Estimation of Time Series Models Using Generalized Spectral Distribution

EEA-ESEM 2022, Milan

Weifeng Jin

Department of Economics, Universidad Carlos III de Madrid

August 25, 2022

Motivation

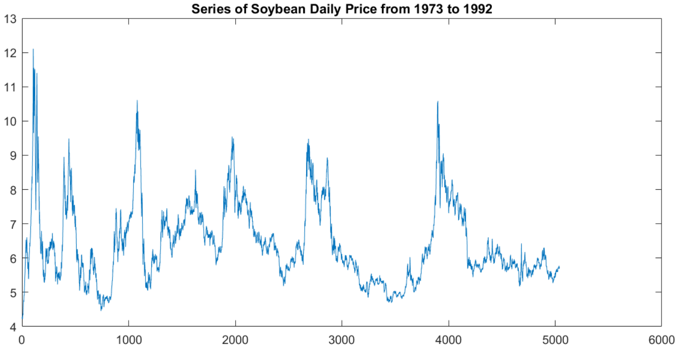


Figure: Path of soybean daily price from 1973 to 1992

- Speculative behaviour leads to asymmetric dynamics like phases of local explosiveness.
- Limitation of conventional linear process in characterizing nonlinear dynamics.

Motivation

- Predictability of future values given private information held by agents $\rightarrow$ Incorporate future components into the model $\Leftrightarrow$ "non-fundamental" time series model (**temporal dependence in both future and past**)
- Autoregressive-moving-average model
  - **Non-causality**, e.g.

$$Y_t = \alpha Y_{t-1} + u_t, |\alpha| > 1$$

    - nonlinear dynamics: **speculative bubbles**, **asymmetric cycles** (Gouriéroux and Zakoïan 2017) illustration and **volatility clustering** (Breidt, Davis, Trindade, et al. 2001)
    - improvement in forecasting accuracy in macroeconomics and finance (Hecq and Voisin 2020; Lanne, Luoto, and Saikkonen 2012)
    - alternative to non-invertible processes in modeling forward-looking behaviour (Lanne and Luoto 2013)

Motivation

- Autoregressive-moving-average model
  - **Non-invertibility**, e.g.

$$Y_t = u_t + \beta u_{t-1}, |\beta| > 1$$

    - economically sensible impulse-response functions (Lippi and Reichlin 1993)
    - information flow with unusual discounting patterns due to the fiscal foresight of tax policy (Leeper, Walker, and Yang 2013)
    - non-revealing equilibrium when agents have heterogeneous information (Kasa, Walker, and Whiteman 2006; Walker 2007)
  - Classical time series analysis is restricted to causal and invertible ARMA models
- In this paper, I investigate a **novel estimation technique** of general time series models which accommodate **non-invertibility** and **non-causality**.

Trajectory of causal/non-causal autoregressions

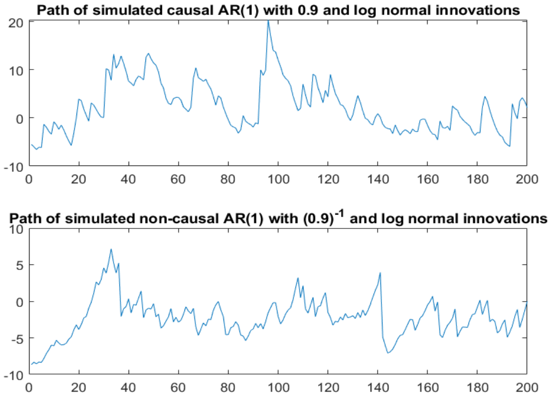


Figure: Comparison of trajectories of causal v.s. non-causal non-Gaussian AR(1)

- non-causal AR can generate repetitive episodes of upward trends followed by a sharp drop.

Obstacle in the identification

- Consider two AR(1) models with $|\rho| < 1$,

$$
\begin{aligned}
Y_t^{(1)} &= \rho Y_{t-1}^{(1)} + u_t, u_t \sim IID(0, \sigma^2) \\
Y_t^{(2)} &= \rho^{-1} Y_{t-1}^{(2)} + v_t \Leftrightarrow Y_{t-1}^{(2)} = \rho Y_t^{(2)} - \rho v_t, v_t \sim IID(0, \sigma^2)
\end{aligned}
$$

- They have the same autocorrelation function,

$$
\begin{aligned}
\mathbb{C}\mathrm{orr}\left(Y_t^{(1)}, Y_{t-1}^{(1)}\right) &= \frac{\mathbb{C}\mathrm{ov}\left(\rho Y_{t-1}^{(1)} + u_t, Y_{t-1}^{(1)}\right)}{\mathbb{V}\mathrm{ar}\left(Y_{t-1}^{(1)}\right)} = \rho \\
\mathbb{C}\mathrm{orr}\left(Y_t^{(2)}, Y_{t-1}^{(2)}\right) &= \frac{\mathbb{C}\mathrm{ov}\left(Y_t^{(2)}, \rho Y_t^{(2)} - \rho v_t\right)}{\mathbb{V}\mathrm{ar}\left(Y_t^{(2)}\right)} = \rho
\end{aligned}
$$

- Failure of estimation methods based on second-order structure for $Y_t^{(2)}$
  - Ordinary Least Square(**OLS**) or **Gaussian Likelihood Estimation**
- $\left(Y_t^{(1)}, Y_t^{(2)}\right)$ display distinct dynamics, see Figure.2

Non-Gaussian Estimation

In the **Gaussian** framework, uncorrelatedness is equivalent to iid, but not in the non-Gaussian case.
So what **other information** from iid model innovations can be employed for identification?

