Justifying conditional inference in time series models

Eric Beutner
joint work with Alexander Heinemann and Stephan Smeekes

based on publication
Electronic Journal of Statistics (2021), 15, 2517-2565

Mathematics colloquium VU Amsterdam, March 30, 2022

A simple yet instructive model

A simple yet instructive model

- One of the simplest time series models is the AR(1) process given by
$$X_t = \beta X_{t-1} + \varepsilon_t$$
with $|\beta| < 1$ and $\varepsilon_t \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$.

A simple yet instructive model

- One of the simplest time series models is the AR(1) process given by
$$X_t = \beta X_{t-1} + \varepsilon_t$$
with $|\beta| < 1$ and $\varepsilon_t \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$.

- **Of interest could be:**
  - Prediction interval for $\beta X_T (= \mathbb{E}[X_{T+1}|X_T])$, i.e. (*unconditional inference*);
  - Or an interval containing the conditional expectation of X_{T+1} given the past, i.e. βx_T (*conditional inference*).

Example 1: AR(1)

- Which one is better?

Example 1: AR(1)

- Which one is better? Intuitively, the latter is more appropriate because it uses more information and hence will lead to more informative intervals.

Example 1: AR(1)

- Which one is better? Intuitively, the latter is more appropriate because it uses more information and hence will lead to more informative intervals.
- This is indeed true.

level $1-\alpha$	length for βX_T	expected length for βx_T
0.90	3.193	2.625
0.95	4.374	3.123
0.99	7.210	4.110

Example 1: AR(1)

- Which one is better? Intuitively, the latter is more appropriate because it uses more information and hence will lead to more informative intervals.
- This is indeed true.

level $1-\alpha$	length for βX_T	expected length for βx_T
0.90	3.193	2.625
0.95	4.374	3.123
0.99	7.210	4.110

- Calculations based on treating $\hat{\beta} X_T$ to be a product of two independent normals (second column).
- Ignoring the influence of x_T on the distribution of $\hat{\beta}_T$ (third column).

Example 1: AR(1)

- Conditional Confidence interval (CCI) for $\mathbb{E}[X_{T+1}|X_T = x_T] = \beta x_T$ based on the approximation

$$\frac{\sqrt{T}(\hat{\beta}(\mathbf{X}_T)\, x_T - \beta\, x_T)}{\sqrt{1 - \hat{\beta}(\mathbf{X}_T)^2}} = x_T \frac{\sqrt{T}(\hat{\beta}(\mathbf{X}_T) - \beta)}{\sqrt{1 - \hat{\beta}(\mathbf{X}_T)^2}} \approx N(0, x_T^2),$$

where $\mathbf{X}_T = (X_T, X_{T-1}, \ldots, X_1)$.

Example 1: AR(1)

- Conditional Confidence interval (CCI) for $\mathbb{E}[X_{T+1}|X_T = x_T] = \beta x_T$ based on the approximation

$$\frac{\sqrt{T}(\hat{\beta}(\mathbf{X}_T)\, x_T - \beta\, x_T)}{\sqrt{1 - \hat{\beta}(\mathbf{X}_T)^2}} = x_T \frac{\sqrt{T}(\hat{\beta}(\mathbf{X}_T) - \beta)}{\sqrt{1 - \hat{\beta}(\mathbf{X}_T)^2}} \approx N(0, x_T^2),$$

  where $\mathbf{X}_T = (X_T, X_{T-1}, \ldots, X_1)$.
- Points of attention:
  - No limit on the rhs, $N(0, x_T^2)$ will not approach a fixed distribution;
  - Inconsistency on the lhs as the obsered x_T and the random X_T appear;
  - Because of x_T lhs needs to be treated as a conditional distribution.

Example 2: GARCH(1,1)

- $X_t = \sigma_t \varepsilon_t$ with $\varepsilon_t \sim i.i.d.(0,1)$ and

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2.$$

- **Goal:** CCI for $\sigma^2_{T+1|T} = \mathbb{E}\big[X^2_{T+1} | \mathbf{X}_T = \mathbf{x}_T\big]$.

Example 2: GARCH(1,1)

- $X_t = \sigma_t \varepsilon_t$ with $\varepsilon_t \sim i.i.d.(0,1)$ and

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2.$$

- **Goal:** CCI for $\sigma^2_{T+1|T} = \mathbb{E}\big[X^2_{T+1} | \mathbf{X}_T = \mathbf{x}_T\big]$.

