

Prediction of Non-Negative Variables under Misspecification and Nonstationarity*

Genaro Sucarrat

<https://www.sucarrat.net/>

Department of Economics
BI Norwegian Business School

Conference on Model Invariance and Constancy in the Face of Large
Shocks (Oslo)
17 June 2026

*Joint with Christian Francq (CREST, Université de Lille)

On misspecification:

On misspecification:

- Let y_t denote a variable of interest to predict, say, inflation
- In empirical applications, it is frequently of interest to build a prediction of y_t , say, $f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$
- A leading example is the linear prediction

$$f(\mathbf{x}_{t-1}, \boldsymbol{\theta}) = \boldsymbol{\theta}' \mathbf{x}_{t-1} = \theta_1 x_{1,t-1} + \cdots + \theta_K x_{K,t-1}$$

On misspecification:

- Let y_t denote a variable of interest to predict, say, inflation
- In empirical applications, it is frequently of interest to build a prediction of y_t , say, $f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$
- A leading example is the linear prediction

$$f(\mathbf{x}_{t-1}, \boldsymbol{\theta}) = \boldsymbol{\theta}' \mathbf{x}_{t-1} = \theta_1 x_{1,t-1} + \cdots + \theta_K x_{K,t-1}$$

- Often, it is assumed that $E(y_t | \mathcal{F}_{t-1}) = f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$,

On misspecification:

- Let y_t denote a variable of interest to predict, say, inflation
- In empirical applications, it is frequently of interest to build a prediction of y_t , say, $f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$
- A leading example is the linear prediction

$$f(\mathbf{x}_{t-1}, \boldsymbol{\theta}) = \boldsymbol{\theta}' \mathbf{x}_{t-1} = \theta_1 x_{1,t-1} + \cdots + \theta_K x_{K,t-1}$$

- Often, it is assumed that $E(y_t | \mathcal{F}_{t-1}) = f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$, and that the DGP is

$$y_t = E(y_t | \mathcal{F}_{t-1}) + u_t \quad (*)$$

On misspecification:

- Let y_t denote a variable of interest to predict, say, inflation
- In empirical applications, it is frequently of interest to build a prediction of y_t , say, $f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$
- A leading example is the linear prediction

$$f(\mathbf{x}_{t-1}, \boldsymbol{\theta}) = \boldsymbol{\theta}' \mathbf{x}_{t-1} = \theta_1 x_{1,t-1} + \cdots + \theta_K x_{K,t-1}$$

- Often, it is assumed that $E(y_t | \mathcal{F}_{t-1}) = f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$, and that the DGP is

$$y_t = E(y_t | \mathcal{F}_{t-1}) + u_t \quad (*)$$

- This is clearly unrealistic!

On misspecification:

- Let y_t denote a variable of interest to predict, say, inflation
- In empirical applications, it is frequently of interest to build a prediction of y_t , say, $f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$
- A leading example is the linear prediction

$$f(\mathbf{x}_{t-1}, \boldsymbol{\theta}) = \boldsymbol{\theta}' \mathbf{x}_{t-1} = \theta_1 x_{1,t-1} + \cdots + \theta_K x_{K,t-1}$$

- Often, it is assumed that $E(y_t | \mathcal{F}_{t-1}) = f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$, and that the DGP is

$$y_t = E(y_t | \mathcal{F}_{t-1}) + u_t \quad (*)$$

- This is clearly unrealistic! It will only be true if you are tremendously lucky

On misspecification:

- Let y_t denote a variable of interest to predict, say, inflation
- In empirical applications, it is frequently of interest to build a prediction of y_t , say, $f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$
- A leading example is the linear prediction

$$f(\mathbf{x}_{t-1}, \boldsymbol{\theta}) = \boldsymbol{\theta}' \mathbf{x}_{t-1} = \theta_1 x_{1,t-1} + \cdots + \theta_K x_{K,t-1}$$

- Often, it is assumed that $E(y_t | \mathcal{F}_{t-1}) = f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$, and that the DGP is

$$y_t = E(y_t | \mathcal{F}_{t-1}) + u_t \quad (*)$$

- This is clearly unrealistic! It will only be true if you are tremendously lucky
- So, in practice, $f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$ is a prediction under misspecification: $f(\mathbf{x}_{t-1}, \boldsymbol{\theta}) \neq E(y_t | \mathcal{F}_{t-1})$, (*) not the DGP

On misspecification:

- Let y_t denote a variable of interest to predict, say, inflation
- In empirical applications, it is frequently of interest to build a prediction of y_t , say, $f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$
- A leading example is the linear prediction

$$f(\mathbf{x}_{t-1}, \boldsymbol{\theta}) = \boldsymbol{\theta}' \mathbf{x}_{t-1} = \theta_1 x_{1,t-1} + \cdots + \theta_K x_{K,t-1}$$

- Often, it is assumed that $E(y_t | \mathcal{F}_{t-1}) = f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$, and that the DGP is

$$y_t = E(y_t | \mathcal{F}_{t-1}) + u_t \quad (*)$$

- This is clearly unrealistic! It will only be true if you are tremendously lucky
- So, in practice, $f(\mathbf{x}_{t-1}, \boldsymbol{\theta})$ is a prediction under misspecification: $f(\mathbf{x}_{t-1}, \boldsymbol{\theta}) \neq E(y_t | \mathcal{F}_{t-1})$, (*) not the DGP
- In linear regression, under suitable assumptions, it is well-known that the linear prediction $\boldsymbol{\theta}'_0 \mathbf{x}_{t-1}$ exists, and that

Our paper:

Our paper:

- Let ϕ_t denote the non-negative variable we would like to predict. Note: ϕ_t is observable

Our paper:

- Let ϕ_t denote the non-negative variable we would like to predict. Note: ϕ_t is observable
- Our setup:

Our paper:

- Let ϕ_t denote the non-negative variable we would like to predict. Note: ϕ_t is observable
- Our setup:
 - ▷ $(\phi_t)_{t \in \mathbb{N}}$ can be stationary or nonstationary

Our paper:

- Let ϕ_t denote the non-negative variable we would like to predict. Note: ϕ_t is observable
- Our setup:
 - ▷ $(\phi_t)_{t \in \mathbb{N}}$ can be stationary or nonstationary
 - ▷ The specification of $E(\phi_t | \mathcal{F}_{t-1})$ can be unknown

Our paper:

- Let ϕ_t denote the non-negative variable we would like to predict. Note: ϕ_t is observable
- Our setup:
 - ▷ $(\phi_t)_{t \in \mathbb{N}}$ can be stationary or nonstationary
 - ▷ The specification of $E(\phi_t | \mathcal{F}_{t-1})$ can be unknown
 - ▷ The DGP of ϕ_t can be unknown

Our paper:

- Let ϕ_t denote the non-negative variable we would like to predict. Note: ϕ_t is observable
- Our setup:
 - ▷ $(\phi_t)_{t \in \mathbb{N}}$ can be stationary or nonstationary
 - ▷ The specification of $E(\phi_t | \mathcal{F}_{t-1})$ can be unknown
 - ▷ The DGP of ϕ_t can be unknown
- The general form of our predictor of ϕ_t :

$$\mu_t(\boldsymbol{\theta}; \phi_{t-1}, \phi_{t-2}, \dots, \phi_1), \quad \boldsymbol{\theta} \in \Theta,$$

Our paper:

- Let ϕ_t denote the non-negative variable we would like to predict. Note: ϕ_t is observable
- Our setup:
 - ▷ $(\phi_t)_{t \in \mathbb{N}}$ can be stationary or nonstationary
 - ▷ The specification of $E(\phi_t | \mathcal{F}_{t-1})$ can be unknown
 - ▷ The DGP of ϕ_t can be unknown
- The general form of our predictor of ϕ_t :

$$\mu_t(\boldsymbol{\theta}; \phi_{t-1}, \phi_{t-2}, \dots, \phi_1), \quad \boldsymbol{\theta} \in \Theta,$$

- ▷ $\mu_t : \Theta \times [0, \infty)^{t-1} \rightarrow [\underline{\mu}, \infty)$ is a measurable function for some constant $\underline{\mu} > 0$
- ▷ μ_t is continuous in $\boldsymbol{\theta}$ for all $\phi_{t-1}, \dots, \phi_1$
- ▷ Θ is a compact subset of $\mathbb{R}^d$

Our paper:

- Let ϕ_t denote the non-negative variable we would like to predict. Note: ϕ_t is observable
- Our setup:
 - ▷ $(\phi_t)_{t \in \mathbb{N}}$ can be stationary or nonstationary
 - ▷ The specification of $E(\phi_t | \mathcal{F}_{t-1})$ can be unknown
 - ▷ The DGP of ϕ_t can be unknown
- The general form of our predictor of ϕ_t :

$$\mu_t(\boldsymbol{\theta}; \phi_{t-1}, \phi_{t-2}, \dots, \phi_1), \quad \boldsymbol{\theta} \in \Theta,$$

- ▷ $\mu_t : \Theta \times [0, \infty)^{t-1} \rightarrow [\underline{\mu}, \infty)$ is a measurable function for some constant $\underline{\mu} > 0$
- ▷ μ_t is continuous in $\boldsymbol{\theta}$ for all $\phi_{t-1}, \dots, \phi_1$
- ▷ Θ is a compact subset of $\mathbb{R}^d$
- Example of μ_t , the GARCH(1,1): $\mu_t = c + a\phi_{t-1} + b\mu_{t-1}$

Our paper (cont.):

Our paper (cont.):

Outline:

Our paper (cont.):

Outline:

- Key assumption of our framework: Uniqueness.

Our paper (cont.):

Outline:

- Key assumption of our framework: Uniqueness. That is, there exists a unique “pseudo-true” parameter θ_0 that minimises the limit of the expected loss (QLIKE)

Our paper (cont.):

Outline:

- Key assumption of our framework: Uniqueness. That is, there exists a unique “pseudo-true” parameter θ_0 that minimises the limit of the expected loss (QLIKE)
- Consistency: Under assumptions, $\hat{\theta} \rightarrow_p \theta_0$ where $\hat{\theta}$ is the QMLE

Our paper (cont.):

Outline:

- Key assumption of our framework: Uniqueness. That is, there exists a unique “pseudo-true” parameter θ_0 that minimises the limit of the expected loss (QLIKE)
- Consistency: Under assumptions, $\hat{\theta} \rightarrow_p \theta_0$ where $\hat{\theta}$ is the QMLE
- Asymptotic Normality (AN): Under assumptions, $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(\mathbf{0}, \Sigma)$
- Consistent estimation: Under assumptions, $\hat{\Sigma} \rightarrow_p \Sigma$

Our paper (cont.):

Outline:

- Key assumption of our framework: Uniqueness. That is, there exists a unique “pseudo-true” parameter θ_0 that minimises the limit of the expected loss (QLIKE)
- Consistency: Under assumptions, $\hat{\theta} \rightarrow_p \theta_0$ where $\hat{\theta}$ is the QMLE
- Asymptotic Normality (AN): Under assumptions, $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(\mathbf{0}, \Sigma)$
- Consistent estimation: Under assumptions, $\hat{\Sigma} \rightarrow_p \Sigma$
- Monte Carlo simulations

Our paper (cont.):

Outline:

- Key assumption of our framework: Uniqueness. That is, there exists a unique “pseudo-true” parameter θ_0 that minimises the limit of the expected loss (QLIKE)
- Consistency: Under assumptions, $\hat{\theta} \rightarrow_p \theta_0$ where $\hat{\theta}$ is the QMLE
- Asymptotic Normality (AN): Under assumptions, $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(\mathbf{0}, \Sigma)$
- Consistent estimation: Under assumptions, $\hat{\Sigma} \rightarrow_p \Sigma$
- Monte Carlo simulations
- Empirical illustration

Related literature:

Related literature:

- “Philosophical”:
 - ▷ [Box and Draper \(1987\)](#), p. 424): “all models are wrong, but some are useful”

Related literature:

- “Philosophical”:
 - ▷ [Box and Draper \(1987\)](#), p. 424): “all models are wrong, but some are useful”
- Empirical:
 - ▷ It is well-known that the GARCH(1,1) model is difficult to beat in volatility prediction in a variety of contexts
 - ▷ Example: [Hansen and Lunde \(2005\)](#): “Does anything beat a GARCH(1,1)? A forecast comparison of volatility models”, *Journal of Applied Econometrics* 20, pp. 873–889

Related literature:

- “Philosophical”:
 - ▷ [Box and Draper \(1987\)](#), p. 424: “all models are wrong, but some are useful”
- Empirical:
 - ▷ It is well-known that the GARCH(1,1) model is difficult to beat in volatility prediction in a variety of contexts
 - ▷ Example: [Hansen and Lunde \(2005\)](#): “Does anything beat a GARCH(1,1)? A forecast comparison of volatility models”, *Journal of Applied Econometrics* 20, pp. 873–889
- Theoretical:
 - ▷ Li and Sucarrat (2025): “Volatility prediction under misspecification” (working paper), a heterogeneous log-ARCH-X prediction under misspecification and stationarity

Theory

Assumptions:

Assumptions:

A1: Assume the existence, for all $\theta \in \Theta$, of
 $L(\theta) = \lim_{n \rightarrow \infty} L_n(\theta)$,

$$L_n(\theta) = \frac{1}{n} \sum_{t=2}^n E \ell_t(\theta) \quad \text{and} \quad \ell_t(\theta) = \frac{\phi_t}{\mu_t(\theta)} + \ln \mu_t(\theta),$$

and assume also the existence of a unique $\theta_0 \in \Theta$ such that

$$\theta_0 = \arg \min_{\theta \in \Theta} L(\theta)$$

Assumptions:

A1: Assume the existence, for all $\theta \in \Theta$, of
 $L(\theta) = \lim_{n \rightarrow \infty} L_n(\theta)$,

$$L_n(\theta) = \frac{1}{n} \sum_{t=2}^n E \ell_t(\theta) \quad \text{and} \quad \ell_t(\theta) = \frac{\phi_t}{\mu_t(\theta)} + \ln \mu_t(\theta),$$

and assume also the existence of a unique $\theta_0 \in \Theta$ such that

$$\theta_0 = \arg \min_{\theta \in \Theta} L(\theta)$$

Note: θ_0 is the “pseudo-true” parameter

Assumptions:

A1: Assume the existence, for all $\theta \in \Theta$, of
 $L(\theta) = \lim_{n \rightarrow \infty} L_n(\theta)$,

$$L_n(\theta) = \frac{1}{n} \sum_{t=2}^n E \ell_t(\theta) \quad \text{and} \quad \ell_t(\theta) = \frac{\phi_t}{\mu_t(\theta)} + \ln \mu_t(\theta),$$

and assume also the existence of a unique $\theta_0 \in \Theta$ such that

$$\theta_0 = \arg \min_{\theta \in \Theta} L(\theta)$$

Note: θ_0 is the “pseudo-true” parameter

A2: The sequence $\{\phi_t, \mu_t(\theta)\}_{t \geq 1}$ is α -mixing uniformly in $\theta \in \Theta$

Assumptions:

A1: Assume the existence, for all $\theta \in \Theta$, of
 $L(\theta) = \lim_{n \rightarrow \infty} L_n(\theta)$,

$$L_n(\theta) = \frac{1}{n} \sum_{t=2}^n E \ell_t(\theta) \quad \text{and} \quad \ell_t(\theta) = \frac{\phi_t}{\mu_t(\theta)} + \ln \mu_t(\theta),$$

and assume also the existence of a unique $\theta_0 \in \Theta$ such that

$$\theta_0 = \arg \min_{\theta \in \Theta} L(\theta)$$

Note: θ_0 is the “pseudo-true” parameter

A2: The sequence $\{\phi_t, \mu_t(\theta)\}_{t \geq 1}$ is α -mixing uniformly in $\theta \in \Theta$

A3: There exists a $\nu > 0$ such that $\sup_t E(\phi_t^{1+\nu}) < \infty$

Assumptions (cont.):

Assumptions (cont.):

A4: For all $\boldsymbol{\theta}, \boldsymbol{\theta}_1 \in \Theta$, there exist $K_1, K_2 > 0$ and $\rho, \rho_1 \in [0, 1)$ such that

$$\begin{aligned} & |\mu_t(\boldsymbol{\theta}) - \mu_t(\boldsymbol{\theta}_1)| \\ & \leq K_1 \|\boldsymbol{\theta} - \boldsymbol{\theta}_1\|_1 \sum_{i=0}^{t-2} \rho_1^i \{1 + \phi_{t-i-1} + \mu_{t-i-1}(\boldsymbol{\theta})\}, \\ & \mu_t(\boldsymbol{\theta}) \leq K_1 \sum_{i=0}^{t-2} \rho^i (1 + \phi_{t-i-1}), \\ & \mu_t(\boldsymbol{\theta}_1) \geq K_2 \left[1 + \rho_1^i \{\phi_{t-i-1} + \mu_{t-i-1}(\boldsymbol{\theta}_1)\}\right], \quad 0 \leq i \leq t-2. \end{aligned}$$