- **Non-Gaussian MLE**: Breidt, Davis, Lh, et al. 1991, Lii and Rosenblatt 1992, Lii and Rosenblatt 1996, Gouriéroux and Zakoïan 2017.
- **Minimum distance estimation**: Gospodinov and Ng 2015,Velasco and Lobato 2018,Velasco 2021, Cabello 2021.

This paper

- **Methodology**: A new minimum distance estimation method based on a dependence measure using the cumulative distribution function of the residuals.
- **Advantages**:
  - Identification of general time series models robust to non-causality and non-invertibility without prior information on the distribution of the innovations.
  - Mild conditions on higher order moments for asymptotic analysis of the estimates.
  - Extension to different dependence structures, e.g. conditional mean independence and conditional quantile independence.

Dependence measure: generalized spectral distribution

- Given a sequence of random variable $\{X_t\}$,

$$X_t \perp X_{t-j} \iff P(X_t \leq x, X_{t-j} \leq y) = P(X_t \leq x) P(X_{t-j} \leq y)$$
$$\forall (x, y) \in \mathbb{R}^2 \text{ for } j = \pm 1, \pm 2, \ldots$$

- General covariance of (X_t, X_{t-j}):

$$\begin{aligned} \sigma_j(x, y) &= \mathbb{C}\mathrm{ov}\left(I(X_t \leq x), I(X_{t-j} \leq y)\right) \\ &= P(X_t \leq x, X_{t-j} \leq y) - P(X_t \leq x) P(X_{t-j} \leq y) \\ &\forall (x, y) \in \mathbb{R}^2 \text{ for } j = \pm 1, \pm 2, \ldots \end{aligned}$$

- Discrepancy between the joint *cdf* and the product of two marginal *cdf*, see Hoeffding 1948.

Dependence measure: generalized spectral distribution

- Generalized spectral density of the sequence $\{X_t\}$, see Hong 2000:

$$h(x,y,\omega) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_j(x,y)e^{-ij\omega}, \quad \omega \in [-\pi,\pi], i=\sqrt{-1}$$

- Generalized spectral distribution function:

$$\begin{aligned} H(x,y,\lambda) =& 2\int_0^{\lambda\pi} h(x,y,\omega)d\omega \\ =& \lambda\sigma_0(x,y) + 2\sum_{j=1}^{\infty}\sigma_j(x,y)\frac{\sin(j\pi\lambda)}{j\pi}, \quad \lambda \in [0,1] \end{aligned}$$

- Take into account all pairwise dependence in $\{X_t\}$.
- Under iid assumption of $\{X_t\}$,

$$\begin{aligned} h_0(x,y,\omega) =& \frac{1}{2\pi}\sigma_0(x,y) \\ H_0(x,y,\lambda) =& \lambda\sigma_0(x,y), \quad \forall(x,y) \in \mathbb{R}^2 \quad \lambda \in [0,1] \end{aligned}$$

The Model

- Consider a general linear time series model

$$Y_t = \sum_{j=-\infty}^{\infty} \varphi_j u_{t-j} \tag{1}$$

where u_t is a sequence of iid innovations with $\mathbb{E}\left|u_t\right| < \infty$, and $\sum_{j=-\infty}^{\infty} \left|\varphi_j\right| < \infty$.

- Two-sided moving average in (1) allows non-causality and non-invertibility in time series.
- **Goal**:identification and estimation of $\theta \in \Theta \subset \mathbb{R}^m$

$$\varphi\left(\theta; L\right) = \sum_{j=-\infty}^{\infty} \varphi_j(\theta) L^j$$

- ARMA(p, q): $\alpha\left(L\right) Y_t = \beta\left(L\right) u_t$, where $\varphi(L) = \alpha^{-1}(L)\beta(L)$ with polynomials $\alpha\left(L\right) = 1 - \sum_{j=1}^{p} \alpha_j L^j$ and $\beta\left(L\right) = 1 + \sum_{j=1}^{q} \beta_j L^j$ having no common roots and all roots away from unity.

