest rate forecasting, stochastic volatility modeling, risk management and other applications requiring statistical robustness in complex economic and financial environments.

Keywords: econometric modeling; autoregressive models; times series analysis; maximum trimmed likelihood estimation; statistical robustness; outliers

check for updates

Academic Editor: Tapas Mishra

Received: 14 January 2025
Revised: 18 February 2025
Accepted: 20 February 2025
Published: 6 March 2025

Citation: Fullerton, T. M., Jr., Pokojovy, M., Anum, A. T., & Nkum, E. (2025). Maximum Trimmed Likelihood Estimation for Discrete Multivariate Vasicek Processes. *Economies*, *13*(3), 68. https://doi.org/10.3390/economies13030068

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Policymakers, central banks, financial institutions, etc., rely on econometric modeling to guide decision making. By utilizing statistical techniques and economic theory, many econometric models have been developed to understand, for example, how stock market performance relates to macroeconomic indicators and to quantify the relationships that shape market behavior and predict future economic outcomes. Many macroeconomic and financial processes are empirically observed to follow the multivariate Vasicek model (Vasicek, 1977). This model is commonly applied to short-term interest rates, log-volatilities and other macroeconomic variables that exhibit tendencies to revert toward a long-term average over time. By incorporating multiple variables into a unified framework, the multivariate Vasicek model provides a powerful tool for analyzing the co-movements of economic indicators (Campbell et al., 1996; Egorov et al., 2011; Lütkepohl, 2005; Sims, 1980).

The multivariate Vasicek model is characterized by three fundamental parameters: the long-term mean vector, the mean-reversion speed matrix and the squared volatility matrix. The long-term mean vector represents the average levels around which the vectorial process oscillates, and is expected to revert to, in the long run under appropriate conditions. The mean-reversion speed matrix governs the speed at which the vectorial process moves toward the long-term mean; larger absolute eigenvalues indicate higher speed and vice versa. The covariance matrix, on the other hand, represents the squared volatility in the fluctuations around the long-term mean. While all model parameters play an important role in statistical inference and economic decision making, financial analysts and economists are particularly interested in estimating the long-term mean for various purposes, including, but not limited to, risk management and hedging (Fabozzi & Mann, 2011), forecasting and planning, etc. The long-term mean is a crucial indicator in risk management and hedging, as it provides a stable reference point to assess expected returns, detect new trends and optimize strategies to mitigate potential losses over time. With regard to planning, the long-term mean provides a reliable benchmark to set realistic goals, make informed decisions and assess potential risks over extended periods. At the same time, estimating the long-term mean can be particularly challenging due to identifiability issues.

For multivariate continuous-time models, when the data are only observed over a discrete time grid, the parameters can be estimated using numerous parameter estimation approaches, including Bayesian methods (Eraker, 2001), multivariate least squares (MLS) and methods based on moments (Merton, 1980). Moreover, simulation-based techniques (Gallant & Tauchen, 1996; Gourieroux et al., 1993) as well as nonparametric techniques (Aït-Sahalia, 1995; Stanton, 1997) also exist. Nonetheless, to obtain accurate estimates of the parameters, simulation-based methods typically require many replications and higher computational time, especially in higher dimensions. Prior knowledge about the parameters, which are usually not available, is additionally expected for Bayesian approaches. Moment-based estimators are less applicable in practice since they heavily rely on theoretical assumptions and are susceptible to outliers and heavy tails.

In statistical literature, the discrete Vasicek process is typically reparametrized as a VAR(1) model and calibrated using the MLS method, whose estimation power can be significantly compromised in the presence of outliers (Chang & Shi, 2024; Hamilton, 2020). Financial time series may be prone to volatility shocks or unusual market events (Brockwell & Davis, 2002), leading to the introduction of anomalous values in the data (Ji et al., 2020; Pokojovy & Anum, 2022; Spelta et al., 2023). These values may emanate from a shift in the innovation process, affecting current observations and subsequent ones, or may come from observation(s) whose own value(s) are contaminated (Chang & Shi, 2024). The presence of outliers in a dataset can cause many estimators to produce biased or inefficient parameter estimates. To account for outliers in multivariate time series, MM-estimation (Kudraszow & Maronna, 2011; Yohai, 1987) and multivariate least trimmed squares (MLTS) (Agulló et al., 2008; Croux & Joossens, 2008) are proposed in the literature as robust alternatives to the MLS approach. A somewhat lower efficiency of the MLTS estimator motivated the development of reweighted multivariate least trimmed squares (RMLTS) (Croux & Joossens, 2008). MM-estimators, like most iterative techniques, require a choice of initial estimates or "warmstarts". Improper specification of this input can affect the robustness and affine equivariance of the estimator. Additionally, when the mean-reversion speed matrix is ill-conditioned, the MLTS approach fails to estimate the long-term mean, while the recently proposed modified MLE (Pokojovy et al., 2024) approach can be unduly influenced by outliers and other model violations. Furthermore, Bayesian methods (Eraker, 2008) and simulation-based techniques (Gallant & Tauchen, 2010) have been applied in econometric

modeling in order to mitigate estimation biases introduced by non-Gaussian shocks and market anomalies.

In this paper, we adopt a maximum trimmed likelihood estimation (MTLE) approach instead of MM-estimation or Bayesian alternatives. While MM-estimators are appealing as they offer a balance between robustness and efficiency, they tend to be more computationally expensive as they typically involve costly floating-point operations arising from computing transcendental functions. Also, they are vulnerable to numerical instability in high-dimensional or near-nonstationary settings (Boudt et al., 2020). This limitation is particularly acute in financial time series with clustered outliers (e.g., during market crises). Bayesian procedures tend to involve even higher computational costs. We therefore used the maximum trimmed likelihood estimation (MTLE) approach to estimate the parameters while accounting for possible outliers and put forth a system of nonlinear estimating equations with an iterative procedure to solve the system. Not only does our formulation account for outliers, but it also provides regularization through the utilization of the Moore–Penrose pseudoinverse when the mean-reversion speed matrix is ill-conditioned. While being on par with the state-of-the-art MLTS method at estimating the mean-reversion speed matrix and the squared volatility matrix, our new estimator is also able to effectively estimate the long-term mean of the Vasicek process, especially over large time horizons.

In addition to methodological developments that underpin the proposed MTLE estimator, we also illustrate how this new econometric instrument can be used to facilitate robust analysis of multivariate time-series data. Focusing on short-rate and stochastic volatility modeling as two common application examples, we have evaluated the discrete Vasicek model for a wide range of synthetic datasets in Section 3.1 corresponding to multiple hypothetical outlier configurations and two-real world datasets in Section 3.2, respectively. In the context of short-rate modeling and forecasting, our findings are consistent with those of Pokojovy et al. (2024) and clearly demonstrate that our MTLE not only can serve as a robust and efficacious alternative to MLE for general model calibration and forecasting purposes but significantly outperforms MLTS as its sole robust competitor at estimating the long-term mean. As reported by Chang and Shi (2024), traditional estimators of VAR models may significantly overestimate some parameters. Conversely, robust VAR models deliver more reliable results. In the context of stochastic volatility modeling, our simulation study and real-world example confirm this conclusion while additionally providing compelling evidence of MTLE's superiority over all competitors considered in estimating the long-term mean from contaminated data. In conclusion, irrespective of application domain, our MTLE methodology offers an attractive econometric instrument for robust and efficient estimation of key parameters necessary for statistical and econometric inference, forecasting and simulating future dynamics, statistical process monitoring and change point detection, risk management, development of investment strategies, etc.

Our results align with and extend prior work on robust estimation for multivariate time series, particularly in the context of short-rate modeling. Traditional robust methods, such as multivariate least trimmed squares (MLTS) (Agulló et al., 2008; Croux & Joossens, 2008), have demonstrated success in estimating parameters of VAR(1) models under contamination. However, these methods were not explicitly designed for the discrete multivariate Vasicek model, where the long-term mean plays a critical role in financial applications such as risk management and portfolio optimization (Egorov et al., 2011; Vasicek, 1977). In line with Chang and Shi (2024), who highlighted the limitations of classical VAR estimators in contaminated settings, our simulations confirm that MLTS struggles to reliably estimate long-term mean when the mean-reversion matrix is ill-conditioned or the calibration horizon is short. This deficiency is particularly pronounced in empirical environments, where small samples and low-frequency data worsen numerical instability

(Pokojovy et al., 2024). Our MTLE methodology addresses this gap by integrating MLTS's robustness with a regularized fixed-point iteration. The improvement is consistent with the findings in robust portfolio optimization (Korn & Koziol, 2006), where regularization enhances stability in ill-conditioned systems. Results from our analysis of US/Euro short rates extend the work of Pokojovy et al. (2024), who demonstrated the non-robust MLE's efficacy in uncontaminated data. Importantly, to the best of our knowledge, our study is the first to systematically validate a robust estimator's ability to recover long-term mean in finite samples—a crucial advancement for applications like yield curve modeling and stress testing, where accurate long-term mean estimates are paramount (Fabozzi & Mann, 2011). In summary, while MLTS remains a gold standard for robust VAR estimation, our MTLE bridges a critical gap in Vasicek-specific applications, offering a tailored solution for short-rate models plagued by contamination and numerical challenges due to ill-conditioned matrices. This advancement aligns with the broader tradition in financial econometrics that prioritizes robustness without sacrificing interpretability (Aït-Sahalia, 1995; Stanton, 1997).

The remainder of this paper is organized as follows. In Section 2, we introduce the multivariate Vasicek model, give a review of VAR processes and discuss MLTS estimation and present a discrete-continuous optimization problem underlying our new MTLE estimator along with a system of nonlinear estimating equations and an iterative procedure for solving the latter. Section 3.1 protocols the results of an extensive simulation study performed in this paper to benchmark and compare the performance of MTLE with MLTS, MLE and MLS at estimating multivariate Vasicek parameters. Furthermore, we analyze two real-world datasets and discuss the results in Section 3.2. Finally, Section 4 summarizes the paper and provides concluding remarks.

2. Materials and Methods

The multivariate Vasicek model (Platen & Rendek, 2009) is given by the system of stochastic differential equations (SDEs)

$$\mathrm{d}\boldsymbol{R}_t = \boldsymbol{A}(\boldsymbol{R}^* - \boldsymbol{R}_t)\,\mathrm{d}t + \boldsymbol{\Sigma}^{1/2}\,\mathrm{d}\boldsymbol{W}_t \text{ for } t \in [0,T], \quad \boldsymbol{R}_0 = \boldsymbol{R}^0, \tag{1}$$

for a p-variate Markovian diffusion process $(\boldsymbol{R}_t)_{t\geq 0}$, where $\boldsymbol{R}^* \in \mathbb{R}^p$ denotes the long-term mean vector, $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ stands for the mean-reversion speed matrix and $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ is the positive definite squared volatility matrix, while $(\boldsymbol{W}_t)_{t\geq 0}$ is a p-variate standard Wiener process and the root matrix is defined via the spectral theorem, i.e.,

$$\boldsymbol{\Sigma}^{1/2} = \sum_{k=1}^{p} \lambda_k^{1/2} \boldsymbol{e}_k \boldsymbol{e}_k'$$

with λ_k and $\boldsymbol{e}_k$, $k = 1,\ldots,p$, denoting the k-th eigenvalue and eigenvector of $\boldsymbol{\Sigma}$, respectively.

When studying low-frequency processes, e.g., daily short rates, Pokojovy et al. (2024) consider a discrete version of Equation (1) obtained by applying the explicit Euler–Maruyama discretization scheme on an equispaced time grid with a constant time step $\Delta t = \frac{T}{n}$. This results in the difference equation

$$\tfrac{1}{\Delta t}(\boldsymbol{R}_{t_{j+1}} - \boldsymbol{R}_{t_j}) = \boldsymbol{A}\big(\boldsymbol{R}^* - \boldsymbol{R}_{t_j}\big) + \tfrac{1}{\Delta t}\boldsymbol{\Sigma}^{1/2}(\boldsymbol{W}_{t_{j+1}} - \boldsymbol{W}_{t_j}), \quad j = 0,\ldots,n-1 \tag{2}$$

which can equivalently be expressed as a vector autoregressive model

$$\boldsymbol{R}_{t_{j+1}} = (\Delta t)\boldsymbol{A}\boldsymbol{R}^* + \big(\boldsymbol{I}_{p\times p} - (\Delta t)\boldsymbol{A}\big)\boldsymbol{R}_{t_j} + \boldsymbol{\varepsilon}_{t_j} \text{ for } j = 0,\ldots,n-1 \tag{3}$$

with innovations $\boldsymbol{\varepsilon}_{t_j} := \boldsymbol{\Sigma}^{1/2}(\boldsymbol{W}_{t_{j+1}} - \boldsymbol{W}_{t_j})$ satisfying

$$\boldsymbol{\varepsilon}_{t_j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_p, (\Delta t)\boldsymbol{\Sigma}). \tag{4}$$

See Section 2.1 for details.

From a formal standpoint, model violations for Equations (3) and (4) can be described as a Huber-type contamination

$$\begin{aligned} \boldsymbol{\varepsilon}_{t_j} | \delta_{t_j}, \boldsymbol{\xi}_{t_j} &\sim (1 - \delta_{t_j})\boldsymbol{\xi}_{t_j} + \delta_{t_j} G_{t_j} \\ \delta_{t_j} \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\varepsilon), &\quad \boldsymbol{\xi}_{t_j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_p, (\Delta t)\boldsymbol{\Sigma}) \end{aligned} \tag{5}$$

where $\varepsilon \in [0, \frac{1}{2})$ is the population-level fraction of outliers and $(G_{t_j})_{j=0,\ldots,n-1}$ is some unknown (potentially nonstationary and autocorrelated) contamination process.

Turning back to the uncontaminated case, using mathematical induction and invoking the law of large numbers, the following well-known large-time asymptotics result holds. It is interesting to observe that the limit in Equation (7) explains why $\boldsymbol{R}^*$ is referred to as the long-term mean.

Theorem 1 (Large-time asymptotics)**.** *The discrete Vasicek process is explicitly given as*

$$\boldsymbol{R}_{t_j} = \big(\boldsymbol{I}_{p\times p} - (\Delta t)\boldsymbol{A}\big)^j \boldsymbol{R}^0 + \big(\boldsymbol{I}_{p\times p} - \big(\boldsymbol{I}_{p\times p} - (\Delta t)\boldsymbol{A}\big)^j\big)\boldsymbol{R}^* + \sum_{l=0}^{j} \big(\boldsymbol{I}_{p\times p} - (\Delta t)\boldsymbol{A}\big)^{j-l} \boldsymbol{\varepsilon}_{t_l}. \tag{6}$$

Assuming the matrix $\big(\boldsymbol{I}_{p\times p} - (\Delta t)\boldsymbol{A}\big)$ *in Equation* (2) *is strictly contracting, i.e., the (possibly complex) eigenvalues of* $\boldsymbol{A}$ *satisfy*

$$\big|1 - (\Delta t)\lambda\big| < 1 \text{ for } \lambda \in \sigma(\boldsymbol{A}),$$

We additionally have

$$\mathbb{E}\Big[\Big\| \frac{1}{T} \sum_{j=0}^{n-1} (\Delta t)\boldsymbol{R}_{t_j} - \boldsymbol{R}^* \Big\|^2\Big] = O(T^{-1/2}) \text{ uniformly in } \Delta t \to 0 \text{ as } T \to \infty$$

and, thus,

$$\frac{1}{T} \sum_{j=0}^{n-1} (\Delta t)\boldsymbol{R}_{t_j} = \boldsymbol{R}^* + O_{\mathbb{P}}(T^{-1/2}) \text{ uniformly in } \Delta t \to 0 \text{ as } T \to \infty \tag{7}$$

where $T = (\Delta t)n$.

Given a (single) batch of data $\{\boldsymbol{R}_{t_0}, \ldots, \boldsymbol{R}_{t_{n-1}}\}$ from a contaminated model, the goal is to estimate the model parameters $\boldsymbol{R}^*$, $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$ while attaining a prescribed breakdown point of $\alpha \in [0, \frac{1}{2})$. Empirically, existing estimators of $\boldsymbol{R}$ tend to fail even if $\boldsymbol{A}$ is just slightly ill-conditioned, which can be further exacerbated by the presence of outliers. All estimators based on VAR(1) parametrization, including the multivariate least trimmed squares (MLTS) estimator (see Section 2.1), fail to be applicable in this situation.

