Econometric Theory

Non-Causal Econometrics

Gilles de Truchis - Université d'Orléans

Elena Dumitrescu - Université Paris-Panthéon-Assas

Ecole Doctorale 396 (Université Paris Nanterre)

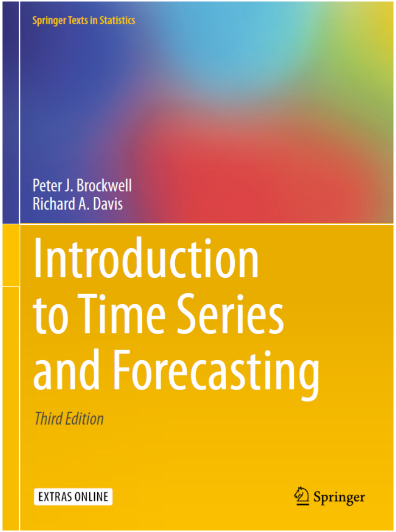
Last update: October 25, 2025

Les chapitres du cours

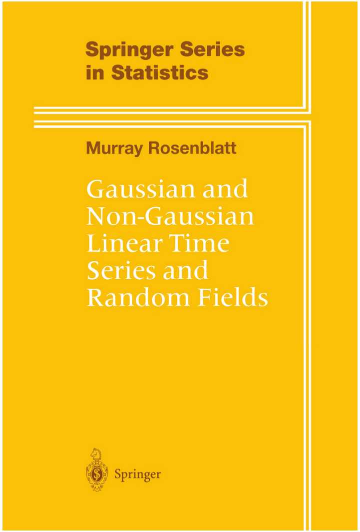
Outline

1. The linear model
2. Mixed-Phase processes
3. Identification
4. α-stable MAR
5. Prediction
6. Simulation of MAR
7. Estimation of MAR
8. Applications

Références


Springer Texts in Statistics
Peter J. Brockwell
Richard A. Davis
Introduction to Time Series and Forecasting
Third Edition
EXTRAS ONLINE
Springer

Références

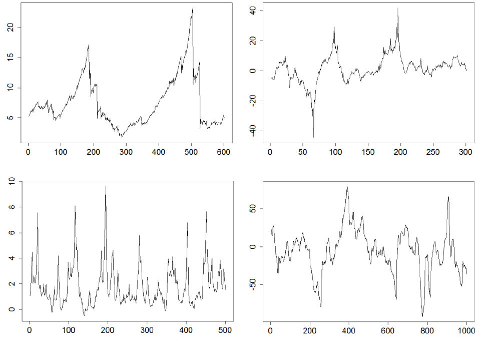

Springer Series
in Statistics
Murray Rosenblatt
Gaussian and
Non-Gaussian
Linear Time
Series and
Random Fields
Springer

Back to basics

- Consider the linear model X_t with $s \in (0,1]$ and $\psi_{(.)} \in \mathbb{R}$ a linear filter :

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \quad \sup_t \mathbb{E}|\varepsilon_t|^s < \infty, \quad \sum_{j=-\infty}^{\infty} |\psi_j|^s < \infty \tag{1}$$

- Can this model generate the following dynamics ?



$\Rightarrow$ Yes ! Even if X_t is strictly stationary !

Linear Time Series

$\Rightarrow$ Let start with some definitions and notations

Definition (1)

A stochastic process $\{X_t\}_{t \in \mathbb{Z}}$ is a linear time series if

$$X_t = f(\ldots, \varepsilon_{t-1}, \varepsilon_t, \varepsilon_{t+1}, \ldots)$$

where ε_t is a white noise and $f(.)$ is a linear function

- In the following we will consider the particular class of famous linear processes : ARMA processes
- We will see that even if ARMA process are linear, they can generate non-linear dynamics

White noises

Definition (2)

A Gaussian White Noise ε_t *is a sequence of* $\mathrm{i.i.d.}$ *random variable with* $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$

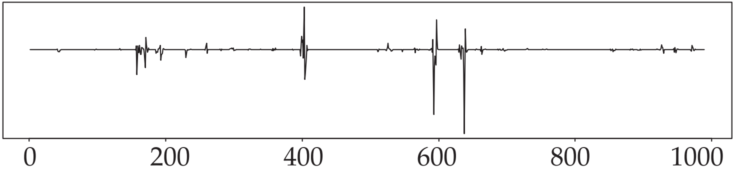
Definition (3)

A Strong White Noise ε_t *is a sequence of* $\mathrm{i.i.d.}$ *random variable with* $\mathbb{E}(\varepsilon_t) = 0$ *and* $\mathbb{E}(\varepsilon_t^2) = \sigma_\varepsilon^2$

Definition (4)

A Weak White Noise $\tilde{\varepsilon}_t$ *is a sequence of uncorrelated random variable with* $\mathbb{E}(\tilde{\varepsilon}_t) = 0$ *and* $\mathbb{E}(\tilde{\varepsilon}_t^2) = \sigma_{\tilde{\varepsilon}}^2$

Example $\tilde{\varepsilon}_t = u_t u_{t+1} \cdots u_{t+k}$ is a Weak White Noise



Strong linear process

Definition (5)

X_t is a strong linear stochastic process if it has an $MA(\infty)$ representation

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \quad \sum_{j=-\infty}^{\infty} |\psi_j|^s < \infty$$

with $s \in (0, 1]$ and $\varepsilon_t \sim SWN(0, \sigma_\varepsilon^2)$ a Strong White Noise

Strict stationarity

Definition (6)

A stochastic process $\{X_t\}_{t\in\mathbb{Z}}$ is a strictly stationary time series if

$$(X_1, X_2, \dots, X_n) \stackrel{d}{=} (X_{1+h}, X_{2+h}, \dots, X_{n+h})$$

for $h, n \geq 1$

Remark The term stationarity, without further qualification, will however stand for weak stationarity in the following

Remark Obviously, in the Gaussian case strict stationarity and weak stationarity are equivalent

Wold decomposition

Definition (7)

*Any purely stochastic (centered) weakly stationary time series has a unique weak linear **causal** representation*

$$X_t = \sum_{j=0}^{\infty} a_j \tilde{\varepsilon}_{t-j}, \quad \sum_{j=0}^{\infty} a_j^2 < \infty$$

where $\tilde{\varepsilon}_t$ is a weak white noise

- The representation is **causal** in the sense that only past shocks describe the dynamic of X_t

ARMA processes

Definition (8)

A linear time series $\{X_t\}_{t\in\mathbb{Z}}$ is an ARMA(p,q) if

$$X_t = \sum_{j=1}^{p} \varphi_j X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \Longleftrightarrow \varphi(B)X_t = \theta(B)\varepsilon_t$$

where ε_t is a SWN, $\varphi(B) = I - \sum_{j=1}^{p} \varphi_j B^j$ and $\theta(B) = I - \sum_{j=1}^{q} \theta_j B^j$

- X_t is stationary if $\varphi(B)$ and $\theta(B)$ have no common roots and

$$\varphi(z) = 0 \Rightarrow |z| \neq 1, \quad \forall z \in \mathbb{C}$$

$\Rightarrow$ no root on the unit circle or equivalently $\varphi(z) \neq 0$ for all $|z| = 1$

- Obviously, $X_t B^j = X_{t-j}$. Later we will also use $X_t F^j = X_{t+j}$, $F = B^{-1}$

Causal ARMA processes

Definition (9)

Let X_t be a stationary ARMA(p, q). Then, X_t is causal if $\varphi(z) \neq 0$ for all $|z| \leq 1$ or equivalently $\varphi(z) = 0 \Rightarrow |z| > 1, \quad \forall z \in \mathbb{C}$.

Illustration with a simple AR(1) : $X_t = \varphi_1 X_{t-1} + \varepsilon_t$

- If $|\varphi_1| < 1$, X_t has a linear Wold representation because

$$\sum_{j=0}^{\infty} \varphi_1^j B^j$$

has absolutely summable coefficients and

$$X_t = \sum_{j=0}^{\infty} \varphi_1^j \varepsilon_{t-j}$$

is the unique stationary solution

$\Rightarrow$ X_t has a strong **causal** MA(∞) representation which is a Wold representation

Non-Causal ARMA processes

Definition (10)

Let X_t be a stationary ARMA(p,q). Then, X_t is non-causal if $\varphi(z) \neq 0$ for all $|z| \geq 1$ or equivalently $\varphi(z) = 0 \Rightarrow |z| < 1, \quad \forall z \in \mathbb{C}$.

Illustration with a simple AR(1) : $X_t = \breve{\varphi}_1 X_{t-1} + \breve{\varepsilon}_t, \quad \breve{\varepsilon}_t \sim SWN$

- If $|\breve{\varphi}_1| > 1$, $\sum_{j=0}^{\infty} \breve{\varphi}_1^j B^j$ does not converge... however
  - Consider $X_{t+1} = \breve{\varphi}_1 X_t + \breve{\varepsilon}_{t+1}$
  - It follows that $X_t = \breve{\varphi}_1^{-1} \breve{\varphi}_1 X_t = \breve{\varphi}_1^{-1} X_{t+1} - \breve{\varphi}_1^{-1} \breve{\varepsilon}_{t+1}$ and iterating

$$X_t = -\sum_{j=0}^{\infty} \breve{\varphi}_1^{-j-1} \breve{\varepsilon}_{t+j+1} = \sum_{j=0}^{\infty} \varphi_1^j \varepsilon_{t+j}, \quad \varphi_1 = \breve{\varphi}_1^{-1}, \quad \varepsilon_t = -\varphi_1 \breve{\varepsilon}_{t+1}$$

where $\sum_{j=0}^{\infty} \varphi_1^j F^j$ has absolutely summable coefficients

$\Rightarrow$ Future shocks drive the dynamics : $X_t = \varphi_1 X_{t+1} + \varepsilon_t$ is **non-causal**

Non-Causal AR and stationary solution

Let X_t be a non-causal AR(1) defined by $X_t = \breve{\varphi}_1 X_{t-1} + \breve{\varepsilon}_t$

- Recall that $X_t = \breve{\varphi}_1^{-1} X_{t+1} - \breve{\varphi}_1^{-1} \breve{\varepsilon}_{t+1}$ and that iterating we have

$$X_t = -\breve{\varphi}_1^{-1} \breve{\varepsilon}_{t+1} - \cdots - \breve{\varphi}_1^{-k-1} \breve{\varepsilon}_{t+k+1} + \breve{\varphi}_1^{-k-1} X_{t+k+1} = \sum_{j=0}^{\infty} \varphi_1^j \varepsilon_{t+j}$$

- If X_t is stationary, $\mathbb{E}(X_t^2)$ is finite and time independent so that

$$\mathbb{E}\Big(\big(X_t - \sum_{j=0}^{k} \varphi_1^j \varepsilon_{t+j}\big)^2\Big) = \varphi_1^{2k+2} \mathbb{E}\Big(\big(X_{t+k+1}\big)^2\Big) \to 0 \text{ as } k \to \infty$$

which arises from the condition $|\varphi_1| = |\breve{\varphi}_1^{-1}| < 1$

$\Rightarrow$ $\sum_{j=0}^{k} \varphi_1^j \varepsilon_{t+j}$ is the unique stationary solution of X_t

- However, X_t depends on future shocks : not a Wold representation

Remark In $X_t = \breve{\varphi}_1 X_{t-1} + \breve{\varepsilon}_t$ the roots are **ill-located** whereas in $X_t = \varphi_1 X_{t+1} + \varepsilon_t$ the roots are **well-located**

Invertible ARMA

Definition (11)

Let X_t be a stationary ARMA(p,q). Then, X_t is invertible if $\theta(z) \neq 0$ for all $|z| \leq 1$ or equivalently $\theta(z) = 0 \Rightarrow |z| > 1, \quad \forall z \in \mathbb{C}$.

Illustration with an ARMA$(1,1)$: $X_t - \varphi_1 X_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}$, $|\theta_1| < 1$

- The power expansion of θ is $\zeta(z) = \sum_{j=0}^{\infty} -\theta_1^j z^j$ and hence

$$\varepsilon_t = \zeta(B)\varphi(B)X_t = \pi(B)X_t$$

with $\sum_{j=-\infty}^{\infty} |\pi_j| < \infty$ as $|\theta_1| < 1$

$\Rightarrow$ The invertible solution of X_t is

$$\varepsilon_t = X_t - (\varphi_1 + \theta_1) \sum_{j=1}^{\infty} -\theta_1^{j-1} X_{t-j}$$

as ε_t is expressed in terms of present and past values of X_s, $s \leq t$

Non-invertible ARMA

Definition (12)

Let X_t be a stationary ARMA(p, q). Then, X_t is non-invertible if $\theta(z) \neq 0$ for all $|z| \geq 1$ or equivalently $\theta(z) = 0 \Rightarrow |z| < 1, \quad \forall z \in \mathbb{C}$.