The Model

- Residual $\{u_t(\theta)\}$ at any given θ is calculated by

$$u_t(\theta) = \varphi(\theta, L)^{-1} Y_t = \varphi(\theta, L)^{-1}\ \varphi(\theta_0, L) u_t \equiv \phi(\theta, L) u_t$$

- Generalized spectral density and distribution of $\{u_t(\theta)\}$

$$h_\theta(x, y; \omega) = \frac{1}{2\pi} \sum_{-\infty}^{\infty} \sigma_{\theta,j}(x, y) e^{-ij\omega}$$

$$H_\theta(x, y; \lambda) = \lambda \sigma_{\theta,0}(x, y) + 2 \sum_{j=1}^{\infty} \sigma_{\theta,j}(x, y) \frac{\sin j\pi\lambda}{j\pi}$$

where

$$\sigma_{\theta,j}(x, y) = \mathbb{C}\mathrm{ov}\left(I(u_t(\theta) \leq x), I(u_{t-j}(\theta) \leq y)\right)$$

- θ_0 is the true value of the parameters,i.e. $u_t(\theta_0) = u_t$

Model Estimation under *iid* Assumption

- $\theta = \theta_0$: from iid-ness of the innovation, we have

$$\sigma_{\theta_0,j}(x,y) = 0 \quad \forall j \neq 0, \forall (x,y) \in \mathbb{R}^2$$

- Generalized spectral density and distribution :

$$h_{\theta_0}(x,y;\omega) = \frac{1}{2\pi}\sigma_{\theta_0,0}(x,y)$$
$$H_{\theta_0}(x,y;\lambda) = \lambda\sigma_{\theta_0,0}(x,y) \quad \lambda \in [0,1]$$

- Population loss criterion: L_2 distance between $H_\theta(x,y;\lambda)$ and $H_{\theta_0}(x,y;\lambda)$

$$\begin{aligned}\mathcal{Q}_0(\theta) &= \int_{\mathbb{R}^2}\int_0^1 |H_\theta(x,y;\lambda) - H_{\theta_0}(x,y;\lambda)|^2 d\lambda dW(x,y) \\ &= 2\sum_{j=1}^{\infty} \frac{1}{(j\pi)^2}\int_{\mathbb{R}^2} \sigma_{\theta,j}^2(x,y)dW(x,y). \text{ (Parseval's identity)}\end{aligned}$$

Identification

- Under Assumption 1-3 Details, if $\theta \neq \theta_0$, $\{u_t(\theta)\}$ will not satisfy full or pairwise independence.

$$\begin{aligned} &\sigma_{\theta,j}(x,y) \neq 0 \quad \exists j \neq 0 \quad \exists (x,y) \in \mathbb{R}^2 \\ \rightarrow &\int_{\mathbb{R}^2} |\sigma_{\theta,j}(x,y)|^2 dW(x,y) > 0 \\ \rightarrow &\mathcal{Q}_0(\theta) > 0 \end{aligned}$$

and

$$\mathcal{Q}_0(\theta) = 0, \text{ when } \theta = \theta_0,$$

since $\sigma_{\theta_0,j}(x,y) = 0$ for $j \neq 0$. Plots

Consistency

Theorem 1

Assume $\{u_t\}$ *is iid with zero mean and* $\mathbb{E}\left(u_t\right)^2 < \infty$, $\theta_0 \in \Theta$, *under Assumptions 1-5* Details, $\mu_0 > 3$ *as* $T \to \infty$,

$$\hat{\theta}_T \longrightarrow_p \theta_0.$$

Asymptotic normality: smooth approximation

- The approach: approximate indicator function by smoothed cumulative distribution function.

$$\tilde{\sigma}_{\theta,j;h}(x,y) = \tilde{F}_{\theta,j}(x,y;h) - \tilde{F}_{\theta,j}(x,\infty;h)\tilde{F}_{\theta,j}(\infty,y;h)$$

where $\tilde{F}_{\theta,j}(x,y;h)$ is a smoothed empirical joint cdf of residuals $u_t(\theta)$,

$$\tilde{F}_{\theta,j}(x,y;h) = \frac{1}{T-j}\sum_{t=j+1}^{T} \Lambda\left(\frac{x-u_t(\theta)}{h}\right)\Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right),$$

where $\Lambda(\cdot)$ is a distribution defined on unbounded support with differentiable $pdf\,\lambda(\cdot)$ away from zero, making sure it is a total-revealing transformation, see Stinchcombe and White 1998.

h is a smoothing parameter which tends to 0^+ so as to

$$\Lambda(\frac{z}{h}) \to I(z>0) \text{ for } |z|>0$$

Asymptotic normality: $h \to 0, \quad T \to \infty$

Theorem 2

Under Assumptions 1-3 and Assumptions 6-7 ‣ Details, $\{u_t\}$ *iid with zero mean,* $\mathbb{E}\left|u_t\right|^3 < \infty, \theta_0 \in \Theta$, $\mu_0 > 3, \mu_1 > 1$, *as* $T \to \infty, h \to 0$

$$T^{1/2}\left(\tilde{\theta}_T^h - \theta_0\right) \longrightarrow_p \mathcal{N}\left(0, H_0^{-1} H_1 H_0^{-1}\right)$$

To maintain the $\sqrt{T}$-rate of the asymptotic distribution, the convergence rate of h to zero needs to be

- slower than T^{-1} and faster than $T^{-1/4}$.
- H_0, H_1 have components of local identification. ‣ Details
- Efficiency comparison with Gaussian MLE.
- Standard error can be computed by replacement of each components by analogy. ‣ Details

Simulation

- AR(1)

$$Y_t = \theta_0 Y_{t-1} + u_t$$

where $\{u_t\}$ are iid non-Gaussian innovations.