Unlike existing methods, instead of treating $\boldsymbol{R}^*$ as a meaningless *nominal* model parameter and attempting to indirectly reconstruct it from an estimate of $\boldsymbol{A}\boldsymbol{R}^*$, we rather leverage the large-time convergence to the *actual* long-term mean in accordance with Equation (7). It is important to emphasize that the latter holds even if no reliable estimates of $\boldsymbol{A}$ and/or $\boldsymbol{\Sigma}$ are available. In light of this fact, our approach (see Section 2.2) leverages an

intrinsic connection between likelihood maximization and the large-time behavior of the process $(\boldsymbol{R}_{t_j})_{t_j}$ (cf. Theorem A2 in Appendix A) to recover the long-term mean $\boldsymbol{R}^*$. Unlike conventional statistical identifiability, our argumentation has a degree of reminiscence with both structural identifiability and detectability for dynamical systems.

2.1. Multivariate Least Trimmed Squares Estimation

The multivariate least trimmed squares (MLTS) estimator (Agulló et al., 2008) is a robust alternative to the multivariate least squares method for estimating the multivariate regression model. The MLTS estimator is also applicable to estimating the parameters of VAR(k) models. For a p-variate discrete-time process $(\boldsymbol{y}_{t_j})_{t_j}$ with $t_j = j(\Delta t)$ for some $\Delta t > 0$, the latter reads as

$$\boldsymbol{y}_{t_{j+1}} = \boldsymbol{\beta}_0' + \boldsymbol{\beta}_1' \boldsymbol{y}_{t_j} + \ldots + \boldsymbol{\beta}_k' \boldsymbol{y}_{t_{j-k+1}} + \boldsymbol{\varepsilon}_{t_j} \quad \text{for } j = k-1, \ldots, n-k \tag{8}$$

where $k \in \mathbb{N}$ is the lag size, $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ is the intercept parameter, $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k \in \mathbb{R}^{p \times p}$ are the partial slopes, and $\boldsymbol{\varepsilon}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}_p, \boldsymbol{\Sigma})$ are the errors. For lag length 1, the VAR($k = 1$) model becomes

$$\boldsymbol{y}_{t_{j+1}} = \boldsymbol{\beta}_0' + \boldsymbol{\beta}_1' \boldsymbol{y}_{t_j} + \boldsymbol{\varepsilon}_{t_j} \quad \text{for } j = 0, 1, \ldots, n-1 \tag{9}$$

which can be expressed in the form of the multivariate regression model

$$\boldsymbol{y}_{t_{j+1}} = \boldsymbol{\beta}' \boldsymbol{x}_{t_j} + \boldsymbol{\varepsilon}_{t_j} \tag{10}$$

where $\boldsymbol{x}_{t_j} = (1, \boldsymbol{y}_{t_j}')' \in \mathbb{R}^{p+1}$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_0', \boldsymbol{\beta}_1')' \in \mathbb{R}^{(p+1) \times p}$ is the matrix containing the regression coefficients. Note that general VAR(k) models can be reduced to the Markovian case of VAR(1) by extending the phase space to include k previous process values.

Letting $\boldsymbol{X} = (\boldsymbol{x}_{t_0}, \ldots, \boldsymbol{x}_{t_{n-1}})' \in \mathbb{R}^{n \times (p+1)}$ and $\boldsymbol{Y} = (\boldsymbol{y}_{t_1}, \ldots, \boldsymbol{y}_{t_n})' \in \mathbb{R}^{n \times p}$ denote the design matrix and the response matrix, respectively, the usual non-robust multivariate least squares estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$

$$\hat{\boldsymbol{\beta}}_{\text{MLS}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}, \quad \hat{\boldsymbol{\Sigma}}_{\text{MLS}} = \frac{1}{n-p-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{MLS}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{MLS}})$$

follow.

Akin to the Minimum Covariance Determinant (MCD) of Rousseeuw (1984), robust MLTS estimation involves finding a subset of h observations such that the MLS fit to these observations minimizes the determinant of the residual covariance. For an index set $\mathcal{I} \subset \{0, 1, \ldots, n-1\}$ with $|\mathcal{I}| = h$, the MLS estimators based on a subsample indexed by $\mathcal{I}$ are given as

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\text{MLS}}(\mathcal{I}) &= (\boldsymbol{X}_{\mathcal{I}}'\boldsymbol{X}_{\mathcal{I}})^{-1}\boldsymbol{X}_{\mathcal{I}}'\boldsymbol{Y}_{\mathcal{I}} \\
\hat{\boldsymbol{\Sigma}}_{\text{MLS}}(\mathcal{I}) &= \frac{1}{h-p-1}\left(\boldsymbol{Y}_{\mathcal{I}} - \boldsymbol{X}_{\mathcal{I}}\hat{\boldsymbol{\beta}}_{\text{MLS}}(\mathcal{I})\right)'\left(\boldsymbol{Y}_{\mathcal{I}} - \boldsymbol{X}_{\mathcal{I}}\hat{\boldsymbol{\beta}}_{\text{MLS}}(\mathcal{I})\right).
\end{aligned}$$

The MLTS estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are then expressed as

$$\hat{\boldsymbol{\beta}}_{\text{MLTS}} = \hat{\boldsymbol{\beta}}_{\text{MLS}}(\mathcal{I}^*), \quad \hat{\boldsymbol{\Sigma}}_{\text{MLTS}} = c_{p,\alpha}\hat{\boldsymbol{\Sigma}}_{\text{MLS}}(\mathcal{I}^*) \tag{11}$$

where

$$\mathcal{I}^* = \underset{\substack{\mathcal{I} \in \{0,1,\ldots,n-1\} \\ |\mathcal{I}| = h}}{\arg\min} \log\det\left(\hat{\boldsymbol{\Sigma}}_{\text{MLS}}(\mathcal{I})\right) \tag{12}$$

and $c_{p,\alpha}$ is a consistency factor defined in Equation (21) below.

The estimators $\hat{\boldsymbol{\beta}}_{\text{MLTS}}$ and $\hat{\boldsymbol{\Sigma}}_{\text{MLTS}}$ serve as a robust alternative to MLS estimation. The MLTS estimators have been shown to be high-breakdown and Fisher-consistent (Agulló et al., 2008), while asymptotic- and $\sqrt{n}$-consistency are only known for scalar-on-vector LTS regression (Víšek, 2006a, 2006b; Zuo, 2024).

The combinatorial optimization problem in Equation (12) is typically solved with a sampling-based algorithm that involves iterative application of the concentration step (C-step) (Agulló et al., 2008; Rousseeuw & Van Driessen, 1999). From a theoretical standpoint, $\det(\cdot)$ or $\log\det(\cdot)$ can equivalently be used in Equation (12); however, computational implementations typically prefer the latter due to better numerical stability properties.

The multivariate Vasicek model in Equation (2) can be expressed as a VAR(1)-process

$$\boldsymbol{R}_{t_{j+1}} = \boldsymbol{\beta}_0' + \boldsymbol{\beta}_1' \boldsymbol{R}_{t_j} + \boldsymbol{\varepsilon}_{t_j} \tag{13}$$

where $\boldsymbol{\beta}_0' = (\Delta t)\boldsymbol{A}\boldsymbol{R}^*$, $\boldsymbol{\beta}_1' = \boldsymbol{I}_{p\times p} - (\Delta t)\boldsymbol{A}$ and $\operatorname{Cov}[\boldsymbol{\varepsilon}_{t_j}] \equiv \boldsymbol{\Xi} = (\Delta t)\boldsymbol{\Sigma}$. This reparametrization is equivalent if and only if $\boldsymbol{A}$ is invertible since the original Vasicek parameters can be reconstructed from the VAR(1) parameters via

$$\boldsymbol{R}^* = \frac{1}{\Delta t}\boldsymbol{A}^{-1}\boldsymbol{\beta}_0', \quad \boldsymbol{A} = \frac{1}{\Delta t}(\boldsymbol{I}_{p\times p} - \boldsymbol{\beta}_1'), \quad \boldsymbol{\Sigma} = \frac{1}{\Delta t}\boldsymbol{\Xi}, \tag{14}$$

which leads to eponymous MLTS estimators of Vasicek parameters. Theoretically, maximum likelihood estimation for the Vasicek model is asymptotically equivalent with that for VAR(1) on the strength of the invariance principle. In turn, maximum likelihood estimation for the Vasicek model is then also equivalent with the least squares estimation of the VAR(1) model (up to the Bessel correction factor for the estimate of $\boldsymbol{\Sigma}$). However, if $\boldsymbol{A}$ is even slightly ill-conditioned, Equation (14) fails to furnish a reasonable estimate of the long-term mean $\boldsymbol{R}^*$ even in relatively large samples.

2.2. Maximum Trimmed Likelihood Estimation

Instead of reparametrizing the Vasicek model in Equation (2) as a VAR(1) model and applying maximum likelihood estimation to the latter, similar to Pokojovy et al. (2024) in the non-robust situation, we put forth a system of estimating equations for the maximum trimmed likelihood estimator (MTLE) and solve it with a numerical scheme based on iterative application of the C-step. Replacing the inverse of $\boldsymbol{A}$ with a suitably truncated Moore–Penrose pseudoinverse, a numerically stable estimate of the long-term mean $\boldsymbol{R}^*$ can be obtained.

In the absence of outliers, Pokojovy et al. (2024) discuss maximum likelihood estimation for the discrete Vasicek model (2). Introducing the log-likelihood function (scaled by Δt)

$$\ell\big(\boldsymbol{\theta}|\boldsymbol{R}\big) = (\Delta t)\sum_{j=0}^{n-1} \log\varphi\Big(\Delta\boldsymbol{R}_{t_j} - \boldsymbol{A}(\boldsymbol{R}^* - \boldsymbol{R}_{t_j})\Delta t \,\Big|\, \boldsymbol{0}_p, (\Delta t)\boldsymbol{\Sigma}\Big) \tag{15}$$

where $\boldsymbol{\theta} = (\boldsymbol{R}^*, \boldsymbol{A}, \boldsymbol{\Sigma})$, $\boldsymbol{R} = (\boldsymbol{R}_{t_0}, \boldsymbol{R}_{t_1}, \ldots, \boldsymbol{R}_{t_n})'$, $\Delta\boldsymbol{R}_{t_j} := \boldsymbol{R}_{t_{j+1}} - \boldsymbol{R}_{t_j}$ and $\varphi(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the p-variate Gaussian density, the maximum likelihood estimator is defined as

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} \equiv (\hat{\boldsymbol{R}}^*_{\text{MLE}}, \hat{\boldsymbol{A}}_{\text{MLE}}, \hat{\boldsymbol{\Sigma}}_{\text{MLE}}) = \operatorname*{arg\,max}_{\boldsymbol{\theta}\in\Theta} \ell\big(\boldsymbol{\theta}|\boldsymbol{R}\big) \equiv \operatorname*{arg\,max}_{(\boldsymbol{R}^*, \boldsymbol{A}, \boldsymbol{\Sigma})\in\Theta} \ell\big(\boldsymbol{R}^*, \boldsymbol{A}, \boldsymbol{\Sigma}|\boldsymbol{R}\big) \tag{16}$$

with open parameter set $\Theta = \big\{(\boldsymbol{R}^*, \boldsymbol{A}, \boldsymbol{\Sigma}) \,|\, \boldsymbol{R}^* \in \mathbb{R}^p, \boldsymbol{A} \text{ regular}, \boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}, \boldsymbol{\Sigma}' = \boldsymbol{\Sigma}, \boldsymbol{\Sigma} \succ 0\big\}$.

Fixing an integer h and letting $\alpha := 1 - h/n$, in lieu of the likelihood function in Equation (15), we consider the trimmed scaled log-likelihood function

$$\begin{aligned}\ell_{\mathcal{I}}(\boldsymbol{\theta}|\boldsymbol{R}) &= (\Delta t)\sum_{j\in\mathcal{I}}\log\varphi\Big(\Delta\boldsymbol{R}_{t_j} - \boldsymbol{A}(\boldsymbol{R}^* - \boldsymbol{R}_{t_j})\Delta t \,\Big|\, \boldsymbol{0}_p, (\Delta t)\boldsymbol{\Sigma}\Big)\\ &= -\tfrac{(1-\alpha)pT}{2}\log\left(2\pi(\Delta t)\right) - \tfrac{(1-\alpha)T}{2}\log|\boldsymbol{\Sigma}|\\ &\quad -\frac{1}{2}\sum_{j\in\mathcal{I}} d^2\big(\Delta\boldsymbol{R}_{t_j} - (\Delta t)\boldsymbol{A}(\boldsymbol{R}^* - \boldsymbol{R}_{t_j})|\boldsymbol{0}_p, (\Delta t)\boldsymbol{\Sigma}\big)\end{aligned} \tag{17}$$

for arbitrary $\mathcal{I} \subset \{0, 1, \ldots, n-1\}$ with $|\mathcal{I}| = h$, where

$$d^2(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \tag{18}$$

is the usual squared Mahalanobis distance function.

We define the maximum trimmed likelihood estimator (MTLE) as

$$\hat{\boldsymbol{R}}^*_{\text{MTLE}} := \hat{\boldsymbol{R}}^*_{\dagger}, \quad \hat{\boldsymbol{A}}_{\text{MTLE}} := \hat{\boldsymbol{A}}_{\dagger}, \quad \hat{\boldsymbol{\Sigma}}_{\text{MTLE}} := c^2_{p,\alpha}\hat{\boldsymbol{\Sigma}}_{\dagger} \quad \text{for } \alpha = 1 - \frac{h}{n} \tag{19}$$

with

$$(\mathcal{I}_{\dagger}, \hat{\boldsymbol{\theta}}_{\dagger}) := \underset{\substack{\boldsymbol{\theta}\in\Theta \\ \mathcal{I}\subset\{0,1,\ldots,n-1\},|\mathcal{I}|=h}}{\arg\max} \ell_{\mathcal{I}}(\boldsymbol{\theta}|\boldsymbol{R}) \tag{20}$$

where the covariance estimate is adjusted using the asymptotic bias correction factor

$$c^2_{p,\alpha} = \frac{1-\alpha}{\mathbb{P}\{\chi^2_{p+2} \le \chi^2_{p,1-\alpha}\}}. \tag{21}$$

In this paper, we primarily focus on the robustness of our estimator. Therefore, we solely consider unreweighted or "raw" estimators. However, the usual reweighting approach (cf. Agulló et al. (2008); Croux and Joossens (2008) for reweighted MLTS) can be adopted to obtain reweighted counterparts should one be interested in increasing the statistical efficiency of the estimator.

The double maximization problem in Equation (20) is a mixed programming problem involving discrete optimization in $\mathcal{I}$ and continuous optimization in $\boldsymbol{\theta}$. Fixing $\boldsymbol{\theta} \in \Theta$, the optimum of $\ell(\cdot)$ with respect to $\mathcal{I}$ is attained at $\mathcal{I} = \{i_1, i_2, \ldots, i_h\}$, where $d_{i_1} \le d_{i_2} \le \cdots \le d_{i_n}$ are the sorted squared Mahalanobis distances

$$d^2_j = d^2\big(\Delta\boldsymbol{R}_{t_j} - (\Delta t)\boldsymbol{A}(\boldsymbol{R}^* - \boldsymbol{R}_{t_j})|\boldsymbol{0}_p, (\Delta t)\boldsymbol{\Sigma}\big).$$

Conversely, fixing $\mathcal{I}$, on the strength of (Pokojovy et al., 2024, Theorem 1), the optimum of $\ell(\cdot)$ with respect to $\boldsymbol{\theta}$ is attained at a solution of the estimating equations

$$\boldsymbol{R}^* = \frac{1}{h}\sum_{j\in\mathcal{I}}\boldsymbol{R}_{t_j} + \frac{1}{(1-\alpha)T}\boldsymbol{A}^{-1}\sum_{j\in\mathcal{I}}(\boldsymbol{R}_{t_{j+1}} - \boldsymbol{R}_{t_j}), \tag{22}$$

$$\boldsymbol{A} = \frac{1}{\Delta t}\left(\sum_{j\in\mathcal{I}}(\boldsymbol{R}_{t_{j+1}} - \boldsymbol{R}_{t_j})(\boldsymbol{R}^* - \boldsymbol{R}_{t_j})'\right)\left(\sum_{j\in\mathcal{I}}(\boldsymbol{R}^* - \boldsymbol{R}_{t_j})(\boldsymbol{R}^* - \boldsymbol{R}_{t_j})'\right)^{-1}, \tag{23}$$

$$\boldsymbol{\Sigma} = \frac{1}{(1-\alpha)T}\sum_{j\in\mathcal{I}}\Big(\Delta\boldsymbol{R}_{t_j} - \boldsymbol{A}\big(\boldsymbol{R}^* - \boldsymbol{R}_{t_j}\big)(\Delta t)\Big)\Big(\Delta\boldsymbol{R}_{t_j} - \boldsymbol{A}\big(\boldsymbol{R}^* - \boldsymbol{R}_{t_j}\big)(\Delta t)\Big)'. \tag{24}$$

Instead of the usual matrix inverse, a suitable truncated Moore–Penrose pseudoinverse

$$\boldsymbol{A}^{-1} = \sum_{k=1}^{p}\mathbb{1}_{[\epsilon,\infty)}(\sigma_k/\|\boldsymbol{A}\|)\frac{1}{\sigma_k}\boldsymbol{u}_k\boldsymbol{v}_k' \quad \text{with } \|\boldsymbol{A}\| = \max_{k=1,\ldots,k}\sigma_k$$

with a proper choice truncation threshold $\epsilon > 0$ at respective occurrences (see Supplementary Materials, MATLAB® code) is used throughout the rest of the paper, where σ_k, $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$, $k = 1, \ldots, p$, are the singular values as well as left- and right-singular vectors of $\boldsymbol{A}$.