- The recursive structure implies

$$\sigma^2_{T+1|T} = \psi_n(\mathbf{x}_T; \theta) = \omega \frac{1-\beta^T}{1-\beta} + \alpha \sum_{k=0}^{\infty} \beta^k x^2_{T-k}.$$

- With an estimator $\hat{\theta}(\mathbf{X}_T)$ for $\theta = (\omega, \alpha, \beta)'$ we have

$$\hat{\sigma}^2_{T+1|T} = \psi_n\big(\mathbf{x}_T; \hat{\theta}(\mathbf{X}_T)\big).$$

  Problem even more severe because now $\mathbf{x}_T = (x_T, x_{T-1}, \ldots)$ and $\mathbf{X}_T$ appear.

General set-up

- Object of interest $\psi_{T+1|T}(x_T, x_{T-1}, \ldots; \theta)$.
- If infeasible look at $\psi^s_{T+1|T}(x_T, x_{T-1}, \ldots, x_1, s_0, s_{-1} \ldots; \theta)$.
- Still infeasible because of θ.
- With an estimator $\hat{\theta}(\mathbf{X}_T)$ the problematic version would be

$$\hat{\psi}^s_{T+1|T}\left(x_T, x_{T-1}, \ldots, x_1, s_0, s_{-1} \ldots; \hat{\theta}(\mathbf{X}_T)\right).$$

The problem in the literature

- For the AR(1) Kreiss points out that researchers approximate the distribution of

$$\hat{\beta}(\mathbf{X}_T)x_T - \beta x_T$$

rather than the distribution of

$$\hat{\beta}(\mathbf{X}_T)X_T - \beta X_T \text{ given } X_T = x_T$$

and that approximating the latter seems to be rather cumbersome because even the rather simple condition $X_T = x_T$ has an influence on the whole series $X_1, \ldots, X_T$.

- Phillips (1979) approximates the conditional distribution by Edgeworth expansion; works only under $\varepsilon_t \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$.

The problem in the literature

- The standard approach takes a shortcut as follows: ”*... the series used for estimation of parameters and the series used for prediction are generated from two independent processes which have the same stochastic structure.*” - Lewis and Reinsel (1985), Lütkepohl (2005), Dufour and Taamouti (2010).

The problem in the literature

- The standard approach takes a shortcut as follows: *"... the series used for estimation of parameters and the series used for prediction are generated from two independent processes which have the same stochastic structure."* - Lewis and Reinsel (1985), Lütkepohl (2005), Dufour and Taamouti (2010).
- The standard approach then proceeds by using the distribution theory for

$$\hat{\psi}^s_{T+1|T}\left(x_T, x_{T-1}, \ldots, x_1, s_0, s_{-1} \ldots; \hat{\theta}(\mathbf{Y}_T)\right),$$

with $\mathbf{Y}_T$ independent of (X_t) and applies it to

$$\hat{\psi}^s_{T+1|T}\left(x_T, x_{T-1}, \ldots, x_1, s_0, s_{-1} \ldots; \hat{\theta}(\mathbf{X}_T)\right).$$

The problem in the literature

- The standard approach takes a shortcut as follows: *"... the series used for estimation of parameters and the series used for prediction are generated from two independent processes which have the same stochastic structure."* - Lewis and Reinsel (1985), Lütkepohl (2005), Dufour and Taamouti (2010).
- The standard approach then proceeds by using the distribution theory for

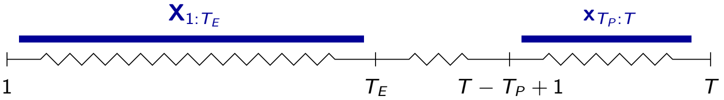
$$\hat{\psi}^{s}_{T+1|T}\left(x_T, x_{T-1}, \ldots, x_1, s_0, s_{-1} \ldots ; \hat{\theta}(\mathbf{Y}_T)\right),$$

with $\mathbf{Y}_T$ independent of (X_t) and applies it to

$$\hat{\psi}^{s}_{T+1|T}\left(x_T, x_{T-1}, \ldots, x_1, s_0, s_{-1} \ldots ; \hat{\theta}(\mathbf{X}_T)\right).$$

- $\Rightarrow$ not consistent (or like Pesaran phrases (2015) it "the particular assumptions that underlie the standard approach are not fully recognized.")
- Can one find another way to justify these CCIs?