Illustration with an ARMA$(1, 1)$: $X_t - \varphi_1 X_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}$, $|\theta_1| > 1$

- Applying a similar reasoning to noncausal ARMA we obtain

$$\varepsilon_t = -\varphi_1 \theta_1^{-1} X_t + (\varphi_1 + \theta_1) \sum_{j=1}^{\infty} -\theta_1^{-j-1} X_{t+j}$$

$\Rightarrow$ X_t is noninvertible as ε_t is expressed in terms of present and **future** values of X_s, $s \geq t$

Mixed-Phase processes

1. The linear model
2. Mixed-Phase processes
3. Identification
4. α-stable MAR
5. Prediction
6. Simulation of MAR
7. Estimation of MAR
8. Applications

Minimum phase and Non-Minimum phase processes

- Let X_t be a stationary ARMA(p, q)

Definition (13)

If $\varphi(z) = 0 \Rightarrow |z| > 1, \quad \forall z \in \mathbb{C}$ and $\theta(z) = 0 \Rightarrow |z| > 1, \quad \forall z \in \mathbb{C}$, X_t is causal and invertible. Such a process is called ***minimum phase***

Definition (14)

If $\varphi(z) = 0 \Rightarrow |z| < 1, \quad \forall z \in \mathbb{C}$ and $\theta(z) = 0 \Rightarrow |z| < 1, \quad \forall z \in \mathbb{C}$, X_t is non-causal and non-invertible. Such a process is called ***non-minimum phase***

- Let X_t be a stationary AR$(p + q)$

Definition (15)

If $\varphi(z) = 0 \Rightarrow |z| \neq 1, \quad \forall z \in \mathbb{C}$ with some zeros inside and outside the unit circle, X_t is mixed-causal/noncausal. Such a process is called Mixed-AR or MAR(p, q) with p and q denoting the number of non-causal and causal lags respectively. The MAR enters the class of ***mixed phase*** *processes.*

Mixed phase AR processes

Definition (1)

A linear time series $\{X_t\}_{t\in\mathbb{Z}}$ is an AR$(p+q)$ if it is strictly stationary and satisfies the equation

$$\varphi(B)X_t = \varepsilon_t$$

where

$$\varphi(B) = 1 - \sum_{j=1}^{p+q} \varphi_j B^j = 0 \Rightarrow |z| \neq 1$$

and ε_t is a strong white noise

- Here we clearly allow the roots of $\varphi(B)$ to be **outside and/or inside the unit circle**
- When $\varphi(B)$ has zeros both outside and inside the unit circle the process is **mixed phase** and called a MAR(p,q)

The MAR(p,q)

Definition (1)

A linear time series $\{X_t\}_{t\in\mathbb{Z}}$ *is a MAR*(p,q) *if* $\varphi(B)X_t = \varepsilon_t$ *can be decomposed as*

$$\varphi^{\circ}(F)\varphi^{\bullet}(B)X_t = \varepsilon_t$$

where

$$\varphi^{\circ}(F) = 1 - \sum_{j=1}^{p} \varphi_j^{\circ} F^j, \quad \varphi^{\circ}(z) \neq 0 \text{ for all } |z| \leq 1$$

and

$$\varphi^{\bullet}(B) = 1 - \sum_{j=1}^{q} \varphi_j^{\bullet} B^j, \quad \varphi^{\bullet}(z) \neq 0 \text{ for all } |z| \leq 1$$

- Defining ε_t as a Gaussian noise would lead to an identification issue

$\Rightarrow$ In the following, ε_t will be assumed to be **Non-Gaussian**

MAR$(1,1)$

Example (1)

Let X_t be a MAR$(1,1)$ defined by $\varphi(B)X_t = \varepsilon_t$. Then, X_t can be decomposed as

$$(1-\breve{\varphi}^{\circ}B)(1-\varphi^{\bullet}B)X_t = \breve{\varepsilon}_t, \quad |\breve{\varphi}^{\circ}| > 1, \quad |\varphi^{\bullet}| < 1$$

or equivalently

$$(1-\varphi^{\circ}F)(1-\varphi^{\bullet}B)X_t = \varepsilon_t, \quad \varphi^{\circ} = 1/\breve{\varphi}^{\circ}, \quad \varepsilon_t = -\varphi^{\circ}\breve{\varepsilon}_{t+1}$$

- Then, the stationary solution of X_t is given by

$$X_t = \sum_{j=1}^{\infty} \frac{(\varphi^{\circ})^k}{1-\varphi^{\circ}\varphi^{\bullet}}\varepsilon_{t+k} + \frac{\varepsilon_t}{1-\varphi^{\circ}\varphi^{\bullet}} + \sum_{j=1}^{\infty} \frac{(\varphi^{\bullet})^k}{1-\varphi^{\circ}\varphi^{\bullet}}\varepsilon_{t-k}$$

or equivalently by the **two-sided** MA(∞) representation

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \text{ where } \psi(L) = \frac{1}{\varphi(B)} = \frac{1}{\varphi^{\circ}(F)\varphi^{\bullet}(B)}$$

The generalized Mixed Phase ARMA

Definition (18)

A linear time series $\{X_t\}_{t\in\mathbb{Z}}$ *is a generalized ARMA*$(p^\circ, p^\bullet, q^\circ, q^\bullet)$ *if* $\varphi(B)X_t = \theta(B)\varepsilon_t$ *can be decomposed as*

$$\varphi^\circ(F)\varphi^\bullet(B)X_t = \theta^\circ(F)\theta^\bullet(B)\varepsilon_t$$

implying the stationary solution $X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$ *if*

$$\varphi^\circ(F) = 1 - \sum_{j=1}^{p^\circ} \varphi_j^\circ F^j, \quad \varphi^\bullet(B) = 1 - \sum_{j=1}^{p^\bullet} \varphi_j^\bullet B^j, \quad \varphi^\circ(z), \varphi^\bullet(z) \neq 0, \ \forall |z| \leq 1$$

and

$$\theta^\circ(F) = 1 - \sum_{j=1}^{q^\circ} \theta_j^\circ F^j, \quad \theta^\bullet(B) = 1 - \sum_{j=1}^{q^\bullet} \theta_j^\bullet B^j, \quad \theta^\circ(z), \theta^\bullet(z) \neq 0, \ \forall |z| \leq 1$$

where $\psi(B) = \varphi(B)^{-1}\theta(B)$

- X_t is purely causal when $p^\circ = 0$ and purely non-causal when $p^\bullet = 0$

... purely invertible when $q^\circ = 0$ and purely non-invertible when $q^\bullet = 0$

The generalized Mixed Phase ARFIMA

Definition (19)

A linear time series $\{X_t\}_{t\in\mathbb{Z}}$ *is a generalized ARFIMA*$(p^\circ, p^\bullet, q^\circ, q^\bullet)$ *if* $(1-B)^\delta \varphi(B) X_t = \theta(B)\varepsilon_t$ *can be decomposed as*

$$(1-B)^\delta \varphi^\circ(F)\varphi^\bullet(B)X_t = \theta^\circ(F)\theta^\bullet(B)\varepsilon_t$$

implying the stationary solution $X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$ *if*

$$\varphi^\circ(F) = 1 - \sum_{j=1}^{p^\circ} \varphi_j^\circ F^j, \quad \varphi^\bullet(B) = 1 - \sum_{j=1}^{p^\bullet} \varphi_j^\bullet B^j, \quad \varphi^\circ(z), \varphi^\bullet(z) \neq 0, \ \forall |z| \leq 1$$

and

$$\theta^\circ(F) = 1 - \sum_{j=1}^{q^\circ} \theta_j^\circ F^j, \quad \theta^\bullet(B) = 1 - \sum_{j=1}^{q^\bullet} \theta_j^\bullet B^j, \quad \theta^\circ(z), \theta^\bullet(z) \neq 0, \ \forall |z| \leq 1$$

where $\psi(B) = \varphi(B)^{-1}\theta(B)$ *and* $\delta \in (-1/2, 1/2)$

- Generalized Mixed Phase ARMA and ARFIMA models are introduced by Wu and Davis (2010) and Wu (2014)

Mixed phase baseline paths

- For a given t and τ, a strong linear process can be represented as

$$X_t = \sum_{\tau=-\infty}^{\infty} \varepsilon_\tau \mathbb{1}_{\tau \leq t} \psi_{t-\tau}$$

if the process is causal and as

$$X_t = \sum_{\tau=-\infty}^{\infty} \varepsilon_\tau \mathbb{1}_{\tau \geq t} \psi_{t-\tau}$$

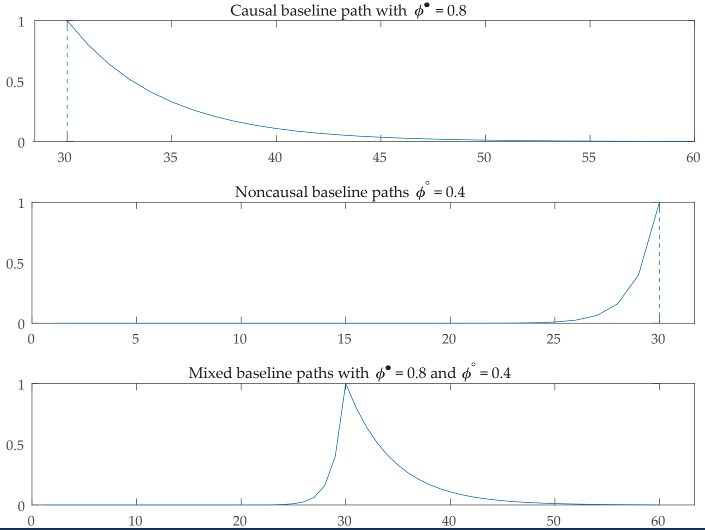
if the process is noncausal, that is a combination of baseline paths with stochastic $i.i.d.$ coefficients

- If the process is mixed phase we have

$$X_t = \sum_{\tau=-\infty}^{\infty} \varepsilon_\tau \mathbb{1}_{\tau \leq t} \psi_{t-\tau} + \sum_{\tau=-\infty}^{\infty} \varepsilon_\tau \mathbb{1}_{\tau \geq t} \psi_{t-\tau}$$

Mixed phase baseline paths

- $X_t = \varphi^\bullet X_{t-1} + \varepsilon_t$ / $X_t = \varphi^\circ X_{t+1} + \varepsilon_t$ / $X_t = \varphi^\circ X_{t+1} + \varphi^\bullet X_{t-1} + \varepsilon_t$, $\tau = 30$



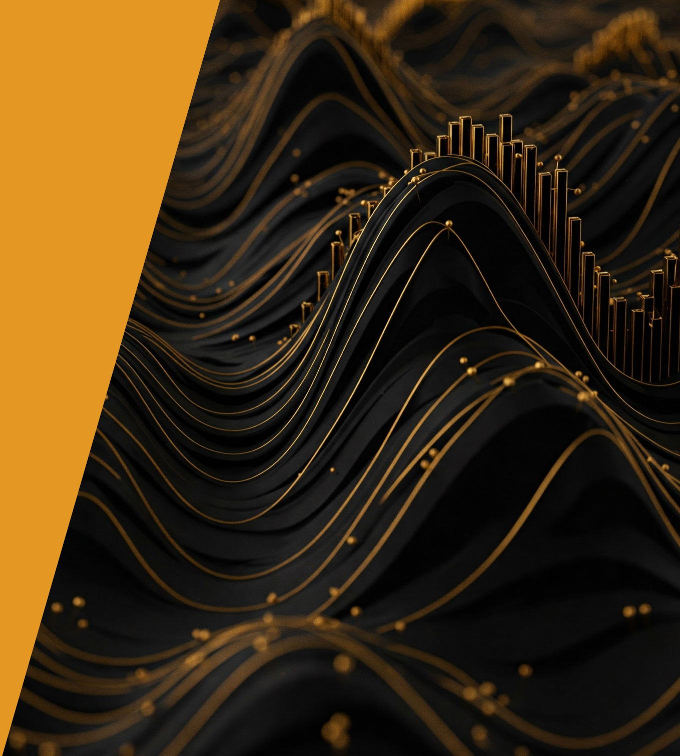
Tracking error in minimum phase

- Let ε_τ a large positive shock occuring at time τ
  - If $t \leq \tau$ the weight of that large shock increases exponentially as t approaches τ
  - At $t = \tau + 1$ it looks like a brutal bubble burst
- For instance, consider a non-causal AR(1) with $|\varphi| = |\breve{\varphi}^{-1}| < 1$

$$
\begin{aligned}
X_t &= \varepsilon_t + \varphi\varepsilon_{t+1} + \ldots + \varphi^{\tau-t-1}\varepsilon_{\tau-1} + \varphi^{\tau-t}\varepsilon_\tau, \quad t < \tau \\
X_{t+1} &= \varepsilon_{t+1} + \varphi\varepsilon_{t+2} + \ldots + \varphi^{\tau-t-1}\varepsilon_\tau + \varphi^{\tau-t}\varepsilon_{\tau+1}, \quad t+1 < \tau \\
&\vdots \\
X_\tau &= \varepsilon_\tau + \varphi\varepsilon_{\tau+1} + \ldots, \quad t = \tau \\
X_{\tau+1} &= \varepsilon_{\tau+1} + \varphi\varepsilon_{\tau+2} + \ldots, \quad t > \tau
\end{aligned}
$$

- In a mixed phase process, if $t > \tau$, the bubble burst is smooth and the weight of that large shock decreases exponentially

Identification



Conditional expectation

- Let X_{t+1} and $X_t = \mu + \rho(1)(X_{t-1} - \mu) + \varepsilon_t$ two stationary variables such that conditional distribution of X_{t+1} given that $X_t = x_t$ is
$$\mathcal{N}(\mu + \rho(1)(x_t - \mu), \sigma^2(1 - \rho(1)^2))$$
with $\rho(h) = \gamma(h)/\gamma(0)$ and $|\rho(1)| < 1$.