Assumptions (cont.):

A4: For all $\boldsymbol{\theta}, \boldsymbol{\theta}_1 \in \Theta$, there exist $K_1, K_2 > 0$ and $\rho, \rho_1 \in [0, 1)$ such that

$$\begin{aligned}
 & |\mu_t(\boldsymbol{\theta}) - \mu_t(\boldsymbol{\theta}_1)| \\
 & \leq K_1 \|\boldsymbol{\theta} - \boldsymbol{\theta}_1\|_1 \sum_{i=0}^{t-2} \rho_1^i \{1 + \phi_{t-i-1} + \mu_{t-i-1}(\boldsymbol{\theta})\}, \\
 & \mu_t(\boldsymbol{\theta}) \leq K_1 \sum_{i=0}^{t-2} \rho^i (1 + \phi_{t-i-1}), \\
 & \mu_t(\boldsymbol{\theta}_1) \geq K_2 [1 + \rho_1^i \{\phi_{t-i-1} + \mu_{t-i-1}(\boldsymbol{\theta}_1)\}], \quad 0 \leq i \leq t-2.
 \end{aligned}$$

Theorem 1 (consistency). Under **A1–A4** and the assumptions on Θ , $\hat{\boldsymbol{\theta}}$ converges in probability to $\boldsymbol{\theta}_0$.

Assumptions (cont.):

Assumptions (cont.):

A5: Assume that, for almost all sequences (ϕ_t) , all the functions $\theta \mapsto \mu_t(\theta)$, for $t \geq 2$, are twice continuously differentiable in a neighbourhood of θ_0

Assumptions (cont.):

A5: Assume that, for almost all sequences (ϕ_t) , all the functions $\theta \mapsto \mu_t(\theta)$, for $t \geq 2$, are twice continuously differentiable in a neighbourhood of θ_0

Let $\phi_t = (\phi_t, \mu_t(\theta_0), \partial \mu_t(\theta_0) / \partial \theta')'$, $\mathcal{V}_{1:t} = \sigma\{\phi_u : 1 \leq u \leq t\}$ and $\mathcal{V}_{t:\infty} = \sigma\{\phi_u : u \geq t\}$ and define

$$\alpha_m^* = \sup_{t \geq 1} \sup_{A \in \mathcal{V}_{1:t}, B \in \mathcal{V}_{t+m:\infty}} |P(A \cap B) - P(A)P(B)|$$

Assumptions (cont.):

A5: Assume that, for almost all sequences (ϕ_t) , all the functions $\theta \mapsto \mu_t(\theta)$, for $t \geq 2$, are twice continuously differentiable in a neighbourhood of θ_0

Let $\phi_t = (\phi_t, \mu_t(\theta_0), \partial \mu_t(\theta_0) / \partial \theta')'$, $\mathcal{V}_{1:t} = \sigma\{\phi_u : 1 \leq u \leq t\}$ and $\mathcal{V}_{t:\infty} = \sigma\{\phi_u : u \geq t\}$ and define

$$\alpha_m^* = \sup_{t \geq 1} \sup_{A \in \mathcal{V}_{1:t}, B \in \mathcal{V}_{t+m:\infty}} |P(A \cap B) - P(A)P(B)|$$

A6: There exists $\nu > 0$ and $\tau > 0$ such that $\sup_{t \geq 1} E \phi_t^{2+\nu+\tau} < \infty$ and $\sum_{m=0}^{\infty} (\alpha_m^*)^{\nu/(2+\nu)} < \infty$, where (α_m^*) is the sequence of α -mixing coefficients

Assumptions (cont.):

Assumptions (cont.):

A7: For all $\varepsilon > 0$, there exists a neighbourhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$, and constants $K_1, K_2 > 0$ and $\rho_1, \rho_2 \in (0, 1)$ such that $\rho_1/\rho_2 < (1 + \varepsilon)$ and for all $\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)$

$$\left\| \frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| + \left\| \frac{\partial^2 \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| \leq K_1 \sum_{i=0}^{t-2} \rho_1^i \{1 + \phi_{t-i-1}\},$$
$$\mu_t(\boldsymbol{\theta}) \geq K_2 \rho_2^i \{1 + \phi_{t-i-1}\}, \quad 0 \leq i \leq t-2.$$

Assumptions (cont.):

A7: For all $\varepsilon > 0$, there exists a neighbourhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$, and constants $K_1, K_2 > 0$ and $\rho_1, \rho_2 \in (0, 1)$ such that $\rho_1/\rho_2 < (1 + \varepsilon)$ and for all $\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)$

$$\left\| \frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| + \left\| \frac{\partial^2 \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| \leq K_1 \sum_{i=0}^{t-2} \rho_1^i \{1 + \phi_{t-i-1}\},$$

$$\mu_t(\boldsymbol{\theta}) \geq K_2 \rho_2^i \{1 + \phi_{t-i-1}\}, \quad 0 \leq i \leq t-2.$$

Lemma 3. The assumptions so far entail the existence of the matrices

$$\mathbf{I}_n(\boldsymbol{\theta}) = \text{Var} \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\}, \quad \mathbf{J}_n(\boldsymbol{\theta}) = E \left\{ \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\}$$

for all $\boldsymbol{\theta}$ belonging to some neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$. Moreover, $\mathbf{J}_n(\cdot)$ is continuous at $\boldsymbol{\theta}_0$ uniformly in n

Assumptions (cont.):

Assumptions (cont.):

