FORECASTING TRENDS WITH ASSET PRICES

Ahmed Belhadjayed, Grégoire Loeper, Frédéric Abergel

FORECASTING TRENDS WITH ASSET PRICES

Ahmed Bel Hadj Ayed[1,2]
Grégoire Loeper [3]
Frédéric Abergel [1]

Abstract. The question of interest in this paper is the estimation of the trend of a financial asset, and the impact of its misspecification on investment strategies. The setting we consider is that of a stochastic asset price model where the trend follows an unobservable Ornstein-Uhlenbeck process. Motivated by the use of Kalman filtering as a forecasting tool, we address the problem of parameters estimation, and measure the effect of parameters misspecification. Numerical examples illustrate the difficulty of trend forecasting in financial time series.

[1]Laboratory MICS, Chair of Quantitative Finance, CentraleSupélec
[2]BNP Paribas Global Markets
[3]School of Mathematical Sciences, Monash University

Introduction

In theory, asset prices could be well described by random walks, according to the Efficient Market Hypothesis. If this were indeed the case, then future returns would not be predictable. Nevertheless, professionals in the finance industry tend to have divergent views on the subject, and trend following strategies are the principal sources of returns for Commodity Trading Advisors (see [33]). In fact, most quantitative strategies are based on the more or less explicit assumption that the trends of assets are known (see [38], [39]) and can be extracted from the asset prices themselves. It is therefore natural to address the question of forecasting asset trends, and to seek to provide reliable statistical estimators.

Unfortunately, the estimation of the trend of an asset is a statistically difficult problem, mainly because of a high measurement noise: consider for example a simple model with a constant trend $\frac{dS_t}{S_t} = \mu dt + \sigma_S dW_t^S$. Then, the best estimate of the trend in the least square sense at time T is given by $\widehat{\mu}_t = \frac{1}{T}\int_0^T \frac{dS_u}{S_u}$. Student's t-test will reject the hypothesis $\mu = 0$ if $|\widehat{\mu}_T| > \frac{1.96\sigma_S}{\sqrt{T}}$ at a 5% significance level. Therefore, with $\sigma_S = 30\%$, the estimate $\widehat{\mu}_T = 1\%$ becomes statistically relevant for observation times $T > 3457$ years!

The purpose of this work is to assess the feasibility of forecasting trends modelled by an unobserved mean-reverting diffusion.

Using a Bayesian approach or maximum likelihood estimation (see e.g. [34],[11], [5], [12] or [15]), inference methods for partially observed processes have been applied to financial time series, mostly in the framework of stochastic volatility models (see [27],[20],[30] or [14]). Closer in spirit to our motivation, several authors have considered the situation of an unobservable stochastic trend, and use filtering methods in this context (most of these methods are introduced in [22]). For example, in [49], [37] and [42] the Wiener-Kolmogorov filter (see [50] for details) is used; in [28], [26] and [23] the trend is supposed to be a random walk and the Kalman filter is used, and in [25] the Butterworth filter (see [10] for details) is applied. As it turns out, most of these filters are based on a parametric stochastic model for the trend, and their usefulness in realistic trading strategies is therefore confronted to the problem of parameters estimation (see [31], [41],[32] or [6] where filtering methods are used in the context of trading strategies).

It is our aim to partly fill the existing gap in the quantitative finance literature, and shed a new light on the feasability of classical trading strategies based on the determination of asset trends. Such a problem is addressed e.g. in [23] and [24], where the emphasis is set on Kalman filtering and the asymptotic behaviour of the maximum likelihood estimator. In this paper, we focus on the question of parameters estimation with a unobserved mean-reverting trend, and measure the

effect of parameters misspecification.
The paper is organized as follows: in Section 1, we present the model and recall some results on Kalman filtering. Section 2 is devoted to parameters inference using discrete time observations. The performance of statistical estimators is evaluated by giving their asymptotic behaviours, and by providing, in closed form, the Cramer-Rao bound. Section 3 introduces the continuous time misspecified Kalman filter. We provide estimates for the impact of parameters misspecification on trend filtering, and compute the probability to have a positive trend, given a positive estimate. Finally, Section 4 contains numerical examples illustrating the relevance of parameters misspecification in trend filtering.

1. Framework

In this section, the model for the asset price and the mean-reverting dynamics of its trend is made precise. Then, the Kalman filtering method, in a time-discretized version, is recalled.

1.1. Model.

1.1.1. *Continuous time model.* Consider a financial market living on a stochastic basis $(\Omega, \mathcal{F}, \mathbf{F}, \mathbb{P})$, where $\mathbf{F} = \{\mathcal{F}_t, t \geqslant 0\}$ is the natural filtration associated to a two-dimensional, uncorrelated Wiener process (W^S, W^μ), and $\mathbb{P}$ is the objective probability measure. The dynamics of the risky asset S is given by

$$\frac{dS_t}{S_t} = \mu_t dt + \sigma_S dW_t^S, \tag{1}$$

$$d\mu_t = -\lambda_\mu \mu_t dt + \sigma_\mu dW_t^\mu, \tag{2}$$

with $\mu_0 = 0$. We also assume that $(\lambda_\mu, \sigma_\mu, \sigma_S) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \times \mathbb{R}_+^*$.
Denote by $\mathbf{F}^S = \left\{\mathcal{F}_t^S\right\}$ be the natural filtration associated to the price process S. An important point is that only $\mathbf{F}^S$-adapted processes are observable, which implies that agents in this market do not observe the trend μ.

Remark 1.1. *The linear and Gaussian framework has well-known shortcomings: in practice, financial asset returns are heavy-tailed (see* [35][13]*) because of jumps and volatility fluctuations. Moreover, interactions between the trend and the volatility processes are possible. In this paper, we focus on a class of simple models so as to extract easily interpreted analytical expressions for various quantities of interest in the context of trend forecasting and investment strategies.*

1.1.2. *Discrete time model.* A discrete-time version of (1)-(2) is now presented. Let then δ be a discrete time step, and denote by the subscript k the value of a process at time $t_k = k\delta$. The discrete time model is:

$$
\begin{aligned}
y_{k+1} = \frac{S_{k+1} - S_k}{\delta S_k} &= \mu_{k+1} + u_{k+1}, & (3)\\
\mu_{k+1} &= e^{-\lambda_\mu \delta}\mu_k + v_k, & (4)
\end{aligned}
$$

where $u_k \sim \mathcal{N}\left(0, \frac{\sigma_S^2}{\delta}\right)$ and $v_k \sim \mathcal{N}\left(0, \frac{\sigma_\mu^2}{2\lambda_\mu}\left(1 - e^{-2\lambda_\mu\delta}\right)\right)$. The system (3)-(4) corresponds to an AR(1) model with noise.

1.2. **Optimal trend estimator.**

1.2.1. *Discrete Kalman filter.* In this subsection, the parameters $\theta = (\lambda_\mu, \sigma_\mu)$ and σ_S are supposed to be known. The discrete time system (3)-(4) corresponds to a Linear Gaussian Space State model where the observation is y and the state, μ (see [7] for details). In this case, the optimal estimator is the conditional expectation $\mathbb{E}\left[\mu_k | y_1, ..., y_k\right]$, given by the Kalman filter (classical results on the discrete Kalman filter are recalled in Appendix A).
Applications of this filter in Finance are numerous. Indeed, it can be used for trend filtering (see [23]), for term structure models (see [36] or [2]), for trading strategies (see [18] or [19]) and many other applications (see [47], [3], [16] or [45]). For simplicity, we let $\widehat{X}_{k/l}$ denote $\mathbb{E}\left[X_k | y_1, ..., y_l\right]$. The Kalman filter is decomposed in two distinct phases:

(1) An *a priori* estimate given $\hat{\mu}_{k+1/k}$ and $\Gamma_{k+1/k} = \mathbb{E}\left[(\mu_{k+1} - \hat{\mu}_{k+1/k})(\mu_{k+1} - \hat{\mu}_{k+1/k})^T\right]$. This estimate is done using the transition equation (4).
(2) An *a posteriori* estimate. When the new observation is available, a correction of the first estimate is done to obtain $\hat{\mu}_{k+1/k+1}$ and $\Gamma_{k+1/k+1} = \mathbb{E}\left[(\mu_{k+1} - \hat{\mu}_{k+1/k+1})(\mu_{k+1} - \hat{\mu}_{k+1/k+1})^T\right]$. The criterion for this correction is the least squares method.

Thus, $\hat{\mu}_{k/k}$ is the minimum variance linear unbiased estimate of the trend μ_k. Formally, the iterative method is given by:

$$
\begin{aligned}
\widehat{\mu}_{k+1/k+1} &= e^{-\lambda_\mu\delta}\widehat{\mu}_{k/k} + K_{k+1}\left(y_{k+1} - e^{-\lambda_\mu\delta}\widehat{\mu}_{k/k}\right), & (5)\\
\Gamma_{k+1/k+1} &= \left(1 - K_{k+1}\right)\Gamma_{k+1/k}, & (6)
\end{aligned}
$$

with

$$
\begin{aligned}
K_{k+1} &= \frac{\Gamma_{k+1/k}}{\Gamma_{k+1/k} + \frac{\sigma_S^2}{\delta}},\\
\Gamma_{k+1/k} &= e^{-2\lambda_\mu\delta}\Gamma_{k/k} + \frac{\sigma_\mu^2}{2\lambda_\mu}\left(1 - e^{-2\lambda_\mu\delta}\right).
\end{aligned}
$$

1.2.2. *Stationary limit and continuous-time representation.* Solving the equation $\Gamma_{k+1/k+1} = \Gamma_{k/k}$ corresponding to the steady-state yields:

$$\begin{aligned}
\Gamma_\infty &= \frac{g(\sigma_S, \lambda_\mu, \sigma_\mu) - f(\sigma_S, \lambda_\mu, \sigma_\mu)}{2e^{-2\lambda_\mu \delta}}, \\
\text{where } f(\sigma_S, \lambda_\mu, \sigma_\mu) &= \left(\frac{\sigma_S^2}{\delta} + \frac{\sigma_\mu^2}{2\lambda_\mu}\right)\left(1 - e^{-2\lambda_\mu \delta}\right), \\
\text{and } g(\sigma_S, \lambda_\mu, \sigma_\mu) &= \sqrt{f(\sigma_S, \lambda_\mu, \sigma_\mu)^2 + \frac{2\sigma_S^2 \sigma_\mu^2}{\lambda_\mu \delta}\left(e^{-2\lambda_\mu \delta} - e^{-4\lambda_\mu \delta}\right)}.
\end{aligned}$$

Using the stationary covariance error Γ_∞, a stationary gain K_∞ is defined and the estimate can be rewritten as a corrected exponential average:

$$\widehat{\mu}_{n+1} = K_\infty \sum_{i=0}^{\infty} e^{-\lambda_\mu \delta i} \left(1 - K_\infty\right)^i y_{n+1-i}. \tag{7}$$

The steady-state Kalman filter has also a continuous-time limit that depends on the asset returns. This result is recalled in the