Proposition (1)

*For $g(.)$ a real function, the **best mean square** predictor of X_{t+1} is*

$$\arg\min_{g(X_t)} \mathbb{E}\left((X_{t+1} - g(X_t))^2\right) = \mu + \rho(1)(X_t - \mu) = \mathbb{E}(X_{t+1}|X_t)$$

Proposition (2)

*For $g(X_t) = aX_t + b$, the best mean square **linear** predictor of X_{t+1} is*

$$\arg\min_{g(X_t)} \mathbb{E}\left((X_{t+1} - g(X_t))^2\right) = \mu + \rho(1)(X_t - \mu) = \mathbb{EL}(X_{t+1}|X_t)$$

$\Rightarrow$ In the Gaussian case, the **best mean square** predictor is **linear**

Best linear predictor and errors

- In Propositions (1) and (2), $X_t = \mu + \rho(1)(X_{t-1} - \mu) + \varepsilon_t$
- It follows that the best linear predictor of X_{t+1} is

$$\mathbb{EL}(X_{t+1}|X_t) = \mu + \rho(1)(X_t - \mu)$$

... leading to the following prediction error

$$\varepsilon_{t+1} = X_{t+1} - \mu + \rho(1)(X_t - \mu)$$

- The predictor $\mathbb{EL}(X_{t+1}|X_t)$ will be uniquely determined if

$$\mathbb{E}(\text{Error} \times \text{Predictor Variable}) = 0$$

$\Rightarrow$ As ε_{t+1} is a SWN, this condition is fulfilled

$$\mathbb{E}(\varepsilon_{t+1} X_t) = 0$$

Note $\mathbb{EL} \equiv$ best approximation of expectation as a linear function

Non-Causal AR and causal representation

$X_t = \varphi_1 X_{t+1} + \varepsilon_t$ or $X_t = \check{\varphi}_1 X_{t-1} + \check{\varepsilon}_t$ leads to 3 types of errors (noises)

1. $\varepsilon_t = X_t - \varphi_1 X_{t+1}$
   - As $\varepsilon_{t-1} = -\varphi_1 \check{\varepsilon}_t \not\perp X_{t-1}$, ε_t is not the innovation of X_t and
   $$\varepsilon_t \not\perp X_t - \mathbb{EL}(X_t | X_s, s < t)$$
2. $\check{\varepsilon}_t = X_t - \check{\varphi}_1 X_{t-1}$ is just an equivalent representation X_t
   - As $-\varphi_1 \check{\varepsilon}_t \not\perp X_{t-1}$, $\check{\varepsilon}_t$ is not the innovation of X_t and
   $$\check{\varepsilon}_t \not\perp X_t - \mathbb{EL}(X_t | X_s, s < t)$$

Note The best predictor in mean square sense, $\mathbb{E}(X_t | X_s, s < t)$, will be nonlinear if $\check{\varepsilon}_t$ is non-Gaussian with finite variance (Rosenblatt, 2000, p. 101)

3. $\tilde{\varepsilon}_t = X_t - \varphi_1 X_{t-1}$ is a **weak causal representation** of X_t
   - $\tilde{\varepsilon}_t \sim WWN$ (except when ε_t is Gaussian, $\tilde{\varepsilon}_t \sim SWN$)
   - Moreover, $\tilde{\varepsilon}_t$ is the linear innovation of X_t because one can show that
   $$\tilde{\varepsilon}_t = X_t - \mathbb{EL}(X_t | X_s, s < t)$$

Non-Causal ARMA and causal representation

Definition (20)

Let X_t be a stationary ARMA(p, q) defined as $\breve{\varphi}(B)X_t = \breve{\theta}(B)\breve{\varepsilon}_t$. Then, it is always possible to find polynomials $\varphi(B)$ and $\theta(B)$ and a weak white noise sequence $\tilde{\varepsilon}_t$ such that $\varphi(B)X_t = \theta(B)\tilde{\varepsilon}_t$.

- Accordingly, any noncausal ARMA has a **weak causal representation**

Remark $\tilde{\varepsilon}_t$ will **not** be a Strong White Noise unless $\breve{\varepsilon}_t$ is Gaussian

- In the Gaussian case, all representations are equivalent (ε_t, $\breve{\varepsilon}_t$, and $\tilde{\varepsilon}_t$)

$\Rightarrow$ Causal and Non-Causal ARMA processes are **indistinguishable**

Non-uniqueness of Gaussian MA processes

- To understand the identification issue in the MA representation

  ... consider the following Gaussian MA(1) process

$$X_t = \check{\varepsilon}_t + \check{\theta}\check{\varepsilon}_{t-1}, \quad \check{\varepsilon}_t \sim \mathcal{N}(0, \sigma^2_{\check{\varepsilon}})$$

$\Rightarrow$ If $\theta \neq 0$ we easily see that **several representations coexist**

$$X_t = \varepsilon_t + \frac{1}{\check{\theta}}\varepsilon_{t-1}, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2_{\check{\varepsilon}}\check{\theta}^2)$$

where both $\check{\varepsilon}_t$ and ε_t are Strong White Noises

$\Rightarrow$ Both MA(∞) representations have the same probability structure

$$X_t = \sum_{j=-\infty}^{\infty} \check{\pi}_j \check{\varepsilon}_{t-j} = \sum_{j=-\infty}^{\infty} \pi_j \varepsilon_{t-j}$$

because $\check{\varepsilon}_t$ and ε_t are two $i.i.d.$ Gaussian random variables

Remark We have a **two-sided** MA(∞) because we do not say which representation is invertible or not

Non-uniqueness of Gaussian MA processes

Proof We use the autocovariance functions that summarize all the probability structure in the Gaussian case:

$$\begin{aligned}\gamma_{X,\breve{\varepsilon}}(h) &= Cov(X_{t+h}, X_t) = Cov(\breve{\varepsilon}_{t+h} + \breve{\theta}\breve{\varepsilon}_{t+h-1}, \breve{\varepsilon}_t + \breve{\theta}\breve{\varepsilon}_{t-1}) \\ &= \gamma_{\breve{\varepsilon}}(h) + \breve{\theta}\gamma_{\breve{\varepsilon}}(h+1) + \breve{\theta}\gamma_{\breve{\varepsilon}}(h-1) + \breve{\theta}^2\gamma_{\breve{\varepsilon}}(h) \\ &= \sigma^2_{\breve{\varepsilon}}(1+\breve{\theta}^2)\mathbb{1}_{h=0} + \sigma^2_{\breve{\varepsilon}}\breve{\theta}\mathbb{1}_{|h|=1}\end{aligned}$$

$$\begin{aligned}\gamma_{X,\varepsilon}(h) &= Cov(X_{t+h}, X_t) = Cov(\varepsilon_{t+h} + \breve{\theta}^{-1}\varepsilon_{t+h-1}, \varepsilon_t + \breve{\theta}^{-1}\varepsilon_{t-1}) \\ &= \gamma_{\varepsilon}(h) + \breve{\theta}^{-1}\gamma_{\varepsilon}(h+1) + \breve{\theta}^{-1}\gamma_{\varepsilon}(h-1) + \breve{\theta}^{-2}\gamma_{\varepsilon}(h) \\ &= \sigma^2_{\breve{\varepsilon}}\breve{\theta}^2(1+\breve{\theta}^{-2})\mathbb{1}_{h=0} + \sigma^2_{\breve{\varepsilon}}\breve{\theta}^2\breve{\theta}^{-1}\mathbb{1}_{|h|=1} \\ &= \sigma^2_{\breve{\varepsilon}}(1+\breve{\theta}^2)\mathbb{1}_{h=0} + \sigma^2_{\breve{\varepsilon}}\breve{\theta}\mathbb{1}_{|h|=1}\end{aligned}$$

with $\mathbb{1}_{(.)}$ the indicatrice function, $\breve{\varepsilon}_t \sim \mathcal{N}(0, \sigma^2_{\breve{\varepsilon}})$ and $\varepsilon_t \sim \mathcal{N}(0, \sigma^2_{\breve{\varepsilon}}\breve{\theta}^2)$

- Hence, $\gamma_{X,\breve{\varepsilon}}(h) = \gamma_{X,\varepsilon}(h)$ and given the Gaussian nature of $\breve{\varepsilon}_t$ and ε_t, the two representations of X_t have a identical probability structure

Non-uniqueness of Gaussian ARMA processes

- Consider the non-causal Gaussian ARMA$(1,1)$

$$X_t - \check{\varphi}_1 X_{t-1} = \check{\varepsilon}_t + \check{\theta}_1 \check{\varepsilon}_{t-1} := \check{\varphi}(B) X_t = \check{\theta}(B)\check{\varepsilon}_t, \quad |\check{\varphi}_1| > 1, \quad |\check{\theta}_1| > 1$$

- The spectral density of X_t is

$$f_X(\lambda) = \left| \frac{\check{\theta}(e^{-i\lambda})}{\check{\varphi}(e^{-i\lambda})} \right|^2 f_{\check{\varepsilon}}(\lambda) = \left| \frac{\check{\theta}(e^{-i\lambda})}{\check{\varphi}(e^{-i\lambda})} \right|^2 \frac{\sigma^2_{\check{\varepsilon}}}{2\pi}$$

- Use $\varphi(B) = 1 - \frac{1}{\check{\varphi}_1} B$ and $\theta(B) = 1 - \frac{1}{\check{\theta}_1} B$ in $\varepsilon_t = \theta(B)^{-1}\varphi(B) X_t$
- Then, the **spectral density** of ε_t is

$$f_\varepsilon(\lambda) = \left| \frac{1 - \check{\theta}_1^{-1} e^{-i\lambda}}{1 - \check{\varphi}_1^{-1} e^{-i\lambda}} \right|^2 \left| \frac{1 - \check{\theta}_1 e^{-i\lambda}}{1 - \check{\varphi}_1 e^{-i\lambda}} \right|^2 \frac{\sigma^2_{\check{\varepsilon}}}{2\pi} = \frac{\check{\theta}_1^2}{\check{\varphi}_1^2} \frac{\sigma^2_{\check{\varepsilon}}}{2\pi} = \frac{\varphi_1^2 \theta_1^{-2} \sigma^2_{\check{\varepsilon}}}{2\pi} = \frac{\sigma^2_\varepsilon}{2\pi},$$

  constant and hence ε_t is also a Gaussian (strong) white noise

$\Rightarrow$ The **probability structure** is the same for

$$\varphi(B) X_t = \theta(B)\varepsilon_t \text{ and } \check{\varphi}(B) X_t = \check{\theta}(B)\check{\varepsilon}_t$$

and their two MA(∞) representations coexist

Uniqueness of linear process representation

Definition (21)

Uniqueness means that the only way for X_t, a linear process, to admit two MA(∞) representations is

the existence of a constant scaling factor $c \in \mathbb{R}$

and a shift in time $l \in \mathbb{Z}$

such that

$$X_t = \sum_{j=-\infty}^{\infty} \check{\psi}_j \check{\varepsilon}_{t-j} = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$$

with

$$\varepsilon_t = c\check{\varepsilon}_{t-l} \text{ and } \psi_j = \frac{1}{c}\check{\psi}_{j+l}$$

- Uniqueness is not found for Gaussian processes... is it similar for Non-Gaussian processes ?

$\Rightarrow$ For finite variance non-Gaussian noises, uniqueness has been demonstrated (Rosenblatt, 2000, Th. 1.3.1)

$\Rightarrow$ For infinite variance noises, we will detail some results

α-stable MAR

How fat should we go

- Non-Gaussian distributions are crucial to identify non-causality
- As non-causality is appropriate to model explosive (bubble) behavior, heavy-tail distributions are of interest
- The literature has primarily focused on the Student's t distribution

$$f_\varepsilon(\sigma,\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)\sqrt{\pi\nu}\sigma}\left(1+\frac{1}{\nu}\left(\frac{\varepsilon}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}$$

with $\sigma > 0$ and $\nu > 0$ but sufficiently small to depart from the Gaussian distribution

$\Rightarrow$ Hecq, Lieb and Teg (2016) extensively discuss this approach and investigate the question "How fat should we go" in terms of heavy tails

Remark The Student's t distribution is symmetric and therefore not necessarily appropriate to model some financial or macro phenomena (typically, bubbles are generally positive, right tail events)

Introduction to Stable distributions

- If the tails of the Student's t distribution are not sufficiently heavy

... the family of Stable laws offers an attractive alternative

Definition (22)

*A real probability measure μ is said to be α-**stable** if $\forall\ k \in \mathbb{N}\ \exists b^k > 0$ such that*

$$X_1 + \ldots + X_k \stackrel{d}{=} b_k X + e_k$$

where $\mathcal{L}(X_1) = \ldots = \mathcal{L}(X_k) = \mathcal{L}(X) = \mu$ and $X_1 + \ldots + X_k$ are independent and where $\exists \alpha \in (0, 2]$ such that $b^k = k^{1/\alpha}$

Some remarks on Stable distributions

- If ε_t is a random variable that follows a stable distribution, then

$$\varepsilon_t \sim \mathcal{S}(\alpha, \beta, \sigma, m)$$

where
- — the stability parameter $\alpha \in (0, 2]$ is also a tail index
- — $\beta \in [-1, 1]$ is the asymmetry parameter
- — $\sigma \in (0, \infty)$ is the scale parameter
- — $m \in \mathbb{R}$ is the location parameter (if $m = 0$, ε_t is strictly stable)

- If $z_t \sim \mathcal{S}(\alpha, \beta, 1, 0)$ then, $\varepsilon_t = z_t\sigma + m \sim \mathcal{S}(\alpha, \beta, \sigma, m)$

Normal If $\varepsilon_t \sim \mathcal{S}(2, \beta, \sigma, m)$ then, $f_\varepsilon(2, \beta, 1, 0) = (\sqrt{2\pi})^{-1}\exp(-\varepsilon^2/2)$

Cauchy If $\varepsilon_t \sim \mathcal{S}(1, 0, \sigma, m)$ then, $f_\varepsilon(1, 0, 1, 0) = \left(\pi(1+\varepsilon^2)\right)^{-1}$

Lévy If $\varepsilon_t \sim \mathcal{S}(1/2, 1, \sigma, m)$ then,

$$f_\varepsilon(1/2, 1, 1, 0) = \frac{1}{\sqrt{2\pi}\varepsilon^{3/2}}\exp\left(\frac{-1}{2\varepsilon}\right)\mathbb{1}_{\varepsilon>0}$$

with $f_\varepsilon(.)$ the density function

Analytic representation of Stable distributions

- In most cases, neither the probability density, $f_\varepsilon(.)$, nor the cumulative distribution, $F_\varepsilon(.)$, functions are analytically expressible
- Fortunately, the characteristic function $\varphi_\varepsilon(.) = \mathbb{E}(e^{-iu\varepsilon})$ has a tractable expression for $\alpha \neq 1$

$$\log \varphi_\varepsilon(u) = -\sigma^\alpha |u|^\alpha \left(1 - i\beta(\operatorname{sign} u) \tan\left(\frac{\pi\alpha}{2}\right)\right) + imu$$

Remark If $f_\varepsilon(.)$ exists, $\varphi_\varepsilon(.) = \mathbb{E}(e^{-iu\varepsilon})$ is also the Fourier transform of $f_\varepsilon(.)$ and one can see that it completely defines its p.d.f.

