**Abstract**

This paper provides first a brief summary of the SEMIFAR (semiparametric fractional autoregressiove) and ESEMIFAR (exponential SEMIFAR) models. Those models are extended slightly to include the moving average part. Under common distribution condition it is shown that the long memory parameter is not affected by the log-transformation. A simple data-driven algorithm is proposed, by which the selected bandwidth and the selected orders of the ARMA model are all consistent. An R package is developed for practical implementation. The application of the proposals are illustrated by different kind of time series.

*Keywords:* Nonparametric regression with long memory, SEMIFAR, ESEMIFAR, bandwidth selection, model selection, implementation in R,

*JEL Codes:* C14, C51

1. Introduction

2. The SEMIFARIMA model

3. The E-SEMIFARIMA model

A theorem about the memory property of the ESEMIFAR (direct after that model)

**Theorem 1.** *Assume that $X_t$ follows an ESEMIFAR model defined above with $d > 0$ and a linear FARIMA error process $\xi_t$ in the log-data, then $Z_t = X_t/v(\tau_t)$ is a stationary long memory process with the same memory parameter $d$.*

This result shows that the level of long memory of the stationary part of the process under consideration is not affected by the log- or exponential transformations. In the case when $\xi_t$ is normal, those results are e.g. obtained in Dittmann and Granger (2002) and Beran et al. (2015). Theorem 2 extends those results to a common innovation distribution in the log-data. The proof of Theorem 2 is given in the appendix, where we will show that the process $Z_t$ is a special case of the general framework defined in Surgailis and Viano (2002). Hence, their results apply to $Z_t$. Note however that the above result does not hold for $d = 0$. Detailed discussion on corresponding properties in case with $d = 0$ is beyond the aim of the current paper.

4. Data-driven estimation

**Theorem 2.** *Under assumptions A1 - A5, the selected AR order $\hat{p}_1$ and MA order $\hat{q}_1$ at the end of the first out-side iteration are both consistent.*

The proof of Theorem 1 is given in the appendix. Note in particular that to achieve consistent model selection only the first out-side iteration is required. Hence, if $p_0$ and $q_0$ happen to be chosen correctly, the procedure will usually be stopped after the first out-side iteration. Otherwise, it will usually be stopped after the second out-side iteration. If $n$ is large enough, in most of the cases the third or further out-side iterations are not required. This ensures that the running time of the proposed algorithm comparable to that required by the original SEMIFAR algorithm.

5. Implementation in R

6. Application to different kinds of time series

6.1 Application of the SEMIFARIMA

6.2 Application of the ESEMIFARIMA

6.3 Application to high-frequency financial data

7. Concluding remarks

Appendix