Computing MTLE for the Vasicek model by solving Equations (19) and (20) offers two advantages over the MLTS approach. While both estimation methodologies are asymptotically equivalent as the sample size $n \to \infty$, our proposed regularized (via adopting the matrix pseudoinverse) estimating Equations (22)–(24) furnish an estimator $\hat{\boldsymbol{R}}^*_{\text{MTLE}}$ converging to the *actual* long-term mean

$$L^2\text{–}\lim_{T\to\infty} \frac{1}{T} \sum_{j=0}^{n-1} (\Delta t) \boldsymbol{R}_{t_j} \text{ with } T = (\Delta t) n$$

in the absence of outliers, where the usual definition of the mean square convergence

$$L^2\text{–}\lim_{j\to\infty} \boldsymbol{X}_j = \boldsymbol{X}^\infty \quad :\Longleftrightarrow \quad \lim_{j\to\infty} \mathbb{E}\big[\|\boldsymbol{X}_j - \boldsymbol{X}^\infty\|^2\big] = 0$$

is adopted. On the strength of Theorems A1 and A2 in Appendix A, our estimator not only preserves the (hypothesized) consistency of MLTS estimation for large samples but also performs well on large time horizons even for moderately sized samples.

In contrast, $\hat{\boldsymbol{R}}^*_{\text{MLTS}}$ usually fails to be consistent for smaller n irrespective of T due to the singularity of $\hat{\boldsymbol{A}}_{\text{MLTS}}$. Our simulation studies and real-world data analyses in Sections 3.1 and 3.2 provide very compelling empirical evidence that our MTLE significantly outperforms the leading competitors, i.e., MLTS, MLS and MLE, at estimating the long-term mean $\boldsymbol{R}^*$ on both short and long calibration time horizons over a range of contamination scenarios. Interestingly, MLTS and MLS fail to accurately estimate the long-term mean $\boldsymbol{R}^*$ on shorter calibration horizons even if the mean-reversion speed matrix $\boldsymbol{A}$ is well-conditioned.

As for a numerical solution of the MTLE problem in Equations (19) and (20), our proposed Algorithm 1 combines an adaptation of the FastMCD algorithm (Rousseeuw & Van Driessen, 1999) to compute the MLTS estimator (Agulló et al., 2008) with a Banach fixed point algorithm for computing MLE (Pokojovy et al., 2024). One major methodological improvement over the latter work is the affine-invariant scale-stable Fréchet-type metric employed in step 2(d) of Algorithm 1 that allows for better error control.

For this iterative procedure (Algorithm 1), at each iteration in step 2, the complexity of computing $\boldsymbol{R}^*$ is $\mathcal{O}(np + p^3)$ because it involves the computation of the Moore–Penrose pseudoinverse. Similarly, for this same reason, the complexity of computing $\boldsymbol{A}$ is $\mathcal{O}(np^2 + p^3)$. The complexity of computing $\boldsymbol{\Sigma}$ is only $\mathcal{O}(np^2)$ since it does not involve computing any matrix multiplication or pseudoinverse. Putting everything together, the overall complexity is $\mathcal{O}\big(\texttt{MaxIter} \cdot (np^2 + p^3)\big)$. Since p is typically low, the cubic complexity in p does not impose major computational restrictions.

The breakdown point of an estimator is typically defined as the smallest proportion of contamination of the sample (with size n) for which the estimator becomes uninformative. This is mathematically represented as, for an estimator $\boldsymbol{T}(\boldsymbol{Z})$ with sample $\boldsymbol{Z}$,

$$\varepsilon^*_n(\boldsymbol{T}) = \min\Big\{ \frac{m}{n} \,\Big|\, \sup_{\tilde{\boldsymbol{Z}}} \|\boldsymbol{T}(\boldsymbol{Z}) - \boldsymbol{T}(\tilde{\boldsymbol{Z}})\| = \infty \Big\},$$

where $\tilde{\boldsymbol{Z}}$ is obtained by replacing m observations in $\boldsymbol{Z}$ by arbitrary points. The MLS and MLE estimators require only $1/n$ to render these estimator useless. Thus, they have a breakdown point of $1/n$ and asymptotic breakdown point of 0. The asymptotic breakdown point for MLTS is 50% (Jung, 2005). The breakdown point for MTLE is equal to that of

MLTS, attaining an asymptotic value of 50%, when $\hat{\boldsymbol{A}}$ is invertible. This suggests, in general, that both MLTS and MTLE are high-breakdown estimators, unlike MLS or MLE.

Algorithm 1 Maximum Trimmed Likelihood Estimator (MTLE)

Select a tolerance level $\varepsilon > 0$ (say, $\varepsilon = 10^{-6}$) and `MaxIter` (say, 5000).

1. Use MLTS estimators (viz., Equation (14)) to form a warmstart $\boldsymbol{R}^*_{(0)}$, $\boldsymbol{A}_{(0)}$ and $\boldsymbol{\Sigma}_{(0)}$.
2. For $k = 1, 2, \ldots,$ `MaxIter`, iterate the following:
   (a) Compute the squared Mahalanobis distances
   $$(d^2_{(k)})_j = d^2\big(\Delta \boldsymbol{R}_{t_j} - (\Delta t)\boldsymbol{A}_{(k-1)}(\boldsymbol{R}^*_{(k-1)} - \boldsymbol{R}_{t_j}) | \boldsymbol{0}_p, (\Delta t)\boldsymbol{\Sigma}_{(k-1)}\big).$$
   and sort
   $$(d^2_{(k)})_{i_1(k)} \le (d^2_{(k)})_{i_2(k)} \le \ldots (d^2_{(k)})_{i_n(k)}.$$
   (b) Update the optimal set
   $$\mathcal{I}_{(k)} = \big\{i_1(k), i_2(k), \ldots, i_h(k)\big\}.$$
   (c) Update parameter estimates
   $$\boldsymbol{R}^*_{(k)} = \frac{1}{h}\sum_{j\in\mathcal{I}_{(k)}} \boldsymbol{R}_{t_j} + \frac{1}{(1-\alpha)T}\boldsymbol{A}^{-1}_{(k-1)}\sum_{j\in\mathcal{I}}(\boldsymbol{R}_{t_{j+1}} - \boldsymbol{R}_{t_j}),$$
   $$\boldsymbol{A}_{(k)} = \frac{1}{\Delta t}\Big(\sum_{j\in\mathcal{I}_{(k)}} (\boldsymbol{R}_{t_{j+1}} - \boldsymbol{R}_{t_j})(\boldsymbol{R}^*_{(k-1)} - \boldsymbol{R}_{t_j})'\Big)\Big(\sum_{j\in\mathcal{I}_{(k)}} (\boldsymbol{R}^*_{(k-1)} - \boldsymbol{R}_{t_j})(\boldsymbol{R}^*_{(k-1)} - \boldsymbol{R}_{t_j})'\Big)^{-1},$$
   $$\boldsymbol{\Sigma}_{(k)} = \frac{1}{(1-\alpha)T}\sum_{j\in\mathcal{I}_{(k)}} \Big(\Delta \boldsymbol{R}_{t_j} - \boldsymbol{A}\big(\boldsymbol{R}^*_{(k-1)} - \boldsymbol{R}_{t_j}\big)(\Delta t)\Big)\Big(\Delta \boldsymbol{R}_{t_j} - \boldsymbol{A}\big(\boldsymbol{R}^*_{(k-1)} - \boldsymbol{R}_{t_j}\big)(\Delta t)\Big)'.$$
   (d) Compute the affine-invariant error
   $$\epsilon_{(k)} = \frac{\big(\frac{1}{p}d^2\big(\boldsymbol{R}^*_{(k)} - \boldsymbol{R}^*_{(k-1)} \,|\, \boldsymbol{0}_p, \frac{1}{2}(\boldsymbol{\Sigma}_{(k-1)} + \boldsymbol{\Sigma}_{(k)})\big)\big)^{1/2}}{1 + \big(\frac{1}{p}d^2\big(\boldsymbol{R}^*_{(k)} - \boldsymbol{R}^*_{(k-1)} \,|\, \boldsymbol{0}_p, \frac{1}{2}(\boldsymbol{\Sigma}_{(k-1)} + \boldsymbol{\Sigma}_{(k)})\big)\big)^{1/2}}$$
   $$+ \frac{\big(\frac{1}{p^2}\|\boldsymbol{A}_{(k)} - \boldsymbol{A}_{(k-1)}\|^2_{\mathcal{F}}\big)^{1/2}}{1 + \big(\frac{1}{p^2}\|\boldsymbol{A}_{(k)} - \boldsymbol{A}_{(k-1)}\|^2_{\mathcal{F}}\big)^{1/2}} + \frac{\big(\frac{1}{p^2}\rho^2(\boldsymbol{\Sigma}_{(k-1)}, \boldsymbol{\Sigma}_{(k)})\big)^{1/2}}{1 + \big(\frac{1}{p^2}\rho^2(\boldsymbol{\Sigma}_{(k-1)}, \boldsymbol{\Sigma}_{(k)})\big)^{1/2}}$$
   with
   $$\rho^2(\boldsymbol{U}, \boldsymbol{V}) = \|\boldsymbol{I}_{p\times p} - \boldsymbol{U}^{-1}\boldsymbol{V}\|^2_{\mathcal{F}} + \|\boldsymbol{I}_{p\times p} - \boldsymbol{U}\boldsymbol{V}^{-1}\|^2_{\mathcal{F}}$$
   and the Frobenius norm
   $$\|\boldsymbol{U}\|^2_{\mathcal{F}} = \mathrm{tr}(\boldsymbol{U}'\boldsymbol{U}).$$
   (e) Terminate the iteration as soon as $\epsilon_{(k)} < \varepsilon$ and return
   $$\big(\boldsymbol{R}^*_{(k)}, \boldsymbol{A}_{(k)}, c^2_{p,\alpha}\boldsymbol{\Sigma}_{(k)}\big)$$
   as an approximation of $(\hat{\boldsymbol{R}}^*_{\mathrm{MTLE}}, \hat{\boldsymbol{A}}_{\mathrm{MTLE}}, \hat{\boldsymbol{\Sigma}}_{\mathrm{MTLE}})$.

3. Results and Discussion

3.1. Simulation Study

We performed extensive simulations for two configurations of parameters $\boldsymbol{R}^*$, $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$ corresponding to two real-world settings of Vasicek dynamics for bivariate (2D) short rates in Setting I and sextivariate (6D) log-volatilities in Setting II under a variety of hypothetical contamination scenarios to benchmark the performance of our MTLE against that of MLTS, MLE and MLS. Under each simulation setting, we generate n_{rep} simulated paths via

Equation (3) with a suitably chosen Huber contamination in Equation (5) for multiple scenarios across several sample sizes and robustness configurations, including trimming percentage α, contamination rate ε and non-centrality parameter values ncp. The choice of a Huber-type contamination model is motivated by its ability to capture deviations from standard Gaussian assumptions while maintaining computational tractability. It provides a flexible framework for handling heavy-tailed distributions as often encountered in financial time series (Brockwell & Davis, 2002). The trimming percentage $\alpha \in [0, \frac{1}{2})$ corresponds to the nominal breakdown point (bdp) and provides information on the proportion of observations that are trimmed from the log-likelihood (cf. Equation (17)). The contamination rate ε is the proportion of outliers introduced, while the non-centrality parameter value (ncp) indicates how far the outlying innovations depart from $\mathbf{0}_p$ measured with the squared Mahalanobis distance.

The rationale behind focusing on what is typically known in the literature as "cluster contamination" (referring to outliers clustered as one group with the same or similar covariance as the "good" bulk of the data) is two-fold. First, this contamination scenario, together with the (less realistic) point contamination, is what is typically studied in the robust statistics literature (Hubert et al., 2012; Pokojovy & Jobe, 2022). Second, "clustered" outliers, especially in the affine-equivariant scenario that we have in our paper, are recognized as probably the most challenging contamination type (Reyen et al., 2009; Rocke & Woodruff, 1996). Therefore, focusing on this "worst-case" scenario gives an honest and simple way to benchmark the methods across multiple nominal bdp choices as well as outlier fractions and ncp values.

Each of the four estimators (MTLE, MLTS, MLE and MLS) is then employed to estimate the parameters. For each scenario, we compare the performance of MTLE to that of its competitors. The error measures utilized in our comparisons are empirical MSE-like affine-invariant expressions

$$\widehat{\mathrm{err}}^2_{\boldsymbol{R}^*} = \frac{1}{n_{\mathrm{rep}}} \sum_{i=1}^{n_{\mathrm{rep}}} \frac{1}{p} d^2(\hat{\boldsymbol{R}}_i^* - \boldsymbol{R}^* \mid \mathbf{0}_p, \hat{\boldsymbol{\Sigma}}), \tag{25}$$

$$\widehat{\mathrm{err}}^2_{\boldsymbol{A}} = \frac{1}{n_{\mathrm{rep}}} \sum_{i=1}^{n_{\mathrm{rep}}} \frac{1}{p^2} \|\hat{\boldsymbol{A}}_i - \boldsymbol{A}\|^2_{\mathcal{F}}, \quad \widehat{\mathrm{err}}^2_{\boldsymbol{\Sigma}} = \frac{1}{n_{\mathrm{rep}}} \sum_{i=1}^{n_{\mathrm{rep}}} \frac{1}{p^2} \rho^2(\hat{\boldsymbol{\Sigma}}_i, \boldsymbol{\Sigma}) \tag{26}$$

with $\rho^2(\cdot, \cdot)$ defined as before

$$\rho^2(\boldsymbol{U}, \boldsymbol{V}) = \|\boldsymbol{I}_{p\times p} - \boldsymbol{U}^{-1}\boldsymbol{V}\|^2_{\mathcal{F}} + \|\boldsymbol{I}_{p\times p} - \boldsymbol{U}\boldsymbol{V}^{-1}\|^2_{\mathcal{F}}$$

We decide to forgo empirical breakdown point computation (Pokojovy & Jobe, 2022) since the theoretical breakdown points of both our proposed MTLE and the conventional robust MLTS are equal assuming the "good" points are in general position. This also explains why both methods are equivalent at estimating the $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$ parameters. The empirically observed difference in MSE for $\boldsymbol{R}^*$ estimation appears to be solely caused by the subpar efficiency of MLTS and should not be mistaken for breakdown (which does occur for MLE or MLS). Indeed, as our extensive simulations reported in the Supplementary Materials suggest, the two methods (MTLE and MLTS) become equivalent as the sample size grows.

3.1.1. Simulation Setting I

This simulation setting is similar to that analyzed by Pokojovy et al. (2024), except that we additionally consider multiple outlier contamination scenarios in our study. A hypothetical short-rate situation with parameters

$$\boldsymbol{R}^* = \begin{pmatrix} 5 \\ 6 \end{pmatrix}, \quad \boldsymbol{A} = \frac{1}{365}\begin{pmatrix} 9 & 4 \\ 2 & 3 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \frac{1}{365^2}\begin{pmatrix} 7 & 3 \\ 3 & 7 \end{pmatrix}$$

was considered, which corresponds to a hypothetical high-interest environment with US/Euro short-rate long-term levels of 5% and 6%, respectively.