Remark When $\alpha = 1$, $\log \varphi_\varepsilon(.)$ simplifies to

$$\log \varphi_\varepsilon(u) = -\sigma|u|\left(1 + i\beta(\operatorname{sign} u)\frac{2}{\pi} \log |u|\right) + imu$$

Numerical evaluation of Stable distributions

- Using the inverse Fourier transform of $\varphi_\varepsilon(.)$, we obtain

$$f_\varepsilon(\alpha,\beta,\sigma,m) = (2\pi)^{-1}\int_{-\infty}^{\infty} \exp\Big(-is(\varepsilon-m)\Big)\varphi_\varepsilon(\sigma s)ds$$

- $f_\varepsilon(.)$ has no closed form solution but a numerical evaluation is possible

$\Rightarrow$ Setting $\sigma = 1$ and $m = 0$ we have

$$f_\varepsilon(\alpha,\beta,1,0) = \pi^{-1}\int_{0}^{\infty} \exp(-s^\alpha)\cos\left(s\varepsilon + \beta\tan\left(\frac{\pi\alpha}{2}\right)(s-s^\alpha)\right)ds$$

when $\alpha \neq 1$ and

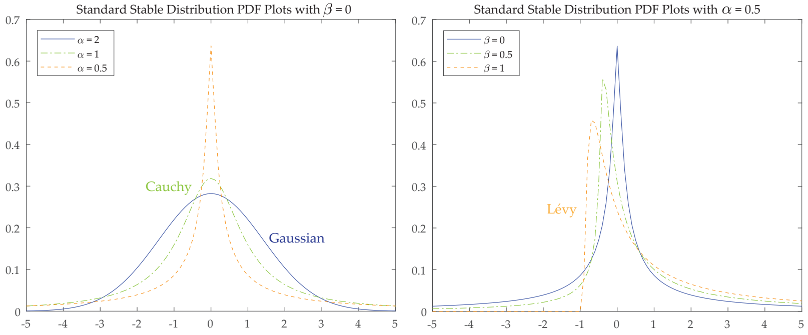
$$f_\varepsilon(\alpha,\beta,1,0) = \pi^{-1}\int_{0}^{\infty} \exp(-s)\cos\left(s\varepsilon + s\beta\frac{2}{\pi}\log s\right)ds$$

when $\alpha = 1$

$\Rightarrow$ Numerical evaluation of these integrals is possible in R and MATLAB

Remark For any $\sigma > 0$ and m we have $f_\varepsilon(\alpha,\beta,\sigma,m) = \sigma^{-1}f_{\sigma^{-1}(\varepsilon-m)}(\alpha,\beta,1,0)$

Numerical evaluation of Stable Laws probability density function



- The tail index $\alpha \in (0, 2)$ is such that for $c_\alpha > 0$ and as $x \to \infty$,

$$\mathbb{P}(X < -x) \sim c_\alpha (1 - \beta) x^{-\alpha} \text{ and } \mathbb{P}(X > x) \sim c_\alpha (1 + \beta) x^{-\alpha}$$

Remark If $X \sim \mathcal{S}(\alpha, \beta, \sigma, m)$ with $\alpha \in (0, 2)$, $\mathbb{E}|X|^s < \infty$ **if and only if** $s < \alpha$

Domains of attraction of α-stable distributions

Definition (23)

For μ an α-stable distribution and ε_t an i.i.d. sequence, we say that $\mathcal{L}(\varepsilon_1)$ belongs to the domain of attraction of μ, also denoted $\mathcal{L}(\varepsilon_1) \in \mathbb{D}_\alpha(\mu)$ if for some $x > 0$

$$\mathbb{P}(\varepsilon_1 > x) \sim x^{-\alpha} L(x)$$

with $L(x)$ is a slowly varying function at infinity and $\alpha \in (0, 2)$ and if there exists a constant $c \in [0, 1]$ such that

$$\lim_{x \to \infty} \frac{\mathbb{P}(\varepsilon_1 > x)}{\mathbb{P}(|\varepsilon_1| > x)} = c$$

Corollary A necessary and sufficient condition for $\mathcal{L}(\varepsilon_t)$ to belong to $\mathbb{D}_\alpha(\mu)$ is

$$\log |\varphi_\varepsilon(s)| \sim -\sigma^\alpha |s|^\alpha L(1/|s|) \text{ as } s \to 0$$

where $\varphi_\varepsilon(s)$ is the characteristic function of ε_t

Absolute convergence of infinite variance sequences

Proposition (3)

Let ε_t be an i.i.d. sequence such that $\mathcal{L}(\varepsilon_1) \in \mathbb{D}_\alpha(\mu)$. If ψ_j is a sequence of constants such that

$$\sum_{j=-\infty}^{\infty} |\psi_j|^s < \infty \text{ for some } s \in (0, \alpha) \cap [0, 1]$$

then the infinite series,

$$\sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$$

converges absolutely with probability one.

- This proposition due to Cline (1983) establishes **strict stationary** conditions for any MA(∞) linear sequence with infinite variance

α-stable MA(∞) process

Proposition (4)

Let ε_t and $\breve{\varepsilon}_t$ be two i.i.d. processes such that

$$\varepsilon_t = \sum_{j=-\infty}^{\infty} \breve{\psi}_j \breve{\varepsilon}_{t-j}$$

where $\mathcal{L}(\varepsilon_1) \in \mathbb{D}_\alpha(\mu)$ with $\alpha \in (0,2)$. Then, if

$$\sum_{j=-\infty}^{\infty} |\breve{\psi}_j|^s < \infty \textit{ for } s \in (0,\alpha) \cap [0,1]$$

the MA representation is trivial (all except one MA coefficients $\breve{\psi}_j$ have to be null)

- Gouriéroux and Zakoian (2015) prove Propositions (4) and (5)

$\Rightarrow$ An α-stable MA(∞) process cannot be i.i.d.

Remark This result is very important in view of proving the uniqueness of the strong MA representation under of stable laws with $\alpha \in (0,2)$

Uniqueness of α-stable strong MA(∞) representation

Proposition (5)

Let ε_t and $\breve{\varepsilon}_t$ be two i.i.d. processes such that

$$\sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} = \sum_{j=-\infty}^{\infty} \breve{\psi}_j \breve{\varepsilon}_{t-j}$$

where $\mathcal{L}(\varepsilon_1) \in \mathbb{D}_\alpha(\mu)$ with $\alpha \in (0,2)$ and

$$\sum_{j=-\infty}^{\infty} |\psi_j|^s < \infty \text{ for } s \in (0,\alpha) \cap [0,1]$$

Now suppose that $\breve{\psi}(B)$ is invertible with

$$\breve{\psi}(B)^{-1} = \sum_{j=-\infty}^{\infty} \tilde{\psi}_j B^j \text{ such that } \sum_{j=-\infty}^{\infty} |\tilde{\psi}_j|^s < \infty$$

Then, for some constants $c \in \mathbb{R}$ and $l \in \mathbb{Z}$, $\varepsilon_t = c\breve{\varepsilon}_{t-l}$ and $\psi_j = \frac{1}{c}\breve{\psi}_{j+l}$

Stable Noncausal AR(1) and unconditional stationary distribution

- Let X_t be a **stable** strong noncausal AR(1)

$$X_t = \varphi X_{t+1} + \varepsilon_t, \quad |\varphi| < 1, \quad \varepsilon_t \sim \mathcal{S}(\alpha, \beta, \sigma, 0), \quad \alpha \in (0, 2)$$

- We know that X_t has the following strictly stationary solution

$$X_t = \sum_{j=0}^{\infty} \varphi^j \varepsilon_{t+j}$$

and we can now compute the unconditional **stable** distribution of X_t

$$X_t \sim \mathcal{S}\left(\alpha, \beta, \frac{\sigma}{(1-|\varphi|^\alpha)^{1/\alpha}}, 0\right), \text{ if } \alpha \neq 1 \text{ and } \varphi \geq 0$$

$$X_t \sim \mathcal{S}\left(\alpha, \beta\frac{1-|\varphi|^\alpha}{1+|\varphi|^\alpha}, \frac{\sigma}{(1-|\varphi|^\alpha)^{1/\alpha}}, 0\right), \text{ if } \alpha \neq 1 \text{ and } \varphi \leq 0$$

$$X_t \sim \mathcal{S}\left(1, \beta\frac{1-|\varphi|}{1+|\varphi|}, \frac{\sigma}{1-|\varphi|}, -\beta\sigma\frac{2}{\pi}\frac{\varphi \log|\varphi|}{(1-\varphi)^2}\right), \text{ if } \alpha = 1$$

and state that $\mathbb{E}(|X_t|^u) < \infty$ if and only if $u < \alpha$

Remark When $\varphi \leq 0$, X_t is **less asymmetric** than ε_t because φ affects β

Stable Noncausal AR(1) and conditional moments

- Let X_t be a strong **stable** noncausal AR(1)

$$X_t = \varphi X_{t+1} + \varepsilon_t, \quad |\varphi| < 1, \quad \varepsilon_t \sim \mathcal{S}(\alpha, \beta, \sigma, 0), \quad \alpha \in (0, 2)$$

- Gouriéroux & Zakoian (2017): for the backward conditional density

$$\mathbb{E}(|X_t|^b | X_{t+1}) < \infty \text{ iff } b < \alpha$$

- Gouriéroux & Zakoian (2017): for the forward conditional density

$$\mathbb{E}(|X_{t+h}|^c | X_{t-1}) < \infty, \ a.s., \text{ iff } c < 2\alpha + 1, \quad \alpha \in (0, 2),$$

  for any $h \geq 0$ and $\beta \neq 1$, or $|\beta| = 1$ if $\varphi^{h+1} < 0$. If $\beta \neq 1$ and $\varphi^{h+1} > 0$

$$\mathbb{E}(|X_{t+h}|^c | X_{t-1}) < \infty \ a.s. \ \forall c > -1$$

- X_t is also a causal homogeneous Markov process as

$$\mathcal{L}(X_t | X_{t-1}, X_{t-2}, \ldots) = \mathcal{L}(X_t | X_{t-1})$$

Remark The number of finite forward conditional moments is $c > (b = u)$

Stable Noncausal AR(1) and forward conditional expectation

- The forward (causal) conditional expectation always exists...