A8: There exist matrices $\mathbf{I}$ and $\mathbf{J}$ such that $\mathbf{I}_n(\boldsymbol{\theta}_n) \rightarrow \mathbf{I}$ and $\mathbf{J}_n(\boldsymbol{\theta}_n) \rightarrow \mathbf{J}$ in probability when $\boldsymbol{\theta}_n \rightarrow \boldsymbol{\theta}_0$ in probability

Assumptions (cont.):

A8: There exist matrices $\mathbf{I}$ and $\mathbf{J}$ such that $\mathbf{I}_n(\boldsymbol{\theta}_n) \rightarrow \mathbf{I}$ and $\mathbf{J}_n(\boldsymbol{\theta}_n) \rightarrow \mathbf{J}$ in probability when $\boldsymbol{\theta}_n \rightarrow \boldsymbol{\theta}_0$ in probability

A2*: The sequence $\{\phi_t, \mu_t(\boldsymbol{\theta}), \partial \mu_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}, \partial^2 \mu_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'\}_{t \geq 1}$ is α -mixing uniformly in $\boldsymbol{\theta} \in \Theta$

Assumptions (cont.):

A8: There exist matrices $\mathbf{I}$ and $\mathbf{J}$ such that $\mathbf{I}_n(\boldsymbol{\theta}_n) \rightarrow \mathbf{I}$ and $\mathbf{J}_n(\boldsymbol{\theta}_n) \rightarrow \mathbf{J}$ in probability when $\boldsymbol{\theta}_n \rightarrow \boldsymbol{\theta}_0$ in probability

A2*: The sequence $\{\phi_t, \mu_t(\boldsymbol{\theta}), \partial \mu_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}, \partial^2 \mu_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'\}_{t \geq 1}$ is α -mixing uniformly in $\boldsymbol{\theta} \in \Theta$

Lemma 3. Under the assumptions so far, for all sequences $\boldsymbol{\theta}_n$ which converges in probability to $\boldsymbol{\theta}_0$, we have

$$\hat{\mathbf{J}}_n := \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \ell_t(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \rightarrow \mathbf{J} \text{ in probability}$$

[Note: This means $\mathbf{J}$ can be estimated consistently]

Assumptions (cont.):

A8: There exist matrices $\mathbf{I}$ and $\mathbf{J}$ such that $\mathbf{I}_n(\boldsymbol{\theta}_n) \rightarrow \mathbf{I}$ and $\mathbf{J}_n(\boldsymbol{\theta}_n) \rightarrow \mathbf{J}$ in probability when $\boldsymbol{\theta}_n \rightarrow \boldsymbol{\theta}_0$ in probability

A2*: The sequence $\{\phi_t, \mu_t(\boldsymbol{\theta}), \partial \mu_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}, \partial^2 \mu_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'\}_{t \geq 1}$ is α -mixing uniformly in $\boldsymbol{\theta} \in \Theta$

Lemma 3. Under the assumptions so far, for all sequences $\boldsymbol{\theta}_n$ which converges in probability to $\boldsymbol{\theta}_0$, we have

$$\hat{\mathbf{J}}_n := \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \ell_t(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \rightarrow \mathbf{J} \text{ in probability}$$

[Note: This means $\mathbf{J}$ can be estimated consistently]

A9: There exists a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that L and L_n are differentiable on $V(\boldsymbol{\theta}_0)$ and

$$\lim_{n \rightarrow \infty} \sqrt{n} \left\{ \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\} = 0, \text{ for all } \boldsymbol{\theta} \in V(\boldsymbol{\theta}_0).$$

[Note: This assumption holds under stationarity]

Assumptions (cont.):

Assumptions (cont.):

A10: The pseudo-true parameter θ_0 belongs to the interior of Θ

Assumptions (cont.):

A10: The pseudo-true parameter θ_0 belongs to the interior of Θ

Theorem 2 (asymptotic normality). Under the assumptions of Theorem 1 and **A6–10**, $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(\mathbf{0}, \mathbf{J}^{-1} \mathbf{I} \mathbf{J}^{-1})$

Assumptions (cont.):

A10: The pseudo-true parameter θ_0 belongs to the interior of Θ

Theorem 2 (asymptotic normality). Under the assumptions of Theorem 1 and **A6–10**, $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(\mathbf{0}, \mathbf{J}^{-1} \mathbf{I} \mathbf{J}^{-1})$

The matrix $\mathbf{I}$ can be estimated with the HAC-type estimator proposed by [De Jong and Davidson \(2000\)](#):

$$\hat{\mathbf{I}}_n = \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n \dot{\mathbf{l}}_t(\hat{\theta}) \dot{\mathbf{l}}_s(\hat{\theta})' k((t-s)/\gamma_n),$$

where the $k(\cdot)$'s are kernel weights, and γ_n is the bandwidth

Assumptions (cont.):

A10: The pseudo-true parameter θ_0 belongs to the interior of Θ

Theorem 2 (asymptotic normality). Under the assumptions of Theorem 1 and **A6–10**, $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(0, J^{-1}IJ^{-1})$

The matrix I can be estimated with the HAC-type estimator proposed by [De Jong and Davidson \(2000\)](#):

$$\hat{I}_n = \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n \dot{l}_t(\hat{\theta}) \dot{l}_s(\hat{\theta})' k((t-s)/\gamma_n),$$

where the $k(\cdot)$'s are kernel weights, and γ_n is the bandwidth

A11 (kernel) $k(\cdot)$ belongs to the kernel class detailed in [De Jong and Davidson \(2000\)](#)

Assumptions (cont.):

A10: The pseudo-true parameter θ_0 belongs to the interior of Θ

Theorem 2 (asymptotic normality). Under the assumptions of Theorem 1 and **A6–10**, $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(0, J^{-1} I J^{-1})$