Choosing $\Delta t = 1$ day and a grid of time horizons $T = 60, 90, 120, \ldots, 330, 360$ days, $n_{\text{rep}} = 5000$ daily paths $\{\boldsymbol{R}_{t_j,i}\}_{j=0,\ldots,n}$ for $i = 1, \ldots, n_{\text{rep}}$ were sampled from the bivariate ($p = 2$) discrete Vasicek process in Equation (3) under multiple configurations of a Huber contamination

$$\boldsymbol{\varepsilon}_{t_j} \overset{\text{i.i.d.}}{\sim} (1-\varepsilon)\mathcal{N}(\boldsymbol{0}_p, (\Delta t)\boldsymbol{\Sigma}) + \varepsilon\mathcal{N}(\sqrt{\text{ncp}/p}\,\mathbb{1}_p, (\Delta t)\boldsymbol{\Sigma})$$

over multiple combinations of the nominal breakdown point bdp $= 0.25, 0.35$, with outlier fraction

$$\varepsilon \in \begin{cases} \{0.0, 0.05, 0.10, 0.20\} & \text{if bdp} = 0.25, \\ \{0.0, 0.1, 0.2, 0.3, 0.4\} & \text{if bdp} = 0.35 \end{cases}$$

and non-centrality parameter values ncp $= 25, 50$. For each ε, $\lfloor n\varepsilon \rfloor$, outlying innovations were chosen to follow $\mathcal{N}(\sqrt{\text{ncp}/p}\,\mathbb{1}_p, (\Delta t)\boldsymbol{\Sigma})$ instead of $\mathcal{N}(\boldsymbol{0}_p, (\Delta t)\boldsymbol{\Sigma})$. The initial value was selected as $\boldsymbol{R}^0 = (4, 7)'$. Applying each of the four estimators to these n_{rep} simulated paths, respective MSE-like errors ($\widehat{\text{err}}^2$) were obtained via Equations (25) and (26). The estimated root-"MSEs" for a total of 18 scenarios considered are reported in Tables S1–S18 in Supplementary Materials Section S2.1 and plotted in Figures S1–S18 in Supplementary Materials Section S1.1.

Two representatively selected situations are presented next. Figure 1 and Table 1 display empirical root-"MSEs" ($\widehat{\text{err}}$) from estimating $\boldsymbol{R}^*$, $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$ for $\varepsilon = 0.20$, ncp $= 50$ and bdp $= 0.25$. The following observations are immediately apparent. When estimating $\boldsymbol{R}^*$, MTLE and MLE exhibit smaller errors compared to MLTS and MLS for smaller time horizons T, viz., 60 to 90. For intermediate T, viz., 120 to 150, MTLE, MLTS and MLE perform comparably and outperform MLS. As T increases further, the performance of MLS heavily deteriorates, leading to larger errors. The errors for MLE also grow in magnitude. MTLE and MLTS perform comparably and outperform MLE and MLS, with MTLE exhibiting the best performance for $T = 360$. Low variability of the empirical MTLE error also suggests that the error distribution has the desirable property of being light-tailed. In addition, MTLE exhibits better consistency since the errors decrease as T becomes large. When estimating $\boldsymbol{A}$, the estimators exhibit similar performance, while MTLE and MLTS outperform non-robust estimators (MLE and MLS) at estimating $\boldsymbol{\Sigma}$.

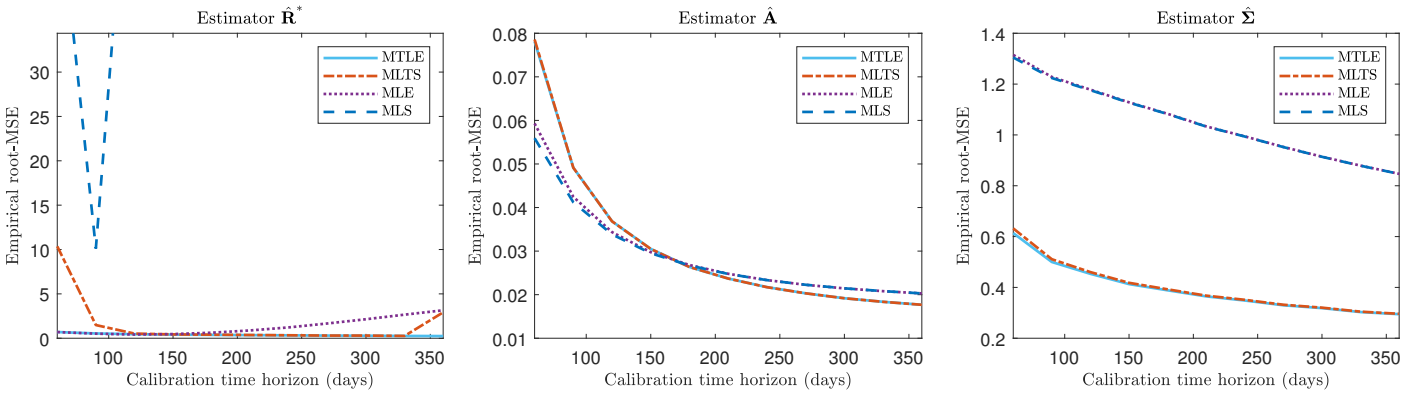


Figure 1. Simulated $\widehat{\text{err}}$ values for $\varepsilon = 0.20$, ncp $= 50$ and bdp $= 0.25$.

Table 1. Simulated $\widehat{\text{err}}$ values for $\varepsilon = 0.20$, ncp $= 50$ and bdp $= 0.25$.

	$\hat{R}^*$				$\hat{A}$				$\hat{\Sigma}$			
T	**MTLE**	**MLTS**	**MLE**	**MLS**	**MTLE**	**MLTS**	**MLE**	**MLS**	**MTLE**	**MLTS**	**MLE**	**MLS**
60	0.6977	10.3566	0.6966	51.6243	0.0781	0.0786	0.0593	0.0559	0.6136	0.6320	1.3148	1.3045
90	0.5393	1.4855	0.5143	10.1155	0.0491	0.0491	0.0425	0.0412	0.5000	0.5108	1.2281	1.2239
120	0.4550	0.5247	0.4139	65.0015	0.0368	0.0368	0.0344	0.0338	0.4533	0.4602	1.1781	1.1760
150	0.4115	0.4298	0.4675	71.1606	0.0305	0.0305	0.0297	0.0295	0.4130	0.4181	1.1283	1.1271
180	0.3779	0.4021	0.6239	262.6636	0.0264	0.0264	0.0268	0.0267	0.3886	0.3924	1.0830	1.0822
210	0.3492	0.3652	0.8790	146.7073	0.0237	0.0237	0.0248	0.0248	0.3647	0.3676	1.0335	1.0329
240	0.3252	0.3334	1.2227	76.7703	0.0217	0.0217	0.0233	0.0233	0.3485	0.3512	0.9942	0.9937
270	0.2997	0.3004	1.6639	93.7638	0.0203	0.0203	0.0223	0.0223	0.3297	0.3318	0.9526	0.9521
300	0.2809	0.2996	2.1418	85.6906	0.0192	0.0192	0.0214	0.0215	0.3185	0.3203	0.9143	0.9139
330	0.2595	0.2599	2.6581	109.3933	0.0184	0.0184	0.0208	0.0208	0.3030	0.3046	0.8792	0.8788
360	0.2432	2.9185	3.1489	3568.5472	0.0177	0.0177	0.0203	0.0203	0.2947	0.2962	0.8469	0.8466

Likewise, Figure 2 and Table 2 report empirical root-"MSEs" from the estimation of model parameters. With 30% of outliers with ncp $= 25$ introduced into simulated data, letting bdp $= 0.25$ for MTLE and MLTS, we observe the following. MLTS, MLE and MLS are unduly influenced by the presence of outliers. This is indicated by inordinately larger errors obtained from the estimation of $\boldsymbol{R}^*$. MTLE, on the other hand, is not adversely influenced and clearly outperforms the competitors. Again, as Table 1 suggests, MTLE is more consistent at estimating $\boldsymbol{R}^*$, and the errors associated with it exhibit lighter tails compared to other estimators. Consequently, MTLE appears to be more robust and efficient. For the estimation of $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$, the reported errors do not vary markedly, suggesting the estimators perform comparably, although MTLE and MLTS have smaller errors.

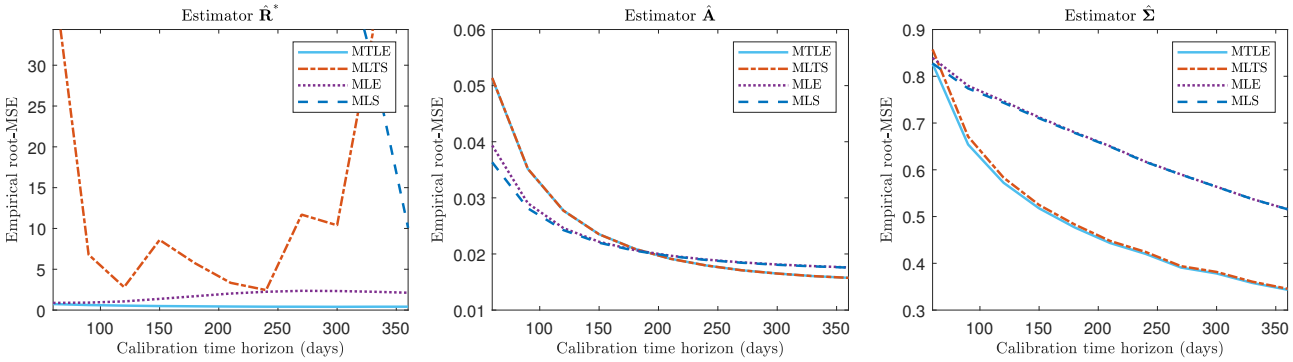


Figure 2. Simulated $\widehat{\text{err}}$ values for $\varepsilon = 0.30$, ncp $= 25$ and bdp $= 0.35$.

Table 2. Simulated $\widehat{\text{err}}$ values for $\varepsilon = 0.30$, ncp $= 25$ and bdp $= 0.35$.

	$\hat{R}^*$				$\hat{A}$				$\hat{\Sigma}$			
T	**MTLE**	**MLTS**	**MLE**	**MLS**	**MTLE**	**MLTS**	**MLE**	**MLS**	**MTLE**	**MLTS**	**MLE**	**MLS**
60	0.7372	41.6813	0.8773	97.4101	0.0513	0.0514	0.0393	0.0363	0.8269	0.8573	0.8392	0.8275
90	0.6147	6.7926	0.9129	134.0426	0.0352	0.0352	0.0290	0.0281	0.6542	0.6711	0.7797	0.7739
120	0.5444	2.8118	1.0641	60.1218	0.0277	0.0277	0.0246	0.0242	0.5722	0.5832	0.7466	0.7434
150	0.4920	8.5968	1.3709	154.1987	0.0235	0.0235	0.0222	0.0220	0.5176	0.5248	0.7122	0.7102
180	0.4625	5.7356	1.6988	133.6683	0.0209	0.0209	0.0207	0.0206	0.4768	0.4831	0.6802	0.6789
210	0.4346	3.3389	2.0262	243.3592	0.0192	0.0191	0.0197	0.0196	0.4434	0.4479	0.6504	0.6496
240	0.4141	2.4478	2.2433	63.9042	0.0180	0.0180	0.0190	0.0189	0.4201	0.4243	0.6172	0.6167
270	0.4100	11.6945	2.3570	133.4085	0.0171	0.0171	0.0185	0.0184	0.3909	0.3939	0.5899	0.5896
300	0.4000	10.3891	2.3338	46.9511	0.0165	0.0165	0.0181	0.0181	0.3783	0.3816	0.5637	0.5635
330	0.4115	33.8369	2.2471	30.4496	0.0161	0.0161	0.0178	0.0178	0.3581	0.3608	0.5376	0.5375
360	0.4140	508.9225	2.1260	10.0797	0.0157	0.0157	0.0176	0.0176	0.3434	0.3455	0.5156	0.5156

These findings are representative of what was typically observed across all scenarios considered. In general, when estimating $\boldsymbol{R}^*$ when there are no outliers ($\varepsilon = 0$), MLE and MTLE demonstrate a comparable performance; however, MLE tends to have smaller errors, which is consistent with the findings reported in Pokojovy et al. (2024). Applied to $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$, the estimators perform comparably well. However, as theoretically expected, MLS and MLE have the best performance. As the percentage of contamination increases, MTLE shows better robustness and mostly outperforms competing estimators, sometimes being on par with MTLE and MLE. Thus, MTLE performs on par with non-robust estimators (MLE and MLS) when there is no contamination, while its advantage becomes especially apparent in the presence of outliers. The performance of MTLE emphasizes its advantageousness as a robust estimator and also demonstrates general applicability under broader circumstances. In addition, simulation results clearly indicate the superiority of MTLE at estimating the long-term mean $\boldsymbol{R}^*$, a parameter playing an important role in risk management and financial decision making. Thus, MTLE provides an attractive robust alternative for estimating the long-term mean, while competing estimators produce subpar results that can lead to unreliable conclusions.

3.1.2. Simulation Setting II

Another simulation study was performed in a sextivariate (six-dimensional) scenario associated with the stochastic volatility data studied in Section 3.2.2. For this situation, theoretical parameters were selected as

$$\boldsymbol{R}^* = \begin{pmatrix} -0.5566 \\ -0.0391 \\ 0.3934 \\ 0.4322 \\ -1.0302 \\ -1.0531 \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} 0.6383 & -0.0029 & 0.0503 & -0.1402 & -0.0285 & 0.0599 \\ -0.0757 & 0.7442 & -0.0803 & 0.0281 & -0.0602 & 0.0973 \\ 0.0079 & -0.0273 & 0.7957 & -0.2526 & -0.0333 & 0.0086 \\ -0.0254 & -0.0172 & -0.0354 & 0.5785 & 0.0047 & -0.0029 \\ 0.0584 & 0.0133 & -0.1031 & 0.0685 & 0.6711 & -0.1734 \\ 0.0902 & -0.0054 & 0.0110 & -0.0608 & -0.1643 & 0.5691 \end{pmatrix},$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 0.1246 & 0.0884 & 0.0297 & 0.0317 & 0.0458 & 0.0628 \\ 0.0884 & 0.1124 & 0.0252 & 0.0267 & 0.0344 & 0.0444 \\ 0.0297 & 0.0252 & 0.1197 & 0.1131 & 0.0118 & 0.0246 \\ 0.0317 & 0.0267 & 0.1131 & 0.1310 & 0.0106 & 0.0262 \\ 0.0458 & 0.0344 & 0.0118 & 0.0106 & 0.0935 & 0.0622 \\ 0.0628 & 0.0444 & 0.0246 & 0.0262 & 0.0622 & 0.1241 \end{pmatrix}.$$

The initial value was chosen as $\boldsymbol{R}^0 = (-0.5, -0.2, 0.5, 0.6, -1.0, -1.3)'$. The same configurations as in Section 3.1.1 were considered under this simulation setting. For each of the 18 configurations, the parameters of the Vasicek model were estimated using the four estimators for each of the n_{rep} paths simulated, and the estimation errors $\widehat{\text{err}}^2$ were obtained. In an analogous fashion, the full range of root-"MSEs" are reported in Tables S19–S36 and Figures S19–S36 in the Supplementary Materials (Sections S2.2 and S1.2, respectively). Note that the vertical axes in some of the figures were adjusted to make the errors more distinguishable for easier comparison. This was warranted due to the consequentially large errors obtained by some of the competing estimators. The root-errors from two representatively selected sets of simulations are presented in Figure 3 and Table 3 ($\varepsilon = 0.20$, ncp $= 25$ and bdp $= 0.25$) as well as Figure 4 and Table 4 ($\varepsilon = 0.10$, ncp $= 25$ and bdp $= 0.35$).

In Figures 3 and 4, the left, middle and right panels benchmark the root-errors from the estimation of $\boldsymbol{R}^*$, $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$, respectively. Tables 3 and 4 tabulate the same information. In both figures, the left panels indicate that MLTS produces larger errors, although the errors tend to become smaller as T increases. MLS is less competitive for smaller T, and as T increases, with the amount of uncontaminated data also increasing, MLS becomes competitive and comparable to MLE. On the other hand, MTLE performs best among these estimators at estimating $\boldsymbol{R}^*$. Since the empirical errors have less variation, this indicates that

the estimates obtained with MTLE are very stable irrespective of the presence or absence of extreme observations. Thus, MTLE is both more robust and less biased. These observations are consistent with Section 3.1.1. When estimating $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$, non-robust estimators, MLE and MLS, have smaller errors compared to MLTS and MTLE, although the difference is not very large. These results highlight what was observed across the different configurations considered under this simulation setting.