Sample split approach



- SPL estimator:

$$\hat{\psi}^{SPL}_{T+1|T}\left(x_T, \ldots, x_{T_P}, s_{T_P-1}, \ldots, \hat{\theta}(X_{1:T_E})\right)$$

- Results in meaningful probabilistic statements because

$$\sqrt{T_E}\Big[\hat{\psi}^{SPL}_{T+1|T}\left(x_T, \ldots, x_{T_P}, s_{T_P-1}, \ldots, \hat{\theta}(X_{1:T_E})\right) - \psi^{SPL}_{T+1|T}\left(x_T, \ldots, x_{T_P}, s_{T_P-1}, \ldots, \theta\right)\Big].$$

does have non-degenerate (conditional) distributions.

Merging and Metrics

- Remains the issue with 'varying limit'.

Merging and Metrics

- Remains the issue with 'varying limit'.
- Usual definition of convergence in distribution $F_T(x) \to F(x)$ for all continuity points: not convenient to generalize.

Merging and Metrics

- Remains the issue with 'varying limit'.
- Usual definition of convergence in distribution $F_T(x) \to F(x)$ for all continuity points: not convenient to generalize.
- However, we know that convergence in distribution can be metricized as it is equivalent to $d_{BL}(F_T, F) \to 0$ with

$$d_{BL}(F, G) = \sup \left\{ \left| \int f d(F - G) \right| : ||f||_{BL} \leq 1 \right\}.$$

- This can be generalized to two sequences: Two sequences of probability measures F_T, G_T *merge* if and only if $d_{BL}(F_T, G_T) \to 0$ (Dudley (1968)).

Merging and Metrics

- Only one more thing needs to be taken into account.

Merging and Metrics

- Only one more thing needs to be taken into account.
- The distributions we look at are conditional distributions not fixed ones as in the merging definition on the previous slide.
- May remind the statistics group of the bootstrap where inference is also conditionally on the sample.

Merging and Metrics

- Only one more thing needs to be taken into account.
- The distributions we look at are conditional distributions not fixed ones as in the merging definition on the previous slide.
- May remind the statistics group of the bootstrap where inference is also conditionally on the sample.
- This suggests the following: Two sequences of random cdfs F_T, G_T *merge* in probability if and only if $d_{BL}(F_T, G_T) \xrightarrow{\mathbb{P}} 0$.

Merging and Metrics

Theorem Under some regularity conditions the conditional distributions of

$$\sqrt{T_E}\Big[\hat{\psi}^{SPL}_{T+1|T}\left(x_T, \ldots, x_{T_P}, s_{T_P-1}, \ldots, \hat{\theta}(X_{1:T_E})\right) \\ -\psi^{SPL}_{T+1|T}\left(x_T, \ldots, x_{T_P}, s_{T_P-1}, \ldots, \theta\right)\Big].$$

and

$$\sqrt{T}\Big[\hat{\psi}^{s}_{T+1|T}\left(x_T, x_{T-1}, \ldots, x_1, s_0, s_{-1} \ldots; \hat{\theta}(\mathbf{Y}_T)\right) \\ \hat{\psi}^{s}_{T+1|T}\left(x_T, x_{T-1}, \ldots, x_1, s_0, s_{-1} \ldots; \theta\right)\Big]$$

merge in probability.

Interval construction

Theorem Under some regularity conditions the intervals

$$\left[\hat{\psi}^{s}_{T+1|T} - \frac{(\hat{F}_T^{STA})^{-1}(1-\gamma_2)}{\sqrt{T}}\,,\, \hat{\psi}^{s}_{T+1|T} - \frac{(\hat{F}_T^{STA})^{-1}(\gamma_1)}{\sqrt{T}}\right]$$

and

$$\left[\hat{\psi}^{SPL}_{T+1|T} - \frac{(\hat{F}_T^{SPL})^{-1}(1-\gamma_2)}{\sqrt{T_E}}\,,\, \hat{\psi}^{SPL}_{T+1|T} - \frac{(\hat{F}_T^{SPL})^{-1}(\gamma_1)}{\sqrt{T_E}}\right]$$

are asymptotically equivalent in the sense that the centers and the lengths of these intervals converge in probability.