... but the unconditional and backward (noncausal) conditional expectations exist only if $\alpha > 1$

- Gouriéroux & Zakoian (2017) : if $\beta = 0$ (symmetric stable laws : SαS)

$$\mathbb{E}(X_{t+h}|X_{t-1}) = |\varphi|^{(h+1)(\alpha-1)}X_{t-1}, \quad \forall h \geq 0, \quad \alpha \in (0,2),$$

with $|\varphi|^0 = \mathrm{sign}(\varphi)$ so that in the Cauchy case ($\alpha = 1$)

$$\mathbb{E}(X_{t+h}|X_{t-1}) = \mathrm{sign}(\varphi)X_{t-1}, \quad \forall h \geq 0$$

and when $\varphi > 0$, X_t behaves as a **stationary martingale !**

Remark when $\varphi > 0$ and $\alpha \in (0,1)$, X_t is a **stationary submartingale**

- Fries & Zakoian (2019) : if $\alpha \in (0,2)$ and $\beta \in (-1,1)$

$$\mathbb{E}(X_{t+h}|X_{t-1}) = |\varphi|^{(h+1)(\alpha-1)}X_{t-1} - \mathbb{1}_{\alpha=1}(h+1)\frac{2}{\pi}\beta\sigma\frac{\varphi\log\varphi}{1-\varphi}, \quad \forall h \geq 0$$

Cauchy Noncausal AR(1) and forward conditional moments

- Let X_t be a strong **Cauchy** noncausal AR(1)
- Gouriéroux & Zakoian (2017) : the causal predictive density is

$$f_X(X_t|X_{t-h}) = \frac{1}{\sigma_h \pi} \frac{\sigma_h^{-2}}{(X_{t-h} - \varphi^h X_t)^2} \frac{\sigma^2 + (1-|\varphi|)^2 X_{t-h}^2}{\sigma^2 + (1-|\varphi|)^2 X_t^2}, \sigma_h = \frac{1-|\varphi|^h}{1-|\varphi|}$$

$\Rightarrow$ the second order conditional moment is heteroscedastic

$$\mathbb{E}(X_t^2|X_{t-1}) = \frac{1}{|\varphi|} X_{t-1}^2 + \frac{\sigma^2}{|\varphi|(1-|\varphi|)}$$

$\Rightarrow$ X_t admits a **semi-strong** representation (à la Drost & Nijman, 1993)

$$X_t = \mathrm{sign}(\varphi) X_{t-1} + \tilde{\varepsilon}_t \text{ where } \tilde{\varepsilon}_t = \varsigma_t \tilde{\eta}_t$$

$$\varsigma_t^2 = (|\varphi|^{-1} - 1) X_{t-1}^2 + \frac{\sigma^2}{|\varphi|(1-|\varphi|)}$$

where $\mathbb{E}(\tilde{\eta}_t|X_{t-1}) = 0$, $\mathbb{E}(\tilde{\eta}_t^2|X_{t-1}) = 1$ and $\tilde{\eta}_t$ is a Weak White Noise

Remark $X_t =$ unit-root (if $\varphi > 0$) + ARCH (based on X_{t-1}^2 rather than $\tilde{\varepsilon}_{t-1}^2$)

- Fries & Zakoian (2019) extend these results to the Cauchy MAR(p, q)

Strong causal representation for the Cauchy Noncausal AR(1)

- As mentioned in (20), a **weak causal** representation exists...

... but, it is also possible to derive a **strong causal** representation

- Use some Gaussian innovations given by $\varepsilon_t^* = \Phi^{-1}\big(F(X_t|X_{t-1})\big)$ with

... $\Phi^{-1}(.)$ the Normal c.d.f. and $F(X_t|X_{t-1})$ the conditional c.d.f. of X_t :

$$F(X_t|X_{t-1}) = \frac{\Lambda}{\pi}\log\left(\frac{1+(1-|\varphi|)^2X_t^2}{1+(X_{t-1}-\varphi X_t)^2}\,\frac{\breve{\varphi}^2}{(1-|\varphi|)^2}\right)$$
$$+\frac{\Upsilon}{\pi}\left(\frac{\pi}{2}-\operatorname{sign}(\varphi)\tan^{-1}(X_{t-1}-\varphi X_t)\right)+\frac{1-\Upsilon}{\pi}\left(\tan^{-1}((1-|\varphi|)X_t)+\frac{\pi}{2}\right)$$
$$\Lambda = \frac{\varphi(1-|\varphi|)^2X_{t-1}}{(1-2|\varphi|)^2+(1-|\varphi|)^2X_{t-1}^2} \text{ and } \Upsilon = \frac{|\varphi|\big((1-|\varphi|)^2X_{t-1}^2-(1-2|\varphi|)\big)}{(1-2|\varphi|)^2+(1-|\varphi|)^2X_{t-1}^2}$$

- Inverting the relation $\varepsilon_t^* = \Phi^{-1}\big(F(X_t|X_{t-1})\big)$, one can derive

$$X_t = G(X_{t-1},\varepsilon_t^*), \quad G(X_{t-1},.) = F^{-1}\big(\Phi(.)|X_{t-1}\big), \quad \varepsilon_t^* \sim \mathcal{N}(0,1)$$

$\Rightarrow$ the **strong causal** AR(1) representation is highly **nonlinear**

The Lévy Noncausal AR(1)

- Let X_t be a strong **Lévy** noncausal AR(1)

$$X_t = \varphi X_{t+1} + \varepsilon_t, \quad 0 < \varphi < 1, \quad \varepsilon_t \sim \mathcal{S}(1/2, 1, \sigma, 0), \quad \alpha \in (0,2)$$

- As for the Cauchy case, Gouriéroux & Zakoian (2017) prove that
  - X_t has a causal predictive density given by

$$f_X(X_t|X_{t-1}) = \frac{1}{\sqrt{2\pi}} \left(\frac{X_{t-1}}{X_t(X_{t-1} - \varphi X_t)} \right)^{3/2}$$
$$\times \exp\left(\frac{-(X_{t-1} - \sqrt{\varphi} X_t)}{2X_{t-1}X_t(1-\sqrt{\varphi})^2(X_{t-1} - \varphi X_t)} \right) \mathbb{1}_{0<\varphi X_t < X_{t-1}}$$

$\Rightarrow$ All the causal forward conditional moments exist whereas even the unconditional first moment does not

  - X_t admits an ARCH type (probably) **semi-strong** representation

$$X_t = \varphi^{-1/2} X_{t-1} + \tilde{\varepsilon}_t, \quad \mathbb{E}(\tilde{\varepsilon}_t | X_{t-1}) = 0$$

albeit the form of $\tilde{\varepsilon}_t$ is tedious to derive

  - The strong causal (non-linear) representation X_t is untracktable

What we known on the Stable Noncausal AR(1): summary

Law	Representation	Equation	Properties
Cauchy	Strong noncausal	$X_t = \varphi X_{t+1} + \varepsilon_t$	$\varepsilon_t/\sigma \sim i.i.\mathcal{C}(0,1)$
$\lvert\varphi\rvert < 1$	Weak causal linear	$X_t = \varphi X_{t-1} + \tilde{\varepsilon}_t$	$\tilde{\varepsilon}_t \sim WWN(0,\sigma^2)$
	Semi-strong causal linear	$X_t = \mathrm{sign}(\varphi) X_{t-1} + \varsigma_t \tilde{\eta}_t$	$\tilde{\eta}_t \sim WWN$
	Strong causal nonlinear	$X_t = G(X_{t+1}, \varepsilon_t^*)$	$\varepsilon_t^* \sim \mathcal{N}(0,1)$
Lévy	Strong noncausal	$X_t = \varphi X_{t+1} + \varepsilon_t$	$\varepsilon_t \sim i.i.\mathcal{L}\left(0, \frac{\sigma}{(1-\sqrt{\varphi})^2}\right)$
$0 < \varphi < 1$	Weak causal linear	$X_t = \varphi X_{t-1} + \tilde{\varepsilon}_t$	$\varepsilon_t \sim WWN(0,\sigma^2)$
	Semi-strong causal linear	$X_t = \varphi^{-1/2} X_{t-1} + \tilde{\nu}_t$	$\tilde{\nu}_t \sim WWN$
	Strong causal nonlinear	no closed form expression	
Stable	Strong noncausal	$X_t = \varphi X_{t+1} + \varepsilon_t$	$\varepsilon_t \sim i.i.d(0,\sigma^2)$
$\lvert\varphi\rvert < 1$	Weak causal linear	$X_t = \varphi X_{t-1} + \tilde{\varepsilon}_t$	$\tilde{\varepsilon}_t \sim WWN(0,\sigma^2)$
	Semi-strong causal linear	no closed form expression	
	Strong causal nonlinear	no closed form expression	

α-stable MAR

1. The linear model
2. Mixed-Phase processes
3. Identification
4. α-stable MAR
5. Prediction
6. Simulation of MAR
7. Estimation of MAR
8. Applications

Prediction of linear causal processes with infinite variance

- The result in P(2) can be generalized to the non-Gaussian L^2 framework

$$\arg\min_{\mathcal{P}(X_{t+h})} \mathbb{E}\Big((X_{t+h} - \mathcal{P}(X_{t+h}))^2 \Big) = \mathbb{EL}(X_{t+h}|\mathscr{F}_{t-1})$$

  where $\mathcal{P}(X_{t+h})$ is a predictor of X_{t+h} and $\mathscr{F}_t = X_t, \dots, X_1$

- Cline & Brockwell (1985) deal with the causal symmetric α-stable case

$\Rightarrow$ the best predictor cannot be defined in a mean square sense... however

$$\arg\min_{\mathcal{P}(X_{t+h})} \mathbb{E}\Big(|X_{t+h} - \mathcal{P}(X_{t+h})|^\alpha \Big) = \mathbb{PL}(X_{t+h}|\mathscr{F}_{t-1})$$

  where $\mathbb{PL}(.)$ is a linear projection defined by the **minimum dispersion**

- If $\alpha \in (1, 2]$, the $\mathbb{PL}(.)$ is unique and $\mathbb{EL}(X_{t+h}|\mathscr{F}_{t-1}) = \mathbb{PL}(X_{t+h}|\mathscr{F}_{t-1})$
- If $\alpha \in (0, 1]$, the $\mathbb{PL}(.)$ is not unique

Prediction of linear stable non-causal processes

- In the non-causal L^2 framework the optimal predictor is more tricky

$\Rightarrow$ $\mathbb{E}(X_{n+h}|\mathscr{F}_{t-1})$ is generally a non-linear function of the observed values

- We know that $\mathbb{E}(|X_t|^b|X_{t+1}) < \infty$ only for $b < \alpha$

$\Rightarrow$ fortunately, $\mathbb{E}(|X_{t+h}|^c|\mathscr{F}_{t-1}) < \infty$ for $c < 2\alpha + 1$ so that

$$\mathbb{E}(X_{t+h}|\mathscr{F}_{t-1}) < \infty \quad \forall \alpha \in (0, 2)$$

- For some specific non-causal models we have a closed form expression :
  - strong non-causal AR(1) with stable errors for $\forall h \geq 0$:

$$\mathbb{E}(X_{t+h}|X_{t-1}) = |\varphi|^{(h+1)(\alpha-1)}X_{t-1} - \mathbb{1}_{\alpha=1}(h+1)\frac{2}{\pi}\beta\sigma\frac{\log\varphi}{1-\varphi}$$

  - strong SαS mixed-causal MAR$(1, 1)$ for $\forall h \geq 0$:

$$\mathbb{E}(X_{t+h}|\mathscr{F}_{t-1}) = (\varphi^{\bullet})^{h+1}\left(X_{t-1} + (X_{t-1} - \varphi^{\bullet}X_{t-2})\sum_{j=1}^{h+1}(\varphi^{<\alpha-1>}\varphi^{\bullet})^{-j}\right)$$

where $\varphi^{<x>} = \mathrm{sign}(\varphi^{\circ})|\varphi^{\circ}|^x$ (see Fries & Zakoian, 2019 for the MAR$(1, q)$)

Bubble prediction in stable non-causal processes

- Consider $X_t = \varphi X_{t+1} + \varepsilon_t$, $\varepsilon_t \sim \mathcal{S}(\alpha, 0, \sigma, 0)$ and $\varphi \in (0, 1)$

$\Rightarrow$ multiple bubbles are likely to occur for large values of ε_t

Remark $\mathbb{E}(X_{t+h}|X_{t-1}) = |\varphi|^{(h+1)(\alpha-1)}X_{t-1}$ always predicts an exponential decay to the central values at rate $\varphi^{\alpha-1}$

$\Rightarrow$ $\mathbb{E}(X_{t+h}|X_{t-1})$ describes paths that depart from the realized trajectory

- If the predictive density is known, once estimated, one can compute

$$\mathbb{P}(X_{t+h} - \varphi X_{t+h+1} > c|X_t),$$

the probability of bubble collapse at $t + h$ for some critical level c

$\Rightarrow$ As the predictive density is generally unknown Gouriéroux & Jasiak (2016) suggest a non-parametric approach

$\Rightarrow$ Fries (2022) investigates higher moments through the joint density of (X_{t+h}, X_t)

Fries (2022) : α-stable random vectors and spectral measure

- Samorodnitsky and Taqqu (1991) show that SαS random vectors $\boldsymbol{X}$ are defined by a unique pair $(\boldsymbol{\Gamma}, \boldsymbol{\mu})$ such that $\forall \boldsymbol{u} \in \mathbb{R}^d$ and $s_1, \ldots, s_d \in \boldsymbol{s}$

$$\varphi_{\boldsymbol{X}}(\boldsymbol{u}) = \exp\left(-\int_{S_d} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^{\alpha} \Big(1 - i(\operatorname{sign}\langle \boldsymbol{u}, \boldsymbol{s}\rangle)\omega(\alpha, \langle \boldsymbol{u}, \boldsymbol{s}\rangle)\Big)\boldsymbol{\Gamma}(d\boldsymbol{s}) + i\langle \boldsymbol{s}, \boldsymbol{\mu}\rangle \right)$$

  is the characteristic function of $\boldsymbol{X}$ with $\langle ., . \rangle$ the scalar product,

$$\omega(\alpha, x) := \tan\left(\frac{\pi\alpha}{2}\right) \text{ if } \alpha \neq 1 \text{ and } \omega(1, x) := -\frac{2}{\pi}\log|x|$$

  and $\boldsymbol{\Gamma}$ a spectral measure on the unit sphere S_d and $\boldsymbol{\mu}^0 \in \mathbb{R}^d$

- $\boldsymbol{\Gamma}$:= scale, asymmetry and dependence between $X_{(1)}, \ldots, X_{(d)} \in \boldsymbol{X}$ and $\boldsymbol{\mu}$ is a non-random shift vector

Results For some $\nu \geq 0$, $\mathbb{E}\left(X_{(2)}^{\gamma} | X_{(1)}\right) < \infty$ if

$$\int_{S_2} |s_1|^{-\nu}\boldsymbol{\Gamma}(d\boldsymbol{s}) < \infty \text{ and } \gamma < \min(\alpha + \nu, 2\alpha + 1)$$

so that if $\alpha > 3/2$ and $\nu \geq 4 - \alpha$, the fourth moment exists

Fries (2022) : non-causal α-stable AR(1) process and spectral measure

- Fries (2022) shows that for X_t a non-causal α-stable AR(1) process...