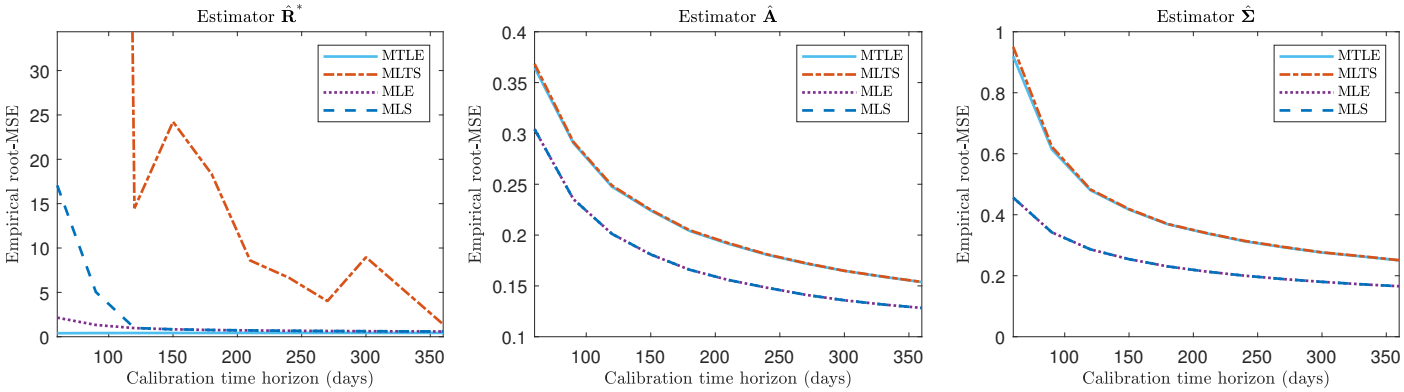


Figure 3. Simulated $\widehat{\text{err}}$ values for $\varepsilon = 0.20$, ncp $= 25$ and bdp $= 0.25$.

Table 3. Simulated $\widehat{\text{err}}$ values for $\varepsilon = 0.20$, ncp $= 25$ and bdp $= 0.25$.

	$\hat{R}^*$				$\hat{A}$				$\hat{\Sigma}$			
T	**MTLE**	**MLTS**	**MLE**	**MLS**	**MTLE**	**MLTS**	**MLE**	**MLS**	**MTLE**	**MLTS**	**MLE**	**MLS**
60	0.39	62.41	2.15	17.06	0.3648	0.3684	0.3041	0.3042	0.9192	0.9501	0.4559	0.4559
90	0.41	405.71	1.33	5.04	0.2906	0.2918	0.2354	0.2354	0.6142	0.6232	0.3422	0.3422
120	0.42	14.48	0.97	0.99	0.2475	0.2487	0.2011	0.2011	0.4804	0.4832	0.2865	0.2865
150	0.42	24.23	0.83	0.84	0.2242	0.2248	0.1809	0.1809	0.4167	0.4187	0.2542	0.2542
180	0.42	18.34	0.76	0.76	0.2041	0.2048	0.1659	0.1659	0.3690	0.3698	0.2306	0.2306
210	0.43	8.59	0.71	0.71	0.1915	0.1922	0.1559	0.1559	0.3393	0.3401	0.2136	0.2136
240	0.43	6.66	0.67	0.67	0.1807	0.1810	0.1482	0.1482	0.3128	0.3141	0.2001	0.2001
270	0.43	4.02	0.64	0.65	0.1721	0.1724	0.1412	0.1412	0.2934	0.2944	0.1890	0.1890
300	0.43	8.96	0.62	0.62	0.1648	0.1648	0.1358	0.1358	0.2761	0.2767	0.1800	0.1800
330	0.43	5.18	0.61	0.61	0.1590	0.1592	0.1317	0.1317	0.2636	0.2644	0.1721	0.1721
360	0.43	1.40	0.59	0.59	0.1538	0.1538	0.1283	0.1283	0.2503	0.2510	0.1656	0.1656

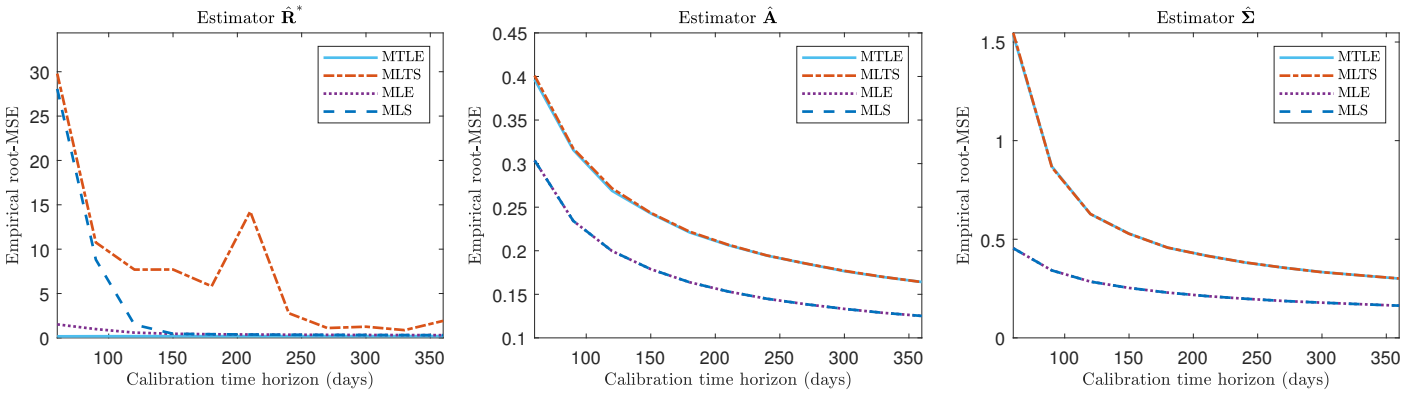


Figure 4. Simulated $\widehat{\text{err}}$ values for $\varepsilon = 0.10$, ncp $= 25$ and bdp $= 0.35$.

In summary, when there are no outliers in the data ($\varepsilon = 0$), the simulation results suggest that MTLE performs on par with the non-robust estimator MLE and, to some extent, MLS, while the advantage of MTLE becomes most pronounced for $\varepsilon > 0$, especially when estimating $\boldsymbol{R}^*$. These observations highlight the need for robust estimators, in particular those able to reliably estimate the long-term mean, as is the case with our MTLE.

Table 4. Simulated $\widehat{\text{err}}$ values for $\varepsilon = 0.10$, ncp $= 25$ and bdp $= 0.35$.

	$\hat{R}^*$				$\hat{A}$				$\hat{\Sigma}$			
T	**MTLE**	**MLTS**	**MLE**	**MLS**	**MTLE**	**MLTS**	**MLE**	**MLS**	**MTLE**	**MLTS**	**MLE**	**MLS**
60	0.18	29.79	1.52	28.06	0.3966	0.4010	0.3037	0.3037	1.5352	1.5462	0.4546	0.4546
90	0.18	10.75	0.99	8.82	0.3156	0.3172	0.2342	0.2342	0.8691	0.8649	0.3420	0.3420
120	0.19	7.70	0.59	1.49	0.2688	0.2716	0.1997	0.1997	0.6275	0.6280	0.2852	0.2852
150	0.20	7.71	0.46	0.46	0.2429	0.2435	0.1788	0.1788	0.5286	0.5276	0.2532	0.2532
180	0.20	5.82	0.41	0.41	0.2211	0.2219	0.1639	0.1639	0.4571	0.4577	0.2296	0.2296
210	0.20	14.24	0.38	0.38	0.2065	0.2069	0.1532	0.1532	0.4173	0.4165	0.2121	0.2121
240	0.20	2.78	0.36	0.36	0.1944	0.1946	0.1449	0.1449	0.3820	0.3828	0.1985	0.1985
270	0.20	1.11	0.34	0.34	0.1853	0.1853	0.1386	0.1386	0.3559	0.3560	0.1872	0.1872
300	0.20	1.26	0.33	0.33	0.1766	0.1770	0.1333	0.1333	0.3336	0.3330	0.1786	0.1786
330	0.20	0.87	0.32	0.32	0.1699	0.1700	0.1287	0.1287	0.3170	0.3170	0.1708	0.1708
360	0.20	1.92	0.31	0.31	0.1641	0.1640	0.1252	0.1252	0.3007	0.3010	0.1634	0.1634

3.2. Examples

Two real-world datasets are studied in this section. In Section 3.2.1, we analyze a bivariate time series of simultaneous US and EU short rates originally studied by Pokojovy et al. (2024). The Vasicek model is calibrated with each of the four methods, MTLE, MLTS, MLS and MLE, on one year's worth of historical data and used to perform a 3-month forecast. Section 3.2.2 builds upon the work of Chang and Shi (2024), who analyze a dataset composed of log-volatilities of six common assets. While the latter paper used a VAR(1) model in their analyses, we rather use the discrete-time Vasicek model and show how MTLE can offer major advantages at estimating the long-term mean parameter.

3.2.1. Daily Treasury and Euro Par Yields

This example combines two publicly available financial datasets, the Daily Treasury Par Yield Curve Rates and the Daily Euro Par Yield Curve Rates, originally presented by Pokojovy et al. (2024). These datasets are available from the US Department of the Treasury (US Department of the Treasury, 2024) and the European Central Bank (European Central Bank, 2024), respectively. The Daily Treasury Par Yield Curve Rates dataset contains daily estimates of US Treasury securities' yield curves, with maturities ranging from 1 month to 30 years, spanning from January 2023 to December 2023. This yield curve serves as a key indicator of the relationship between yield and maturity for US Treasury securities, which are regarded as risk-free investments. Similarly, the Daily Euro Par Yield Curve Rates dataset provides yield data for Euro-denominated government bonds, covering the same maturity range and time period. This dataset includes yields on secondary market-traded bonds, with separate curves for AAA-rated Euro-area central government bonds and all Euro-area central government bonds. The data are updated on target business days, with additional details on zero-coupon, forward and Par Yield curves available on the European Central Bank's website.

In our analysis, we calibrated a multivariate discrete Vasicek model on a bivariate time series constructed from the Daily Treasury and AAA-rated Euro-area 3-month Par Yield data, spanning from 1 January 2023 to 31 December 2023. The 3-month maturity was chosen as a proxy for the short rate. As pointed out by Pokojovy et al. (2024), during this period, the FED Funds rate and the European Central Bank's main refinancing rate remained relatively stable, allowing us to apply the model without adjustments for potential change points. We then assessed the performance of the maximum trimmed likelihood estimator (MTLE), multivariate least trimmed squares (MLTS), maximum likelihood estimator (MLE) and multivariate least squares (MLS) estimators using 3-month maturities from January 2024 to March 2024. For consistency across business days, holidays and weekends were

imputed through linear interpolation and extrapolation. The MTLE and MLTS methods were specifically chosen to protect against potential outliers due to irregularities, sudden shifts and/or continuous drifts in financial data.

In all four panels of Figure 5, historical data are displayed to the left of the vertical dotted line, marking the calibration period (1 January 2023–31 December 2023). To the right side of the dotted line, forecasting results generated based on each of the estimators are plotted, displaying both the predicted mean and the 90% prediction intervals. These projections serve to compare each estimator's capacity to obtain tight prediction bounds while maintaining a desired confidence level. As can be seen, unlike their non-robust competitors, both MTLE and MLTS produced tighter and more upwards-pointing confidence regions due to their ability to filter out outliers and change(s) in trend. Using short rates to price bonds of arbitrary maturities (Mamon, 2004), one could similarly compute value at risk (VaR) or other conventional risk management metrics for these bonds.

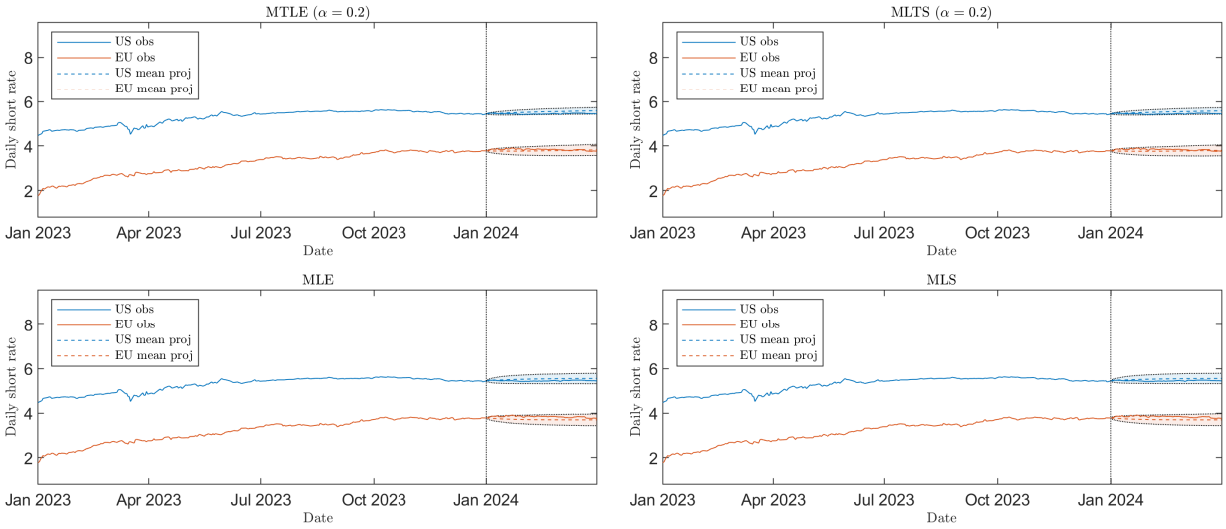


Figure 5. Historic US/EU 3-month rates (1 January 2023–31 12 December 2023) as well as forecasted mean and 90% projection bands (1 January 2024–31 March 2024).

Figure 6 displays contour plots of bivariate density functions of the forecasted short rate distribution on 31 March 2024 under the Vasicek model calibrated using each of the four estimators under comparison. The plots indicate that the distribution of the forecasted short rate appears to be Gaussian, as Equation (6) in Theorem 1 predicts. The figure also suggests that the distributions arising from MLS and MLE estimators are more spread out, indicating greater uncertainty in comparison with the robust competitors (MTLE and MLTS), which perform head-to-head.

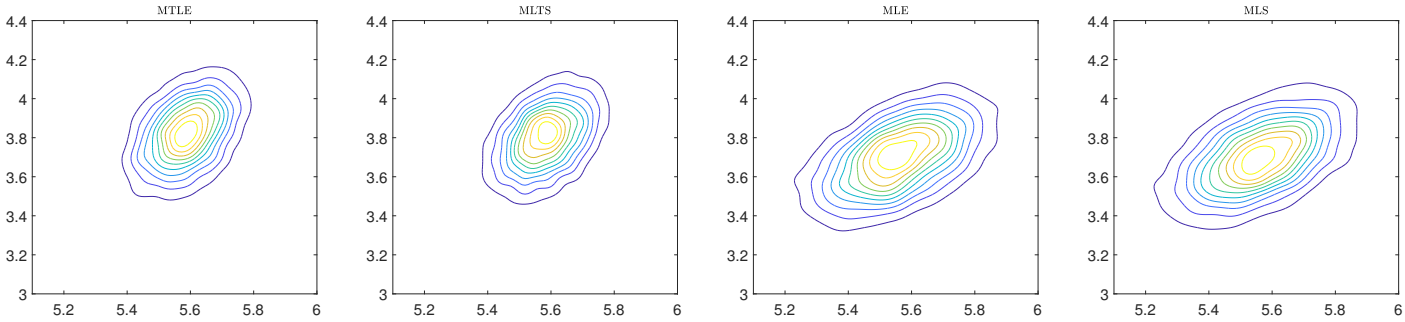


Figure 6. The contour plots of the probability density function of the forecasted short rate R_t distribution on 31 March 2024.

To further evaluate and benchmark model performance, we computed sphered empirical residuals based on parameter estimates given in Table 5. The residuals, after being decorrelated and standardized, are displayed in Figure 7 for each estimator accompanied by 95% prediction circles. Both MLE and MLS identify up to five empirical sphered residuals as outliers and appear to have an identical performance. A comparable performance of these two non-robust estimators can also be seen from the estimates of the parameters reported in Table 5. MTLE and MLTS, on the other hand, labeled more sphered residuals (up to 14 points) as outliers. These two estimators also appear to have similar performance for this example, which is also confirmed by Table 5. The large number of residuals flagged by these methods may indicate the existence of possible model inadequacies, e.g., a heavier-than-Gaussian tail, or could be caused by what is known as "swamping" effects (Jobe & Pokojovy, 2015).

Table 5. Parameter estimates using MTLE (bdp = 0.2), MLTS (bdp = 0.2), MLE and MLS estimators.