The matrix I can be estimated with the HAC-type estimator proposed by [De Jong and Davidson \(2000\)](#):

$$\hat{I}_n = \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n \dot{l}_t(\hat{\theta}) \dot{l}_s(\hat{\theta})' k((t-s)/\gamma_n),$$

where the $k(\cdot)$'s are kernel weights, and γ_n is the bandwidth

A11 (kernel) $k(\cdot)$ belongs to the kernel class detailed in [De Jong and Davidson \(2000\)](#)

A12 (bandwidth) $\gamma_n \rightarrow \infty$ and $\gamma_n = o(n^{1/2})$

Assumptions (cont.):

Assumptions (cont.):

Theorem 3 (consistent estimation of I). Under the assumptions of Theorem 2 and **A11–A12**, $\hat{I}_n \rightarrow_p I$

Examples

Example: The ARMA(p, q) predictor:

Example: The ARMA(p, q) predictor:

$$\mu_t(\boldsymbol{\theta}) = c + \sum_{i=1}^p a_i \phi_{t-i} + \sum_{j=1}^q b_j \mu_{t-j}, \quad t \geq 2$$

Example: The ARMA(p, q) predictor:

$$\mu_t(\boldsymbol{\theta}) = c + \sum_{i=1}^p a_i \phi_{t-i} + \sum_{j=1}^q b_j \mu_{t-j}, \quad t \geq 2$$

Note: The ARMA(p, q) predictor = GARCH(p, q)

Example: The ARMA(p, q) predictor:

$$\mu_t(\boldsymbol{\theta}) = c + \sum_{i=1}^p a_i \phi_{t-i} + \sum_{j=1}^q b_j \mu_{t-j}, \quad t \geq 2$$

Note: The ARMA(p, q) predictor = GARCH(p, q)

If:

- Θ is a compact subset of $(0, \infty)^{p+q+1}$
- $\forall \boldsymbol{\theta} \in \Theta: \sum_{j=1}^q b_j < 1$

Example: The ARMA(p, q) predictor:

$$\mu_t(\boldsymbol{\theta}) = c + \sum_{i=1}^p a_i \phi_{t-i} + \sum_{j=1}^q b_j \mu_{t-j}, \quad t \geq 2$$

Note: The ARMA(p, q) predictor = GARCH(p, q)

If:

- Θ is a compact subset of $(0, \infty)^{p+q+1}$
- $\forall \boldsymbol{\theta} \in \Theta: \sum_{j=1}^q b_j < 1$

Then:

- **A4**, **A5** and **A7** hold

Example: The AR(1) predictor

Example: The AR(1) predictor

- Recall **A1**: $\theta_0 \in \Theta$ uniquely minimises $L(\theta)$

Example: The AR(1) predictor

- Recall **A1**: $\theta_0 \in \Theta$ uniquely minimises $L(\theta)$
- Assume the DGP is

$$\phi_t = h_t \eta_t, \quad h_t = 0.1 + 0.1\phi_{t-1} + 0.8h_{t-1}, \quad \eta_t \stackrel{iid}{\sim} \chi^2(1)$$

Example: The AR(1) predictor

- Recall **A1**: $\theta_0 \in \Theta$ uniquely minimises $L(\theta)$
- Assume the DGP is

$$\phi_t = h_t \eta_t, \quad h_t = 0.1 + 0.1\phi_{t-1} + 0.8h_{t-1}, \quad \eta_t \stackrel{iid}{\sim} \chi^2(1)$$

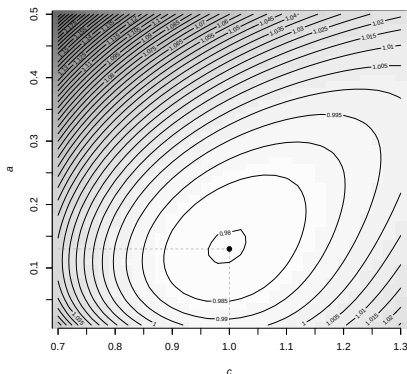
- Consider the AR(1) predictor: $\mu_t = c + a\phi_{t-1}$, $\theta = (c, a)'$

Example: The AR(1) predictor

- Recall **A1**: $\theta_0 \in \Theta$ uniquely minimises $L(\theta)$
- Assume the DGP is

$$\phi_t = h_t \eta_t, \quad h_t = 0.1 + 0.1\phi_{t-1} + 0.8h_{t-1}, \quad \eta_t \stackrel{iid}{\sim} \chi^2(1)$$

- Consider the AR(1) predictor: $\mu_t = c + a\phi_{t-1}$, $\theta = (c, a)'$
- Simulations suggest $\theta_0 = (1.0, 0.13)'$:



Example: The seasonal predictor

Example: The seasonal predictor

- Recall **A1**: $\theta_0 \in \Theta$ uniquely minimises $L(\theta)$

Example: The seasonal predictor

- Recall **A1**: $\theta_0 \in \Theta$ uniquely minimises $L(\theta)$
- Let $(\phi_t)_{t \in \mathbb{N}}$ be a seasonal process of periodicity s ,
 $k \in \{1, \dots, s\}$

Example: The seasonal predictor

- Recall **A1**: $\theta_0 \in \Theta$ uniquely minimises $L(\theta)$
- Let $(\phi_t)_{t \in \mathbb{N}}$ be a seasonal process of periodicity s ,
 $k \in \{1, \dots, s\}$
- Note: $k(t) = 1 + ((t - 1) \bmod s)$, $k(t) \in \{1, \dots, s\}$