Proposition 1. *The steady-state Kalman filter $\widehat{\mu}$ solves the following stochastic differential equation*

$$d\widehat{\mu}_t = -\lambda_\mu \beta(\lambda_\mu, \sigma_\mu, \sigma_S) \widehat{\mu}_t dt + \lambda_\mu \left(\beta(\lambda_\mu, \sigma_\mu, \sigma_S) - 1\right) \frac{dS_t}{S_t}, \tag{8}$$

where

$$\beta(\lambda_\mu, \sigma_\mu, \sigma_S) = \left(1 + \frac{\sigma_\mu^2}{\lambda_\mu^2 \sigma_S^2}\right)^{\frac{1}{2}}. \tag{9}$$

Proof. Based on [31], the Kalman filter is given by:

$$\begin{aligned}
E\left[\mu_t | \mathcal{F}_t^S\right] &= \phi(t)\left(\widehat{\mu}_0 + \frac{1}{\sigma_S^2}\int_0^t \frac{P(u)}{\phi(u)} \frac{dS_u}{S_u}\right), \\
\phi(t) &= e^{-\lambda_\mu t - \frac{1}{\sigma_S^2}\int_0^t P(u) du},
\end{aligned}$$

where the estimation error variance P is the solution of the following Riccati equation:

$$P'(t) = \frac{-1}{\sigma_S^2} P(t)^2 - 2\lambda_\mu P(t) + \sigma_\mu^2.$$

In this steady-state regime, we have $P'(t) = 0$. Then, the positive solution of this equation is given by

$$P^\infty = \sigma_S^2 \lambda_\mu \left(\beta(\lambda_\mu, \sigma_\mu, \sigma_S) - 1\right),$$

and there holds:

$$\begin{aligned}
\widehat{\mu}_t &= \phi^\infty(t)\left(\widehat{\mu}_0 + \frac{1}{\sigma_S^2}\int_0^t \frac{P^\infty}{\phi^\infty(u)} \frac{dS_u}{S_u}\right), \\
\phi^\infty(t) &= e^{-\lambda_\mu \beta(\lambda_\mu, \sigma_\mu, \sigma_S) t}.
\end{aligned}$$

Since:

$$\frac{d\phi^{\infty}(t)}{\phi^{\infty}(t)} = -\lambda_\mu \beta(\lambda_\mu, \sigma_\mu, \sigma_S)\, dt,$$

Equation (8) follows. □

The Kalman filter is the optimal estimator for linear systems with Gaussian uncertainty, and such a continuous-time representation can be used for risk/return analysis of trend following strategies (see [9] for details). Of course, in practice, the parameters $\theta = (\lambda_\mu, \sigma_\mu)$ are unknown and must be estimated. This important question is addressed in the next section.

2. Inference of the trend parameters

In this section, the problem of parameters inference is tackled, based on the use of statistical estimators such as Maximum Likelihood or Bayesian estimators. We analyze the asymptotic behaviours of statistical estimators and provide the Cramer-Rao bound in closed form. Since these classical estimators are based on the likelihood, two on-line computations of this function are presented in Appendix B.

2.1. **Asymptotic behaviour of statistical estimators.** The discrete time model (3)-(4) can be reformulated using the following proposition:

Proposition 2. *Consider the model* (3)-(4) *with* $(\lambda_\mu, \sigma_\mu, \sigma_S) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \times \mathbb{R}_+^*$. *In this case, the process* (y_i) *is* $ARMA(1,1)$.

The asymptotic behaviour of the classical estimators follows. Indeed, the identifiability property and the asymptotic normality of the maximum likelihood estimator are well known for stationary ARMA Gaussian processes (see [8], section 10.8). Moreover, the asymptotic behaviour of the Bayesian estimators are also guaranteed by the ARMA$(1,1)$ property of the process (y_i). If the prior density function is continuous and positive in an open neighbourhood of the real parameters, the Bayesian estimators are asymptotically normal (see [46] in which a generalized Bernstein-Von Mises theorem for stationary "short memory" processes is given, or [40] for a discussion on the Bayesian analysis of ARMA processes).

2.2. **Cramer-Rao bound.** This bound is the lowest variance of the unbiased estimators. We recall the following result, providing a formal description of the Cramer-Rao bound (CRB in short).

Corollary 2.1. *Consider the model* (3)-(4) *and* N *observations* $(y_1, \cdots, y_N)^T$. *Suppose that* $(\lambda_\mu, \sigma_\mu, \sigma_S) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \times \mathbb{R}_+^*$. *If* $\hat{\theta}_N$ *is an unbiased estimator of* $\theta = (\lambda_\mu, \sigma_\mu)$, *we have:*

$$\mathbb{C}ov_\theta\left(\hat{\theta}_N\right) \geqslant CRB(\theta).$$

This bound is given by $CRB(\theta) = I_N^{-1}(\theta)$*, where* $I_N(\theta)$ *is the Fisher Information matrix:*

$$\left(I_N(\theta)\right)_{i,j} = -\mathbb{E}\left[\frac{\partial^2 \log f(y_1, ...y_N|\theta)}{\partial\theta_i \partial\theta_j}\right],$$

and $I_N(\theta) = NI_1(\theta)$*. Moreover, the maximum likelihood estimator* $\widehat{\theta}_N^{ML}$ *attains this bound:*

$$\sqrt{N}\left(\widehat{\theta}_N^{ML} - \theta\right) \to \mathcal{N}\left(0, I_1^{-1}(\theta)\right).$$

This result is a consequence of Proposition 2 (see [8], section 10.8). The following result is an analytic representation of the Fisher information matrix:

Theorem 2.2. *For the model* (3)-(4)*, if* $(\lambda_\mu, \sigma_\mu, \sigma_S) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \times \mathbb{R}_+^*$*, we have:*

$$I_1(\theta) = \left(\frac{1}{4\Pi}\int_{-\Pi}^{\Pi} f_\theta^{-2}(\omega)\frac{\partial f_\theta}{\partial\theta_i}(\omega)\frac{\partial f_\theta}{\partial\theta_j}(\omega)\,d\omega\right)_{1\leq i,j\leq 2},$$

where f_θ *is the spectral density of the process* (y_i)*:*

$$f_\theta(\omega) = \frac{\frac{\sigma_\mu^2}{2\lambda_\mu}\left(1 - e^{-2\lambda_\mu\delta}\right) + \frac{\sigma_S^2}{\delta}\left(1 + e^{-2\lambda_\mu\delta}\right) - \frac{2e^{-\lambda_\mu\delta}\sigma_S^2}{\delta}\cos(\omega)}{1 + e^{-2\lambda_\mu\delta} - 2e^{-\lambda_\mu\delta}\cos(\omega)}.$$

Proof. Whittle's formula (see [48] for details) gives the integral representation of the Fisher information matrix. Since the process (y_i) is ARMA$(1,1)$, the expression of its spectral density follows (see [8], section 4.4). □

Finally, the Cramer-Rao Bound of the trend parameters can be computed using Theorem 2.2.

3. Impact of parameters misspecification

In this section, we consider the continuous-time Kalman filter with a bad calibration in the steady-state regime. The law of the residuals between the filter (misspecified or not) and the hidden process is characterized, and the impact of parameters misspecification on the detection of a positive trend is studied.

3.1. Context.

Suppose that the risky asset S is given by the model (1)-(2) with $\theta^* = \left(\sigma_\mu^*, \lambda_\mu^*\right)$, and suppose that an agent thinks that the parameters are equal to $\theta = (\sigma_\mu, \lambda_\mu)$. Assuming the steady-state regime and using these estimates and Proposition 1, the agent implements the continuous-time misspecified Kalman filter:

$$d\widehat{\mu}_t = -\lambda_\mu\beta\widehat{\mu}_t dt + \lambda_\mu(\beta - 1)\frac{dS_t}{S_t}, \tag{10}$$

where $\beta = \beta(\lambda_\mu, \sigma_\mu, \sigma_S)$ (see Equation (9)) and $\widehat{\mu}_0 = 0$. The following lemma gives the law of the misspecified Kalman filter:

Lemma 3.1. *Consider the model* (1)-(2) *with* $\theta^* = \left(\sigma_\mu^*, \lambda_\mu^*\right)$. *The misspecified, continuous-time filter of Equation (10) is given by:*

$$\widehat{\mu}_t = \lambda_\mu \left(\beta - 1\right) e^{-\lambda_\mu \beta t} \left(\int_0^t e^{\lambda_\mu \beta s} \mu_s^* ds + \sigma_S \int_0^t e^{\lambda_\mu \beta s} dW_s^S \right). \qquad (11)$$

Moreover, $\widehat{\mu}$ *is a centered Gaussian process and its variance is given by:*

$$\begin{aligned}\mathbb{V}ar\left[\widehat{\mu}_t\right] = \mathbb{E}\left[\widehat{\mu}_t^2\right] = &\frac{\lambda_\mu^2 \left(\beta - 1\right)^2 \left(\sigma_\mu^*\right)^2}{\lambda_\mu^* \left(\lambda_\mu \beta - \lambda_\mu^*\right)} \left[\frac{1 - e^{-\left(\lambda_\mu \beta + \lambda_\mu^*\right)t}}{\lambda_\mu \beta + \lambda_\mu^*} \right.\\ &\left. + \frac{2e^{-\left(\lambda_\mu \beta + \lambda_\mu^*\right)t} - e^{-2\lambda_\mu^* t} - e^{-2\lambda_\mu \beta t}}{\lambda_\mu \beta - \lambda_\mu^*} + \frac{e^{-2\lambda_\mu \beta t} - 1}{2\lambda_\mu \beta} \right]\\ &+ \frac{\lambda_\mu \left(\beta - 1\right)^2 \sigma_S^2}{2\beta} \left(1 - e^{-2\lambda_\mu \beta t}\right).\end{aligned}$$

Proof. Applying Itō's lemma to the function $f\left(\widehat{\mu}, t\right) = e^{\lambda_\mu \beta t} \widehat{\mu}_t$, and integrating from 0 to t yields Equation (11). Therefore, $\widehat{\mu}$ is also a Gaussian process. Its mean is zero (because $\mu_0^* = 0$). Since the processes μ^* and W^S are supposed to be independent, the variance of $\widehat{\mu}$ is given by the sum of the variances of the terms in Equation (11). Moreover:

$$\begin{aligned}\mathbb{V}ar\left[\int_0^t e^{\lambda_\mu \beta} dW_s^S\right] &= \frac{e^{2\lambda_\mu \beta t} - 1}{2\lambda_\mu \beta},\\ \mathbb{V}ar\left[\int_0^t e^{\lambda_\mu \beta s} \mu_s^* ds\right] &= \int_0^t \int_0^t e^{\lambda_\mu \beta (s_1 + s_2)} \mathbb{C}ov\left(\mu_{s_1}^*, \mu_{s_2}^*\right) ds_1 ds_2,\end{aligned}$$

and $\mathbb{C}ov\left(\mu_{s_1}^*, \mu_{s_2}^*\right)$ is given by Equation (23). The variance of the process $\widehat{\mu}_t$ follows. □

3.2. **Filtering with parameters misspecification.** The impact of parameters misspecification on trend filtering can be measured using the difference between the filter and the hidden process.
The following theorem gives the law of the residuals.

