# Performance Guarantees for Misspecified Score-Driven Filters

EEA-ESEM 2024

**Simon Donker van Heel** **Rutger-Jan Lange**

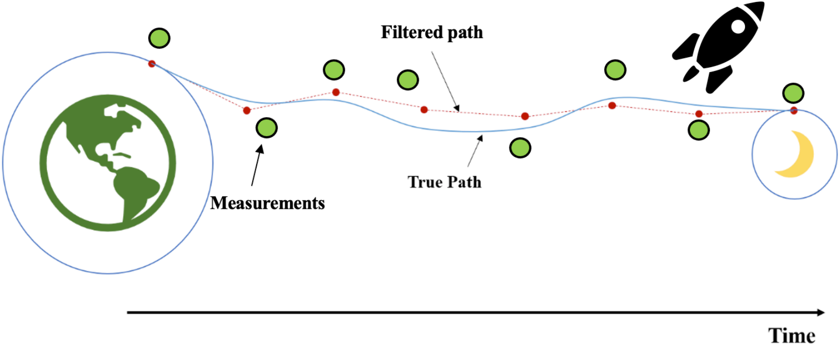
**Dick van Dijk** **Bram van Os**

August 26, 2024

Erasmus University Rotterdam

tinbergen institute

# Filtering: Tracking latent states in real-time using noisy measurements



# Motivation

## Applications

- Finance: Volatility modelling
- Climate science: Earthquake detection, sea ice cover sizing

## Related literature

- Filtering using the score, i.e. gradient of log-likelihood (Harvey, 2013; Creal et al., 2013; Artemova et al., 2022).
- $>$ 300 academic articles on **explicit score-driven (ESD)** filters available on `www.gasmodel.com`.
- 1 on **implicit score-driven (ISD)** filters (Lange et al., 2024).
- Related to Stochastic Gradient Descent (SGD), used for training neural networks, e.g. implicit SGD (Toulis & Airoldi, 2017).

# Contributions

## Contributions

1. Derive performance guarantees, i.e. (non-)asymptotic MSE bounds for score-driven filters,
   - while making almost no assumptions on the true state process the filter aims to track over time.
2. Reveal that **implicit score-driven** filters are more stable than **explicit score-driven** filters;
   - ESD filters may overshoot or diverge to infinity without a certain regularity condition (Lipschitz gradient).

# Problem setting

- Observe data vector $\boldsymbol{y}_t$ at times $t = 1, \ldots, T$ from true conditional observation density $p^\dagger(\boldsymbol{y}_t|\boldsymbol{\theta}_t^\dagger)$.
- True latent state vector $\boldsymbol{\theta}_t^\dagger$.
- Researcher-postulated density $p(\boldsymbol{y}_t|\boldsymbol{\theta}_t)$.
- Objective: Track true density $p^\dagger(\boldsymbol{y}_t|\boldsymbol{\theta})$ at true path $\{\boldsymbol{\theta}_t^\dagger\}$ by postulated density $p(\boldsymbol{y}_t|\boldsymbol{\theta}_t)$ at filtered path $\{\boldsymbol{\theta}_{t|t}\}$, $\{\boldsymbol{\theta}_{t|t-1}\}$.
- Loss function: $\text{MSE}_{t|t} := \text{E}[\|\boldsymbol{\theta}_{t|t} - \boldsymbol{\theta}_t^\star\|^2]$, where $\boldsymbol{\theta}_t^\star$ pseudo-truth.

# Filtering multivariate states using the score

- Implicit score-driven (ISD) filter, $\{\boldsymbol{\theta}^{\text{im}}_{t|t}\}$ and $\{\boldsymbol{\theta}^{\text{im}}_{t+1|t}\}$.
- Explicit score-driven (ESD) filter, $\{\boldsymbol{\theta}^{\text{ex}}_{t|t}\}$ and $\{\boldsymbol{\theta}^{\text{ex}}_{t+1|t}\}$.

- Given initializations $\boldsymbol{\theta}^{j}_{0|0} \in \mathbb{R}^d$, with $j \in \{\text{im}, \text{ex}\}$, $\forall t \in \mathbb{N}$:

$$\text{prediction step: } \boldsymbol{\theta}^{j}_{t|t-1} = (\boldsymbol{I}_d - \boldsymbol{\Phi})\,\boldsymbol{\omega} + \boldsymbol{\Phi}\,\boldsymbol{\theta}^{j}_{t-1|t-1},\; j \in \{\text{im}, \text{ex}\} \quad (1)$$

$$\text{implicit-gradient update: } \quad \boldsymbol{\theta}^{\text{im}}_{t|t} = \boldsymbol{\theta}^{\text{im}}_{t|t-1} + \boldsymbol{H}_t \nabla \ell(\boldsymbol{y}_t \mid \boldsymbol{\theta}^{\text{im}}_{t|t}) \quad (2)$$

$$\text{explicit-gradient update: } \quad \boldsymbol{\theta}^{\text{ex}}_{t|t} = \boldsymbol{\theta}^{\text{ex}}_{t|t-1} + \boldsymbol{H}_t \nabla \ell(\boldsymbol{y}_t \mid \boldsymbol{\theta}^{\text{ex}}_{t|t-1}) \quad (3)$$