Sample splitting in practice

DGPs in the simulation study

$$x_t = \mu_t + \varepsilon_t, \quad \mu_t = \sum_{j=1}^{p} \beta_j x_{t-j}, \qquad t = 1, \dots, T, \tag{1}$$

with $T = 50, 75, 100, 150, 200$ and $\varepsilon_t \sim N(0,1)$.

Table 1: AR models considered in the simulation study

DGP	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
A	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B	0.20	-0.50	0.40	0.40	0.00	0.00	0.00	0.00
C	1.20	-0.96	0.77	-0.61	0.49	-0.39	0.31	-0.25
D	0.80	-0.64	0.51	-0.41	0.33	-0.26	0.21	-0.17

Sample splitting in practice: Conditional approach

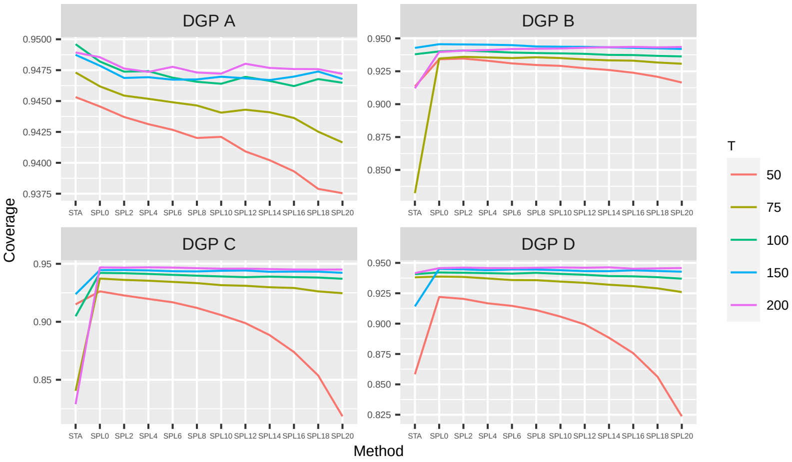


Figure 1: (Conditional) mean coverage

Sample splitting in practice: Conditional approach

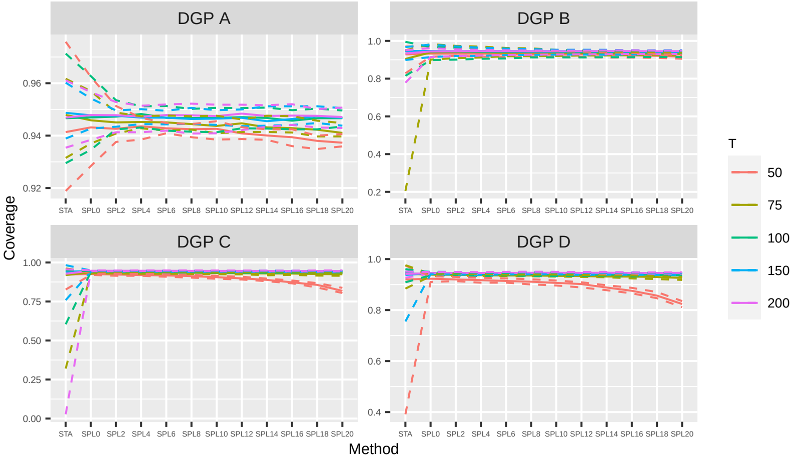


Figure 2: Conditional median (solid), minimum and maximum (dashed) coverage

Sample splitting in practice: Conditional approach

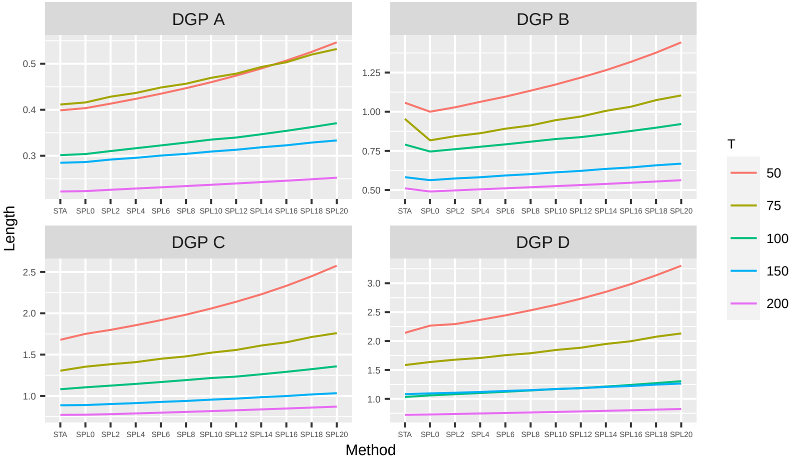


Figure 3: Mean interval length

More on coverage: Unconditional

Table 2: Coverage probability of *STA* and *SPL* for the AR(8) process in row C. DGP 1: shifted gamma innovations, DGP 2 innovations are a mixture of shifted gamma and normal

	DGP 1		DGP 2	
	STA	*SPL*	*STA*	*SPL*
$T_E = 40,\ T = 50$	89.8	91.1	90.2	91.2
$T_E = 50,\ T = 60$	90.8	92.1	91.3	92.5
$T_E = 60,\ T = 70$	91.5	92.8	91.9	92.9
$T_E = 70,\ T = 80$	92.1	93.2	92.4	93.5
$T_E = 80,\ T = 90$	92.6	93.6	92.7	93.5

Thank you for your attention!

References (merging)

- D'Aristotile, A., Diaconis, P. and Freedman, D. (1988). On merging of probabilities, Sankhyā: The Indian Journal of Statistics, Series A, 50, 363–380.
- Davydov, Y. and Rotar, V. (2009). On asymptotic proximity of distributions, Journal of Theoretical Probability, 22, 82–98,
