Estimation of the Long-run Error Variance in Nonparametric Regression with Time Series Errors

Marina Khismatullina [1] Michael Vogt [2]

EcoSta 2023

[1]Erasmus University Rotterdam
[2]Ulm University

Introduction

Model

We observe a single time series $\{Y_t : 1 \leq t \leq T\}$ of length T. The observations come from the following model:

$$Y_t = m\Big(\frac{t}{T}\Big) + \varepsilon_t$$

- m is an unknown trend function on $[0, 1]$;
- $\{\varepsilon_t : 1 \leq t \leq T\}$ is a zero-mean stationary and causal error process.

Problem

Estimate the long-run error variance $\sigma^2 = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\varepsilon_0, \varepsilon_\ell)$.

Literature

Residual-based approach: estimate σ^2 from the residuals

$$\widehat{\varepsilon}_t = Y_t - \widehat{m}\left(\frac{t}{T}\right)$$

- AR(p) error processes (Truong, 1991; Shao and Yang, 2011; Qiu et al., 2013)

Difference-based approach: estimate σ^2 from the ℓ-th differences $Y_t - Y_{t-\ell}$.

- AR(p) error processes (Hall and Van Keilegom, 2003)
- MA(m) error processes (Müller and Stadtmüller, 1988; Herrmann et al., 1992; Tecuapetla-Gómez and Munk, 2017)

Model

Setting

Estimate the long-run error variance $\sigma^2 = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\varepsilon_0, \varepsilon_\ell)$ of the error terms $\{\varepsilon_t\}$ in the model

$$Y_t = m\Big(\frac{t}{T}\Big) + \varepsilon_t,$$

where $m(\cdot)$ is Lipshitz and $\{\varepsilon_t\}$ is an AR($p^\star$) process of the form

$$\varepsilon_t = \sum_{j=1}^{p^\star} a_j \varepsilon_{t-j} + \eta_t.$$

- $a_1, a_2, a_3, \ldots$ are the unknown parameters;
- η_t are i.i.d. innovations with $\mathbb{E}[\eta_t] = 0$ and $\mathbb{E}[\eta_t^2] = \nu^2$;
- $p^\star \in \mathbb{N} \cup \{\infty\}$ is unknown.Two possible cases:
  (A) $p^\star$ is not known but we know an upper bound p on it;
  (B) or we neither know $p^\star$ nor an upper bound on it.

Setting

We assume that

$$A(z) := 1 - \sum_{j=1}^{\infty} a_j z^j \neq 0$$

for all complex numbers $|z| \leq 1 + \delta$ with some small $\delta > 0$.

Therefore,

- the error process $\{\varepsilon_t\}$ is stationary and causal;
- the coefficients $a_1, a_2, a_3, \ldots$ decay to zero exponentially fast;
- $\{\varepsilon_t\}$ has an MA(∞) representation of the form $\varepsilon_t = \sum_{k=0}^{\infty} c_k \eta_{t-k}$.

Estimation

Motivation for the estimator

If $\{\varepsilon_t\}$ is an AR($p^\star$) process, then the time series $\{\Delta_q \varepsilon_t\}$ of the differences $\Delta_q \varepsilon_t = \varepsilon_t - \varepsilon_{t-q}$ is an ARMA($p^\star, q$) process of the form

$$\Delta_q \varepsilon_t - \sum_{j=1}^{p^\star} a_j \Delta_q \varepsilon_{t-j} = \eta_t - \eta_{t-q}.$$

Then, since the trend function $m(\cdot)$ is Lipshitz, $\Delta_q Y_t = Y_t - Y_{t-q}$ is approximately an ARMA($p^\star, q$) process.

Yule-Walker equations

For any differencing order $q \geq 1$, we have

$$\gamma_q(\ell) - \sum_{j=1}^{p^\star} a_j \gamma_q(\ell - j) = \begin{cases} -\nu^2 c_{q-\ell} & \text{for } 1 \leq \ell < q+1, \\ 0 & \text{for } \ell \geq q+1. \end{cases}$$

where

- $\boldsymbol{c}_q = (c_{q-1}, \ldots, c_{q-p})^\top$ are the coefficients from the MA(∞) expansion of $\{\varepsilon_t\}$;
- $\boldsymbol{\gamma}_q = (\gamma_q(1), \ldots, \gamma_q(p))^\top$ with $\gamma_q(\ell) = \mathrm{Cov}(\Delta_q \varepsilon_t, \, \Delta_q \varepsilon_{t-\ell})$;
- and Γ_q is the $p \times p$ covariance matrix $\Gamma_q = (\gamma_q(i-j) : 1 \leq i, j \leq p)$.

In vector notation

$$\Gamma_q \boldsymbol{a} = \boldsymbol{\gamma}_q + \nu^2 \boldsymbol{c}_q$$

Estimator, first stage

Note

$\Gamma_q \boldsymbol{a} \approx \gamma_q$ for large values of q.