- Two approaches of choosing $W(\cdot)$
  - empirical distribution of the residuals
  - standard normal cdf after standardization of the residuals : $u_t^*(\theta) = \frac{u_t(\theta)}{\sqrt{\widehat{\mathbb{V}\mathrm{ar}}(u_t(\theta))}}$
- $\theta_0 = 0.4(0.4^{-1})$ and $0.9(0.9^{-1})$ in the causal (noncausal) case.
- $\{u_t\}$ follows U, t_3 and $\chi_5^2 - 5$.
- Sample size is 100 and 200 with 100 Monte Carlo replications.
- Percentage of correct root identification is reported in both approaches.

Simulaton result 1: comparison of estimates using two approaches in AR(1)

Table: Proportion of Correct Identification of AR(1)

				W: empirical *cdf*				W: standard normal *cdf*			
u_t	T		θ_0:	0.4	0.4^{-1}	0.9	0.9^{-1}	0.4	0.4^{-1}	0.9	0.9^{-1}
$U_{[-5,5]}$	100	PCI		67.00%	84.00%	53.00%	54.00%	62.00%	77.00%	45.00%	41.00%
	200	PCI		87.00%	86.00%	64.00%	63.00%	82.00%	78.00%	50.00%	54.00%
t_3	100	PCI		60.00%	69.00%	69.00%	57.00%	81.00%	86.00%	75.00%	72.00%
	200	PCI		76.00%	75.00%	63.00%	59.00%	90.00%	84.00%	71.00%	76.00%
$\chi^2_5 - 5$	100	PCI		94.00%	94.00%	71.00%	69.00%	94.00%	93.00%	78.00%	73.00%
	200	PCI		99.00%	98.00%	80.00%	76.00%	99.00%	99.00%	81.00%	77.00%

Simulation result 1: comparison of estimates using two approaches in AR(1)

- Both approaches with different weighting functions work well when the parameter is not close to unity.
- When θ_0 approaches unity, it becomes more difficult to distinguish causality and non-causality.
- As sample size increases, PCI gets higher.
- Skewness and excess kurtosis provide information in identification.
- In general, the approach with standard normal cdf is recommended.
- More complicated case when the order of AR is higher. Simulation

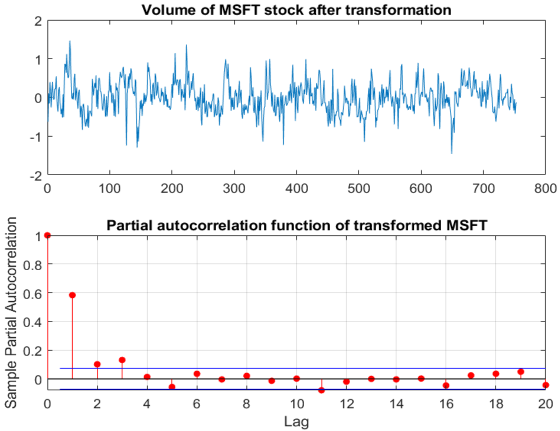


Figure: Microsoft daily trading volume from 6/3/1993 to 5/26/1999

- Our proposed method: $\hat{u}_t^{gsd} = Y_t - 1.7953Y_{t-1}$
- Noncausal autoregression is chosen.
- Compared to AR-ARCH model (

$$\begin{cases} Y_t = 0.5854Y_{t-1} + u_t \\ u_t = \sigma_t \epsilon_t \\ \sigma_t^2 = 0.088 + 0.1667u_{t-1}^2 \end{cases}$$

), both clustering volatility and asymmetric patterns are captured by non-causal AR model.

  - Fewer parameters to be estimated.
  - Dependence across quantiles: $\mathbb{C}\text{ov}\left(I\left(u_t < Q_u(\tau_1)\right), I\left(u_{t-j} < Q_u(\tau_2)\right)\right) \neq 0$.
  - Time irreversibility: $\mathbb{C}\text{ov}\left(I\left(Y_{t_1} \leq x\right), I\left(Y_{t_2} \leq y\right)\right) \neq \mathbb{C}\text{ov}\left(I\left(Y_{t_2} \leq x\right), I\left(Y_{t_1} \leq y\right)\right)$.

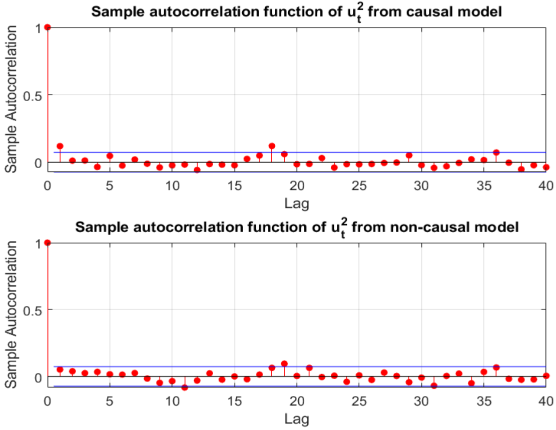
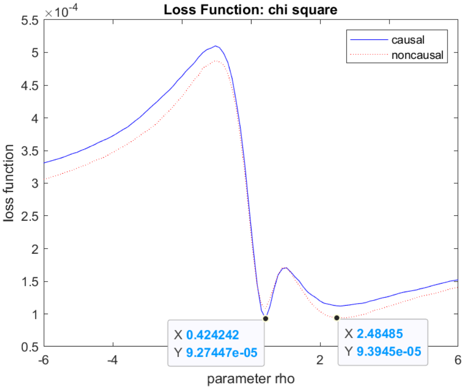