	$\hat{R}^*$	$\hat{A}$	$\hat{\Sigma}$
MTLE	$\begin{pmatrix} 5.6890 \\ 4.0105 \end{pmatrix}$	$\begin{pmatrix} 0.0171 & -0.0050 \\ -0.0050 & 0.0167 \end{pmatrix}$	$10^{-3}.\begin{pmatrix} 0.2526 & -0.0396 \\ -0.0396 & 0.6459 \end{pmatrix}$
MLTS	$\begin{pmatrix} 5.6706 \\ 3.9619 \end{pmatrix}$	$\begin{pmatrix} 0.0170 & -0.0049 \\ -0.0049 & 0.0175 \end{pmatrix}$	$10^{-3}.\begin{pmatrix} 0.2548 & -0.0427 \\ -0.0427 & 0.6359 \end{pmatrix}$
MLE	$\begin{pmatrix} 5.5689 \\ 3.7141 \end{pmatrix}$	$\begin{pmatrix} 0.0299 & -0.0116 \\ -0.0116 & 0.0233 \end{pmatrix}$	$10^{-3}.\begin{pmatrix} 0.9193 & -0.0435 \\ -0.0435 & 0.7551 \end{pmatrix}$
MLS	$\begin{pmatrix} 5.5705 \\ 3.7171 \end{pmatrix}$	$\begin{pmatrix} 0.5269 & -0.0116 \\ -0.0116 & 0.0232 \end{pmatrix}$	$10^{-3}.\begin{pmatrix} 0.9193 & -0.0435 \\ -0.0435 & 0.7551 \end{pmatrix}$

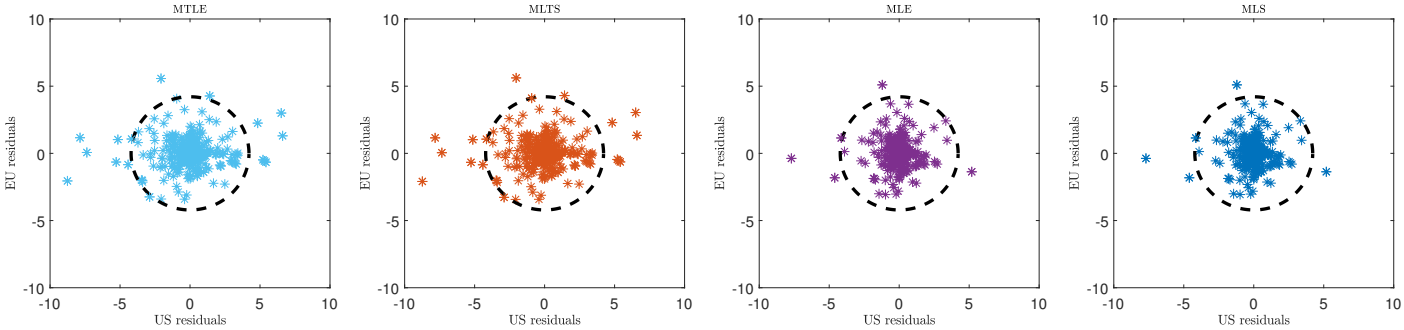


Figure 7. Sphered empirical residuals for MTLE (bdp = 0.2), MLTS (bdp = 0.2), MLE and MLS estimators with respective 95% prediction circles.

For additional performance assessment, we report empirical root mean squared error (RMSE)

$$\widehat{\mathrm{MSE}}(t_j) = \frac{1}{10,000} \sum_{i=1}^{10,000} \left(\left| \hat{R}^{\mathrm{US}}_{t_j,i} - R^{\mathrm{US}}_{t_j,\mathrm{obv}} \right|^2 + \left| \hat{R}^{\mathrm{EU}}_{t_j,i} - R^{\mathrm{EU}}_{t_j,\mathrm{obv}} \right|^2 \right), \tag{27}$$

in the left-hand panel and the mean absolute percentage error (MAPE)

$$\widehat{\mathrm{MAPE}}(t_j) = \frac{1}{10,000} \sum_{i=1}^{10,000} \left(\frac{\left| \hat{R}^{\mathrm{US}}_{t_j,i} - R^{\mathrm{US}}_{t_j,\mathrm{obv}} \right|}{\left| R^{\mathrm{US}}_{t_j,\mathrm{obv}} \right|} + \frac{\left| \hat{R}^{\mathrm{EU}}_{t_j,i} - R^{\mathrm{EU}}_{t_j,\mathrm{obv}} \right|}{\left| R^{\mathrm{EU}}_{t_j,\mathrm{obv}} \right|} \right), \tag{28}$$

in the right-hand panel of Figure 8, calculated from a Monte Carlo simulation of size $N = 10,000$, where $R_{t_j,\mathrm{obv}}$ represents the observed US and EU rates and $\hat{R}_{t_j,i}$ denotes forecasted rates from each simulation run.

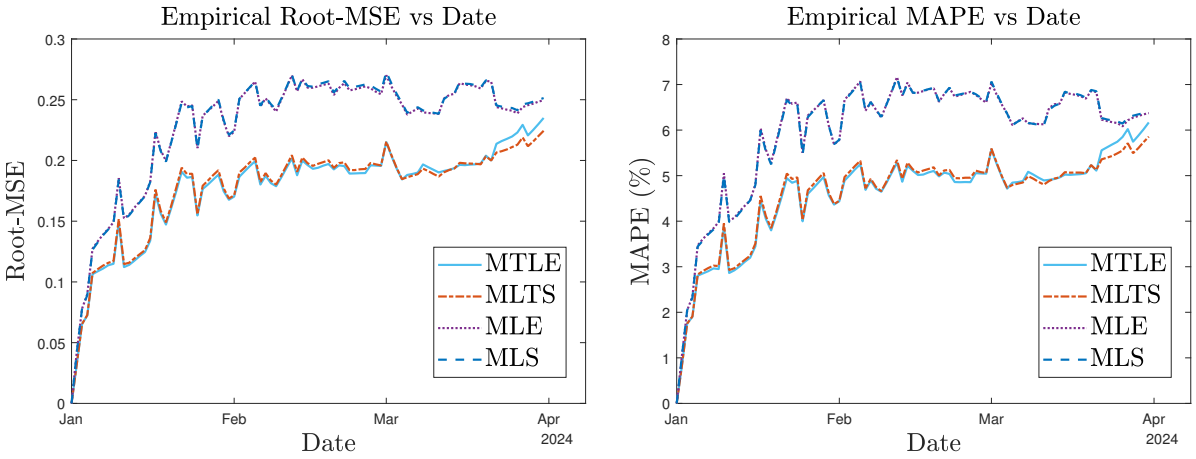


Figure 8. Empirical backtesting root-MSE and MAPE using MTLE, MLTS, MLE and MLS estimators.

Both MTLE and MLTS estimators were tuned at a nominal breakdown point of bdp $= 0.2$ and used, alongside MLE and MLS, to compute the projected US and EU rates. The results from this backtesting analysis, which are reported in Figure 8, further buttress the advantage of the robust estimators over the classical ones seen in the previous figure. MTLE and MLTS perform head-to-head and outperform MLE and MLS, however, not tremendously in this example.

As briefly described in Section 1, with robustly estimated Vasicek parameters at hand, a wide variety of practically relevant problems can be addressed. For example, starting in a steady state, a robust multivariate Shewhart $\boldsymbol{X}$-chart (Mason & Young, 2002) can be constructed to detect shifts in the long-term mean based on a Hotelling-like signal statistic

$$T^2_{t_k} = (\boldsymbol{R}_{t_k} - \hat{\boldsymbol{R}}^*)' \hat{\boldsymbol{\Psi}}^{-1} (\boldsymbol{R}_{t_k} - \hat{\boldsymbol{R}}^*)$$

with the asymptotic covariance matrix

$$\begin{aligned} \hat{\boldsymbol{\Psi}} &= (\Delta t) \sum_{n=0}^{\infty} \Big((\boldsymbol{I}_{p\times p} - (\Delta t)\hat{\boldsymbol{A}}) \hat{\boldsymbol{\Sigma}} (\boldsymbol{I}_{p\times p} - (\Delta t)\hat{\boldsymbol{A}})' \Big)^n \\ &= (\Delta t) \Big(\boldsymbol{I}_{p\times p} - \big(\boldsymbol{I}_{p\times p} - (\Delta t)\hat{\boldsymbol{A}}\big) \hat{\boldsymbol{\Sigma}} \big(\boldsymbol{I}_{p\times p} - (\Delta t)\hat{\boldsymbol{A}}'\big) \Big)^{-1} \end{aligned}$$

where $\Delta t > 0$ is assumed sufficiently small so that the latter geometric series (cf. Theorem 1) converges with respect to a suitable $\hat{\boldsymbol{\Sigma}}^{1/2}$-weighted operator norm. The well-known Stahel–Donoho estimator has recently been studied in the context of i.i.d. data by Raji et al. (2021). Since $\boldsymbol{R}_t$s are autocorrelated, a covariance correction due to Grimshaw (2023) can be adopted to further enhance the performance of the chart. Fixing the in-control average run length (IC ARL) at a desired level, e.g., IC ARL $= 365$ days, the chart would issue a false alarm every IC ARL time period, e.g., 365 days, on average provided the long-term mean remains unchanged, while being sensitive to actual shifts in the latter parameter. This robust monitoring scheme can be used to detect credit regimes for portfolio rebalancing or used as a model risk management tool, etc. With MTLE outperforming all competitors at estimating $\boldsymbol{R}^*$ as demonstrated in this paper, we expect this chart to be superior to its MLS-, MLE- and MLTS-based competitors.

Another promising application field lies in optimal portfolio allocation. Multivariate robust portfolio optimization extends the classical mean–variance framework (Markowitz, 1952) by incorporating statistical robustness into portfolio selection, particularly for bond portfolios influenced by multi-factor term structures (Korn & Koziol, 2006). In this approach, the expected portfolio return is modeled as a vector $\boldsymbol{R}_r = \boldsymbol{w}'\boldsymbol{\mu}$, where $\boldsymbol{w}$ is the weight

allocation across bonds and $\boldsymbol{\mu}_r$ is the vector of expected returns. Similarly, the portfolio risk is captured by a full covariance matrix $\boldsymbol{\Sigma}_r = \boldsymbol{w}'\boldsymbol{\Sigma}_r\boldsymbol{w}$ preserving the multi-dimensional dependencies among different bonds and interest rate factors. A key innovation in this robust framework is the incorporation of a robust estimator, such as our proposed MTLE, to estimate the multivariate Vasicek model parameters, ensuring that outlier contamination and model violations do not distort the optimization process. Unlike traditional estimation methods that are sensitive to anomalies in financial data, the robust approach leverages a trimmed likelihood technique to provide more reliable estimates of the long-term mean and covariance structure. This ensures that the portfolio optimization remains stable and efficient even in the presence of extreme market conditions. The final optimization problem seeks to maximize the robust Sharpe ratio while accounting for these robustly estimated parameters, leading to a more resilient and theoretically grounded bond portfolio allocation.

As a simple illustration, we treat the aforedescribed short-term US and EU bonds as a multi-currency cash account. After a daily compounding ($\Delta t = 1$ day) until the time period $t_n = (\Delta t)n$, the originally deposited USD 1 and EUR 1 will produce the returns

$$\Big(\prod_{l=0}^{n}\big(1+(\Delta t)R^{\mathrm{US}}_{t_l}\big)-1\Big) \quad \text{and} \quad \Big(\prod_{l=0}^{n}\big(1+(\Delta t)R^{\mathrm{EU}}_{t_l}\big)-1\Big),$$

respectively. For simplicity, we assume the exchange rate is nearly constant (should this assumption be violated, the Vasicek model can be recalibrated on historical short data after adjusting them for variable exchange rate) and construct a 90-day portfolio (starting on 1 January 2024) with a minimum return rate of 1.1% maximizing the Sharpe rate. Using Equation (3) to simulate future returns, the mean vector and covariance matrix of the returns can be computed under each of the four choices of the estimators. Excluding borrowing and short-selling, the usual quadratic programming problem

$$\boldsymbol{w}'\boldsymbol{\Sigma}_r\boldsymbol{w} \to \min \text{ over } \boldsymbol{w} \in [0,1]^2 \text{ such that } \mathbb{1}_2'\boldsymbol{w} = 1 \text{ and } \boldsymbol{w}'\boldsymbol{\mu}_r \geq 0.011 \tag{29}$$

can be easily solved to obtain the optimal weights (Goldfarb & Idnani, 1983). Optimal portfolio weights, estimated mean vector and covariance of returns as well as rate of returns are reported in Table 6. As can be seen, even despite the very short 90-day investment horizon, the robust methodologies (MTLE and MLTS) offer a rate superior to non-robust counterparts, with our proposed MTLE outperforming all competitors. Though beyond the scope of this work, studying inherently riskier bonds of longer maturities (Korn & Koziol, 2006) in lieu of a cash account, one would naturally expect robust portfolios to exhibit even stronger dominance.

Table 6. Markowitz-style optimal portfolios consisting of US Treasuries and 3-month Euro bonds based on returns obtained from MTLE, MLTS, MLE and MLS estimators with a target rate of return of 1.1% over a 90-day time period starting on 1 January 2024.

	Estimated Mean Vector of Returns μ_r	**Estimated Covariance Matrix of Returns Σ_r**	**Optimal US/EU Weights**	**Optimal Rate of Return**
MTLE	$\begin{pmatrix}0.0137\\0.0094\end{pmatrix}$	$10^{-7}.\begin{pmatrix}0.1964 & 0.1338\\0.1338 & 0.5286\end{pmatrix}$	$\begin{pmatrix}0.5036\\0.4964\end{pmatrix}$	1.1571%
MLTS	$\begin{pmatrix}0.0137\\0.0093\end{pmatrix}$	$10^{-7}.\begin{pmatrix}0.1991 & 0.1340\\0.1340 & 0.5059\end{pmatrix}$	$\begin{pmatrix}0.5036\\0.4964\end{pmatrix}$	1.1550%
MLE	$\begin{pmatrix}0.0137\\0.0092\end{pmatrix}$	$10^{-7}.\begin{pmatrix}0.4588 & 0.3192\\0.3192 & 0.5900\end{pmatrix}$	$\begin{pmatrix}0.5037\\0.4963\end{pmatrix}$	1.1476%
MLS	$\begin{pmatrix}0.0137\\0.0092\end{pmatrix}$	$10^{-7}.\begin{pmatrix}0.4548 & 0.3170\\0.3170 & 0.5815\end{pmatrix}$	$\begin{pmatrix}0.5037\\0.4963\end{pmatrix}$	1.1472%

3.2.2. Commodities and Currencies

This example aims to revisit the analysis of six common asset (commodities and currencies) prices (expressed in US dollars), viz., the futures of gold (XAU), silver (XAG), Brent oil (BRE) and West Texas Intermediate oil (WTI) and the currencies of Swiss Francs (CHF) and Japanese Yen (JPY), ranging from July 2017 to June 2020, originally performed by Chang and Shi (2024). There are 771 records for each asset. To demonstrate how the multivariate Vasicek model can be applied to stochastic volatilities, we analyzed the (transformed) daily log-volatilities of the aforementioned asset prices directly provided by the authors of Chang and Shi (2024). The daily volatility is calculated as the root of summation of squared hourly close prices (Chang & Shi, 2024). Figure 9 plots the daily logged volatilities over the entire range considered.

The sextivariate time series appears to be stationary from July 2017 until the early months of 2020, when the volatilities (i.e., volatilities-of-(log-)volatility) start to rise through the third month of 2020. There is a strong indication of a potential shift in the "long-term" mean of the multivariate process. The shift can probably be attributed to a spike in volatility associated with the onset of the COVID-19 pandemic. One can also observe from the plots that the increased volatilities in all of the six time series appear to revert or head down toward their normal levels. In general, the currencies appear to have lower volatilities than the oil prices.

While Chang and Shi (2024) thoroughly investigated how robust estimation using MLTS can improve estimation and inference for the mean-reversion speed $\boldsymbol{A}$ and the squared volatility matrix $\boldsymbol{\Sigma}$ in the presence of outliers, we solely focus on long-term mean estimation in our analyses as our MTLE is known to perform head-to-head with MLTS when applied to the other two parameters. Thus, all advantages of MLTS reported by Chang and Shi (2024) are shared by our MTLE.