# Reformulation as optimization-based filters

- Trade-off between *improving goodness of fit* and *maintaining stability over time*
- Reformulate gradient-based updates (2) and (3) in terms of an optimization framework:

$$\boldsymbol{\theta}^{\text{im}}_{t|t} = \underset{\boldsymbol{\theta}\in\mathbb{R}^d}{\text{argmax}} \left\{ \ell\left(\boldsymbol{y}_t \mid \boldsymbol{\theta}\right) - \frac{1}{2} \left\| \boldsymbol{\theta} - \boldsymbol{\theta}^{\text{im}}_{t|t-1} \right\|^2_{\boldsymbol{P}_t} \right\}, \tag{4}$$

$$\boldsymbol{\theta}^{\text{ex}}_{t|t} = \underset{\boldsymbol{\theta}\in\mathbb{R}^d}{\text{argmax}} \left\{ \underbrace{\ell(\boldsymbol{y}_t \mid \boldsymbol{\theta}^{\text{ex}}_{t|t-1}) + (\boldsymbol{\theta} - \boldsymbol{\theta}^{\text{ex}}_{t|t-1})' \, \nabla\ell(\boldsymbol{y}_t \mid \boldsymbol{\theta}^{\text{ex}}_{t|t-1})}_{\approx \ell(\boldsymbol{y}_t|\boldsymbol{\theta})} - \frac{1}{2} \left\| \boldsymbol{\theta} - \boldsymbol{\theta}^{\text{ex}}_{t|t-1} \right\|^2_{\boldsymbol{P}_t} \right\}. \tag{5}$$

# We distinguish four different cases

**Assumption**

1. **ISD filter:** $\ell(\boldsymbol{y}_t|\boldsymbol{\theta})$ *is strongly concave.*
2. **ESD filter:** $\ell(\boldsymbol{y}_t|\boldsymbol{\theta})$ *is strongly concave + Lipschitz c. gradient.*

(a) **Misspecification:** *Pseudo-truth* $\boldsymbol{\theta}_t^\star$ *exists and is unique, and the increments* $\{\boldsymbol{\theta}_t^\star - \boldsymbol{\theta}_{t-1}^\star\}$ *have a finite second moment.*

(b) **Correct specification:** *Observation density is correctly specified. State equation is linear, Gaussian with known coefficients.*

# Deriving error bounds

Linear system of inequalities to bound $\{\mathrm{MSE}_{t|t}\}_{t\geq 1}$, $\{\mathrm{MSE}_{t|t-1}\}_{t\geq 1}$:

$$\text{filtering-error bound : } \quad \mathrm{MSE}_{t|t} \leq a\,\mathrm{MSE}_{t|t-1} + \underbrace{b}_{\text{noise}}, \tag{6}$$

$$\text{prediction-error bound: } \quad \mathrm{MSE}_{t+1|t} \leq c\,\mathrm{MSE}_{t|t} + \underbrace{d}_{\text{drift}}. \tag{7}$$

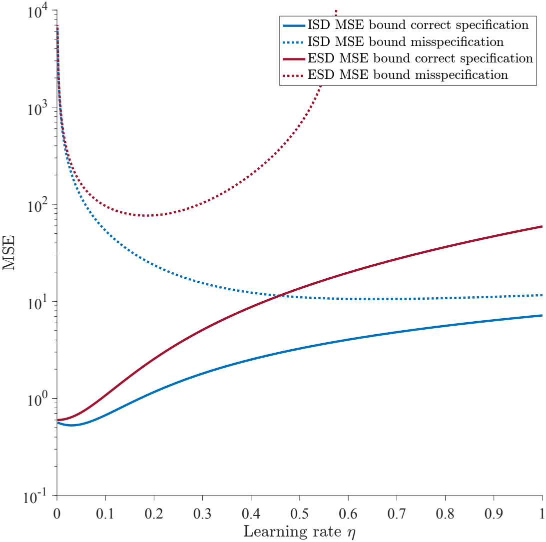
**Finite-sample MSE bounds**, if $ac < 1$:

$$\mathrm{MSE}_{t|t} \leq \underbrace{a^t c^{t-1}\,\mathrm{MSE}_{1|0}}_{\text{initialisation}} + \underbrace{\frac{1 - a^t c^t}{1 - a\,c} b}_{\text{noise}} + \underbrace{\frac{1 - a^{t-1} c^{t-1}}{1 - a\,c} a\,d}_{\text{drift}}, \tag{8}$$