Figure: Sample autocorrelation function of squared residuals from causal/non-causal AR(1) models

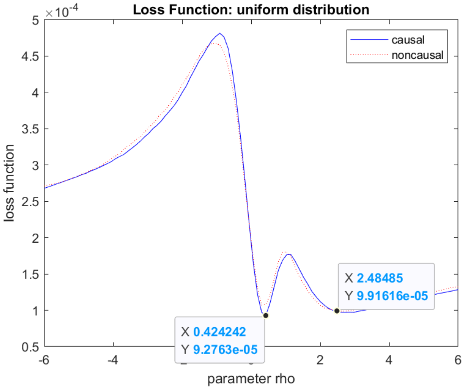
Discussion

- Choice of smoothing parameter h.
- Working on the application of modelling speculative bubbles.
- Extension to dependence under martingale difference innovations.
- Comparison of this method with other alternatives in terms of efficiency, like Gaussian PMLE and MLE assuming the distribution of innovation is known.
- Efficiency improvement by adding higher order dependence, i.e. triple dependence (u_t, u_{t-j}, u_{t-i}).

Local identification and global identification



(a) errors follow chi distribution



(b) errors follow uniform distribution

Figure: Loss function for both causal and noncausal AR(1)

Back to Identification

Two-step Estimation

- Noncausal Autoregression AR(p) model:

$$
\begin{aligned}
\alpha\left(L^{-1}\right) Y_t = u_t, &\quad u_t \sim IID(0, \sigma^2) \\
\Leftrightarrow \alpha(L)\, \alpha\left(L^{-1}\right) Y_t = \alpha(L)\, u_t & \\
\Leftrightarrow \alpha(L)\, Y_t = \frac{\alpha(L)}{\alpha\left(L^{-1}\right)} u_t = \epsilon_t &\quad \epsilon_t \sim WN(0, \tilde{\sigma}^2)
\end{aligned}
$$

- Causal AR(p) driven from ϵ_t
- $\{\epsilon_t\}$ is a white noise sequence but not independent (all-pass time series model), where the roots to the AR polynomial are the reciprocals of the roots to the MA polynomial.

Back to main page

Related literature

- **Test for serial dependence of a sequence**:
  - **Generalized covariance** $\sigma_j(x, y)$:
    - Nonparametric test for independence of two random variables (Hoeffding 1948)
    - First-order (Skaug and Tjøstheim 1993) and p serial dependence (Delgado 1996) in time series context
  - **Generalized spectral density and distribution**:
    - Serial dependence of residuals (Hong 2000, Du and Escanciano 2015)
    - Martingale difference hypothesis (Escanciano and Velasco 2006)
- **Capture dynamic features of variable of interest**:
  - conditional shape, time irreversibility and dependence in extremes or across quantiles (Kley et al. 2016, Lee and Rao 2011)
  - cyclical behaviours (Hagemann 2011)

Model Estimation under iid Assumption: Identification

Assumption 1

1. *Given a compact* Θ, *for any*
   $\theta \neq \theta_0, \phi(\theta, z) \neq a_0 z^{j_0}$, *in a subset of* $\mathbb{C}$ *with positive measure such that* $|z| = 1$.
2. *If* $|\phi(\theta, z)|^2 = 1$ *a.e.for* $z \in \mathbb{C}$ *such that* $|z| = 1$ *for some* $\theta \neq \theta_0$, *then* u_t *is non-Gaussian.*

- This assumption guarantees that the true innovation is only recovered at $\theta = \theta_0$.

Assumption 2

For compact Θ *and* $\mu_0 > 1$,

$$\sup_{\theta \in \Theta} |\varphi_j(\theta)| + \sup_{\theta \in \Theta} \left| \varphi_j^{-1}(\theta) \right| \leq C|j|^{-\mu_0}, \quad j = \pm 1, \pm 2, \ldots$$

- The Assumption.2 imposes weak dependence structure on the residual sequence $u_t(\theta)$.
- For ARMA(p, q), Assumption 1.1 is fulfilled and Assumption 2 is satisfied for any $\mu_0 > 0$.

Model Estimation under iid Assumption: Identification

Assumption 3

$W(x,y) = W(x,\infty)W(\infty,y)$ where $W(\cdot)$ is a probability distribution defined on $\mathbb{R}^2$, continuous and strictly increasing.

Identification

Model Estimation under iid Assumption: Consistency

Assumption 4

u_t admits a density function $f(u)$ with the first order derivative $f^{(1)}(u)$ that is Lebesgue integrable and has a^{th} order bounded moment, i.e. $\int_{\mathbb{R}} \left| f^{(1)}(u) \right| du < \infty$ and $\int_{\mathbb{R}} \left| u^a f^{(1)}(u) \right| du < \infty, a \geq 2$.