Theorem 3.2. *Consider the model* (1)-(2) *with* $\theta^* = \left(\sigma_\mu^*, \lambda_\mu^*\right)$ *and the trend estimate defined in Equation (11). The process* $\widehat{\mu} - \mu^*$ *is a centered Gaussian process and its variance has a stationary limit:*

$$\lim_{t \to \infty} \mathbb{V}ar\left[\widehat{\mu}_t - \mu_t^*\right] = \frac{\sigma_S^2}{2\beta} \left(\lambda_\mu \left(\beta - 1\right)^2 + \lambda_\mu^* \left(\left(\beta^*\right)^2 - 1\right) \frac{\lambda_\mu^* \beta + \lambda_\mu}{\lambda_\mu \beta + \lambda_\mu^*} \right), \qquad (12)$$

where $\beta = \beta\left(\lambda_\mu, \sigma_\mu, \sigma_S\right)$ *and* $\beta^* = \beta\left(\lambda_\mu^*, \sigma_\mu^*, \sigma_S\right)$ *as given in Equation (9).*

Moreover, if $\left(\sigma_\mu, \lambda_\mu\right) = \left(\sigma_\mu^*, \lambda_\mu^*\right)$, *Equation (12) becomes:*

$$\lim_{t \to \infty} \mathbb{V}ar\left[\widehat{\mu}_t^* - \mu_t^*\right] = \lambda_\mu^* \sigma_S^2 \left(\beta^* - 1\right). \qquad (13)$$

Proof. Equation (11) implies that the process $\widehat{\mu} - \mu^*$ is a centered Gaussian process. Its variance can be computed in closed form:

$$\mathbb{V}ar\left[\widehat{\mu}_t - \mu_t^*\right] = \mathbb{V}ar\left[\widehat{\mu}_t\right] + \mathbb{V}ar\left[\mu_t^*\right] - 2 * \mathbb{C}\mathrm{ov}\left[\widehat{\mu}_t, \mu_t^*\right],$$

where $\mathbb{V}ar\left[\mu_t^*\right] = \frac{\left(\sigma_\mu^*\right)^2}{2\lambda_\mu^*}\left(1 - e^{-2\lambda_\mu^* t}\right)$, and $\mathbb{V}ar\left[\widehat{\mu}_t\right]$ is given by Lemma 3.1. Since the processes W^S and μ^* are supposed to be independent, there holds:

$$\begin{aligned}\mathbb{C}\mathrm{ov}\left[\widehat{\mu}_t, \mu_t^*\right] &= \frac{\lambda_\mu\left(\beta - 1\right)\left(\sigma_\mu^*\right)^2}{2\lambda_\mu^*}\left(\frac{1 - e^{-\left(\lambda_\mu\beta + \lambda_\mu^*\right)t}}{\lambda_\mu\beta + \lambda_\mu^*}\right. \\ &\left. - \frac{e^{-2\lambda_\mu^* t} - e^{-\left(\lambda_\mu\beta + \lambda_\mu^*\right)t}}{\lambda_\mu\beta - \lambda_\mu^*}\right).\end{aligned}$$

The asymptotic variance is obtained by letting $t \to \infty$:

$$\begin{aligned}\lim_{t\to\infty} \mathbb{V}ar\left[\widehat{\mu}_t - \mu_t^*\right] &= \frac{\lambda_\mu\left(\beta - 1\right)}{2\beta}\left[\left(\beta - 1\right)\sigma_S^2 - \frac{\left(\sigma_\mu^*\right)^2\left(\beta + 1\right)}{\lambda_\mu^*\left(\lambda_\mu\beta + \lambda_\mu^*\right)}\right] \\ &+ \frac{\left(\sigma_\mu^*\right)^2}{2\lambda_\mu^*},\end{aligned}$$

and Equation (12) follows. Finally, Equation (13) is obtained by letting $\theta \to \theta^*$. $\square$

Remark 3.3. *Consider the well-specified case* $\left(\sigma_\mu, \lambda_\mu\right) = \left(\sigma_\mu^*, \lambda_\mu^*\right)$. *Using Equation (13), it follows that:*

$$\lim_{t\to\infty} \frac{\mathbb{V}ar\left[\widehat{\mu}_t^* - \mu_t^*\right]}{\mathbb{V}ar\left[\mu_t^*\right]} = \frac{2}{1 + \sqrt{1 + \frac{\left(\sigma_\mu^*\right)^2}{\left(\lambda_\mu^*\right)^2\sigma_S^2}}}. \tag{14}$$

Then, the asymptotic relative variance of the well-specified residuals is an increasing function of λ_μ^* *and a decreasing function of* σ_μ^*.

3.3. **Detection of a positive trend.** In practice, the trend estimate (misspecified or not) will be used to make an investment decision. For example, a positive estimate leads to a long position. So, it is interesting to estimate the probability of a **positive** trend knowing a **positive** estimate. We derive this probability in closed form, based on the following proposition giving the asymptotic conditional law of the trend $\left(\mu_t^* | \hat{\mu}_t = x\right)$:

Proposition 3. *Consider the model* (1)-(2) *with* $\theta^* = \left(\sigma_\mu^*, \lambda_\mu^*\right)$ *and the trend estimate defined in Equation (11). Then, there holds:*

$$\left(\mu_t^* | \hat{\mu}_t = x\right) \xrightarrow[t\to\infty]{\mathcal{L}} \mathcal{N}\left(\mathbb{M}_{\mu^*|\hat{\mu}}^{\infty}, \mathbb{V}ar_{\mu^*|\hat{\mu}}^{\infty}\right), \tag{15}$$

with:

$$\mathbb{M}^{\infty}_{\mu^*|\hat{\mu}} = \frac{\lambda^*_\mu \beta \left((\beta^*)^2 - 1 \right)}{(\beta - 1)\left(\lambda_\mu \beta + \lambda^*_\mu (\beta^*)^2 \right)} x, \tag{16}$$

$$\mathbb{V}ar^{\infty}_{\mu^*|\hat{\mu}} = \mathbb{V}ar^{\infty}_{\mu^*} \left(1 - \frac{\lambda^*_\mu \lambda_\mu \beta \left((\beta^*)^2 - 1 \right)}{\left(\lambda^*_\mu + \lambda_\mu \beta \right) \left(\lambda_\mu \beta + \lambda^*_\mu (\beta^*)^2 \right)} \right), \tag{17}$$

where $\mathbb{V}ar^{\infty}_{\mu^*} = \frac{(\sigma^*_\mu)^2}{2\lambda^*_\mu}$.

Moreover, if $(\sigma_\mu, \lambda_\mu) = \left(\sigma^*_\mu, \lambda^*_\mu \right)$*, Equation (15) becomes:*

$$(\mu^*_t | \hat{\mu}^*_t = x) \underset{t \to \infty}{\xrightarrow{\mathcal{L}}} \mathcal{N} \left(x, \frac{2\mathbb{V}ar^{\infty}_{\mu^*}}{\beta^* + 1} \right), \tag{18}$$

where $\beta^* = \beta \left(\lambda^*_\mu, \sigma^*_\mu, \sigma_S \right)$ *(see Equation (9)).*

Proof. Since the estimate $\hat{\mu}$ and the trend μ^* are two centered and correlated Gaussian processes (see Lemma 3.1 and the proof of Theorem 3.2), the conditional law $(\mu^*_t | \hat{\mu}_t = x)$ is Gaussian with a mean and a variance given by:

$$\mathbb{M}_{\mu^*_t|\hat{\mu}_t} = \frac{\mathbb{C}ov\left(\hat{\mu}_t, \mu^*_t \right)}{\mathbb{V}ar\left[\hat{\mu}_t \right]} x,$$

$$\mathbb{V}ar_{\mu^*_t|\hat{\mu}_t} = \mathbb{V}ar\left[\mu^*_t \right] - \frac{\mathbb{C}ov\left(\hat{\mu}_t, \mu^*_t \right)^2}{\mathbb{V}ar\left[\hat{\mu}_t \right]}.$$

Using Lemma 3.1 and the expression of $\mathbb{C}ov\left(\hat{\mu}_t, \mu^*_t \right)$ in the proof of Theorem 3.2 yields

$$\lim_{t \to \infty} \mathbb{M}_{\mu^*_t|\hat{\mu}_t} = \mathbb{M}^{\infty}_{\mu^*|\hat{\mu}},$$

$$\lim_{t \to \infty} \mathbb{V}ar_{\mu^*_t|\hat{\mu}_t} = \mathbb{V}ar^{\infty}_{\mu^*|\hat{\mu}},$$

and Equation (15) follows. Finally, Equation (18) is obtained by letting $\theta \to \theta^*$. □

The following proposition is a consequence of the previous proposition. It gives the asymptotic probability to have a positive trend, knowing a positive estimate equal to x.

Proposition 4. *Consider the model* (1)-(2) *with* $\theta^* = \left(\sigma^*_\mu, \lambda^*_\mu \right)$ *and the trend estimate defined in Equation (11). In this case:*

$$\lim_{t \to \infty} \mathbb{P}\left(\mu^*_t > 0 | \hat{\mu}_t = x \right) = \mathbb{P}_\infty \left(\mu^* > 0 | \hat{\mu} = x \right), \tag{19}$$

where

$$\mathbb{P}_\infty \left(\mu^* > 0 | \hat{\mu} = x \right) = 1 - \Phi \left(\frac{-\mathbb{M}^{\infty}_{\mu^*|\hat{\mu}=x}}{\sqrt{\mathbb{V}ar^{\infty}_{\mu^*|\hat{\mu}=x}}} \right), \tag{20}$$

where $\mathbb{M}^{\infty}_{\mu^*|\hat{\mu}=x}$ and $\mathbb{V}ar^{\infty}_{\mu^*|\hat{\mu}=x}$ are defined in Equations (16) and (17), and Φ is the cumulative distribution function of the standard normal law.

Moreover, if $x > 0$ and $(\sigma_\mu, \lambda_\mu) = \left(\sigma^*_\mu, \lambda^*_\mu\right)$, this asymptotic probability becomes an increasing function of σ^*_μ and a decreasing function of λ^*_μ.