**Asymptotic MSE bounds**, if $ac < 1$:

$$\limsup_{t\to\infty} \mathrm{MSE}_{t|t} \leq \frac{b + a\,d}{1 - a\,c}. \tag{9}$$

# Asymptotic MSE bounds in four scenarios



# When is the bound tight?

## Example

Consider data generated from a **local level model**. Suppose we are correctly specified, then the *ISD filter asymptotic MSE bound*

$$\limsup_{t\to\infty} \mathrm{MSE}_t \le \frac{\sigma_\varepsilon^{\star 2} + \sigma_\eta^{\star 2}\sigma_\varepsilon^{\star 4}\gamma^2}{2\gamma\sigma_\varepsilon^{\star 2} + 1}.$$

is minimized for a *learning rate*

$$\hat{\eta} := \hat{\gamma}^{-1} = \frac{2}{\left(\sigma_\eta^{\star 2} + \sigma_\eta^{\star 2}\sqrt{4\sigma_\varepsilon^{\star 2} + \sigma_\eta^{\star 2}}\right)} =: \bar{P},$$

where $\bar{P}$ is the steady-state Kalman filter covariance for the local level model in Durbin & Koopman (2012). Since the Kalman filter is optimal in the MSE sense, the bound cannot be improved.

# New hyperparameter tuning strategy

## Choosing update- and prediction parameters optimally

- We propose to select $\boldsymbol{\omega}^j$, $\boldsymbol{\Phi}^j$, $\boldsymbol{H}^j$ for $j \in \{\text{ex}, \text{im}\}$ to minimize (upper bound on) the asymptotic MSE.
- This strategy provides an alternative to maximum likelihood:
  1. Does not require many observations.
  2. Requires less parameters to be 'determined'.
  3. Has analytical expressions available.
  4. Provides strong (worst-case) asymptotic performance.
- Similar to Toulis & Airoldi (2017).

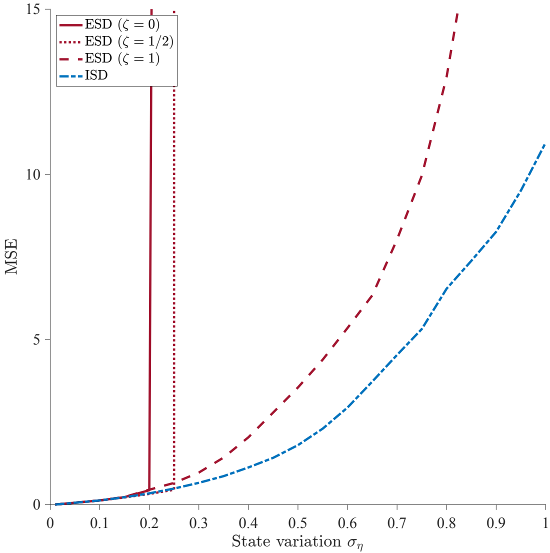
# Simulation: Time-varying Poisson

- Simulate state space model:
- Observation equation is Poisson with time-varying rate:

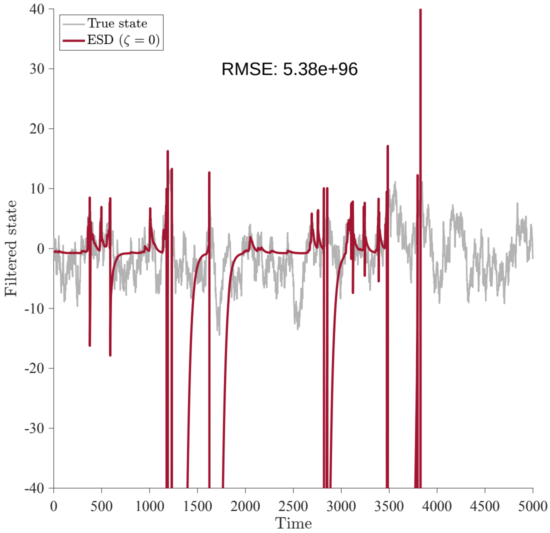
  $y_t \sim p^\dagger(y_t|\theta_t^\dagger) = \lambda_t^{\dagger y_t} \exp(-\lambda_t^\dagger)/y_t!, \quad \lambda_t^\dagger = \exp(\theta_t^\dagger)$
- State equation is linear Gaussian AR(1):