Assumption 5

The filter $\phi(\theta; z)$ is differentiable with the first order derivative $\phi^{(1)}(\theta; z) := \frac{\partial}{\partial \theta} \phi^{-1}(\theta; z) = \sum_{j=-\infty}^{\infty} \phi_j^{(1)}(\theta) z^j$ such that there exists a $\mu_1 > 1$,

$$\sup_{\theta \in \Theta} \left\| \phi_j^{(1)}(\theta) \right\| \leq C |j|^{-\mu_1}, \quad j = \pm 1, \pm 2, \ldots$$

Back to Theorem

Model Estimation under iid Assumption: Asymptotic Normality

Assumption 6

The filter $\phi(\theta) = \sum_{j=-\infty}^{\infty} \phi_j(\theta)$ *with three derivatives* $\phi^{(a)}(\theta)$ *satisfies following condition:*

$$\sup_{\theta \in \Theta} \left\| \phi_j^{(a)}(\theta) \right\| < C\,|j|^{-\eta_a} \text{ with } \eta_a > 1$$

for a=1,2,3.

Assumption 7

1. $\{u_t\}$ *admits uniformly bounded probability density function* $f(u)$ *with differentiable derivatives* $f^{(a)}(u)$ *of order a uniformly bounded by some constants C for a =1,2.*
2. H_0 *is positive definite (local identification).*

Back to Theorem

Asymptotic normality: smooth approximation

- Sample loss function:

$$\tilde{\mathcal{Q}}_T(\theta;h) = 2\sum_{j=1}^{T-1}(1-\frac{j}{T})\frac{1}{(j\pi)^2}\int_{\mathbb{R}^2}\tilde{\sigma}^2_{\theta,j}(x,y;h)dW(x,y)$$

- Smoothed estimator: $\tilde{\theta}_T^h = \text{argmin}_{\theta\in\Theta}\,\tilde{\mathcal{Q}}_T\left(\theta;h\right)$.
- Consistency of $\tilde{\theta}_T^h$ is guaranteed for any $h>0$ as $\Lambda\left(\frac{\cdot}{h}\right)$ is a total-revealing transformation, see Stinchcombe and White 1998.
- Asymptotic properties can be developed based on $\frac{\partial}{\partial\theta}\tilde{\mathcal{Q}}_T(\theta_0;h)$ and Hessian matrix $\frac{\partial^2}{\partial\theta\partial\theta'}\tilde{\mathcal{Q}}_T(\theta_0;h)$, see Velasco 2021. ▸ Details

Asymptotic normaliy: h fixed

Assumption 8

1. $\Lambda(u)$ *admits uniformly bounded positive probability density function* $\lambda(u)$ *with differentiable first order and second order derivatives* $\dot{\lambda}(u)$ *and* $\ddot{\lambda}(u)$ *uniformly bounded by some constants C.*
2. $H_{0,h}$ *is positive definite (local identification).*

Theorem 3

Let $\{u_t\}$ *be* iid *with zero mean,* $\mathbb{E}\left|u_t\right|^3 < \infty$ *and ,* $\mu_0 > 3, \mu_1 > 1, \theta_0 \in \Theta$*, Under Assumptions 1-3, Assumption 6 and 8, as* $T \to \infty$*,*

$$T^{1/2}\left(\tilde{\theta}_T^h - \theta_0\right) \longrightarrow_d \mathcal{N}\left(0, H_{0,h}^{-1} H_{1,h} H_{0,h}^{-1}\right)$$

Back

Asymptotic normality

- $$\Sigma_{0,1} := \sum_{j=1}^{\infty} j^{-2}\phi_j^{(1)}(\theta_0)\phi_j^{(1)}(\theta_0)' \quad \Sigma_{0,1}^{*} := \sum_{j=1}^{\infty} j^{-2}\phi_{-j}^{(1)}(\theta_0)\phi_{-j}^{(1)}(\theta_0)'$$
  and $\Sigma_{0,1}^{\dagger} := \sum_{j=1}^{\infty} j^{-2}\phi_j^{(1)}(\theta_0)\phi_{-j}^{(1)}(\theta_0)'$.
- $$\Sigma_{0,2} := \sum_{j=1}^{\infty} j^{-4}\phi_j^{(1)}(\theta_0)\phi_j^{(1)}(\theta_0)' \quad \Sigma_{0,2}^{*} := \sum_{j=1}^{\infty} j^{-4}\phi_{-j}^{(1)}(\theta_0)\phi_{-j}^{(1)}(\theta_0)'$$
  and $\Sigma_{0,2}^{\dagger} := \sum_{j=1}^{\infty} j^{-4}\phi_j^{(1)}(\theta_0)\phi_{-j}^{(1)}(\theta_0)'$.