We construct the first-stage estimator by

$$\widetilde{\boldsymbol{a}}_q = \widehat{\Gamma}_q^{-1} \widehat{\gamma}_q,$$

where $\widehat{\Gamma}_q$ and $\widehat{\gamma}_q$ are constructed from the sample autocovariances

$$\widehat{\gamma}_q(\ell) = (T-q)^{-1} \sum_{t=q+\ell+1}^{T} \Delta_q Y_t \Delta_q Y_{t-\ell}.$$

Estimator, second stage

Problem

If the trend m is pronounced, the estimator $\widetilde{\boldsymbol{a}}_q$ will have a strong bias.

Solution:

- Compute estimators $\widetilde{c}_k$ of c_k based on $\widetilde{\boldsymbol{a}}_q$.
- Estimate the innovation variance ν^2 by $\widetilde{\nu}^2 = (2T)^{-1} \sum_{t=p+2}^{T} \widetilde{r}_t^2$, where $\widetilde{r}_t = \Delta_1 Y_t - \sum_{j=1}^{p} \widetilde{a}_j \Delta_1 Y_{t-j}$.
- Estimate $\boldsymbol{a}$ by
$$\widehat{\boldsymbol{a}}_r = \widehat{\Gamma}_r^{-1} (\widehat{\gamma}_r + \widetilde{\nu}^2 \widetilde{\boldsymbol{c}}_r).$$
- Average the estimators $\widehat{\boldsymbol{a}}_r$: $\widehat{\boldsymbol{a}} = \frac{1}{\overline{r} - \underline{r} + 1} \sum_{r=\underline{r}}^{\overline{r}} \widehat{\boldsymbol{a}}_r$.
- Estimate the long-run variance σ^2 by
$$\widehat{\sigma}^2 = \frac{\widehat{\nu}^2}{(1 - \sum_{j=1}^{p} \widehat{a}_j)^2}.$$

Tuning parameters

$$\widetilde{\boldsymbol{a}}_q = \widehat{\Gamma}_q^{-1} \widehat{\gamma}_q$$

Problem

How to choose q?

(i) q should be large enough so that $\boldsymbol{c}_q = (c_{q-1}, \ldots, c_{q-p})^\top$ is close to zero;

(ii) q should not be too large to sufficiently eliminate the trend.

In case of AR(1), $q = 20$ is enough.

For the consistency, we need $\log T \ll q \ll \sqrt{T}$.

Tuning parameters

$$\widehat{\boldsymbol{a}}_r = \widehat{\Gamma}_r^{-1}(\widehat{\boldsymbol{\gamma}}_r + \widetilde{\nu}^2 \widetilde{\boldsymbol{c}}_r)$$

(i) r is a much smaller differencing order than q;

(ii) $r \geq 1$ is sufficient.

$$\widehat{\boldsymbol{a}} = \frac{1}{\overline{r} - \underline{r} + 1} \sum_{r=\underline{r}}^{\overline{r}} \widehat{\boldsymbol{a}}_r$$

Problem

How to choose $\underline{r}$ and $\overline{r}$?

We choose them to be fixed (small) natural numbers. Simulations in the paper.

Theoretical properties

Performance:

- Our estimator $\widehat{\boldsymbol{a}}$ produces accurate estimation results even when the AR polynomial $A(z) = 1 - \sum_{j=1}^{p^\star} a_j z^j$ has a root close to the unit circle.
- Our pilot estimator $\widetilde{\boldsymbol{a}}_q$ tends to have a substantial bias when the trend m is pronounced. Our estimator $\widehat{\boldsymbol{a}}$ reduces this bias considerably.

Proposition

Our estimators $\widetilde{\boldsymbol{a}}_q$, $\widehat{\boldsymbol{a}}$ and $\widehat{\sigma}^2$ are $\sqrt{T}$-consistent.

Simulations

Simulations

Setting:

- data from the model $Y_t = m(t/T) + \varepsilon_t$, where ε_t is an AR(1) process of the form $\varepsilon_t = a_1 \varepsilon_{t-1} + \eta_t$;
- $a_1 \in \{-0.95, -0.75, -0.5, -0.25, 0.25, 0.5, 0.75, 0.95\}$;
- sample size T is 500;
- the trend function is linear $m(u) = \beta u$;
- we generate $S = 1000$ data samples;
- $q = 25, \underline{r} = 1, \overline{r} = 10$;
- tuning parameters for the estimators from Hall and Van Keilegom (2003) are $m_1 = 20$ and $m_2 = 30$.

Simulations

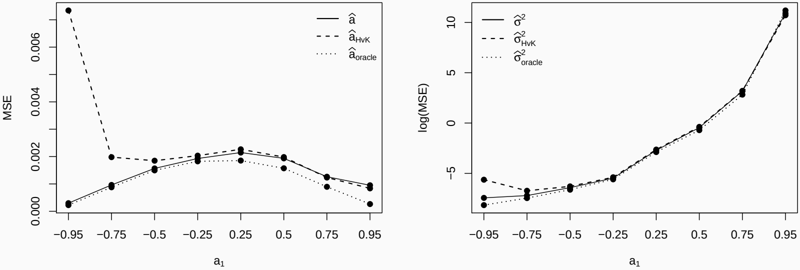

Figure 1: MSE values for the estimators $\widehat{a}$, $\widehat{a}_{\text{HvK}}$, $\widehat{a}_{\text{oracle}}$ and $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\text{HvK}}$, $\widehat{\sigma}^2_{\text{oracle}}$ in the simulation scenarios for AR(1) with a moderate trend.