Example: The seasonal predictor

- Recall **A1**: $\theta_0 \in \Theta$ uniquely minimises $L(\theta)$
- Let $(\phi_t)_{t \in \mathbb{N}}$ be a seasonal process of periodicity s ,
 $k \in \{1, \dots, s\}$
- Note: $k(t) = 1 + ((t - 1) \bmod s)$, $k(t) \in \{1, \dots, s\}$
- Assume $(\phi_t)_{t \in \mathbb{N}}$ is nonstationary periodic in the sense that

$$E(\phi_t) \in \{m_1, \dots, m_s\} \quad \text{for all } t$$

and $m_{k_1} \neq m_{k_2}$ for some pair k_1, k_2

Example: The seasonal predictor

- Recall **A1**: $\theta_0 \in \Theta$ uniquely minimises $L(\theta)$
- Let $(\phi_t)_{t \in \mathbb{N}}$ be a seasonal process of periodicity s ,
 $k \in \{1, \dots, s\}$
- Note: $k(t) = 1 + ((t - 1) \bmod s)$, $k(t) \in \{1, \dots, s\}$
- Assume $(\phi_t)_{t \in \mathbb{N}}$ is nonstationary periodic in the sense that

$$E(\phi_t) \in \{m_1, \dots, m_s\} \quad \text{for all } t$$

and $m_{k_1} \neq m_{k_2}$ for some pair k_1, k_2

- Let $\mu_t(\theta)$ with $\theta = (c_1, \dots, c_s)'$ be the seasonal predictor

$$\mu_t(\theta) = c_1 I(k(t) = 1) + \dots + c_s I(k(t) = s),$$

where the $I(A)$'s are indicator functions

Example: The seasonal predictor

- Recall **A1**: $\theta_0 \in \Theta$ uniquely minimises $L(\theta)$
- Let $(\phi_t)_{t \in \mathbb{N}}$ be a seasonal process of periodicity s , $k \in \{1, \dots, s\}$
- Note: $k(t) = 1 + ((t - 1) \bmod s)$, $k(t) \in \{1, \dots, s\}$
- Assume $(\phi_t)_{t \in \mathbb{N}}$ is nonstationary periodic in the sense that

$$E(\phi_t) \in \{m_1, \dots, m_s\} \quad \text{for all } t$$

and $m_{k_1} \neq m_{k_2}$ for some pair k_1, k_2

- Let $\mu_t(\theta)$ with $\theta = (c_1, \dots, c_s)'$ be the seasonal predictor

$$\mu_t(\theta) = c_1 I(k(t) = 1) + \dots + c_s I(k(t) = s),$$

where the $I(A)$'s are indicator functions

- There exists a θ_0 that uniquely minimises $L(\theta)$

Example: The seasonal predictor

Example: The seasonal predictor

- The (purely) seasonal predictor can straightforwardly be combined with ARMA specifications

Example: The seasonal predictor

- The (purely) seasonal predictor can straightforwardly be combined with ARMA specifications
- Example (simple):

$$\mu_t(\boldsymbol{\theta}) = \underbrace{c_1 I(k(t) = 1) + \cdots + c_s I(k(t) = s)}_{\text{seasonal part}} + \underbrace{a\phi_{t-1} + b\mu_{t-1}}_{\text{ARMA}(1,1)},$$

Example: The seasonal predictor

- The (purely) seasonal predictor can straightforwardly be combined with ARMA specifications
- Example (simple):

$$\mu_t(\boldsymbol{\theta}) = \underbrace{c_1 I(k(t) = 1) + \cdots + c_s I(k(t) = s)}_{\text{seasonal part}} + \underbrace{a\phi_{t-1} + b\mu_{t-1}}_{\text{ARMA}(1,1)},$$

- Example (more elaborate):

$$\begin{aligned} \mu_t(\boldsymbol{\theta}) = & \underbrace{c_1 I(k(t) = 1) + \cdots + c_s I(k(t) = s)}_{\text{seasonal part}} \\ & + \underbrace{\sum_{i=1}^p a_i \phi_{t-i} + \sum_{j=1}^q b_j \mu_{t-j}}_{\text{ARMA}(p,q)}, \end{aligned}$$

Simulations

Simulations:

- Recall: The asymptotic coefficient covariance in Theorem 2 (Asymptotic Normality) is of the form $\Sigma \equiv J^{-1} I J^{-1}$
- A main objective of our simulations is to study the ordinary estimator of Σ , which we denote $\hat{\Sigma}_n^{\text{Ord}}$, and the HAC-robust estimator, which we denote $\hat{\Sigma}_n^{\text{HAC}}$

Simulations:

- Recall: The asymptotic coefficient covariance in Theorem 2 (Asymptotic Normality) is of the form $\Sigma \equiv J^{-1} I J^{-1}$
- A main objective of our simulations is to study the ordinary estimator of Σ , which we denote $\hat{\Sigma}_n^{\text{Ord}}$, and the HAC-robust estimator, which we denote $\hat{\Sigma}_n^{\text{HAC}}$
- We study two versions of the HAC-estimator, HAC1 (“short”) and HAC2 (“long”): In HAC2 the bandwidth lag-length is three times higher (when $n > 100$) than in HAC1

Simulations:

- Recall: The asymptotic coefficient covariance in Theorem 2 (Asymptotic Normality) is of the form $\Sigma \equiv J^{-1} I J^{-1}$
- A main objective of our simulations is to study the ordinary estimator of Σ , which we denote $\hat{\Sigma}_n^{\text{Ord}}$, and the HAC-robust estimator, which we denote $\hat{\Sigma}_n^{\text{HAC}}$
- We study two versions of the HAC-estimator, HAC1 (“short”) and HAC2 (“long”): In HAC2 the bandwidth lag-length is three times higher (when $n > 100$) than in HAC1
- Main take-away: HAC2 (“long”) performs best across experiments, since its estimates are close to the actual standard errors under both correct and incorrect specification