Asymptotic normality: h fixed

- $H_{0,h} = \left(\Sigma_{0,1} + \Sigma^*_{0,1}\right) \rho^h_1 \rho^h_2 + \left(\Sigma^\dagger_{0,1} + \Sigma^{\dagger'}_{0,1}\right) \left(\rho^h_{12}\right)^2$, where

$$\rho^h_1 = \int_{\mathbb{R}} \left(\mu^h(x)\right)^2 dW(x), \;\; \rho^h_2 = \int_{\mathbb{R}} \left(\lambda^h(x)\right)^2 dW(x)$$

  and $\rho^h_{12} = \int_{\mathbb{R}} \mu^h(x)\lambda^h(x)dW(x)$
- $\varphi^h(x) := \mathbb{E}\left(\Lambda\left(\frac{x-u_t}{h}\right)\right), \;\; \lambda^h(x) := \frac{1}{h}\mathbb{E}\left(\lambda\left(\frac{x-u_t}{h}\right)\right), \;\; \mu^h(x) := \mathbb{E}\left(u_t\Lambda\left(\frac{x-u_t}{h}\right)\right).$
  $H_{1,h} = \left(\Sigma_{0,2} + \Sigma^*_{0,2}\right) \sigma^2_{e;h}\sigma^2_{\nu;h} + \left(\Sigma^\dagger_{0,2} + \Sigma^{\dagger'}_{0,2}\right) \sigma^2_{e\nu;h}$
- $\{\sigma^2_{e;h}, \sigma^2_{\nu;h}\}$ and $\sigma^2_{e\nu;h}$ are the variance and covariance of $\{e^h_t\}, \{\nu^h_t\}$ respectively.
-

$$e^h_t := \int_{\mathbb{R}} \left(\Lambda\left(\frac{x-u_t}{h}\right) - \varphi^h(x)\right) \lambda^h(x)dW(x)$$
$$\nu^h_t := \int_{\mathbb{R}} \left(\Lambda\left(\frac{x-u_t}{h}\right) - \varphi^h(x)\right) \mu^h(x)dW(x)$$

Asymptotic distribution: $h \to 0$

- $H_0 = \left(\Sigma_{0,1} + \Sigma_{0,1}^{*}\right)\rho_1\rho_2 + \left(\Sigma_{0,1}^{\dagger} + \Sigma_{0,1}^{\dagger\prime}\right)(\rho_{12})^2$, where

$$\rho_1 = \int_{\mathbb{R}} (\mu(x))^2\, dW(x),\ \ \rho_2 = \int_{\mathbb{R}} f^2(x) dW(x)$$

  and $\rho_{12} = \int_{\mathbb{R}} f(x)\lambda(x)dW(x), \mu(x) = \mathbb{E}\left(u_t I\left(u_t \leq x\right)\right)$.
- $H_1 = \left(\Sigma_{0,2} + \Sigma_{0,2}^{*}\right)\sigma_e^2\sigma_\nu^2 + \left(\Sigma_{0,2}^{\dagger} + \Sigma_{0,2}^{\dagger\prime}\right)\sigma_{e\nu}^2$
- $\{\sigma_e^2, \sigma_\nu^2\}$ and $\sigma_{e\nu}^2$ are the variance and covariance of $\{e_t\}, \{\nu_t\}$ respectively.
-

$$e_t := \int_{\mathbb{R}} \left(I\left(u_t \leq x\right) - F(x)\right) f(x) dW(x)$$
$$\nu_t := \int_{\mathbb{R}} \left(I\left(u_t \leq x\right) - F(x)\right) \mu(x) dW(x)$$

Back

Standard error

Following the expressions of H_0 and H_1, their components can be replaced by corresponding sample counterparts:

$$\hat{e}_t := \int_{\mathbb{R}} \left(I\left(\hat{u}_t(\hat{\theta}_T) \leq x\right) - \hat{F}(x) \right) \hat{f}(x) dW(x)$$
$$\hat{\nu}_t := \int_{\mathbb{R}} \left(I\left(\hat{u}_t(\hat{\theta}_T) \leq x\right) - \hat{F}(x) \right) \hat{\mu}(x) dW(x)$$

where

$$\hat{F}(x) = \frac{1}{T}\sum_{t=1}^{T} I\left(\hat{u}_t(\hat{\theta}_T) \leq x\right)$$

$$\hat{f}(x) = \frac{1}{Th}\sum_{t=1}^{T} \lambda\left(\frac{x - \hat{u}_t(\hat{\theta}_T)}{h}\right)$$ for properly chosen h and smooth pdf λ

$$\hat{\mu}(x) = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t(\hat{\theta}_T) I\left(\hat{u}_t(\hat{\theta}_T) \leq x\right).$$

Back

Simulation result 2: estimates of AR(2)

Table: Proportion of Correct Identification in AR(2)

	$\chi^2(5)-5$							
	T= 100				T= 200			
θ_0 :	$(0.4, 0.8)$	$(0.4^{-1}, 0.8^{-1})$	$(0.4^{-1}, 0.8)$	$(0.4, 0.8^{-1})$	$(0.4, 0.8)$	$(0.4^{-1}, 0.8^{-1})$	$(0.4^{-1}, 0.8)$	$(0.4, 0.8^{-1})$
PCI	59.00%	81.00%	84.00%	85.00%	80.00%	94.00%	95.00%	90.00%
PN	41.00%	95.00%	96.00%	85.00%	20.00%	100.00%	98.00%	99.00%

PCI: percentage of correct root identification including the number of roots lying inside unit circle

PN: percentage of detecting existence of noncausality in the process. $i.e.$ There is at least one root lying inside unit circle.