Simulations

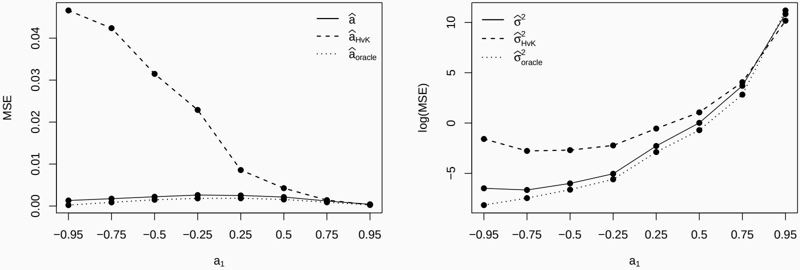

Figure 2: MSE values for the estimators $\widehat{a}$, $\widehat{a}_{\mathrm{HvK}}$, $\widehat{a}_{\mathrm{oracle}}$ and $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\mathrm{HvK}}$, $\widehat{\sigma}^2_{\mathrm{oracle}}$ in the simulation scenarios for AR(1) with a pronounced trend.

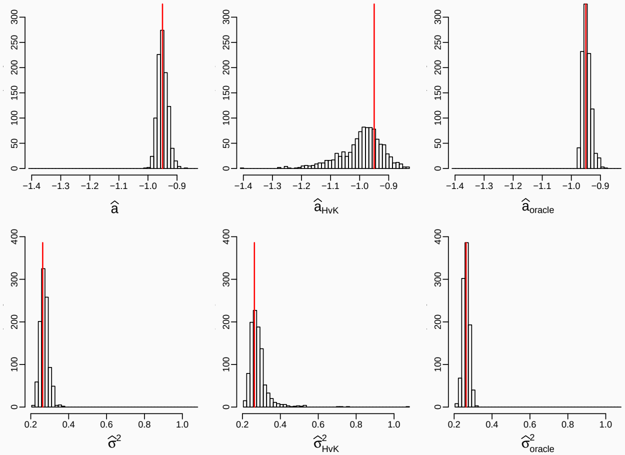

Figure 3: Histograms of the estimators $\widehat{a}$, $\widehat{a}_{\text{HvK}}$, $\widehat{a}_{\text{oracle}}$ and $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\text{HvK}}$, $\widehat{\sigma}^2_{\text{oracle}}$ in the AR(1) model with $a_1 = -0.95$ and moderate trend.

Simulations

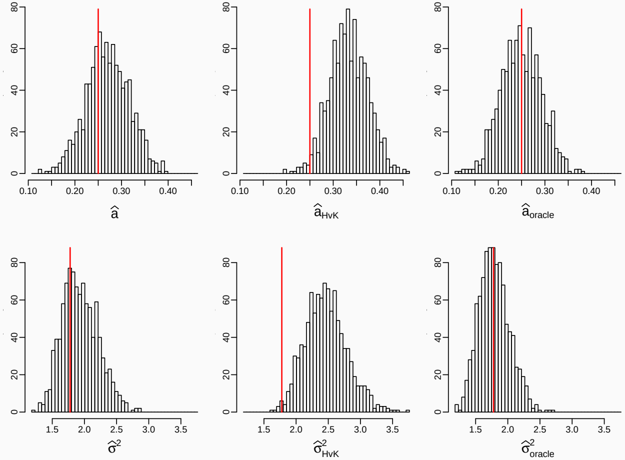

Figure 4: Histograms of the estimators $\widehat{a}$, $\widehat{a}_{\text{HvK}}$, $\widehat{a}_{\text{oracle}}$ and $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\text{HvK}}$, $\widehat{\sigma}^2_{\text{oracle}}$ in the AR(1) model with $a_1 = 0.25$ and pronounced trend.

Conclusion

Conclusion

- We constructed the long-run variance estimator for a wide range of error procesess.
- We proved the $\sqrt{T}$-consistency for our estimators.
- Our estimator produces accurate estimation results even when the AR polynomial has a root close to the unit circle.
- In the simulations our estimators tend to perform well even in the presence of a strong trend.

References

Khismatullina, M., Vogt, M. (2020)
Multiscale inference and long-run variance estimation in nonparametric regression with time series errors.
Journal of the Royal Statistical Society: Series B, 82 5-37.

Hall, P. and Van Keilegom, I. (2003).
Using difference-based methods for inference in nonparametric regression with time series errors.
Journal of the Royal Statistical Society: Series B, 65 443-456.

Thank you!