- Dudley, R. M. (2002). Real Analysis and Probability, Cambridge University Press, Cambridge.
- Lunde, R. (2019). Sample splitting and weak assumption inference for time series, preprint, in revision.

Appendix

Assumption A1: *2IP Estimator*

1.1 (Estimator) $Z_n^{2IP} = m_n\big(\hat{\theta}(\mathbf{Y}_n) - \theta_0\big) \sim P_{Z_n}^{2IP} \to P_{\theta_0,\xi_0}$

1.2 (*Independence*) $\{Y_t\}$ is independent of $\{X_t\}$ and S_0

1.3 (*Differentiability*) $\psi_n(\,\cdot\,,\,\cdot\,;\theta)$ is continuous on Θ and twice differentiable on $\mathring{\Theta} = int(\Theta)$

1.4 (*Initial Cond.*) $\sqrt{n}\big(\psi_n(S_0, \mathbf{X}_n; \theta_0) - \psi_n(s_0^{\circ}, \mathbf{X}_n; \theta_0)\big) = o_p(1)$

1.5 (*Hessian*) $\sup_{\theta\in\mathring{\Theta}} \left|\left|\frac{\partial^2\psi_n(s_0^{\circ},\mathbf{X}_n;\theta)}{\partial\theta\partial\theta'}\right|\right| = O_p(1)$

1.6 (*Gradient*) $\left|\left|\frac{\partial\psi_n(s_0^{\circ},\mathbf{X}_n;\theta_0)}{\partial\theta}\right|\right| = O_p(1)$

Assumption A2: *SPL Estimator*

2.1 (Estimator) $Z_n^{SPL} = m_n\big(\hat{\theta}(\mathbf{X}_n^1) - \theta_0\big) \sim P_{Z_n}^{SPL} \to P_{\theta_0,\xi_0}$

2.2 (*Weak Dependence*) $\int h \, d\Big(P_{Z_n}^{SPL}(\cdot|\mathbf{X}_n^2, S_{n-n_2}) - P_{Z_n}^{SPL}\Big) \xrightarrow{p} 0$ w.r.t. $n_3 = n - n_1 - n_2$ for all h bounded Lipschitz

2.3 (*Differentiability*) $\psi_{n_2}(\,\cdot\,,\,\cdot\,;\theta)$ is continuous on Θ and twice differentiable on $\mathring{\Theta}$

2.4 (*Initial Cond.*) $\sqrt{n}\big(\psi_{n_2}(S_{n-n_2}, \mathbf{X}_n^2; \theta_0) - \psi_{n_2}(s_{n-n_2}^{\circ}, \mathbf{X}_n^2; \theta_0)\big) = o_p(1)$ w.r.t. n_2