... $\boldsymbol{X}_t = (X_t, X_{t+h})$ is α-stable and has a spectral measure

$$\boldsymbol{\Gamma}_h = \frac{\bar{\sigma}^\alpha}{2} \sum_{\vartheta \in S_1} \Bigg(\Big(1 - |\varphi|^{\alpha h} + 1 - (\varphi^{<\alpha>})^h \vartheta \bar{\beta} \Big) \delta(\vartheta, 0) + \Big(1 + |\varphi|^{2h} \Big)^{\alpha/2} (1 + \vartheta \bar{\beta}) \delta(\vartheta \boldsymbol{s}_h) \Bigg)$$

with $\delta(x)$ the Dirac measure at $x \in \mathbb{R}$, $S_1 = \{-1, +1\}$,

$$\bar{\sigma}^\alpha = \frac{\sigma^\alpha}{1 - |\varphi|^\alpha} \text{ and } \bar{\beta} = \beta \frac{1 - |\varphi|^\alpha}{1 - |\varphi|^{<\alpha>}}$$

and finally $\boldsymbol{s}_h = (\varphi^h, 1)/\sqrt{1 + |\varphi|^{2h}} \in S_2$

- The analytical form of $\boldsymbol{\Gamma}_h$ allows to compute $\mathbb{E}\Big(X_{t+h}^\gamma | X_h\Big)$

$\Rightarrow$ see Fries (2018) for the moments $\gamma = 1, \ldots, 4$

Fries (2022) : non-causal α-stable AR(1) process and bubble behavior

- Fries (2022) investigates how the standardized conditional moments

$$m_\gamma(x) = \mathbb{E}\left(X_{t+h}^{\gamma} | X_t = x\right)$$

behave during a bubble driven by large values (i.e. as $x \to \infty$) for $\varphi > 0$

$$\begin{aligned}
m_1(x) &\sim (\varphi^{-h}x)\varphi^{\alpha h}, && \text{if } \alpha \in (0,2) \\
m_2(x) &\sim (\varphi^{-h}x)\varphi^{\alpha h}(1-\varphi^{\alpha h}), && \text{if } \alpha \in (1/2,2) \\
m_3(x) &\to \frac{1-2\varphi^{\alpha h}}{\sqrt{\varphi^{\alpha h}(1-\varphi^{\alpha h})}}, && \text{if } \alpha \in (1,2) \\
m_4(x) &\to \frac{1-6\varphi^{\alpha h}}{\varphi^{\alpha h}}, && \text{if } \alpha \in (3/2,2)
\end{aligned}$$

- These are the moments of a **weighted Bernoulli distribution** !
  - charging probability $\varphi^{\alpha h}$ to the weight $\varphi^{-h}x$ ...
  - ... and probability $1-\varphi^{\alpha h}$ to the value 0

$\Rightarrow$ $\varphi^{\alpha h}$:= probability that the bubble survives at least h more time periods

Limitations of the Spherical Representation

- de Truchis, Fries & Thomas (2025) suggest to extend this approach by considering

$$\boldsymbol{X}_t = (X_{t-m}, \dots, X_t, X_{t+1}, \dots, X_{t+h})$$

  where X_t is defined as a general α-stable infinite moving average

- Samorodnitsky & Taqqu (1991) defines the spectral measure $\boldsymbol{\Gamma}$ on the **unit sphere** $\mathcal{S}_{m+1+h}$

$\Rightarrow$ This representation is based on a **norm** (e.g., Euclidean) of the full trajectory vector $\boldsymbol{X}_t$

- Problem for prediction !

$\Rightarrow$ the conditioning event, $||\boldsymbol{X}_t||_e > x$, **is not observable** at time t as it contains h future values

- de Truchis, Fries & Thomas (2025) propose a new representation theory to solve this

de Truchis, Fries & Thomas (2025): Seminorm Representation

- The key idea: replace the *norm* with a *seminorm* $||\cdot||$ that **only depends on the observed past**.
- The seminorm is defined to ignore all future components

$$||(x_{-m}, \dots, x_0, x_1, \dots, x_h)|| := ||(x_{-m}, \dots, x_0, 0, \dots, 0)||$$

$\Rightarrow$ the **unit sphere** S_{m+1+h} is forced to become a **unit cylinder** $C^{||\cdot||}_{m+1+h}$

- Fundamental result: not all processes admit such a representation as process is representable on this cylinder **if and only if** it is "anticipative enough"

$\Rightarrow$ **Purely causal processes** are **non-representable** in this framework.

$\Rightarrow$ This theory directly links **anticipativeness** to the **predictability** of extreme events.

de Truchis, Fries & Thomas (2025): Prediction Framework

- The prediction problem becomes
  - finding the future, path defined by a Borel set A on the cylinder, that matches a known pattern
  - given an extreme past trajectory, let say the Borel $V \subset S_{m+1}^{\|\cdot\|}$, that defines $B(V) = V \times \mathbb{R}^h$
- The limite behavior of $\boldsymbol{X}_t / \|\boldsymbol{X}_t\|$ reveals the discrete nature of the tail distribution

$$\lim_{x \to \infty} \mathbb{P}\left(\frac{\boldsymbol{X}_t}{\|\boldsymbol{X}_t\|} \in A \,\middle|\, \|\boldsymbol{X}_t\| > x,\ \frac{\boldsymbol{X}_t}{\|\boldsymbol{X}_t\|} \in B(V) \right) = \frac{\Gamma^{\|\cdot\|}(A \cap B(V))}{\Gamma^{\|\cdot\|}(B(V))}$$

where

$\|\boldsymbol{X}_t\| > x$ $\Rightarrow$ The **observed past** is extreme.
$B(V)$ $\Rightarrow$ The **shape** of the observed past matches pattern V.
A $\Rightarrow$ The (unobserved) **future path** we are forecasting.

$\Rightarrow$ For anticipative AR($p \geq 2$), the tail distribution degenerate to a single Dirac masses.

$\Rightarrow$ If the past pattern is identified, the future path becomes **asymptotically deterministic**.

Simulation of MAR

1. The linear model
2. Mixed-Phase processes
3. Identification
4. α-stable MAR
5. Prediction
6. Simulation of MAR
7. Estimation of MAR
8. Applications

Gouriéroux & Jasiak (2016) : partial fraction representation

- Let X_t be a strong MAR$(1,1)$ process

$$(1-\varphi^{\circ}F)(1-\varphi^{\bullet}B)X_t = \Phi^{\circ}(F)\Phi^{\bullet}(B)X_t = \varepsilon_t, \quad |\varphi^{\circ}|<1, \quad |\varphi^{\bullet}|<1$$

- Gouriéroux and Jasiak (2016) suggest a partial decomposition of ε_t

$$\varepsilon_t = \Phi^{\circ}(F)u_t = \Phi^{\bullet}(B)v_t$$

where $u_t = \Phi^{\bullet}(B)X_t$ and $v_t = \Phi^{\circ}(F)X_t$ hence leading to

$$\begin{aligned}
X_t &= \frac{1}{1-\varphi^{\bullet}\varphi^{\circ}}\left(\frac{1}{1-\varphi^{\circ}F} + \frac{\varphi^{\bullet}L}{1-\varphi^{\bullet}L}\right)\varepsilon_t \\
&= \frac{1}{1-\varphi^{\bullet}\varphi^{\circ}}\left(\varphi^{\circ}u_{t+1} + v_t\right) \\
&= \frac{1}{1-\varphi^{\bullet}\varphi^{\circ}}\left(u_t + \varphi^{\bullet}v_{t-1}\right)
\end{aligned}$$

where v_t and u_t are the causal and noncausal components of ε_t

Beyond partial fraction decomposition

- The approach of Gouriéroux & Jasiak (2016) is general...

$\Rightarrow$ ...but analytical solutions are complex and essentially known only for the MAR(1,1) case

- For higher orders (p, q), usual methods rely on **recursive** algorithms
- Problem: these methods introduce a systematic **truncation bias** by approximating the infinite sum
- de Truchis & Thomas (2025) propose an **exact** analytical solution for MAR(p, q)

Exact analytical solution

- We use the MA(∞) representation under the strict stationarity assumption:

$$X_t = [\varphi^\circ(F)\varphi^\bullet(B)]^{-1}\varepsilon_t = \sum_{k\in\mathbb{Z}} \delta_k \varepsilon_{t+k}$$

- By the fundamental theorem of algebra, we have $\varphi^\bullet(B) = \prod_{i=1}^{q}(1-\lambda_i B)$ and $\varphi^\circ(F) = \prod_{j=1}^{p}(1-\zeta_j F)$.
- de Truchis & Thomas (2025) obtain via the contour integral method:
- Future coefficients (non-causal, $k > 0$):

$$\delta_k = \sum_{j=1}^{p} \frac{\zeta_j^{(p-1)+k}}{\prod_{m\neq j}^{p}(\zeta_j-\zeta_m)\cdot\prod_{i=1}^{q}(\lambda_i\zeta_j-1)} \cdot (-1)^q$$

- Past coefficients (causal, $k \leq 0$):

$$\delta_k = \sum_{i=1}^{q} \frac{\lambda_i^{(q-1)+|k|}}{\prod_{l\neq i}^{q}(\lambda_i-\lambda_l)\cdot\prod_{j=1}^{p}(\lambda_i\zeta_j-1)} \cdot (-1)^p$$

Simulation (exact)

- **Step 1: Coefficient computation**
  - Use the exact formulas to compute the δ_k for $k \in [-m, m]$.
  - m is chosen large enough so that $\delta_k \approx 0$ if $|k| > m$.
- **Step 2: Innovation generation**
  - Simulate a long path of i.i.d. innovations ε_t (e.g., α-stable $S(\alpha, \beta, \sigma, 0)$ or Student's t).
- **Step 3: Simulation (Convolution)**
  - Compute x_t by direct convolution (a simple finite MA filter):

$$x_t = \sum_{k=-m}^{m} \delta_k \varepsilon_{t+k}$$

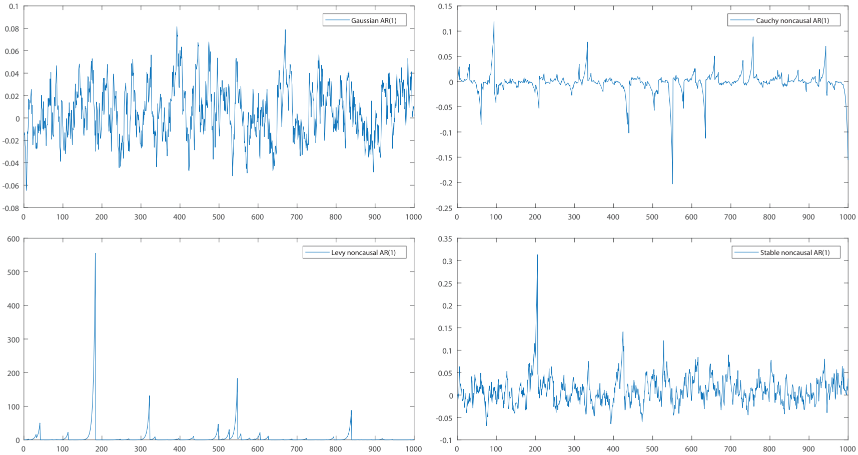
Remarks: $\Rightarrow$ Avoids the recursive approximation of the u_t and v_t components.

$\Rightarrow$ Eliminates the systematic truncation bias of recursive methods.

$\Rightarrow$ The δ_k coefficients are also the basis for new **forecasting** methods ("pattern-based forecasting").