Proof. Equations (19) and (20) follow from Proposition 3. Now, consider the well-specified case $(\sigma_\mu, \lambda_\mu) = \left(\sigma^*_\mu, \lambda^*_\mu\right)$ and $x > 0$. Using Equation (18), it follows that:

$$\mathbb{V}ar^{\infty}_{\mu^*|\hat{\mu}^*=x} = f\left(\sigma^*_\mu, \lambda^*_\mu, \sigma_S\right),$$

where

$$f\left(\sigma^*_\mu, \lambda^*_\mu, \sigma_S\right) = \frac{\left(\sigma^*_\mu\right)^2}{\lambda^*_\mu\left(1+\sqrt{1+\frac{\left(\sigma^*_\mu\right)^2}{\sigma_S^2\left(\lambda^*_\mu\right)^2}}\right)}.$$

Since

$$\begin{aligned}
\frac{\partial f}{\partial \lambda^*_\mu}\left(\sigma^*_\mu, \lambda^*_\mu, \sigma_S\right) &= \frac{-\left(\sigma^*_\mu\right)^2}{\left(\lambda^*_\mu\right)^2\left(1+\sqrt{1+\frac{\left(\sigma^*_\mu\right)^2}{\sigma_S^2\left(\lambda^*_\mu\right)^2}}+\frac{\left(\sigma^*_\mu\right)^2}{\sigma_S^2}\right)} \leq 0,\\
\frac{\partial f}{\partial \sigma^*_\mu}\left(\sigma^*_\mu, \lambda^*_\mu, \sigma_S\right) &= \frac{\lambda^*_\mu\sigma^*_\mu\sigma_S^2\sqrt{1+\frac{\left(\sigma^*_\mu\right)^2}{\sigma_S^2\left(\lambda^*_\mu\right)^2}}}{\left(\sigma^*_\mu\right)^2+\sigma_S^2\left(\lambda^*_\mu\right)^2} \geq 0,
\end{aligned}$$

the asymptotic well-specified probability to have a positive trend, knowing a positive estimate equal to x is an increasing function of σ^*_μ and a decreasing function of λ^*_μ. □

Remark 3.4. *This probability is an increasing function of x. Indeed, it is easier to detect the sign of the real trend with a high estimate than with a low estimate. Moreover, this probability is always superior to 0.5. This is due to the non-zero correlation between the trend and the filter. As shown in the previous sections, trend filtering is easier with a small spot volatility σ_S. Here, the probability to make a good detection is also a decreasing function of σ_S.*

4. Simulations

In this section, numerical simulations are performed, in order to make the reader aware of the trend filtering problem. First, the feasibility of trend forecasting with statistical estimators is illustrated on different trend regimes. Then, the effects of a bad forecast on trend filtering and on the detection of a positive trend are discussed. Finally, we study the numerical maximum likelihood inference of this model using

Kalman filtering and the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm.

4.1. **Feasibility of trend forecasting.** Suppose that only discrete-time observations are available and that the time step is equal to $\delta = 1/252$, corresponding to daily returns.
Note that, if this time step is divided by M, the variance of the observation noise u_k is multiplied by M ($u_k \sim \mathcal{N}\left(0, \frac{\sigma_S^2}{\delta}\right)$), so that increasing the frequency of the observations will not be really helpful to calibrate the trend.
We also assume that the agent uses an unbiased estimator. Given T years of observations, The Cramer-Rao bound is given by:

$$CRB_T(\theta) = \frac{I_1^{-1}(\theta)}{T * 252},$$

where $I_1(\theta)$ is given by Theorem 2.2. The smallest confidence region is obtained with this matrix. In practice, the real values of the parameters θ are unknown and asymptotic confidence regions are computed (replacing θ by the estimates $\hat{\theta}$ in the Fisher information matrix $I_1\left(\hat{\theta}\right)$). Since the goal of this subsection is to evaluated the feasibility of this estimation problem, we suppose that we know the real values of the parameters. In such a case, the exact Cramer-Rao bound can be computed. Suppose that a target standard deviation x_i is fixed for the parameter θ_i. In this case, to reach the precision x_i, the length of the observations must be larger than $T_i^x = \frac{\left(I_1^{-1}(\theta)\right)_{ii}}{252 * x_i^2}$.

We consider a fixed spot volatility $\sigma_S = 30\%$, two target precisions for each parameter θ_i and we compute T_i^x for several configurations. Figures 1, 2, 3 and 4 represent the results. It is well-known that for a high measurement noise, which means a high spot volatility, the problem is harder because of a low signal-to-noise ratio. The higher the volatility, the longer the observations must be. Here, we observe that with a higher drift volatility σ_μ and a lower λ_μ, the problem is easier. Indeed, the drift takes higher values and is more detectable. Moreover, the simulations show that the classical estimators are not adapted to such a weak signal-to-noise ratio: even after a long period of observations, the estimators exhibit high variances. Indeed, the shortest observation period is longer than 29 years. It corresponds to a target standard deviation equal to 0.5 for a parameter $\lambda_\mu = 1$, and a trend standard deviation equal to $\sigma_\mu (2\lambda_\mu)^{-1/2} \approx 63\%$. Therefore, for this configuration, after 30 years of observations, the standard deviation is equal to 50% of the real parameter value λ_μ. After 742 years, this standard deviation is equal to 10%. Even in such a regime, the trend forecast with a good precision is impossible.

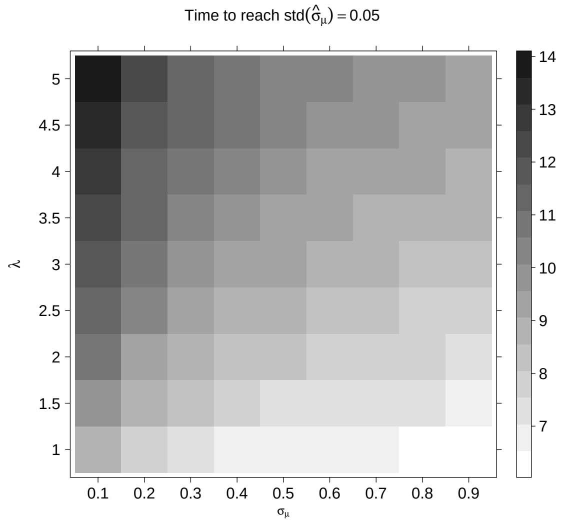

FIGURE 1. Time to reach a target standard deviation on σ_μ equal to 0.05 (ln(years))

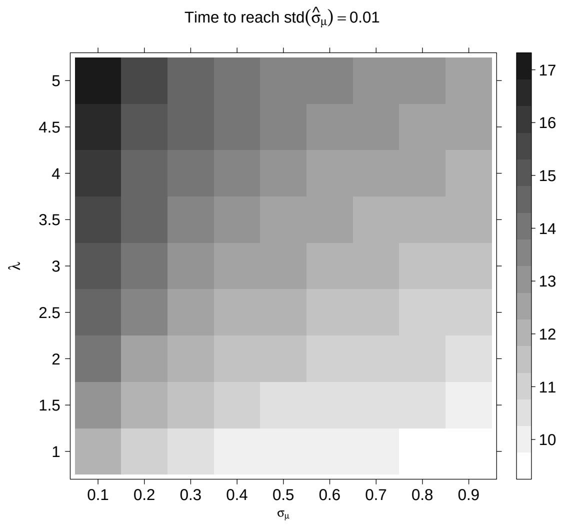

FIGURE 2. Time to reach a target standard deviation on σ_μ equal to 0.01 (ln(years))

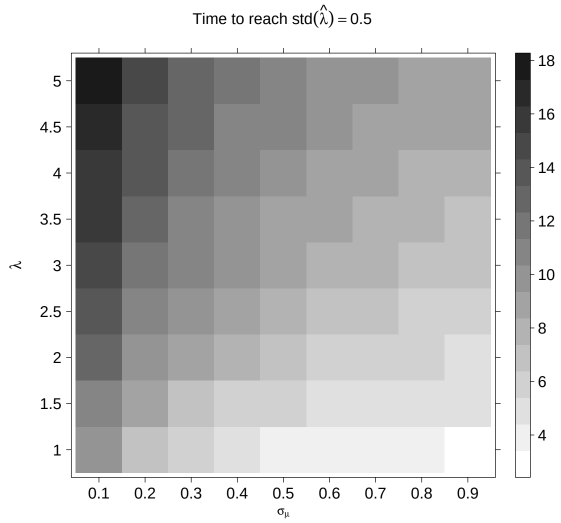

FIGURE 3. Time to reach a target standard deviation on λ_μ equal to 0.5 (ln(years))

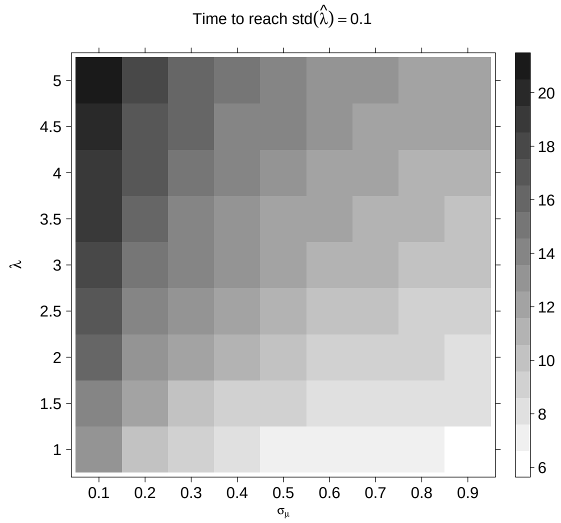

FIGURE 4. Time to reach a target standard deviation on λ_μ equal to 0.1 (ln(years))

4.2. **Impact of parameters misspecification on trend filtering.** This subsection illustrates the impact of parameters misspecification on trend filtering. Using the results of Theorem 3.2, we represent, for different configurations, and for the well- and mis-specified case, the asymptotic standard deviation of the residuals between the trend and the filter. Figures 5 and 6 represent the asymptotic standard deviation of the trend and of the residuals in the well-specified case (the agent uses the real values of the parameters) for different configurations. As seen in Equation (13), the asymptotic standard deviation of the well-specified residuals is an increasing function of the drift volatility σ_μ^* and a decreasing function of the parameter λ_μ^*. For $\lambda_\mu^* = 1$ and $\sigma_\mu^* = 90\%$, the standard deviation of the residuals ($\simeq 44\%$) is inferior to the standard deviation of the trend ($\simeq 64\%$). For a high λ_μ^* and a small drift volatility, the two quantities are approximately equal. This figure leads to the same conclusions than Equation (14). Indeed, as for the calibration problem, the problem of trend filtering is easier with a small λ_μ^* and a high drift volatility σ_μ^*.

Now consider the worst configuration $\sigma_S = 30\%$, $\lambda_\mu^* = 5$ and $\sigma_\mu^* = 10\%$. Figure 7 represents the asymptotic standard deviation of the residuals for different estimates $(\lambda_\mu, \sigma_\mu)$. This regime corresponds to a standard deviation of the trend equal to $\sigma_\mu^* \left(2\lambda_\mu^*\right)^{-1/2} \approx 3.2\%$ and to a standard deviation of the residuals equal to 3.16% in the well-specified case. If the agent implements the Kalman filter with $\lambda_\mu = 1$ and $\sigma_\mu = 90\%$, the standard deviation of the residuals becomes superior to 25%. Finally, consider the best configuration $\sigma_S = 30\%$, $\lambda_\mu^* = 1$ and $\sigma_\mu^* = 90\%$. Figure 8 represents the asymptotic standard deviation of the residuals for different estimates $(\lambda_\mu, \sigma_\mu)$. This regime corresponds to a standard deviation of the trend equal to $\sigma_\mu^* \left(2\lambda_\mu^*\right)^{-1/2} \approx 63\%$ and to a standard deviation of the residuals equal to 44% in the well-specified case. If the agent implements the Kalman filter with $\lambda_\mu = 5$ and $\sigma_\mu = 10\%$, the standard deviation of the residuals becomes larger than 60%. Even in a favourable regime, the impact of parameters misspecification on trend filtering is not negligible.