2.5 (*Hessian*) $\sup_{\theta \in \mathring{\Theta}} \left|\left| \frac{\partial^2 \psi_{n_2}(s_{n-n_2}^{\circ}, \mathbf{X}_n^2; \theta)}{\partial\theta\partial\theta'} \right|\right| = O_p(1)$ w.r.t. n_2

2.6 (*Gradient*) $\left|\left| \frac{\partial \psi_{n_2}(s_{n-n_2}^{\circ}, \mathbf{X}_n^2; \theta_0)}{\partial\theta} \right|\right| = O_p(1)$ w.r.t. n_2

Merging of 2IP and SPL in probability

Assumption A3: *Merging Gradients*

$$\left\| \frac{\partial \psi_n(s_0^\circ, \mathbf{X}_n; \theta_0)}{\partial \theta} - \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta_0)}{\partial \theta} \right\| \xrightarrow{p} 0 \, .$$

Theorem 1:

Under Assumptions **A1** to **A3**,

$$\mathbb{P}\Big[\sqrt{n}\big(\hat{\lambda}_{n+1}^{2IP} - \lambda_{n+1}\big) \in \cdot | \mathbf{X}_n, S_0\Big] \quad \text{and}$$

$$\mathbb{P}\Big[\sqrt{n}\big(\hat{\lambda}_{n+1}^{SPL} - \lambda_{n+1}\big) \in \cdot | \mathbf{X}_n^2, S_{n-n_2}\Big]$$

merge in probability.

Interval construction

Assumption A4: *Plug-in Estimator*

4.1 (*2IP*) $\int h \, dP_{\hat{\theta}(\mathbf{X}_n),\hat{\xi}(\mathbf{X}_n)} \xrightarrow{p} \int h \, dP_{\theta_0,\xi_0}$ $\forall h$ bounded Lipschitz;

4.2 (*SPL*) $\int h \, dP_{\hat{\theta}(\mathbf{X}_n^1),\hat{\xi}(\mathbf{X}_n^1)} \xrightarrow{p} \int h \, dP_{\theta_0,\xi_0}$ $\forall h$ bounded Lipschitz.

$$I_\gamma^{2IP}(\mathbf{Y}_n,\mathbf{X}_n) = \left[\hat{\lambda}_{n+1}^{2IP} - \frac{\widehat{F_n^{2IP}}^{-1}(1-\gamma_2)}{\sqrt{n}} \, , \, \hat{\lambda}_{n+1}^{2IP} - \frac{\widehat{F_n^{2IP}}^{-1}(\gamma_1)}{\sqrt{n}}\right]$$

$$I_\gamma^{SPL}(\mathbf{X}_n^1,\mathbf{X}_n^2) = \left[\hat{\lambda}_{n+1}^{SPL} - \frac{\widehat{F_n^{SPL}}^{-1}(1-\gamma_2)}{\sqrt{n}} \, , \, \hat{\lambda}_{n+1}^{SPL} - \frac{\widehat{F_n^{SPL}}^{-1}(\gamma_1)}{\sqrt{n}}\right]$$

$$I_\gamma^{ILL}(\mathbf{X}_n,\mathbf{X}_n) = I_\gamma^{2IP}(\mathbf{Y}_n,\mathbf{X}_n)|_{\mathbf{Y}_n=\mathbf{X}_n}$$

Theorem 2: *Asymptotic Coverage*

2.1 Under **A1** and **A4**, $P_n^{2IP}(\cdot|\mathbf{X}_n, S_0)$ and $\widehat{P_n^{2IP}}(\cdot)$ merge in probability. If in addition $\widehat{F_n^{2IP}}(\cdot)$ is stochastically uniformly equicontinuous, $\mathbb{P}\big[I_\gamma^{2IP}(\mathbf{Y}_n, \mathbf{X}_n) \ni \lambda_{n+1} \big| \mathbf{X}_n, S_0\big] \xrightarrow{p} 1-\gamma$;

2.2 Under **A2** and **A4**, $P_n^{SPL}(\cdot|\mathbf{X}_n^2, S_{n-n_2})$ and $\widehat{P_n^{SPL}}(\cdot)$ merge in probability. If in addition $\widehat{F_n^{SPL}}(\cdot)$ is stochastically uniformly equicontinuous, $\mathbb{P}\big[I_\gamma^{SPL}(\mathbf{X}_n^1, \mathbf{X}_n^2) \ni \lambda_{n+1} \big| \mathbf{X}_n^2, S_{n-n_2}\big] \xrightarrow{p} 1-\gamma$.