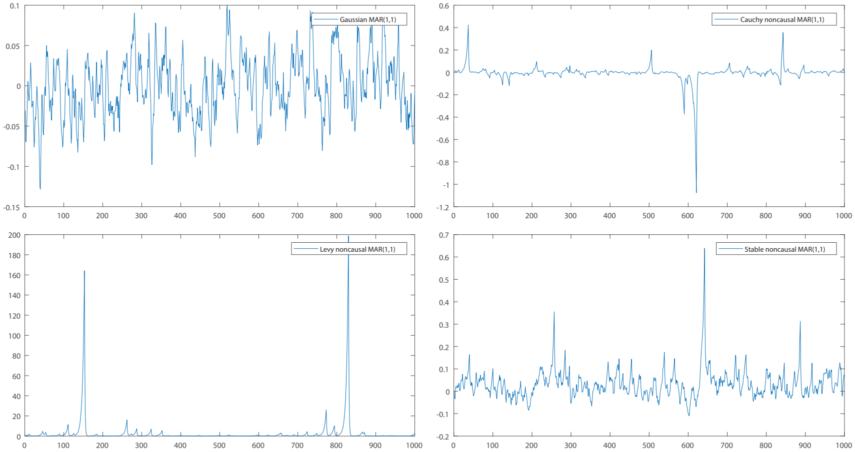
Simulation of noncausal AR(1) processes

- In all cases, $\varphi = 0.8$ / the bottom-right case is $\varepsilon_t \sim \mathcal{S}(1.8, 0, 0.01, 0)$



Simulation of noncausal MAR$(1,1)$ processes

- In all cases, $\varphi^{\circ} = 0.8$, $\varphi^{\bullet} = 0.4$ / the bottom-right case is $\varepsilon_t \sim \mathcal{S}(1.8, 0, 0.01, 0)$



Estimation of MAR

1. The linear model
2. Mixed-Phase processes
3. Identification
4. α-stable MAR
5. Prediction
6. Simulation of MAR
7. Estimation of MAR
8. Applications

Incomplete literature review

- **Ordinary Least Squares** : Davis and Resnick (1986)
  - Distribution assumptions : very general and mild
  - Limit theory : depends on distribution parameters (α if ε_t is α-stable)
  - Other : simple but cannot identify the causal and noncausal components
- Stable Maximum Likelihood : Andrews et al. (2009)
  - Limit theory : converges $n^{1/\alpha}$ if ε_t is α-stable
  - Other : computationally cumbersome
- **Least Absolute Deviation** : Davis and Wu (2010)
  - Limit theory : locally Gaussian and converges $n^{1/2}$ if ε_t is Laplacian
  - Other : simple but sensitive to causal and noncausal misspecification
- Semi-parametric log-concave projected MLE : Davis and Zhang (2017)
  - Powerful and very general albeit tedious to implement
- Frequency domain Minimum Distance : Lobato and Velasco (2018), Velasco (2022)
  - Powerful and general albeit requiring higher moments to exist
- **Generalized Covariance (GCov)** : Gourieroux and Jasiak (2023)
  - Powerful and general albeit requiring higher moments to exist

Autocorrelation of strong noncausal AR(p)

- Assumed that ε_t in $X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$ has Pareto-like tail

$$\mathbb{P}(|\varepsilon_t| > x) = x^{-\alpha} L(x) \text{ and } \sum |\psi_j|^s < \infty, \text{ for } s \in (0,\alpha) \cap [0,1]$$

$\Rightarrow$ X_t is stationary and ε_t enters the domain of attraction of α-stable laws

- Davis and Resnick (1986) investigates the limit theory for sample autocorrelation of strong non-Gaussian linear processes

$$\hat{\rho}_n(h) = \left(\sum_{t=h+1}^{n} X_t X_{t-h} \right) \left(\sum_{t=1}^{n} X_t^2 \right)^{-1}$$

and prove that $n^{1/\alpha} \left(\hat{\rho}_n(h) - \rho(h) \right) \xrightarrow{p} 0$ where

$$\rho(h) = \left(\sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \right) \left(\sum_{j=-\infty}^{\infty} \psi_j^2 \right)^{-1}$$

- Note that for $h = 1$, $\hat{\rho}_n(1)$ is the **OLS estimator** of $\rho(1) = \varphi$

Remark $\rho(h) \neq$ theoretical autocorrelations of X_t as **the latter do not exist**

Limit theory of OLS estimator of strong noncausal AR(p)

- Then, Davis and Resnick (1986) prove that for $h \geq 1$,

$$\hat{\boldsymbol{\rho}}_n = \big(\hat{\rho}_n(1), \ldots, \hat{\rho}_n(h)\big)' \text{ and } \boldsymbol{\rho} = \big(\rho(1), \ldots, \rho(h)\big)'$$

  the limit distribution of $\hat{\boldsymbol{\rho}}_n$ is a function of stable variables

$$\frac{a_n^2}{b_n}(\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}) \xrightarrow{d} (\zeta_1, \ldots, \zeta_h)$$

  with $a_n = \inf\{x : \mathbb{P}(|\varepsilon_1| > x) \leq n^{-1}\}$, $b_n = \inf\{x : \mathbb{P}(|\varepsilon_1 \varepsilon_2| > x) \leq n^{-1}\}$ and

$$\zeta_l = \sum_{j=1}^{\infty} \big(\rho(j+l) + \rho(l-j) - 2\rho(j)\rho(l)\big) \frac{S_j}{S_0},$$

  where S_j are i.i.d. α-stable and S_0 is positive $\alpha/2$-stable

- If the law of $|\varepsilon_t|$ is asymptotically equivalent to a Pareto distribution

$$\frac{a_n^2}{b_n} \equiv \left(\frac{n}{\log(n)}\right)^{\alpha}$$

$\Rightarrow$ the limit theory of the OLS estimator $\hat{\varphi}_n = \hat{\boldsymbol{\rho}}_n$ depends on α

Limit theory of OLS estimator of noncausal Cauchy AR(1)

- Let X_t be a strong Cauchy AR(1) process

$\Rightarrow$ Then, the OLS estimator $\hat{\varphi}$ satisfies

$$\frac{n}{\log(n)}(\hat{\varphi}_n - \varphi) \xrightarrow{d} (1+\varphi)S_1/S_0$$

where $S_1 \sim \mathcal{C}(0,1)$ and $S_0 \sim \mathcal{L}(0,1)$ and hence $1/S_0 \sim \chi^2(1)$

$\Rightarrow$ the Cauchy case is very surprising because

- — limit distribution: ratio of standard Cauchy and Chi-squared variables
- — $\hat{\varphi}_n$ converges faster than $\sqrt{n}$ even if $\mathbb{E}(S_1/S_0) = \infty$

Remark As $\mathbb{E}(X_t|X_{t-1}) = |\varphi|^{(\alpha-1)}X_{t-1} = \mathrm{sign}(\varphi)X_{t-1}$ in the Cauchy case, we show that even if $\mathbb{E}(X_t|X_{t-1}) \neq \varphi X_{t-1}$, $\hat{\varphi}_n$ converges to φ !

$\Rightarrow$ the empirical auto-correlation function reveals serial dependence in reverse time, not in direct time

Limit theory of OLS estimator of noncausal MAR(p, q)

- Let X_t be a strong MAR(p, q) process

$$\Phi(L)X_t = \varphi^\circ(F)\varphi^\bullet(B)X_t = \varepsilon_t \text{ with } \mathbb{P}(|\varepsilon_0| > x) = x^{-\alpha}L(x)$$

and denote by $\hat{\Phi}_n$ the OLS estimator of Φ, the parameters of $\Phi(L)$

- Fries & Zakoian (2018) show that

$$\frac{a_n^2}{b_n}(\hat{\Phi}_n - \Phi) \xrightarrow{d} R^{-1}(\boldsymbol{\zeta} - \boldsymbol{Z}\Phi)$$

with $R = \rho(|i-j|)_{i,j=1,\ldots,p+q}$, $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{p+q})$, $\boldsymbol{Z} = (\zeta_{i-j})_{i,j=1,\ldots,p+q}$

$$\zeta_l = \sum_{j=1}^{\infty} \big(\rho(j + |l|) + \rho(|l| - j) - 2\rho(j)\rho(|l|)\big)\frac{S_j}{S_0},$$

and $\zeta_0 = 0$ where S_j are i.i.d. α-stable and S_0 is positive $\alpha/2$-stable

Remark causal and noncausal parameters are not identifiable in the MAR(p, q)

Least Absolute Deviation (LAD)

- The LAD criterion is derived via a likelihood approximation assuming that the underlying noise ε_t is **Laplacian**

$$f_\varepsilon = (2\sigma)^{-1} \exp\left(- |\varepsilon_t| \sigma^{-1} \right), \quad \sigma > 0$$

$\Rightarrow$ the approximate log-likelihood is

$$L_n(\boldsymbol{\theta}|\varepsilon_t) = -(n-r)\log(2\sigma) - \frac{1}{\sigma}\sum_{t=1}^{n} |\varepsilon_t|$$

and can be maximized with respect to the scale parameter to obtain

$$\hat{\sigma} = (n-r)^{-1} \sum_{t=1}^{n} |\varepsilon_t|$$

and derive the following concentrated likelihood objective function

$$Q_n(\boldsymbol{\theta}|\varepsilon_t) = -(n-r)(1+\log(2)) - (n-r)^{-1}\log\left(2(n-r)^{-1}\right) - \log\left(\sum_{t=1}^{n} |\varepsilon_t|\right)$$

where $\boldsymbol{\theta}$ is a r vector of parameters

Least Absolute Deviation for generalized Mixed Phase ARMA

- As the **Laplace** distribution has fat tails, it is attractive here
- Let X_t be a generalized Mixed Phase ARMA

$$\varphi^\circ(F)\varphi^\bullet(B)X_t = \theta^\circ(F)\theta^\bullet(B)\varepsilon_t$$

  with ε_t a SWN in the domain of attraction of α-stable distributions

- Wu (2011) shows that the LAD objective function

$$Q_n(\boldsymbol{\theta}|\varepsilon_t) \equiv \sum_{t=1}^{n+p-q} \left|\frac{\theta^\circ_{q^\circ}}{\varphi^\circ_{p^\circ}}\varepsilon_t(\boldsymbol{\theta})\right|$$

  with $p = p^\circ + p^\bullet$, $q = q^\circ + q^\bullet$ and $\boldsymbol{\theta}$ a vector of $p + q$ parameters

- In practice, to build $\varepsilon_t(\boldsymbol{\theta})$, Wu and Davis (2010) suggest
  1. to compute forwards $v^\circ_t(\boldsymbol{\theta}) = \Phi(L)X_t - \theta^\bullet_1 v^\circ_{t-1}(\boldsymbol{\theta}) - \ldots - \theta^\bullet_{q^\bullet} v^\circ_{t-q^\bullet}(\boldsymbol{\theta})$ recursively for $t = 1, \ldots, n+p$ and $v^\circ_t(\boldsymbol{\theta}) = 0$ for $t \leq 0$
  2. to compute backwards $\varepsilon_t(\boldsymbol{\theta})$ given that $v^\circ_t(\boldsymbol{\theta}) = \theta^\circ(F)\varepsilon_t$ and hence

$$\varepsilon_t(\boldsymbol{\theta}) = \frac{1}{\theta^\circ_{q^\circ}}\Big(v^\circ_{t+q^\circ}(\boldsymbol{\theta}) - \varepsilon_{t+q^\circ}(\boldsymbol{\theta}) - \theta^\circ_1\varepsilon_{t+q^\circ-1}(\boldsymbol{\theta}) - \ldots - \theta^\circ_{q^\circ-1}\varepsilon_{t+1}(\boldsymbol{\theta})\Big)$$

  for $t = n + p - q^\circ, \ldots, -q^\circ + 1$ and $\varepsilon_t(\boldsymbol{\theta}) = 0$ for $t > n + p - q^\circ$.

Limit theory of the LAD for generalized Mixed Phase ARMA

- Wu (2011) shows that the LAD estimator $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta} Q_n(\boldsymbol{\theta}|\varepsilon_t)$

$$a_n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \boldsymbol{\nu}_{\min}, \quad a_n = \inf\{x : \mathbb{P}(|\varepsilon_1| > x) \leq n^{-1}\}$$

with $\boldsymbol{\nu}_{\min}$ a random variable with no closed form expression

$\Rightarrow$ the LAD estimator is $n^{1/\alpha}$-consistent $> n^{1/2}$ when $\alpha \in (0, 2)$

- As $\boldsymbol{\nu}_{\min}$ is untracktable $\Rightarrow$ bootstrap : Davis and Wu (1997) & Cavaliere et al. (2018)
- Davis and Wu (2010) discuss the finite variance case

$\Rightarrow$ for instance, if X_t is a strong MAR(p, q) with Laplacian errors

$$\sqrt{n}(\hat{\boldsymbol{\varphi}}^{\bullet} - \boldsymbol{\varphi}^{\bullet}) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4f_{\varepsilon}^2(0)\mathbb{E}(v_t^2)}\right), \quad \sqrt{n}(\hat{\boldsymbol{\varphi}}^{\circ} - \boldsymbol{\varphi}^{\circ}) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4f_{\varepsilon}^2(0)\mathbb{E}(u_t^2)}\right)$$

with $v_t = \varphi^{\circ}(F)X_t$, $u_t = \varphi^{\bullet}(L)X_t$ and $\boldsymbol{\varphi}^{\circ} = \varphi_1^{\circ}, \ldots, \varphi_p^{\circ}$ ($\boldsymbol{\varphi}^{\bullet}$ resp.)