Since no "ground truth" value is known for the long-term mean $\boldsymbol{R}^*$ for our time series, we chose to perform a different type of experiment. To this end, we created different subsets of daily log-volatilities for each of the assets and estimated $\boldsymbol{R}^*$ on each subset. The subsets were created in the same fashion moving averages are computed, i.e., by "sliding" across the dataset based on some given window size w. Choosing $w = 50$, we formed the following $(n - w + 1)$ contiguous subsets:

$$\{x_{(1)}, \ldots, x_{(w)}\}, \{x_{(2)}, \ldots, x_{(w+1)}\}, \{x_{(3)}, \ldots, x_{(w+2)}\}, \ldots, \{x_{(n-w+1)}, \ldots, x_{(n)}\}.$$

Both robust estimators, MTLE and MLTS, tuned with bdp $= 0.25$ and non-robust MLE and MLS estimators were employed to estimate $\boldsymbol{R}^*$ from each of the subsets. The resulting moving "long-term means" for the log-volatilities of the assets are plotted vs. subset index in Figure 10. The plots show that MLTS can be seen to be overly volatile and "exploded" on many occasions. The lengthy upward and downward "candles" or spikes observed in the plots are indicative of occasions on which the MLTS fails as an adequate statistical estimator. Our MTLE, on the other hand, exhibits good statistical robustness and moderate variability compared to the other estimators. Through the end of 2019 (up to subset index 595), MTLE is essentially following the other non-robust estimators, suggesting only minor fluctuations in the first two moments of the log-volatility process. In contrast, starting in early 2020 (after subset index 595), MTLE starts to exhibit moderate oscillations, which could be indicative of possible changes in the "long-term" mean and/or the squared volatility of the process. In summary, MTLE is clearly preferred over MLTS. Additionally, some evidence exists that MTLE should be chosen over MLE and MLS starting in early 2020 to better capture a shift in process parameters for this dataset.

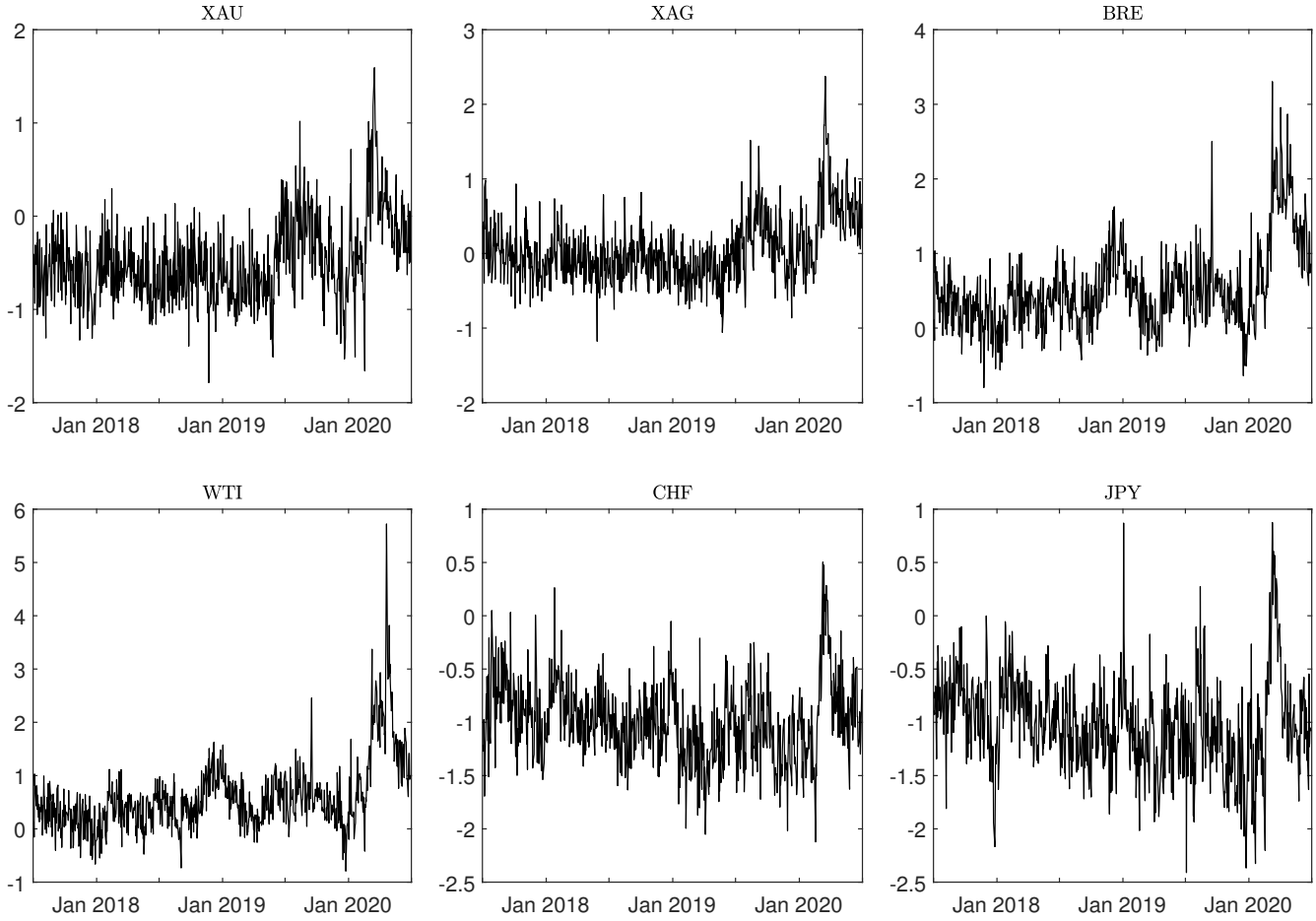


Figure 9. Daily logged volatilities: July 2017–June 2020.

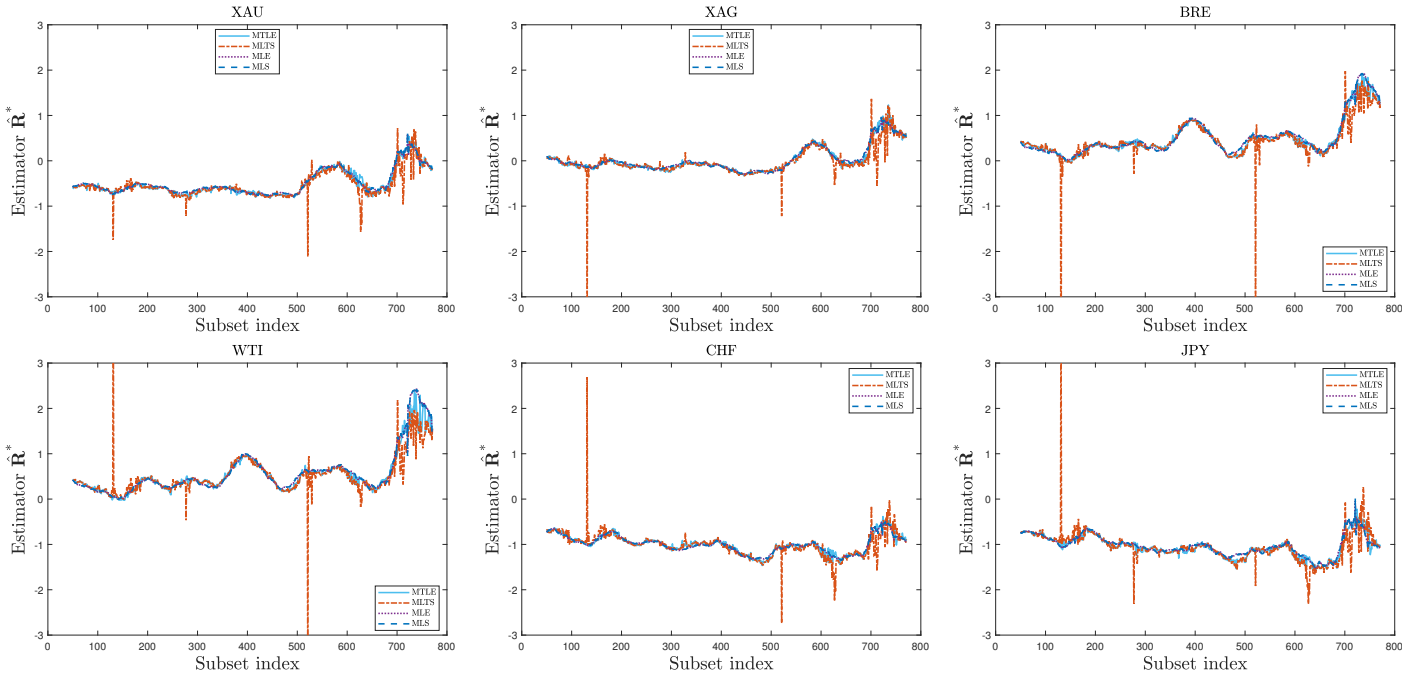


Figure 10. Estimates of $\boldsymbol{R}^*$ for daily log-volatilities with $w = 50$.

4. Conclusions

Modeling econometric and financial processes is essential for analyzing and forecasting dynamic interactions between multiple potentially useful economic and financial variables.

The multivariate Vasicek model, oftentimes reparametrized as a VAR(1) model, is widely used as a quantitative description of the mean-reverting dynamics typical for short rates, stochastic log-volatilities and other macroeconomic processes. Numerous studies have emphasized the importance of statistical robustness in econometric modeling and why the results obtained with the usual MLS can be flawed in the presence of outliers, heavy tails and other model violations.

Adopting the maximum trimmed likelihood estimation (MTLE) approach to account for these type of model inadequacies, we derived a discrete-continuous optimization problem underlying the MTLE estimator and developed an iterative procedure to solve for the maximum of the trimmed likelihood. We applied our method to the multivariate Vasicek model for both synthetic and real-world datasets in the context of short rates and stochastic log-volatilities. Empirical evidence in this study indicates that MTLE, at minimum, performs head-to-head and oftentimes outperforms MLS and MLE on datasets with outliers and MLTS, especially at estimating the long-term mean. Thus, this work provides and thoroughly evaluates a new econometric tool for estimating the parameters of the multivariate Vasicek model. Our proposed methodology is applicable under broad circumstances and can be helpful to investors and analysts across a wide variety of use cases in portfolio optimization, risk management and other types of decision making. With a primary focus on robustness instead of improved statistical efficiency, unreweighted "raw" estimators were compared in this study.

Our findings generally confirm what is known about MLTS in terms of robust calibration of VAR(1) models. However, while MLTS and existing methodologies are successful at estimating the mean-reversion speed matrix $\boldsymbol{A}$ and covariance $\boldsymbol{\Sigma}$, we documented major deficiencies of existing methodologies at estimating the long-term mean, especially in smaller samples. Using MLTS as a "primer" estimator to produce a warm start, we were able to alleviate this deficiency by adopting a regularized trimmed EM-type iteration to compute MTLE estimates. This underpins the robust nature of MLTS manifest in its original ability to find an outlier-free bulk set while explaining the role of our MTLE in improving the estimation accuracy of the long-term mean in the presence of outliers and ill-conditioned mean-reversion speed matrices.

In future work, a reweighted version of MTLE can be developed in a similar fashion as its MLTS counterpart (Agulló et al., 2008; Chang & Shi, 2024). Other potential extensions include high-dimensional time series, missing data imputation, statistical process monitoring and change point detection, among other problems.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/economies13030068/s1 and https://github.com/AndrewsJunior/Vasicek-MTLE (accessed on 25 February 2025), Supplementary MATLAB® codes; supplementary figures and tables (PDF).

Author Contributions: Conceptualization, T.M.F.J., M.P. and A.T.A.; methodology, M.P.; software, M.P. and A.T.A.; validation, M.P., A.T.A. and E.N.; formal analysis, T.M.F.J., M.P. and A.T.A.; investigation, T.M.F.J., M.P., A.T.A. and E.N.; resources, T.M.F.J.; data curation, A.T.A. and E.N.; writing—original draft preparation, M.P., A.T.A. and E.N.; writing—review and editing, T.M.F.J., M.P., A.T.A. and E.N.; visualization, A.T.A. and E.N.; supervision, T.M.F.J. and M.P.; project administration, T.M.F.J. and M.P.; funding acquisition, T.M.F.J. and M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the National Science Foundation (DMS-2210929); (ERC-ASPIRE-1941524); (DUE-2216396) and Department of Education (Award #P116S210004). Additional partial funding support was provided by El Paso Water (EPW 226621069A); Rio Grande Council of Governments (COG 226621071A); RAIZ FCU (RAIZ 2024E); and UTEP Center for the Study of Western Hemispheric Trade (BRMP 2024E).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The short-rate data reported in this work are available at https://github.com/AndrewsJunior/Vasicek-MTLE (accessed on 25 February 2025). The dataset containing the six assets can be requested from the authors of the paper Chang and Shi (2024).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

bdp	Breakdown point
CLT	Central Limit Theorem
COVID	Coronavirus disease
C-step	Concentration step
EU	European Union
LTS	Least trimmed squares
MCD	Minimum covariance determinant
MLE	Maximum likelihood estimator
MLS	Multivariate least squares
MLTS	Multivariate least trimmed squares
MM	Modified maximum likelihood-type
MSE	Mean squared error
MTLE	Maximum trimmed likelihood estimation
ncp	Non-centrality parameter
RMLTS	Reweighted multivariate least trimmed squares
RMSE	Root mean squared error
SDE	Stochastic differential equation
US	United States
VAR	Vector autoregression

Appendix A. Asymptotic Consistency

Two asymptotic consistency results are presented in this Appendix. Theorem A1 describes the large-sample asymptotics of MTLE, assuming respective MLTS estimators are consistent. Exploiting the large-time behavior of the discrete Vasicek process, Theorem A2 gives an alternative statement under milder assumptions, solely relying on the large-time asymptotics of the Vasicek process. This offers a theoretical explanation of the empirically observed efficacy of our approach (cf. Sections 3.1 and 3.2).

To show the latter result, assuming the matrix $\left(\boldsymbol{I}_{p\times p} - (\Delta t)\boldsymbol{A}\right)$ is strictly contractive, the idea is to study the estimating Equation (22)

$$\boldsymbol{R}^*_{\mathrm{MTLE}} = \frac{1}{h}\sum_{j\in\mathcal{I}_\dagger} \boldsymbol{R}_{t_j} + \frac{1}{(1-\alpha)T}\boldsymbol{A}^{-1}_{\mathrm{MTLE}}\sum_{j\in\mathcal{I}_\dagger}\left(\boldsymbol{R}_{t_{j+1}} - \boldsymbol{R}_{t_j}\right) \tag{A1}$$

from Section 2.2 (with $\boldsymbol{A}^{-1}$ denoting the Moore–Penrose pseudoinverse). Indeed, arguing heuristically, if the latter term can be shown to vanish, which one would intuitively expect in view of asymptotic stability, Theorem 1 readily furnishes the convergence of $\boldsymbol{R}^*_{\mathrm{MTLE}}$ to the (nominal and actual) long-term mean $\boldsymbol{R}^*$.

Theorem A1. *Let $T = (\Delta t)n$ be a time horizon. Further, let the matrix $\boldsymbol{A}$ be non-singular. Assume the MLTS estimators $\hat{\boldsymbol{A}}_{\text{MLTS}}$, $\hat{\boldsymbol{\Sigma}}_{\text{MLTS}}$ and $\widehat{\boldsymbol{AR}^*}_{\text{MLTS}}$ (and, thus, their MTLE counterparts) are consistent under the uncontaminated model with*

$$\varepsilon^2 := \text{MSE}(\hat{\boldsymbol{A}}_{\text{MLTS}}) + \text{MSE}(\hat{\boldsymbol{\Sigma}}_{\text{MLTS}}) + \text{MSE}(\widehat{\boldsymbol{AR}^*}_{\text{MLTS}}) \to 0 \text{ as } n \to \infty \text{ and } T \to \infty.$$

Then, the MTLE estimator of the long-term mean satisfies

$$\hat{\boldsymbol{R}}^*_{\text{MTLE}} = \boldsymbol{R}^* + O_{\mathbb{P}}(\varepsilon) + O_{\mathbb{P}}(n^{-1/2}) + O_{\mathbb{P}}(T^{-1/2}) \text{ as } n \to \infty \text{ and } T \to \infty.$$

Proof. The estimating Equation (A1) for $\hat{\boldsymbol{R}}^*$ can be written as

$$\hat{\boldsymbol{R}}^*_{\text{MTLE}} = \frac{1}{h}\sum_{j=0}^{n-1} w_j \boldsymbol{R}_{t_j} + \frac{1}{(1-\alpha)T}\hat{\boldsymbol{A}}^{-1}_{\text{MTLE}} \sum_{j=0}^{n-1} w_j (\boldsymbol{R}_{t_{j+1}} - \boldsymbol{R}_{t_j}) \tag{A2}$$

with

$$w_j := \mathbb{1}_{\{j \in \mathcal{I}_+\}}.$$

Observing that $j \in \mathcal{I}_+$ holds if and only if

$$d^2\big(\Delta \boldsymbol{R}_{t_j} - (\Delta t)\boldsymbol{A}(\boldsymbol{R}^* - \boldsymbol{R}_{t_j}) | \boldsymbol{0}_p, (\Delta t)\boldsymbol{\Sigma}\big) \le \chi^2_{p,1-\alpha} + O_{\mathbb{P}}(\varepsilon)$$

as $\varepsilon \to 0$ with the Portmanteau theorem applied to the estimators[1] of $\boldsymbol{AR}^*$, $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$, the weights

$$\begin{aligned} w_j &:= \mathbb{1}_{[0,\chi^2_{p,1-\alpha}]}\Big(d^2\big(\Delta \boldsymbol{R}_{t_j} - (\Delta t)\boldsymbol{A}(\boldsymbol{R}^* - \boldsymbol{R}_{t_j}) | \boldsymbol{0}_p, (\Delta t)\boldsymbol{\Sigma}\big)\Big) \\ &\overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(1-\alpha) + O_{\mathbb{P}}(\varepsilon) \text{ as } \varepsilon \to 0 \end{aligned}$$

are asymptotically independent of $\boldsymbol{R}_{t_j}$s.