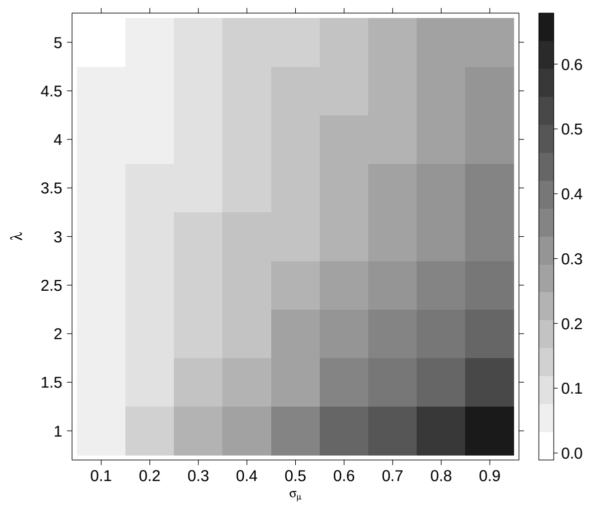

FIGURE 5. Asymptotic standard deviation of the trend as a function of the trend parameters with $\sigma_S = 30\%$.

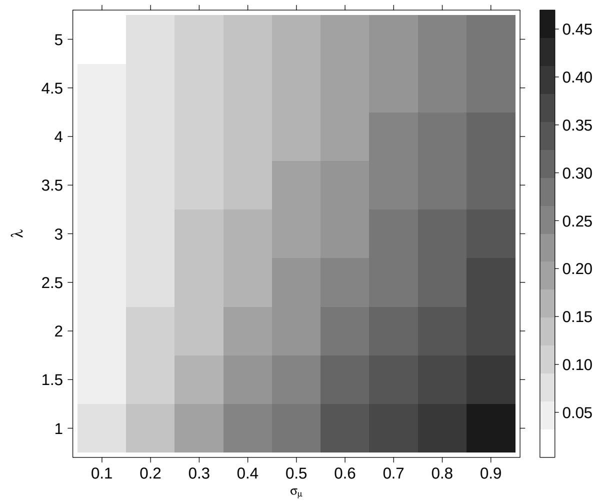

FIGURE 6. Asymptotic standard deviation of the residuals of the well-specified Kalman filter as a function of the trend parameters with $\sigma_S = 30\%$.

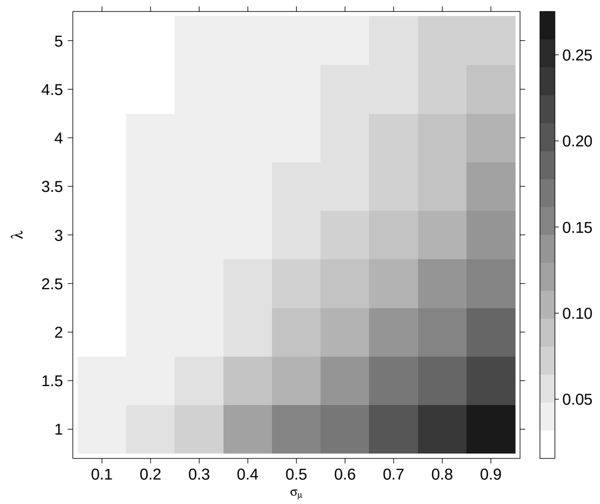

FIGURE 7. Asymptotic standard deviation of the residuals of the misspecified Kalman filter as a function of the trend estimate parameters with $\sigma_S = 30\%$, $\lambda_\mu^* = 5$ and $\sigma_\mu^* = 10\%$.

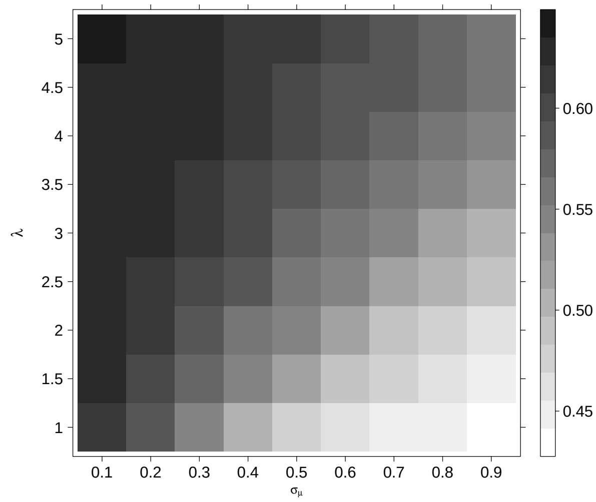

FIGURE 8. Asymptotic standard deviation of the residuals of the misspecified Kalman filter as a function of the trend estimate parameters with $\sigma_S = 30\%$, $\lambda^*_\mu = 1$ and $\sigma^*_\mu = 90\%$.

4.3. **Detection of a positive trend.** In this subsection, Equation (20) - giving the asymptotic probability to have a positive trend, knowing a trend estimate equal to a threshold x - is illustrated.
In order to compare this probability for different trend regimes, we choose a threshold equal to the standard deviation of the filter $\hat{\mu}$: this quantity is tractable in practice and moreover, since the continuous time misspecified filter $\hat{\mu}$ is a centered Gaussian process, the probability that $\hat{\mu}$ becomes larger than its standard deviation is independent of the parameters $\left(\sigma^*_\mu, \lambda^*_\mu, \sigma_\mu, \lambda_\mu, \sigma_S\right)$.

Suppose first that the agent uses the real values of the parameters, and consider the asymptotic probability $\mathbb{P}\left(\mu^* > 0 | \hat{\mu}^* = \sqrt{\mathbb{V}_{\hat{\mu}^*}}\right)$ to have a positive trend, knowing an estimate equal to its standard deviation. Figure 9 represents this probability for different regimes. As seen in Proposition 4, in the well-specified case, this probability is an increasing function of the trend volatility σ^*_μ and a decreasing function of λ^*_μ. Again, as in the calibration and filtering problems, the detection is easier with a small λ^*_μ and a high drift volatility. Now, suppose that the agent uses wrong estimates $(\sigma_\mu, \lambda_\mu)$. In this case, the agent implements the continuous time misspecified Kalman filter. Figures 10 and 11 represent the asymptotic probability $\mathbb{P}\left(\mu^* > 0 | \hat{\mu} = \sqrt{\mathbb{V}_{\hat{\mu}}}\right)$ for the best and the worst configuration of Figure 9. As explained in Remark 3.4, this probability is always superior to 0.5, even with a bad calibration of the parameters. For each case, the probability to have a positive trend, knowing an estimate equal to its standard deviation, does not vary a lot with an error on the parameters. This quantity seems to be robust to parameters misspecifications.

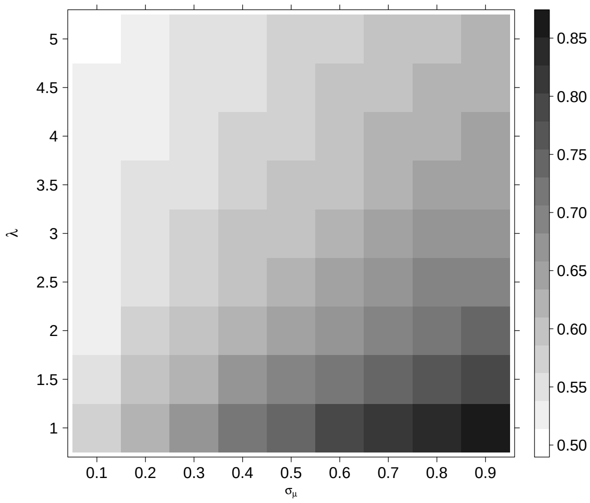

FIGURE 9. Asymptotic probability to have a positive trend given a well-specified estimate equal to its standard deviation with $\sigma_S = 30\%$.

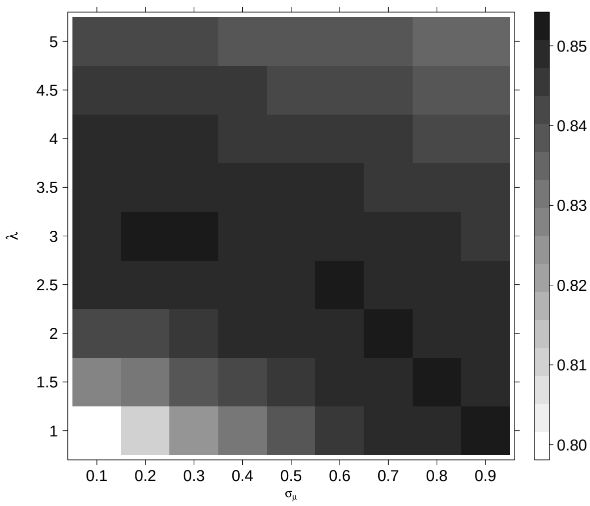

FIGURE 10. Asymptotic probability to have a positive trend given a misspecified estimate equal to its standard deviation with $\sigma_S = 30\%$, $\lambda_\mu^* = 1$ and $\sigma_\mu^* = 90\%$.

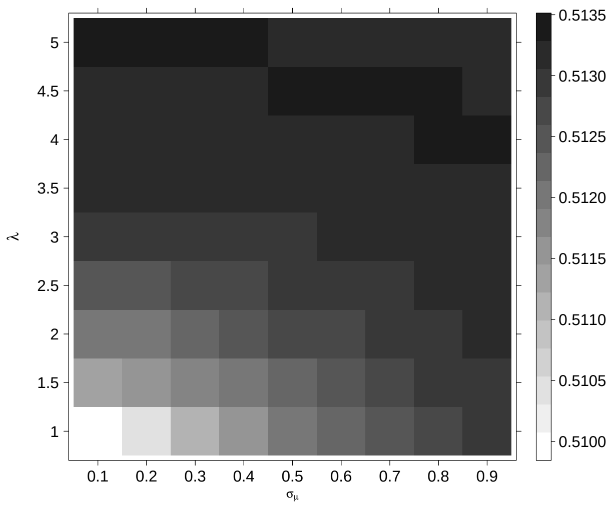

FIGURE 11. Asymptotic probability to have a positive trend given a misspecified estimate equal to its standard deviation with $\sigma_S = 30\%$, $\lambda^*_\mu = 5$ and $\sigma^*_\mu = 10\%$.

4.4. **Kalman filtering in practice.** In this subsection, we focus on the numerical maximum likelihood inference of the model (1)-(2) using Kalman filtering and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. We first illustrate the performance of a parametric bootstrap, and then give a typical application on real market data.