  $\theta_t^\dagger = \omega^\dagger + \phi^\dagger \theta_{t-1}^\dagger + \eta_t, \quad \eta_t \sim \text{i.i.d.}\, \mathrm{N}(0, \sigma_\eta^2), \quad \phi^\dagger < 1,\ T = 5000$
- Goal is to track $\{\theta_t^\dagger\}$ using ISD and ESD[1,2,3] filter
  1. Identity scaling
  2. Inverse square root Fisher scaling
  3. Inverse Fisher scaling

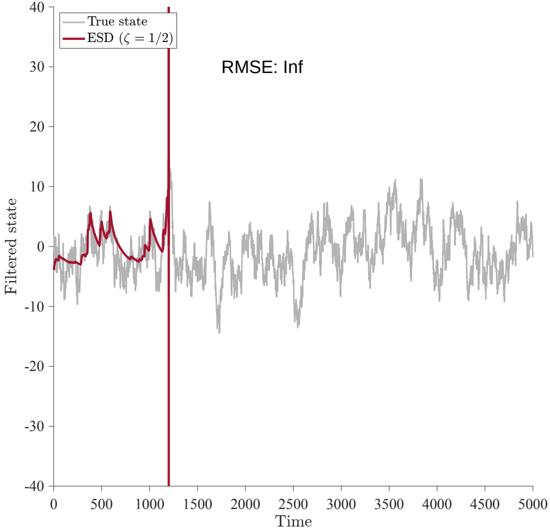
# ISD filter outperforms ESD filter



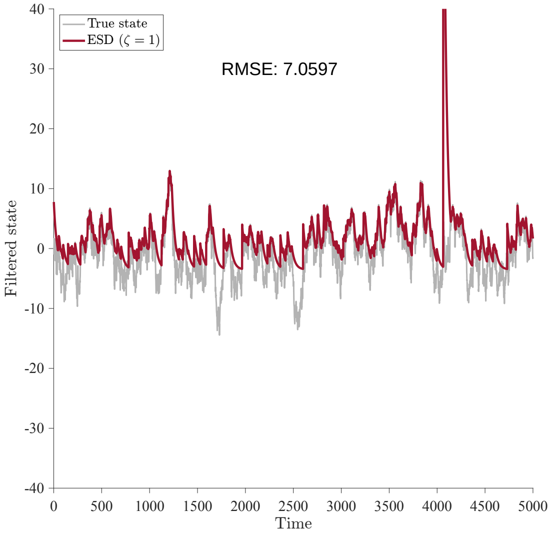
# (a) ESD ($\zeta = 0$) overshoots downward



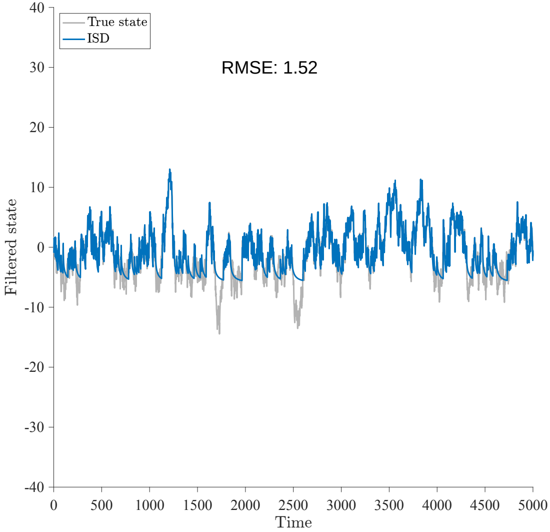
# (b) ESD ($\zeta = 1/2$) diverges to infinity



# (c) ESD ($\zeta = 1$) overshoots upward



# (d) ISD filter performs relatively well



# Discussion

## Summary

- Performance guarantees for ISD and ESD filters
- Minimal assumptions on the DGP
- ESD filters require more regularity conditions
- ISD filters are slightly more computationally expensive
- ISD filters have sharper MSE bounds

## What else did we do in the paper?

- Show how to filter non-stationary latent states.
- Show how to compute the implicit-gradient update.
- Show Kalman filter is a special case of ISD ánd ESD filters.

*Thank you!*

Simon Donker van Heel
donkervanheel@ese.eur.nl

Erasmus
University
Rotterdam

Artemova, M., Blasques, F., van Brummelen, J., & Koopman, S. J. (2022). Score-driven models: Methodology and theory. In *Oxford research encyclopedia of economics and finance.*

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777–795.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (Vol. 38). OUP Oxford.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series* (Vol. 52). Cambridge University Press.

Lange, R.-J., van Os, B., & van Dijk, D. (2024). Implicit score-driven filters for time-varying parameter models. *ResearchGate*.

Toulis, P., & Airoldi, E. M. (2017). Asymptotic and finite-sample properties of estimators based on stochastic gradients. *Annals of Statistics*, *45*, 1694–1727.