- The estimation in noncausal case outperforms the one in causal case in terms of PCI.
- The method is able to detect existence of noncausality but cannot pin down precisely where the noncausal root is. Simulation

Bibliography I

- Breidt, F Jay, Richard A Davis, Keh-Shin Lh, and Murray Rosenblatt (1991). "Maximum likelihood estimation for noncausal autoregressive processes". In: *Journal of Multivariate Analysis* 36.2, pp. 175–198.
- Breidt, F Jay, Richard A Davis, A Alexandre Trindade, et al. (2001). "Least absolute deviation estimation for all-pass time series models". In: *The Annals of Statistics* 29.4, pp. 919–946.
- Cabello, Miguel Angel (2021). "Identification and estimation of linear structural VARMA models using pairwise dependence measures". In: *Working paper.*
- Delgado, Miguel A (1996). "Testing serial independence using the sample distribution function". In: *Journal of Time Series Analysis* 17.3, pp. 271–285.
- Du, Zaichao and Juan Carlos Escanciano (2015). "A nonparametric distribution-free test for serial independence of errors". In: *Econometric Reviews* 34.6-10, pp. 1011–1034.
- Escanciano, Juan Carlos and Carlos Velasco (2006). "Generalized spectral tests for the martingale difference hypothesis". In: *Journal of Econometrics* 134.1, pp. 151–185.

Bibliography II

- Gospodinov, Nikolay and Serena Ng (2015). "Minimum distance estimation of possibly noninvertible moving average models". In: *Journal of Business & Economic Statistics* 33.3, pp. 403–417.
- Gouriéroux, Christian and Jean-Michel Zakoïan (2017). "Local explosion modelling by non-causal process". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79.3, pp. 737–756.
- Hagemann, Andreas (2011). "Robust spectral analysis". In: *arXiv preprint arXiv:1111.1965*.
- Hecq, Alain and Elisa Voisin (2020). "Forecasting bubbles with mixed causal-noncausal autoregressive models". In: *Econometrics and Statistics*.
- Hoeffding, Wassily (1948). "A non-parametric test of independence". In: *The annals of mathematical statistics*, pp. 546–557.
- Hong, Yongmiao (2000). "Generalized spectral tests for serial dependence". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62.3, pp. 557–574.

Bibliography III

- Kasa, Kenneth, Todd B Walker, and Charles H Whiteman (2006). "Asset prices in a time series model with perpetually disparately informed, competitive traders". In.
- Kley, Tobias, Stanislav Volgushev, Holger Dette, Marc Hallin, et al. (2016). "Quantile spectral processes: Asymptotic analysis and inference". In: *Bernoulli* 22.3, pp. 1770–1807.
- Lanne, Markku and Jani Luoto (2013). "Autoregression-based estimation of the new Keynesian Phillips curve". In: *Journal of Economic Dynamics and Control* 37.3, pp. 561–570.
- Lanne, Markku, Jani Luoto, and Pentti Saikkonen (2012). "Optimal forecasting of noncausal autoregressive time series". In: *International Journal of Forecasting* 28.3, pp. 623–631.
- Lee, Junbum and Suhasini Subba Rao (2011). "The quantile spectral density and comparison based tests for nonlinear time series". In: *arXiv preprint arXiv:1112.2759*.
- Leeper, Eric M, Todd B Walker, and Shu-Chun Susan Yang (2013). "Fiscal foresight and information flows". In: *Econometrica* 81.3, pp. 1115–1145.

Bibliography IV

- Lii, Keh-Shin and Murray Rosenblatt (1992). "An approximate maximum likelihood estimation for non-Gaussian non-minimum phase moving average processes". In: *Journal of Multivariate Analysis* 43.2, pp. 272–299.
- — (1996). "Maximum likelihood estimation for nonGaussian nonminimum phase ARMA sequences". In: *Statistica Sinica*, pp. 1–22.
- Lippi, Marco and Lucrezia Reichlin (1993). "The dynamic effects of aggregate demand and supply disturbances: Comment". In: *The American Economic Review* 83.3, pp. 644–652.
- Skaug, Hans Julius and Dag Tjøstheim (1993). "A nonparametric test of serial independence based on the empirical distribution function". In: *Biometrika* 80.3, pp. 591–602.
- Stinchcombe, Maxwell B and Halbert White (1998). "Consistent specification testing with nuisance parameters present only under the alternative". In: *Econometric theory* 14.3, pp. 295–325.

Bibliography V

Velasco, Carlos (2021). "Estimation of time series using residuals dependence measures". In: *Working paper.*

Velasco, Carlos and Ignacio N Lobato (2018). "Frequency domain minimum distance inference for possibly noninvertible and noncausal ARMA models". In: *The Annals of Statistics* 46.2, pp. 555–579.

Walker, Todd B (2007). "How equilibrium prices reveal information in a time series model with disparately informed, competitive traders". In: *Journal of Economic Theory* 137.1, pp. 512–537.