4.4.1. *Parametric bootstrap.* Consider the discrete time model (3)-(4) and suppose that we want to use the maximum likelihood method. This model is Gaussian and the likelihood is thus characterized by the moments of first and second order. When the number of observations N is large, these moments can be difficult to compute numerically (to mitigate these problems, two methods are introduced in Appendix B). Here, we focus on the computation using the Kalman filter. Since we are able to compute recursively this likelihood with a large number of observations, we choose the direct maximization of this function using the BFGS algorithm. An alternative way is to use the EM-algorithm (see [21] or [17] for details). The aim of these tests is to compare the standard deviation of this numerical maximum likelihood estimator to the theoretical results of Theorem 2.2. So, we consider the following parametric bootstrap:

(1) Simulate M paths of length T years with the model (3)-(4).
(2) Consider only one parameter (for example $\theta = \sigma_\mu$), the other one is supposed to be known.
(3) On each path $i \in \{1..M\}$, we compute a numerical maximum likelihood estimate $\hat{\theta}^{ML}_{T,i}$ with Kalman filtering and the BFGS algorithm (initialised with $\sigma^0_\mu = 0.1$ or $\lambda^0_\mu = 0.1$).

(4) The variance of θ_T^{ML} is estimated by:

$$\hat{Var}\left(\theta_T^{ML}\right) = \frac{1}{M-1}\sum_{i=1}^{M}\left(\hat{\theta}_{T,i}^{ML} - \frac{1}{M}\sum_{k=1}^{M}\hat{\theta}_{T,k}^{ML}\right)^2.$$

Figures 12 and 13 illustrate the results for the estimation of the trend mean reversion speed and of the trend volatility with $M = 1000$, $\sigma_\mu^* = 90\%$, $\lambda_\mu^* = 1$ and $\sigma_S = 30\%$.

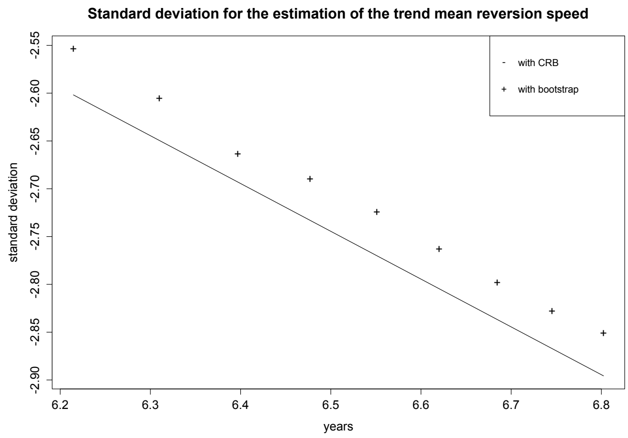

FIGURE 12. Standard deviation for the estimation of $\lambda_\mu^* = 1$ with $\sigma_\mu^* = 90\%$ and $\sigma_S = 30\%$ on a double logarithmic scale.

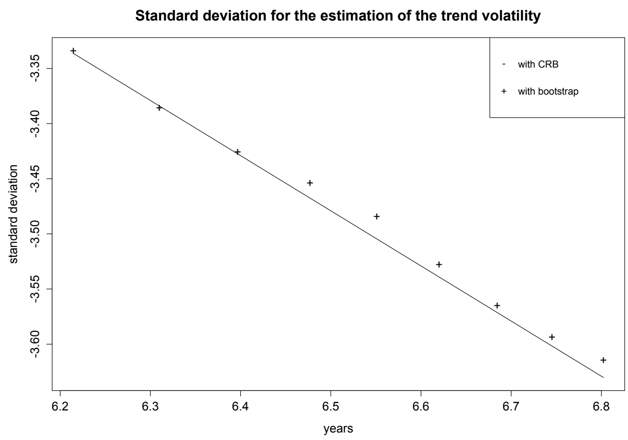

FIGURE 13. Standard deviation for the estimation of $\sigma_\mu^* = 90\%$ with $\lambda_\mu^* = 1$ and $\sigma_S = 30\%$ on a double logarithmic scale.

As a conclusion, one can see that for these particular examples, the asymptotic behaviour of the numerical maximum likelihood estimator is close to the asymptotic behaviour of the theoretical maximum likelihood estimator.

4.4.2. *Example of application.* This paragraph gives an example of application of this estimator on real market data. We consider the components $\{S^j, j = 1..50\}$ of the Eurostoxx 50 on 11/11/2014. The backtest period is from 6/19/2008 to 11/11/2014. In this test, we assume a risk free rate equal to zero and that each stock is tradable at its closing price and without transaction costs. We consider two self-financing portfolios $\{P^i, i = 1, 2\}$:

$$\begin{aligned} \frac{dP_t^i}{P_t^i} &= \sum_{j=0}^{50} \omega_t^{j,i} \frac{dS_t^j}{S_t^j}, \\ P_0^i &= 100, \end{aligned}$$

based on two trend indicators $\{(\alpha_t^i(S^j)), i = 1, 2\}$. Let s be a threshold, for each portfolio $\{P^i, i = 1, 2\}$, the allocation $\left(\omega_t^{j,i}\right)$ is given by the following rules:

(1) At each time t, we define:

$$\begin{aligned} N_t^{i,+} &= \left\{ j \in \{1..50\} \setminus \alpha_t^i\left(S^j\right) > s \right\}, \\ N_t^{i,-} &= \left\{ j \in \{1..50\} \setminus \alpha_t^i\left(S^j\right) < -s \right\}. \end{aligned}$$

(2) If $N_t^{i,+} = \emptyset \bigvee N_t^{i,-} = \emptyset$:

$$\forall j \in \{1..50\}, \ \omega_t^{j,i} = 0,$$

else:

$$\forall j \in \{1..50\}, \ \omega_t^{j,i} = \frac{1}{\# N_t^{i,+}} \mathbf{1}_{j \in N_t^{i,+}} - \frac{1}{\# N_t^{i,-}} \mathbf{1}_{j \in N_t^{i,-}}$$

Since $\sum_{j=0}^{50} \omega_t^{j,i} = 0$, the two portfolios are market neutral. We consider the following trend indicators:

(1) $(\alpha_t^1(S^j))$ is the annualised sliding moving average estimator of daily returns:

$$\alpha_t^1\left(S^j\right) = \frac{252}{T} \sum_{k=1}^{T} \frac{dS_{t-k}^j}{S_{t-k}^j},$$

with $T = 252$ business days.

(2) $(\alpha_t^2(S^j))$ is the Kalman filter estimator of the model (1)-(2) where the parameters $\left(\lambda_\mu^t, \sigma_\mu^t, \sigma_S^t\right)$ are given by the numerical maximum likelihood (introduced above) over the last 252 business days. The initialisation of the BFGS algorithm at time $t > 0$ is given by the previous estimate $\left(\lambda_\mu^{t-1}, \sigma_\mu^{t-1}, \sigma_S^{t-1}\right)$ and the first one is equal to $(0.1, 10\%, 30\%)$.

Figure 14 illustrates the results with $s = 10\%$. In this example, the portfolio with Kalman filtering outperforms the one with the sliding moving average. Moreover, Figure 14 shows that the annualised Sharpe ratio (see [44]) of the Kalman filtering portfolio, computed with the daily returns, is higher and more robust to a threshold variation than the one with the sliding moving average.

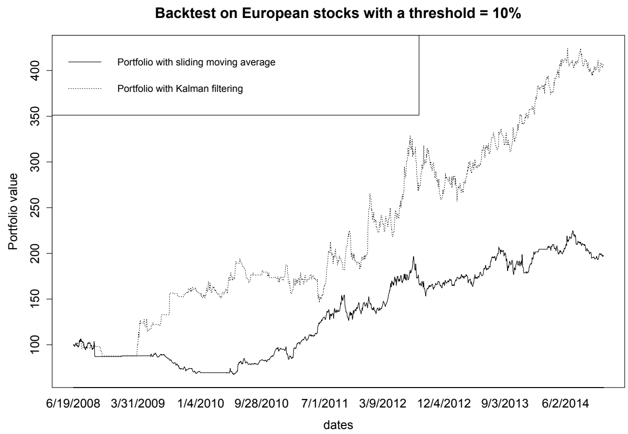

FIGURE 14. Portfolio evolutions with a threshold $s = 10\%$.

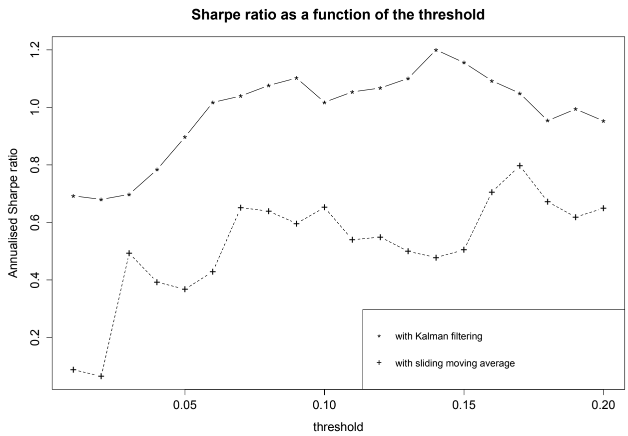

FIGURE 15. Sharpe ratios as a function of the threshold.

Satisfactory as the results may seem on this specific dataset, it is however important to note that the comparison performed in this section concerns only a very specific strategy. In a related work, a thorough study of single asset investment strategies based on various criteria is made, and the robustness of the investment strategy with respect to the model parameters is actually challenged. The interested reader is referred to [4].

5. Conclusion

The present work tries to illustrate the difficulty of trend filtering in a model based on an unobserved mean-reverting diffusion.

This model belongs to the class of Linear Gaussian Space State models. The advantage of this kind of system is to have an on-line method of estimation: the Kalman filter.

In practice, the parameters of the model are unknown, and the calibration of filtering parameters becomes crucial. The linear and Gaussian case allows to compute, in closed form, the likelihood. The Kalman filter can also be used for this analysis. These methods can be generalized to a non-constant volatility, and classical estimators can be easily implemented in practice.

Although this framework is particularly convenient for forecasting, the results of the analysis show that the classical estimators are not adapted to such a weak signal-to-noise ratio. The horizons of observations needed for an acceptable precision are too long. Therefore, the convergence is not guaranteed and the impact of misspecification on trend filtering is not negligible.

Despite these difficulties, the non-zero correlation between the trend and its estimate can be used for trend detection. This fact can explain why trend following strategies work well in practice.