Empirical Illustration

Empirical application:

Empirical application:

- Prediction of intraday USD/EUR exchange rate volatility

Empirical application:

- Prediction of intraday USD/EUR exchange rate volatility
- To keep the illustration simple, we distinguish between two “seasons” only: Main trading session from 8h CET to 17h CET, and overnight session from 17h CET to 8h CET the day after (3 January 2017 – 31 December 2020)

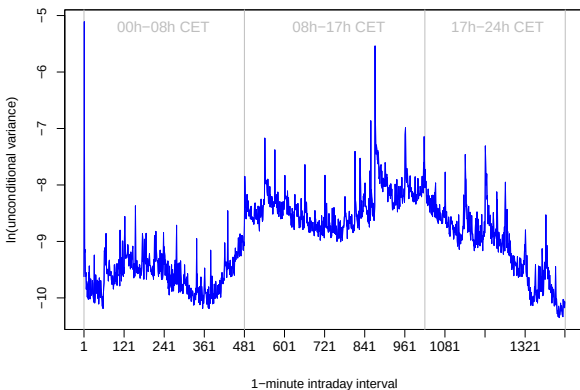


Illustration: Empirical application (cont.)

Illustration: Empirical application (cont.)

- Squared percentage return: $\phi_t \equiv [(\ln P_t - \ln P_{t-1}) \cdot 100]^2$

Illustration: Empirical application (cont.)

- Squared percentage return: $\phi_t \equiv [(\ln P_t - \ln P_{t-1}) \cdot 100]^2$
- The seasonal models we fit:

Seasonal constant: $\mu_t = c_1 + c_2 I(k(t) = 2)$

Seasonal AR(1): $\mu_t = c_1 + c_2 I(k(t) = 2) + a\phi_{t-1}$

Seasonal ARMA(1,1): $\mu_t = c_1 + c_2 I(k(t) = 2) + a\phi_{t-1} + b\mu_{t-1}$

(Note: $k(t) = 2$ is overnight return)

Illustration: Empirical application (cont.)

- Squared percentage return: $\phi_t \equiv [(\ln P_t - \ln P_{t-1}) \cdot 100]^2$
- The seasonal models we fit:

Seasonal constant: $\mu_t = c_1 + c_2 I(k(t) = 2)$

Seasonal AR(1): $\mu_t = c_1 + c_2 I(k(t) = 2) + a\phi_{t-1}$

Seasonal ARMA(1,1): $\mu_t = c_1 + c_2 I(k(t) = 2) + a\phi_{t-1} + b\mu_{t-1}$

(Note: $k(t) = 2$ is overnight return)

- The ordinary (“non-seasonal”) models we fit:

Ordinary constant: $\mu_t = c_1$

Ordinary AR(1): $\mu_t = c_1 + a\phi_{t-1}$

Ordinary ARMA(1,1): $\mu_t = c_1 + a\phi_{t-1} + b\mu_{t-1}$

Illustration: Empirical application (cont.)

Results:

	$\hat{c}_1$ (s.e.)	$\hat{c}_2$ (s.e.)	$\hat{a}$ (s.e.)	$\hat{b}$ (s.e.)	QLIKE	PD	$m(\hat{\eta}_t)$
<u>Seasonal models:</u>							
Constant:	0.0662 (0.0054)	0.0541 (0.0106)			-1.4167	Yes	1.000
AR(1):	0.0539 (0.0049)	0.0548 (0.0099)	0.1127 (0.0364)		-1.4504	Yes	1.000
ARMA(1,1):	0.0001 (0.0002)	0.0019 (0.0009)	0.0433 (0.0085)	0.9448 (0.0064)	-1.4939	No	1.003
<u>Ordinary models:</u>							
Constant:	0.0932 (0.0070)				-1.3727	Yes	1.000
AR(1):	0.0817 (0.0051)		0.1166 (0.0445)		-1.4001	Yes	1.000
ARMA(1,1):	0.0008 (0.0005)		0.0404 (0.0368)	0.9511 (0.0231)	-1.4924	Yes	1.003

s.e., standard error of estimate computed as the square root of the diagonal of $(\hat{\mathbf{J}}_n^{-1} \hat{\mathbf{I}}_n \hat{\mathbf{J}}_n^{-1})/n$. QLIKE, the value of the objective function Q_n in **A3** at the parameter estimate $\hat{\boldsymbol{\theta}}$. PD, “Yes” if the Hessian of the objective function is positive definite at $\hat{\boldsymbol{\theta}}$ (i.e. all eigenvalues are positive), “No” otherwise. $m(\hat{\eta}_t)$, the average of the scaled residuals, i.e. $m(\hat{\eta}_t) = n^{-1} \sum_{t=1}^n \hat{\eta}_t$ where $\hat{\eta}_t = \phi_t / \hat{\mu}_t$

Conclusions

Conclusions

- In empirical practice, models are misspecified: They differ from the conditional expectation and/or the DGP
- We consider a broad class of predictive specifications for non-negative variables (e.g. volatility, spreads, volume, unemployment, duration)
- The predictive specifications need not equal the (unknown) conditional expectation or DGP, and the non-negative variables can be nonstationary. Examples:
 - ARMA/GARCH-specifications
 - Nonstationary seasonality predictors
- CAN is established for the QMLE under high-level assumptions; they are verified for ARMA specifications
- The finite sample properties of the estimator is studied by simulation
- An empirical application illustrates the results

Thanks!

References:

Box, G. E. and N. R. Draper (1987). *Empirical Model-Building and Response Surfaces*. New York: Wiley.

De Jong, R. M. and J. Davidson (2000). Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica* 68(2), 407–423.

Hansen, P. R. and A. Lunde (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* 20, 873–889.