Theorem 3: *Asymptotic Equivalence of Confidence Intervals*

3.1 (Location) If **A1** and **A2** hold, then $\hat{\lambda}_{n+1}^{ILL} - \hat{\lambda}_{n+1}^{SPL} \xrightarrow{p} 0$.

3.2 (Length) Under the assumptions of Theorem 1 and 2 and $\widehat{F_n^{SPL}}^{-1}(\cdot)$ being stochastically pointwise equicontinuous at $u = \gamma_1, 1-\gamma_2$, we have $\widehat{F_n^{ILL}}^{-1}(u) - \widehat{F_n^{SPL}}^{-1}(u) \xrightarrow{p} 0$.

Discussion

"... although [the parameter] is assumed fixed at the estimation stage, it is unknown to the forecaster and, from this perspective, it is best viewed as a random variable at the forecasting stage." - Pesaran (2015)

- Pesaran assigns to θ some posterior distribution.

- Treating θ not fixed but random, the issue solves, e.g. $\mathscr{L}\left(\mu_{n+1} - \hat{\mu}^{ILL}_{n+1} | \mathbf{X}_n = \mathbf{x}_n\right)$ is non-degenerate

Critique

Combining a frequentist view with a Bayesian-akin approach does not seem to be coherent.

φ - Mapping

Example: GARCH(1,1)

$$\underbrace{\begin{pmatrix} \sigma_t^2 \\ X_t^2 \end{pmatrix}}_{S_t} = \underbrace{\begin{pmatrix} \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2 \\ X_t^2 \end{pmatrix}}_{\varphi(S_{t-1}, X_t; \theta)},$$

Example: ARMA(1,1)

$$\underbrace{\begin{pmatrix} \varepsilon_t \\ X_t \end{pmatrix}}_{S_t} = \underbrace{\begin{pmatrix} X_t - \alpha \varepsilon_{t-1} - \beta X_{t-1} \\ X_t \end{pmatrix}}_{\varphi(S_{t-1}, X_t; \theta)},$$

Other nested models

ARMA(p,q), GARCH(p,q), T-GARCH (Zakoïan, 1994), ACD (Engle and Russell, 1998), Score models (Harvey, 2013)

Object of interest

- consider $\lambda_{n+1} = \pi(S_n; \theta_0)$

- The recursive structure implies

$$\lambda_{n+1} = \psi_n(S_0, \mathbf{X}_n; \theta_0)\,, \qquad \text{(random)}$$

where $\psi_n = \pi \circ \underbrace{\varphi \cdots \circ \varphi}_{n\text{-times}}$

- Conditioning on $\mathbf{X}_n = \mathbf{x}_n$ and $S_0 = s_0$, it reduces to

$$\lambda_{n+1|n} = \psi_n(s_0, \mathbf{x}_n; \theta_0) \qquad \text{(non-random)}$$

Taylor expansion

- 2IP estimator

$$\sqrt{n}\Big(\hat{\lambda}_{n+1}^{2IP} - \lambda_{n+1}\Big) = \frac{\partial \psi_n(s_0^\circ, \mathbf{X}_n; \theta_0)}{\partial \theta'} \sqrt{n}\big(\hat{\theta}(\mathbf{Y}_n) - \theta_0\big) + R_n^{2IP}$$

- SPL estimator

$$\sqrt{n}\Big(\hat{\lambda}_{n+1}^{SPL} - \lambda_{n+1}\Big) = \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta_0)}{\partial \theta'} \sqrt{n}\big(\hat{\theta}(\mathbf{X}_n^1) - \theta_0\big) + R_n^{SPL}$$

Assumption 5

Assumption 5: Normality
We assume $P_{\theta_0,\xi_0} = N(0, \Upsilon_0)$ with $\Upsilon_0 = \Upsilon(\theta_0, \xi_0)$ and there exist $\hat{\Upsilon}(\mathbf{X}_n)$ and $\hat{\Upsilon}(\mathbf{X}_n^1)$ converging in probability to Υ_0.