Using Equation (2), we obtain

$$\boldsymbol{R}_{t_{j+1}} - \boldsymbol{R}_{t_j} = (\Delta t)\boldsymbol{A}\big(\boldsymbol{R}^* - \boldsymbol{R}_{t_j}\big) + \boldsymbol{\varepsilon}_{t_j} \quad \text{for } j = 0, \ldots, n-1.$$

with

$$\boldsymbol{\varepsilon}_{t_j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\big(\boldsymbol{0}_p, (\Delta t)\boldsymbol{\Sigma}\big).$$

Plugging this into Equation (A2), we further find

$$\begin{aligned} \hat{\boldsymbol{R}}^*_{\text{MTLE}} &= \frac{1}{h}\sum_{j=0}^{n-1} w_j \boldsymbol{R}_{t_j} - \frac{\Delta t}{(1-\alpha)T}\sum_{j=0}^{n-1} w_j (\boldsymbol{R}_{t_j} - \boldsymbol{R}^*) + \hat{\boldsymbol{A}}^{-1}_{\text{MTLE}} \frac{1}{(1-\alpha)T}\sum_{j=0}^{n-1} w_j \boldsymbol{\varepsilon}_{t_j} \\ &= \Big(\frac{1}{h}\sum_{j=0}^{n-1} w_j\Big)\boldsymbol{R}^* + \hat{\boldsymbol{A}}^{-1}_{\text{MTLE}} \frac{1}{(1-\alpha)T}\sum_{j=0}^{n-1} w_j \boldsymbol{\varepsilon}_{t_j} \end{aligned} \tag{A3}$$

since $(\Delta t)h = (1-\alpha)T$. Applying the Central Limit Theorem (CLT) to the first term in Equation (A3) and observing

$$\hat{\boldsymbol{A}}^{-1} = O_{\mathbb{P}}(1) \text{ as } n \to \infty \text{ and } T \to \infty,$$

taking into account the consistency[2] of $\hat{\boldsymbol{A}}_{\text{MLTS}}$ as well as calculating

$$\text{Cov}\Big[\frac{1}{(1-\alpha)T}\sum_{j=0}^{n-1} w_j \boldsymbol{\varepsilon}_{t_j}\Big] = \frac{1}{(1-\alpha)T}\boldsymbol{\Sigma} = O_{\mathbb{P}}(T^{-1/2})$$

to estimate the second term, the claim follows. □

Theorem A2. *Let $\boldsymbol{A}$ be nonsingular and positive definite. Further, assume the pseudoinverse satisfies $\boldsymbol{A}_{\mathrm{MLTS}}^{-1} = O_{\mathbb{P}}(1)$ uniformly in $\Delta t \to 0$ as $T \to \infty$. Then, for any $\varepsilon > 0$, the MTLE estimator satisfies*

$$\hat{\boldsymbol{R}}_{\mathrm{MTLE}}^{*} = \boldsymbol{R}^{*} + O_{\mathbb{P}}(\varepsilon) \text{ for } T \geq \underline{T}(\varepsilon) \text{ uniformly in } \Delta t \to 0.$$

Proof. Under the conditions of the theorem, the discrete Vasicek process is asymptotically stable uniformly in $\Delta t \to 0$ (Horváth-Bokor & Mitsui, 2004), i.e.,

$$\boldsymbol{R}_{t_j} \to \boldsymbol{R}^{*} \ \mathbb{P}\text{-almost surely as } t_j \to \infty \text{ uniformly in } \Delta t \to 0.$$

Thus, for any $\varepsilon > 0$,

$$\lim_{\tilde{T}\to\infty} \mathbb{P}\Big\{ \sup_{t_j \geq \tilde{T}} \|\boldsymbol{R}_{t_j}\| \leq \varepsilon \Big\} = 1.$$

Further, we can write

$$\begin{aligned}\frac{1}{T}\sum_{j=0}^{n-1} w_j(\boldsymbol{R}_{t_{j+1}} - \boldsymbol{R}_{t_j}) &= \frac{1}{T}\sum_{j=0}^{n-1} \mathbb{1}_{\{t_j<\tilde{T}\}} w_j(\boldsymbol{R}_{t_{j+1}} - \boldsymbol{R}_{t_j}) + \frac{1}{T}\sum_{j=0}^{n-1} \mathbb{1}_{\{t_j\geq\tilde{T}\}} w_j(\boldsymbol{R}_{t_{j+1}} - \boldsymbol{R}_{t_j}) \\ &\equiv S_1 + S_2.\end{aligned}$$

For the former term, S_1, using Equation (6), we estimate

$$\begin{aligned}\mathbb{E}\big[\|S_1\|^2\big] &\leq \frac{1}{T^2}\mathbb{E}\Big[\Big\| \sum_{t_j\leq\tilde{T}} \boldsymbol{R}_{t_j}\Big\|^2\Big] = \frac{\tilde{T}}{T^2}\sum_{t_j\leq\tilde{T}} \mathbb{E}\Big[\big\|\boldsymbol{R}_{t_j}\big\|^2\Big] \\ &\leq \frac{\tilde{T}}{T^2}\sum_{t_j\leq\tilde{T}} \mathbb{E}\Big[\operatorname{tr}\operatorname{Cov}[\boldsymbol{R}_{t_j}] + \big\|\mathbb{E}[\boldsymbol{R}_{t_j}]\big\|^2\Big] \leq \frac{\tilde{T}}{T^2}\big(C + (\Delta t)\operatorname{tr}(\Sigma)\tilde{T}\big) = O\Big(\frac{\tilde{T}^2}{T^2}\Big)\end{aligned}$$

with $C := \big(\|\boldsymbol{R}^{*}\| + \|\boldsymbol{R}^{0}\|\big)^2 > 0$. For the latter term, S_2, due to asymptotic stability,

$$\begin{aligned}\mathbb{E}\big[\|S_2\|^2\big] &\leq \frac{4\varepsilon^2}{T^2}\mathbb{E}\Big[\Big\|\sum_{j=0}^{n-1} \mathbb{1}_{\{t_j\geq\tilde{T}\}} w_j\Big\|^2\Big] \\ &= \frac{4\varepsilon^2}{T^2}\operatorname{Var}\Big[\sum_{j=0}^{n-1} \mathbb{1}_{\{t_j\geq\tilde{T}\}} w_j\Big] + \frac{4\varepsilon^2}{T^2}\mathbb{E}^2\Big[\sum_{j=0}^{n-1} \mathbb{1}_{\{t_j\geq\tilde{T}\}} w_j\Big] = O\Big(\varepsilon^2\frac{(T-\tilde{T})^2}{T^2}\Big) = O(\varepsilon^2)\end{aligned}$$

with probability converging to 1 as $\tilde{T} \to \infty$. Thus, choosing $T > 0$ sufficiently large such that $\frac{\tilde{T}}{T} = O(\varepsilon)$ holds true, we have

$$\mathbb{E}\Big[\Big\|\frac{1}{T}\sum_{j=0}^{n-1} w_j(\boldsymbol{R}_{t_{j+1}} - \boldsymbol{R}_{t_j})\Big\|^2\Big] = O(\varepsilon^2)$$

with probability converging to 1 as $\tilde{T} \to \infty$. With a similar argumentation,

$$\mathbb{E}\Big[\Big\|\frac{1}{h}\sum_{j\in\mathcal{I}^{\dagger}} \boldsymbol{R}_{t_j} - \boldsymbol{R}^{*}\Big\|^2\Big] = O(\varepsilon^2).$$

Plugging the last two equations into Equation (A2), the claim follows. □

Remark A1. *For non-robust MLE (i.e., MTLE with $\alpha = 0$), using the telescope sum (cf. Pokojovy et al. (2024)) and Theorem 1, an explicit convergence rate can be established in Theorem A2:*

$$\hat{\boldsymbol{R}}^*_{\mathrm{MTLE}} = \boldsymbol{R}^* + O_{\mathbb{P}}(T^{-1/2}) \text{ for } T \to \infty \text{ uniformly in } \Delta t \to 0.$$

Notes

[1] The estimator for $\boldsymbol{R}^*$ itself does not need to be consistent.

[2] Note that $\hat{\boldsymbol{A}}^{-1}$ is empirically well behaved since a truncated pseudoinverse is used.

References

Agulló, J., Croux, C., & Van Aelst, S. (2008). The multivariate least-trimmed squares estimator. *Journal of Multivariate Analysis*, *99*(3), 311–338. [CrossRef]

Aït-Sahalia, Y. (1995). Nonparametric pricing of interest rate derivative securities. *Econometrica*, *64*(3), 527–560. [CrossRef]

Boudt, K., Rousseeuw, P. J., Vanduffel, S., & Verdonck, T. (2020). The minimum regularized covariance determinant estimator. *Statistics and Computing*, *30*(1), 113–128. [CrossRef]

Brockwell, P. J., & Davis, R. A. (2002). *Introduction to time series and forecasting* (2nd ed.). Springer-Verlag.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1996). *The econometrics of financial markets*. Princeton University Press.

Chang, L., & Shi, Y. (2024). A discussion on the robust vector autoregressive models: Novel evidence from safe haven assets. *Annals of Operations Research*, *339*(3), 1725–1755. [CrossRef] [PubMed]

Croux, C., & Joossens, K. (2008, August 24–29). *Robust estimation of the vector autoregressive model by a least trimmed squares procedure*. COMPSTAT 2008: Proceedings in Computational Statistics COMPSTAT 2008 (pp. 489–501), Porto, Portugal.

Egorov, A. V., Li, H., & Ng, D. (2011). A tale of two yield curves: Modeling the joint term structure of dollar and euro interest rates. *Journal of Econometrics*, *162*(1), 55–70. [CrossRef]

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *Journal of Business & Economic Statistics*, *19*(2), 177–191.

Eraker, B. (2008). A Bayesian view of temporary components in asset prices. *Journal of Empirical Finance*, *15*(3), 503–517. [CrossRef]

European Central Bank. (2024, November). *Euro area yield curves*. Available online: https://www.ecb.europa.eu/stats/financial_markets_and_interest_rates/euro_area_yield_curves/html/index.en.html (accessed on 1 November 2024).

Fabozzi, F. J., & Mann, S. V. (2011). *Handbook of fixed income securities* (8th ed.). McGraw Hill Professional.

Gallant, A. R., & Tauchen, G. (1996). Which moments to match? *Econometric Theory*, *12*(4), 657–681. [CrossRef]

Gallant, A. R., & Tauchen, G. (2010). Simulated score methods and indirect inference for continuous-time models. In *Handbook of financial econometrics: Tools and techniques* (pp. 427–477). Elsevier.

Goldfarb, D., & Idnani, A. (1983). A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming*, *27*(1), 1–33. [CrossRef]

Gourieroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, *8*(S1), S85–S118. [CrossRef]

Grimshaw, S. D. (2023). Constructing control charts for autocorrelated data using an exhaustive systematic samples pooled variance estimator. *Journal of Quality Technology*, *55*(3), 349–365. [CrossRef]

Hamilton, J. D. (2020). *Time series analysis*. Princeton University Press.

Horváth-Bokor, R., & Mitsui, T. (2004). *On mean-square and asymptotic stability for numerical approximations of stochastic ordinary differential equations*. Nagoya University (Vol. 2004-1, Preprint Series in Mathematical Sciences). Available online: https://www.math.human.nagoya-u.ac.jp/preprint/2004.html (accessed on 20 December 2024).

Hubert, M., Rousseeuw, P. J., & Verdonck, T. (2012). A deterministic algorithm for robust location and scatter. *Journal of Computational and Graphical Statistics*, *21*(3), 618–637. [CrossRef]

Ji, Q., Zhang, D., & Zhao, Y. (2020). Searching for safe-haven assets during the COVID-19 pandemic. *International Review of Financial Analysis*, *71*, 101526. [CrossRef] [PubMed]

Jobe, J. M., & Pokojovy, M. (2015). A cluster-based outlier detection scheme for multivariate data. *Journal of the American Statistical Association*, *110*(512), 1543–1551. [CrossRef]

Jung, K.-M. (2005). Multivariate least-trimmed squares regression estimator. *Computational Statistics & Data Analysis*, *48*(2), 307–316.

Korn, O., & Koziol, C. (2006). Bond portfolio optimization. *The Journal of Fixed Income*, *15*(4), 48–60. [CrossRef]

Kudraszow, N. L., & Maronna, R. A. (2011). Estimates of MM type for the multivariate linear model. *Journal of Multivariate Analysis*, *102*(9), 1280–1292. [CrossRef]

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Mamon, R. S. (2004). Three ways to solve for bond prices in the Vasicek model. *Journal of Applied Mathematics & Decision Sciences*, *8*(1), 1–14.

Markowitz, H. (1952). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Journal of Finance*, *7*(1), 77–91.

Mason, R. L., & Young, J. C. (2002). *Multivariate statistical process control with industrial applications*. SIAM.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, *8*(4), 323–361. [CrossRef]

Platen, E., & Rendek, R. (2009). Exact scenario simulation for selected multi-dimensional stochastic processes. *Communications on Stochastic Analysis*, *3*(3), 443–465. [CrossRef]

Pokojovy, M., & Anum, A. T. (2022). A fast initial response approach to sequential financial surveillance. In H. Han & E. Baker (Eds.), *The recent advances in transdisciplinary data science* (pp. 19–33). Springer Nature.

Pokojovy, M., & Jobe, J. M. (2022). A robust deterministic affine-equivariant algorithm for multivariate location and scatter. *Computational Statistics & Data Analysis*, *172*, 107475.

Pokojovy, M., Nkum, E., & Fullerton, T. M., Jr. (2024). Maximum likelihood estimation for discrete multivariate vasicek processes. In H. Han & E. Baker (Eds.), *Next generation data science. SDSC 2023. Communications in computer and information science* (Vol. 2113, pp. 3–18). Springer Nature.

Raji, I. A., Abbas, N., Abujiya, M. R., & Riaz, M. (2021). Robust multivariate shewhart control chart based on the Stahel-Donoho robust estimator and Mahalanobis distance for multivariate outlier detection. *Mathematics*, *9*(21), 2772. [CrossRef]

Reyen, S. S., Miller, J. J., & Wegman, E. J. (2009). Separating a mixture of two normals with proportional covariances. *Metrika*, *70*(3), 297–314. [CrossRef]

Rocke, D. M., & Woodruff, D. L. (1996). Identification of outliers in multivariate data. *Journal of the American Statistical Association*, *91*(435), 1047–1061. [CrossRef]

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, *79*(388), 871–880. [CrossRef]

Rousseeuw, P. J., & Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, *41*(3), 212–223. [CrossRef]

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, *48*(1), 1–48. [CrossRef]

Spelta, A., Pecora, N., Flori, A., & Giudici, P. (2023). The impact of the SARS-CoV-2 pandemic on financial markets: A seismologic approach. *Annals of Operations Research*, *330*(1), 639–664. [CrossRef]

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance*, *52*(5), 1973–2002. [CrossRef]

US Department of the Treasury. (2024, November). *Daily treasury yield curve rates.* Available online: https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yield (accessed on 1 November 2024).

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, *5*(2), 177–188. [CrossRef]

Víšek, J. Á. (2006a). The least trimmed squares. Part I: Consistency. *Kybernetika*, *42*(1), 1–36.

Víšek, J. Á. (2006b). The least trimmed squares. Part II: $\sqrt{n}$-consistency. *Kybernetika*, *42*(2), 181–202.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, *15*(20), 642–656. [CrossRef]

Zuo, Y. (2024). Large sample behavior of the least trimmed squares estimator. *Mathematics*, *12*(22), 3586. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.