Appendix A: discrete Kalman filter

Framework. This section is based on [29]. The discrete Kalman filter is a recursive method. Consider two objects: the observations $\{y_k\}$ and the states of the system $\{x_k\}$. This filter is based on a Gauss-Markov first order model. Consider the following system:

$$\begin{aligned} x_{k+1} &= F_k x_k + v_k, \\ y_k &= H_k x_k + u_k. \end{aligned}$$

The first equation is an *a priori* model, the transition equation of the system. The matrix F_k is the transition matrix and v_k is the transition noise. The second equation is the measurement equation. The matrix H_k is named the measurement matrix and u_k is the measurement noise. The aim is to identify the underlying process $\{x_k\}$. The two noises are supposed white, Gaussian, centered and decorrelated. In particular:

$$\mathbb{E}\left[\begin{pmatrix} u_k \\ v_k \end{pmatrix}\begin{pmatrix} u_l \\ v_l \end{pmatrix}^T\right] = \begin{pmatrix} R_k^u & 0 \\ 0 & R_k^v \end{pmatrix}\delta_{kl}.$$

The two noises are also supposed independent of x_k and the initial state is Gaussian. So, it can be proved with a recurrence that all states are Gaussian. Therefore, just the mean and the covariance matrix are needed for the characterization of the state. The estimation is given by two steps. The first one is an *a priori* estimation given $\hat{x}_{k+1/k}$ and $\Gamma_{k+1/k} = \mathbb{E}\left[(x_{k+1} - \hat{x}_{k+1/k})(x_{k+1} - \hat{x}_{k+1/k})^T\right]$. When the new observation is available, a correction of the estimation is done to obtain $\hat{x}_{k+1/k+1}$ and $\Gamma_{k+1/k+1} = \mathbb{E}\left[(x_{k+1} - \hat{x}_{k+1/k+1})(x_{k+1} - \hat{x}_{k+1/k+1})^T\right]$. This is the *a posteriori* estimation. The criterion considered for the *a posteriori* estimation is the least squares method, which corresponds to the minimization of the trace of $\Gamma_{k+1/k+1}$.

Filter. The prediction (*a priori* estimation) is given by

$$\begin{aligned} \hat{x}_{k+1/k} &= F_k \hat{x}_{k/k}, \\ \Gamma_{k+1/k} &= F_k \Gamma_{k/k} F_k^T + R_k^v. \end{aligned}$$

The *a posteriori* estimation is a correction of the *a priori* estimation. A gain is introduced to do this correction:

$$\hat{x}_{k+1/k+1} = \hat{x}_{k+1/k} + K_{k+1}\left(y_{k+1} - H_{k+1}\hat{x}_{k+1/k}\right).$$

As explained above, the gain K_{k+1} is found by least squares method, which corresponds to

$$\frac{\partial trace\left(\Gamma_{k+1/k+1}\right)}{\partial K_{k+1}} = 0.$$

Using straightforward matrix differentiation, the gain is found:

$$\begin{aligned} K_{k+1} &= \Gamma_{k+1/k} H_{k+1}^T \left[H_{k+1}\Gamma_{k+1/k}H_{k+1}^T + R_{k+1}^u\right]^{-1}, \\ \Gamma_{k+1/k+1} &= \left(I_d - K_{k+1}H_{k+1}\right)\Gamma_{k+1/k}. \end{aligned}$$

Appendix B: Likelihood computation

The likelihood can be computed using two methods. The first one is based on a direct calculus while the second method uses the Kalman filter.

5.0.3. *Direct computation of the likelihood.* A first approach is to directly compute the likelihood. The vectorial representation of the discrete time model (3)-(4) is:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix},$$

where $(\mu_1, \cdots, \mu_N)^T$ and $(u_1, \cdots, u_N)^T$, knowing $\theta = (\sigma_\mu, \lambda_\mu)$, are two independent Gaussian processes. Therefore the vector $(y_1, \cdots, y_N)^T$, knowing θ, is also a Gaussian process. The likelihood is then characterized by the mean $M_{y_{1:N}|\theta}$ and the covariance $\Sigma_{y_{1:N}|\theta}$:

$$M_{y_{1:N}|\theta} = 0 \ (\mu_0 = 0 \text{ is assumed}), \tag{21}$$

$$\Sigma_{y_{1:N}|\theta} = \Sigma_{\mu_{1:N}|\theta} + \Sigma_{u_{1:N}|\theta}, \tag{22}$$

where $\Sigma_{u_{1:N}|\theta} = \frac{\sigma_S^2}{\delta} I_N$ and $\Sigma_{\mu_{1:N}|\theta} = (\mathbb{C}ov\,(\mu_t, \mu_s))_{1 \le t,s \le N}$. Since the drift μ is an Ornstein Uhlenbeck process, then:

$$\mathbb{C}ov\,(\mu_t, \mu_s) = \frac{\sigma_\mu^2}{2\lambda_\mu} e^{-\lambda_\mu(s+t)} \left(e^{2\lambda_\mu s \wedge t} - 1\right). \tag{23}$$

Finally, the likelihood is given by:

$$f\,(y_1, ...y_N|\theta) = \frac{1}{(2\pi)^{N/2}\sqrt{det\Sigma_{y_{1:N}|\theta}}} e^{\left(\frac{-1}{2}(y_1,...,y_N)\Sigma^{-1}_{y_{1:N}|\theta}(y_1,...,y_N)^T\right)}. \tag{24}$$

Remark 5.1. *When the dimension N is large, it is difficult and numerically unstable to directly invert the covariance matrix $\Sigma_{y_{1:N}|\theta}$ and compute its determinant. An iterative approach, the details of which are given in Appendix C, is preferred.*

5.0.4. *Computation of the likelihood using the Kalman filter.* The likelihood can also be evaluated via the prediction error decomposition (see [43] for details):

$$\begin{aligned} f\,(y_1, ...y_N|\theta) &= f\,(y_N|y_1, ...y_{N-1}, \theta)\, f\,(y_1, ...y_{N-1}|\theta)\,. \\ &= \prod_{n=1}^{N} f\,(y_n|y_1, ...y_{n-1}, \theta)\,, \end{aligned}$$

where the conditional laws are given in the following proposition:

Proposition 5. *The process $(y_n|y_1, ...y_{n-1}, \theta)$ is Gaussian:*

$$(y_n|y_1, ...y_{n-1}, \theta) \sim \mathcal{N}\left(M_{y_{n|n-1}}, \mathbb{V}ar_{y_{n|n-1}}\right),$$

with

$$\begin{aligned}M_{y_{n|n-1}} &= e^{-\lambda\delta}\hat{\mu}_{n-1/n-1},\\ \mathbb{V}ar_{y_{n|n-1}} &= e^{-2\lambda\delta}\Gamma_{n-1/n-1} + \frac{\sigma_\mu^2}{2\lambda}\left(1-e^{-2\lambda\delta}\right) + \frac{\sigma_S^2}{\delta}.\end{aligned}$$

The a posteriori estimate of the trend $\hat{\mu}_{n-1/n-1}$ and the covariance error $\Gamma_{n-1/n-1}$ are given by Kalman filtering (see Equations (5) and (6)).

Proof. Since the process y_n is Gaussian, so is the process $(y_n|y_1, ...y_{n-1}, \theta)$. Moreover, using Equations (3)-(4), we have:

$$\begin{aligned}M_{y_{n|n-1}} &= \hat{\mu}_{n/n-1} + 0,\\ \hat{\mu}_{n/n-1} &= e^{-\lambda\delta}\hat{\mu}_{n-1/n-1} + 0,\end{aligned}$$

and

$$\begin{aligned}\mathbb{V}ar_{y_{n|n-1}} &= \Gamma_{n/n-1} + \frac{\sigma_S^2}{\delta},\\ \Gamma_{n/n-1} &= e^{-2\lambda\delta}\Gamma_{n-1/n-1} + \frac{\sigma_\mu^2}{2\lambda}\left(1-e^{-2\lambda\delta}\right).\end{aligned}$$

□

Remark 5.2. *In practice, the volatility is not constant. However, if the volatility σ_S is $\mathbf{F}^S$-adapted, the two methods can be suitably modified and implemented. This assumption is satisfied if the volatility is a continuous process.*

Appendix C: Iterative methods for the inverse and the determinant of the covariance matrix

We provide here, for the sake of completeness, some standard iterative methods to compute the inverse and the determinant of a covariance matrix.

Inverse of the covariance matrix. The use of the Matrix Inversion Lemma on Equation (21) gives:

$$\Sigma_{y_{1:N}|\theta}^{-1} = \Sigma_{\mu_{1:N}|\theta}^{-1} - \Sigma_{\mu_{1:N}|\theta}^{-1} A_N^{-1} \Sigma_{\mu_{1:N}|\theta}^{-1},$$

where $A_N = \frac{\delta}{\sigma_S^2} I_N + \Sigma_{\mu_{1:N}|\theta}^{-1}$. Then, we have to compute the inverse of the matrices A_N and $\Sigma_{\mu_{1:N}|\theta}$.

Inverse of the matrix A_N. Suppose that A_N^{-1} is computed. The matrix A_{N+1} can be broken into four sub-matrices:

$$A_{N+1} = \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix},$$

where

$$\begin{aligned}
B_1 &= \frac{\delta}{\sigma_S^2} + \frac{2\lambda_\mu\left(e^{\lambda_\mu\delta}+e^{-\lambda_\mu\delta}\right)}{\sigma_\mu^2\left(e^{\lambda_\mu\delta}-e^{-\lambda_\mu\delta}\right)},\\
B_2 &= \begin{pmatrix} \frac{-2\lambda_\mu}{\sigma_\mu^2\left(e^{\lambda_\mu\delta}-e^{-\lambda_\mu\delta}\right)} & 0 & \cdots & 0 \end{pmatrix},\\
B_3 &= B_2^T,\\
B_4 &= A_N.
\end{aligned}$$

Therefore, the matrix A_{N+1} can be inverted blockwise.

Inverse of the matrix $\Sigma_{\mu_{1:N}|\theta}$. The following lemma is used (see [1] for details):

Lemma 5.3. *Let μ be an Ornstein Uhlenbeck process with parameters $\theta = (\lambda_\mu, \sigma_\mu)$. The covariance matrix of $\mu_1, .., \mu_N$ is $\Sigma_{\mu_{1:N}|\theta}$. Then:*

$$\Sigma^{-1}_{\mu_{1:N}|\theta} = \frac{2\lambda_\mu}{\sigma_\mu^2\left(e^{\lambda_\mu\delta}-e^{-\lambda_\mu\delta}\right)} B_N,$$

$$B_N = \begin{pmatrix}
e^{\lambda_\mu\delta}+e^{-\lambda_\mu\delta} & -1 & 0 & \cdots & \cdots & 0\\
-1 & e^{\lambda_\mu\delta}+e^{-\lambda_\mu\delta} & -1 & \ddots & & \vdots\\
0 & -1 & e^{\lambda_\mu\delta}+e^{-\lambda_\mu\delta} & -1 & & \vdots\\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots\\
\vdots & & -1 & e^{\lambda_\mu\delta}+e^{-\lambda_\mu\delta} & -1 & 0\\
\vdots & & \ddots & -1 & e^{\lambda_\mu\delta}+e^{-\lambda_\mu\delta} & -1\\
0 & \cdots & \cdots & 0 & -1 & e^{\lambda_\mu\delta}
\end{pmatrix}.$$

Therefore, the inverse of the matrix $\Sigma_{\mu_{1:N+1}}$ is given by:

$$\Sigma^{-1}_{\mu_{1:N+1}|\theta} = \begin{pmatrix}
\frac{2\lambda_\mu\left(e^{\lambda_\mu\delta}+e^{-\lambda_\mu\delta}\right)}{\sigma_\mu^2\left(e^{\lambda_\mu\delta}-e^{-\lambda_\mu\delta}\right)} & \frac{-2\lambda_\mu}{\sigma_\mu^2\left(e^{\lambda_\mu\delta}-e^{-\lambda_\mu\delta}\right)} \quad 0 \quad \cdots \quad 0\\
\frac{-2\lambda_\mu}{\sigma_\mu^2\left(e^{\lambda_\mu\delta}-e^{-\lambda_\mu\delta}\right)} & \\
0 & \Sigma^{-1}_{\mu_{1:N}|\theta}\\
\vdots & \\
0 &
\end{pmatrix}.$$

Procedure. Finally, at time t, the inverse of the covariance matrix is given by the following protocol:

- Computation of the matrix A_t^{-1} using A_{t-1}^{-1}.
- Computation of the matrix $\Sigma^{-1}_{\mu_{1:t}|\theta}$ using $\Sigma^{-1}_{\mu_{1:t-1}|\theta}$.
- Using $\Sigma^{-1}_{y_{1:t}|\theta} = \Sigma^{-1}_{\mu_{1:t}|\theta} - \Sigma^{-1}_{\mu_{1:t}|\theta} A_t^{-1} \Sigma^{-1}_{\mu_{1:t}|\theta}$, the matrix $\Sigma^{-1}_{y_{1:t}|\theta}$ is obtained.

Determinant of the covariance matrix. The iterative computation of $\det\left(\Sigma_{y_{1:N}|\theta}\right)$ is based on the following lemma:

Lemma 5.4. *The determinant of the matrix $\Sigma_{y_{1:N}|\theta}$ is given by:*

$$\det\left(\Sigma_{y_{1:N}|\theta}\right) = \frac{\det\left(I_N + \frac{\sigma_S^2}{\delta}\Sigma^{-1}_{\mu_{1:N}|\theta}\right)}{\det\left(\Sigma^{-1}_{\mu_{1:N}|\theta}\right)}, \tag{25}$$

and for $N \geq 2$, we have:

$$\begin{aligned}
\det\left(\Sigma^{-1}_{\mu_{1:N+1}|\theta}\right) &= g(\lambda_\mu,\sigma_\mu)\left(e^{\lambda_\mu\delta}+e^{-\lambda_\mu\delta}\right)\det\left(\Sigma^{-1}_{\mu_{1:N}|\theta}\right) \\
&\quad -g(\lambda_\mu,\sigma_\mu)^2\det\left(\Sigma^{-1}_{\mu_{1:N-1}|\theta}\right), \\
\det\left(I_{N+1}+\frac{\sigma_S^2}{\delta}\Sigma^{-1}_{\mu_{1:N+1}|\theta}\right) &= \left(1+\frac{\sigma_S^2}{\delta}g(\lambda_\mu,\sigma_\mu)\left(e^{\lambda_\mu\delta}+e^{-\lambda_\mu\delta}\right)\right)\det\left(I_N+\frac{\sigma_S^2}{\delta}\Sigma^{-1}_{\mu_{1:N}|\theta}\right) \\
&\quad -\left(\frac{\sigma_S^2}{\delta}g(\lambda_\mu,\sigma_\mu)\right)^2\det\left(I_{N-1}+\frac{\sigma_S^2}{\delta}\Sigma^{-1}_{\mu_{1:N-1}|\theta}\right),
\end{aligned}$$

where

$$g\left(\lambda_\mu, \sigma_\mu\right) = \frac{2\lambda_\mu}{\sigma_\mu^2\left(e^{\lambda_\mu\delta} - e^{-\lambda_\mu\delta}\right)}.$$

Proof. The multiplication of Equation (21) by $\Sigma^{-1}_{\mu_{1:N}|\theta}$ gives:

$$\Sigma^{-1}_{\mu_{1:N}|\theta}\Sigma_{y_{1:N}|\theta} = I_N + \frac{\sigma_S^2}{\delta}\Sigma^{-1}_{\mu_{1:N}|\theta}.$$

Equation (25) follows. Using Lemma 5.3, The matrices $\left(I_N + \frac{\sigma_S^2}{\delta}\Sigma^{-1}_{\mu_{1:N}|\theta}\right)$ and $\Sigma^{-1}_{\mu_{1:N}|\theta}$ are tridiagonal. The recursive computation of their determinant is then possible. □

References

[1] F. Akesson and J. Lehoczky. Discrete eigenfunction expansion of multi-dimensional brownian motion and the ornstein-uhlenbeck process. Technical report, 1998.

[2] S. H. Babbs and K. M. Nowman. Kalman filtering of generalised vasicek term structure models. *Journal of Financial and Quantitative Analysis*, 1999.

[3] M. Bebendo and S. D. Hodges. The dynamics of the volatility skew: A Kalman filter approach. *Journal of Banking and Finance*, 2009.

[4] A. Bel Hadj Ayed. *Robustness of the optimal trading strategy.* PhD thesis, Université de Paris Saclay, 2016.

[5] B. Benmiloud and W. Pieczynski. Parameter Estimation in Hidden Markov Chains and Segmentation of Images. Technical report, Institut National des Télécommunications, 1995.

[6] S. Brendle. Portfolio selection under incomplete information. *Stochastic Processes and their Applications*, 116(5):701 – 723, 2006.

[7] P. J. Brockwell and R. A. Davis. *Introduction to time series and forecasting.* Springer-Verlag New York, 2002.

[8] P. J. Brockwell and R. A. Davis. *Time series: Theory and Methods.* Springer-Verlag New York, 2002.

[9] B. Bruder and N. Gaussel. Risk-Return Analysis of Dynamic Investment Strategies. Technical report, Lyxor, 2011.

[10] S. Butterworth. On the theory of filter amplifiers. *Experimental Wireless and the Wireless Engineer*, 1930.

[11] O. Cappé, T. Rydén, and E. Moulines. *Inference in Hidden Markov Models.* Springer-Verlag, 2005.

[12] R. Casarin and J. M. Marin. Online data processing: comparison of Bayesian regularized particle filters. Technical report, INRIA, 2007.

[13] A. Chakraborti, I. Muni Toke, M. Patriarca, and F. Abergel. Econophysics review: I. empirical facts. *Quantitative Finance*, 11(7):991–1012, 2011.

[14] S. Chib, F. Nardari, and N. Shephard. Markov chain monte carlo methods for stochastic volatility models. *Journal of Econometrics*, June 2002.

[15] K. Dahia. *Nouvelles méthodes en filtrage particulaire. Application au recalage de navigation inertielle par mesures altimétriques.* PhD thesis, Université Joseph Fourier, January 2005.

[16] A. Das and T. K. Ghoshal. Market risk beta estimation using adaptive kalman filter. *International Journal of Engineering Science and Technology*, 2010.

[17] A. Dembo and O. Zeitouni. Parameter Estimation of Partially Observed Continious Time Stochastic Processes Via EM Algorithm. *The annals of Applied Probability*, 2007.

[18] J. DiLellio. A Kalman filter control technique in mean-variance portfolio management. *Journal of Economics and Finance*, 2012.

[19] R. J. Elliott, J. Van Der Hoek, and W. P. Malcolm. Pairs trading. *Quantitative Finance*, 2005.

[20] B. Eraker. Mcmc analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics*, 1998.

[21] H. Frydman and P. Lakner. Maximum Likelihood Estimation of Hidden Markov Processes. *The annals of Applied Probability*, 2007.

[22] R. Genay, F. Seluk, and B. Whitcher. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics.* Academic Press, 2002.

[23] A. C. Harvey. *Forecasting, structural time series models and the Kalman filter.* Cambridge Univ. Press, 1990.

[24] A. C. Harvey and S. Peters. Estimation procedures for structural time series models. *Journal of Forecasting*, 1990.

[25] A. C. Harvey and T.M. Trimbur. General model-based filters for extracting cycles and trends in economic time series. *The review of Economics and Statistics*, 2003.

[26] C. C. Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 2004.

[27] E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian analysis of stochastic volatility models, 1994.

[28] F. M. John. Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association*, 1960.

[29] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 1960.

[30] S. Kim, N. Shephard, and S. Chib. Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies*, pages 361–393, 1998.

[31] P. Lakner. Optimal trading strategy for an investor: the case of partial information. *Stochastic Processes and their Applications*, 1998.

[32] J. M. Laskry and P. L. Lions. Contrôle stochastique avec informations partielles et applications la finance. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 328(11):1003 – 1010, 1999.

[33] Y. Lempérière, C. Deremble, P. Seager, M. Potters, and J. P. Bouchaud. Two centuries of trend following. *Journal of Investment Strategies*, 2014.

[34] B. G. Leroux. Maximum-likelihood estimation for hidden markov models. *Stochastic Processes and their Applications*, 40(1):127 – 143, 1992.

[35] B. B. Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 1963.

[36] M. Manoliu and S. Tompaidis. Energy futures prices: term structure models with kalman filter estimation. *Applied Mathematical Finance*, 2002.

[37] A. Maravall. Stochastic linear trends. *Journal of Econometrics*, 1993.

[38] H. Markovitz. Portfolio Selection. *The Journal of Finance*, 1952.

[39] R. C. Merton. *Continuouse-Time Finance*. Blackwell, 1990.

[40] J. F. Monahan. Fully bayesian analysis of arma time series models. *Journal of Econometrics*, 21(3):307 – 331, 1983.

[41] H. Pham and M. C. Quenez. Optimal Portfolio in Partially Observed Stochastic Volatility Models. *Annals of Applied Probability*, 11:210–238, 2001.

[42] D. S. G. Pollock. Wiener-kolmogorov filtering, frequency-selective filtering, and polynomial regression. *Econometric Theory*, 2007.

[43] F. C. Schweppe. Evaluation of likelihood functions for gaussian signals. *IEEE Transactions on Information Theory*, 1965.

[44] W. F. Sharpe. Mutual fund performance. *The Journal of Business*, 1966.

[45] C. Stasinakis, G. Sermpinis, K. Theofilatos, and A. Karathanasopoulos. Forecasting us unemployment with radial basis neural networks, kalman filters and support vector regressions. *Computational Economics*, 2014.

[46] K. Tamaki. The bernstein-von mises theorem for stationary processes. *Journal of the Japan statistical society*, 38(2):311–323, 2008.

[47] C. Wells. *The Kalman Filter in Finance*. Springer Netherlands, 1996.

[48] P. Whittle. The analysis of multiple stationary time series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 15(1):125–139, 1953.

[49] P. Whittle and T. J. Sargent. *Prediction and Regulation by Linear Least-Square Methods*. University of Minnesota Press, ned - new edition edition, 1983.

[50] P. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series With engineering applications*. Technology Press of the Massachusetts Institute of Technology; John Wiley and Sons: New York, 1949.