- To evaluate $f_{\varepsilon}^2(0)$ a logistic kernel can be used (see Hecq et al. 2017)

Generalized Covariance (GCov) Estimator

- **Motivation:** estimate semi-parametric dynamic models with **i.i.d. errors**.
- Applicable to models like non-linear mixed causal-noncausal VAR, Stochastic Volatility...
- Let the model be: $g(\tilde{Y}_t;\theta) = u_t$, where u_t is i.i.d. ($\tilde{Y}_t$ contains current/lagged Y_t).
- Core Idea: Minimize a residual-based multivariate **portmanteau statistic**.
- Define residuals $\hat{u}_t(\theta) = g(\tilde{Y}_t;\theta)$.
- Let $\hat{\Gamma}(h;\theta)$ be the sample autocovariance of $\hat{u}_t(\theta)$ at lag h.
- Define $\hat{R}^2(h,\theta) = \hat{\Gamma}(h;\theta)\hat{\Gamma}(0,\theta)^{-1}\hat{\Gamma}(h;\theta)'\hat{\Gamma}(0;\theta)^{-1}$.
- GCov Estimator of Gourieroux & Jasiak (2023)

$$\hat{\theta}_T(H) = \arg\min_{\theta} \sum_{h=1}^{H} Tr[\hat{R}^2(h,\theta)]$$

Properties of the GCov Estimator

- **Asymptotic Properties:** Under regularity conditions:
  - Consistent: $\hat{\theta}_T(H) \xrightarrow{p} \theta_0$
  - Asymptotically Normal: $\sqrt{T}(\hat{\theta}_T - \theta_0) \sim N[0, \Omega(\theta_0)^{-1}]$
  - Semi-parametrically efficient
- **Identification:**
  - Identifies parameters characterizing **serial dependence**
  - Does **not** identify drift or scale parameters affecting only the marginal distribution of u_t
- **Drawbacks:** relies on higher moment conditions (although simulations support good performance in absence of finite variance)

Applications

Empirical papers

1. Forecasting extreme trajectories using seminorm representations by G. de Truchis, S. Fries, A. Thomas, 2025

2. Autoregression-based estimation of the new Keynesian Phillips curve by M. Lanne and J. Luoto, 2013

Empirical Application: Forecasting El Niño/La Niña

- **Context:** El Niño and La Niña are major climate anomalies with significant societal and economic impacts (extreme weather, agriculture, commodity prices, macroeconomic fluctuations)
- They are often measured by the Southern Oscillation Index (SOI), derived from air pressure differences
- The SOI exhibits patterns resembling explosive bubbles followed by crashes, making it suitable for non-causal modeling
- **Goal:** Apply the seminorm representation theory to forecast:
  - The probability of returning to normal conditions (reversal probabilities)
  - The exact timing of the peak and end of El Niño/La Niña episodes (reversal dates)
- Uses the results from de Truchis, Fries & Thomas (2025)

Data and Model Selection

- **Data:** Monthly SOI index.
- **Sample Period:** 01/1951 to 01/2024, split into:
  - In-sample: 01/1951 - 12/1991 (for estimation)
  - Out-of-sample: 01/1992 - 01/2024 (for forecast evaluation)
- **SOI Time Series Plot:**
- **Model Selection:**
  - ACF and PACF analysis suggests an AR structure
  - Estimation using GCoV estimator favors a purely non-causal AR(2) model

$$(1 - \varphi_1^\circ F - \varphi_2^\circ F^2)X_t = \varepsilon_t$$

  - Errors ε_t are assumed α-stable to capture heavy tails

Forecasting Reversal Probabilities

- **Goal:** Estimate the probability of El Niño and La Niña reversal
- **Methodology:** Approximate $\mathbb{P}(\text{Crash at } t+h | \text{Extreme at } t, \text{Pattern matched})$
  - Condition 1: extreme magnitude $\|\boldsymbol{X}_t\| > x$, approximated by $X_t \geq q_\tau$ (e.g., 90th or 95th quantile)
  - Condition 2: observed pattern
    $$(X_{t-m}, \ldots, X_t)/\|\boldsymbol{X}_t\|$$
    falls in a small neighborhood $B(V_0)$ around a theoretical pattern $\vartheta_0 \boldsymbol{d}_{k_0}/\|\boldsymbol{d}_{k_0}\|$
  - Target Event: future path returns to central values
- **Results (Selected):** Probabilities (in %) of returning to central values ($\delta = 0.5$)

Horizon (h)	In-Sample $p_{0.95}$	In-Sample $\hat{p}_{0.95}$	Out-of-Sample $p_{0.95}$	Out-of-Sample $\hat{p}_{0.95}$
$h = 3$	54.17	75.00	63.46	67.31
$h = 5$	83.33	87.50	92.31	84.62

Notes: p_q = theoretical prob., $\hat{p}_q$ = empirical freq. for quantile $q = 0.95$. $m = 2$.

- High probability of reversal within 5 months: unlikely persistence of very extreme La Niña events

Forecasting Reversal Dates: Methodology

- **Goal:** Predict the exact peak and end date of an El Niño/La Niña episode
- Leverage the **deterministic** nature of asymptotic forecasts for MAR(0,2) when $m \geq 1$
- If $\|\boldsymbol{X}_t\|$ is large and $(X_{t-m}, \ldots, X_t)$ matches a specific MA(∞) segment, the future path is determined
- **4-Step Procedure**:
  1. Compute the observed normalized pattern $(X_{t-m}, \ldots, X_t)/\|\boldsymbol{X}_t\|$ for a chosen $m \geq 1$
  2. Compute theoretical patterns $\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\|$ for a range of k (using the exact δ_k coefficients)

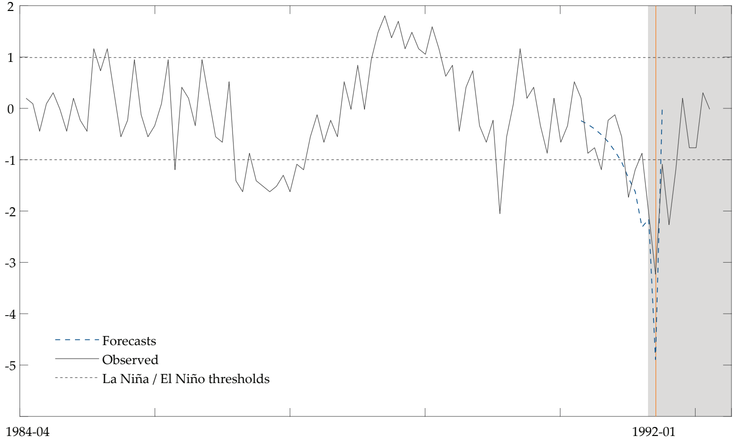
$$\boldsymbol{d}_k = (\delta_{k+m}, \ldots, \delta_k, \ldots, \delta_{k-h})$$

  3. Find the unique k_0 such that $\vartheta_0 \boldsymbol{d}_{k_0}/\|\boldsymbol{d}_{k_0}\|$ is closest to the observed pattern with ϑ_0 the sign of the event
  4. The future path is predicted to follow the shape of $\vartheta_0 \boldsymbol{d}_{k_0}$

Forecasting Reversal Dates: El Niño 1991/1992 Example

- **Scenario:** forecast the El Niño event starting at the end of the in-sample period (Dec 1991) with X_t large and negative ($\vartheta_0 = -1$)
- **Pattern Matching:**
  - Try different past window lengths $m \in [1, 10]$
  - For $m = 1, 2$ and $m \in [5, 10]$, the procedure robustly identifies $k_0 = 1$
  - Retain $k_0 = 1$ with $m = 10$. This means the observed pattern matches the segment $(\delta_{11}, \ldots, \delta_1)$
- **Prediction:**
  - Since $k_0 = 1$, the process is predicted to follow the shape $-\boldsymbol{d}_1 / \|\boldsymbol{d}_1\|$
  - The coefficients δ_k for $k \leq 0$ are zero for MAR(0,2). The peak occurs at $t + k_0 = t + 1$. The reversal (return to zero) occurs shortly after, at $t + k_0 + 1 = t + 2$
  - Predicted peak date: Jan 1992. Predicted reversal date: Feb 1992

Forecasting Reversal Dates: El Niño 1991/1992 Example



Forecasting Reversal Dates: Out-of-Sample Performance

Table: Forecasting out-of-sample El Niño and La Niña anomalies

Type of anomaly	El Niño	El Niño	La Niña	El Niño	La Niña	La Niña	El Niño	La Niña	La Niña	La Niña	La Niña
Start date	12/1991	07/1994	11/2007	12/2009	07/2010	11/2010	07/2015	11/2021	02/2022	08/2022	11/2022
Peak date	01/1992	09/1994	02/2008	02/2010	09/2010	12/2010	10/2015	01/2021	03/2022	10/2022	12/2022
End date	04/1992	10/1994	03/2008	03/2010	11/2010	04/2011	11/2015	03/2021	05/2022	11/2022	02/2023
Forecasted Peak	01/1992	09/1994	02/2008	03/2010	08/2010	01/2011	09/2015	01/2021	04/2022	10/2022	01/2023
Forecasted End	02/1992	10/1994	03/2008	04/2010	09/2010	02/2011	10/2015	02/2021	05/2022	11/2022	02/2023
Peak forecast error	0	0	0	1	-1	1	-1	0	1	0	-1
End forecast error	-2	0	0	1	-1	-2	-1	-1	0	0	0
k_0	1	2	3	3	1	2	2	2	2	2	2
m	10	10	10	9	10	10	10	10	10	10	10

- The 4-step procedure was applied to predict **all** El Niño and La Niña events (14 total) in the out-of-sample period (1992-2024).
- Forecasting starts when SOI first crosses the threshold (± 1)
- **Accuracy Summary:**
  - Average error in predicting the **peak date**: 0.42 months
  - Average error in predicting the **end date** (reversal): 0.57 months

Empirical paper 2 : stylized facts

- New Keynesian Phillips curve (NKPC) :

$$\pi_t = \gamma_f \mathbb{E}_t \pi_{t+1} + \lambda x_t$$

  inflation rate π_t depends linearly on
  ● the expected inflation rate next period, $E_t \pi_{t+1}$, and
  ● a measure of marginal costs, x_t.

- Hybrid NKPC (Gali and Gertler, 1999) : also include lagged inflation

$$\pi_t = \gamma_f \mathbb{E}_t \pi_{t+1} + \gamma_b \pi_{t-1} + \lambda x_t$$

- Inflation is highly persistent $\rightarrow$ debate on the source :
  ● dependence on past inflation in forming expectations ?
  ● agents' forward-looking behavior ?

Empirical paper 2 : stylized facts

- Main issue: the marginal cost variable x_t is not directly observable
  - use real unit labor cost
  - use the output gap
- Following the rational expectations literature

$$\pi_t = \gamma_f \pi_{t+1} + \gamma_b \pi_{t-1} + \eta_{t+1},$$

where $\eta_{t+1} = \xi_{t+1} + \lambda x_t$, with $\xi_{t+1} = \gamma_f \mathbb{E}_t \pi_{t+1} - \gamma_f \pi_{t+1}$

- If η_t were i.i.d., this would be the MAR(1, 1) model of Lanne and Saikkonen (2011a)
- In practice, need to allow η_t to be autocorrelated, hence assume the autocorrelation in the error term η_t to be adequately captured by a (potentially noncausal) MAR$(p-1, q-1)$ process

$$\varphi^{\circ}(F)\varphi^{\bullet}(B)y_t = \varepsilon_t$$

Empirical paper 2 : dataset

- U.S. NKPC with quarterly data from $1955:1$ to $2010:3$
- Inflation is computed as $\pi_t = 400\ln(P_t/P_{t-1})$,
  - π_t^{GDP} : P_t is the implicit price deflator of the GDP or
  - π_t^{CPI} : P_t is the consumer price index for all consumers
- proxies for the marginal cost:
  - the real unit labor cost and
  - linearly detrended logarithmic real GDP per capita

Empirical paper 2 : classical GMM estimation

- The estimated coefficients, their statistical significance and even their signs vary from one instrument set to another
- The results vary depending on the marginal cost proxy being used
- Overall it appears to be difficult to obtain general results concerning the issue of forward-looking vs. backward-looking inflation dynamics using the GMM

Empirical paper 2: noncausal autoregressions

Step 1 : Autocorrelation and non-normality of residuals

- Specify a Gaussian autoregression with serially uncorrelated errors and check whether the residuals are normally distributed
- Use Ljung–Box autocorrelation and Jarque–Bera normality tests
  - Normality is rejected, excess kurtosis points towards a student distribution
  - Identify the need for 5 AR lags for π^{GDP} and 4 lags for π^{CPI}

Empirical paper 2 : noncausal autoregressions

Step 2: Find the correct orders of causal and noncausal lag polynomials, p and q

- Estimate all MAR(p, q) models with t-distributed errors where the sum of p and q equals 5 for π^{GDP} and 4 for π^{CPI}
- Choose the specification that maximizes the log-likelihood function
- Finding: For both series, a mixed model involving both leads and lags is selected : MAR$(2, 3)$ and MAR$(3, 1)$ respectively

Empirical paper 2 : noncausal autoregressions

Step 3: Estimation of the hibrid NKPC by ML

Table 3
Estimation results of the new Keynesian Phillips curves based on the U.S. inflation series.

	π_t^{GDP}	π_t^{CPI}
AR Model	MAR(2,3)	MAR(3,1)
γ_b	0.302	0.189
	(0.099)	(0.060)
γ_f	0.675	0.768
	(0.086)	(0.057)
σ	1.154	1.917
	(0.108)	(0.356)
ν	4.527	3.010
	(1.490)	(0.706)

The row labeled AR Model gives the best-fitting AR(p, q) model that the estimation of the NKPC is based on. σ and ν are the scale and degree-of-freedom parameters of the error distribution, respectively. The figures in parentheses are ML standard errors based on the Hessian matrix.

Empirical paper 2 : further results

- Finding the correct variable driving the process of inflation is crucial for identification in conventional GMM and ML estimation approaches put forth in the previous literature
- Because here we have no estimate of λ, the deep parameters cannot be uniquely solved, but λx_t can be solved as

$$\lambda x_t = \pi_t - \hat{\gamma}_f \mathbb{E}_t \pi_{t+1} - \hat{\gamma}_b \pi_{t-1}$$

  once the NKPC is estimated and the obtained time series is informative about the properties of the implied drivers of the inflation series.
- The authors find that the driving processes of the two inflation series exhibit relatively low persistence
- This indicates that persistence is mostly intrinsic instead of being inherited from a persistent driving process

Empirical paper 2 : Conclusion

- The results lend support to both forward-looking and backward-looking dynamics, with the former clearly dominating
- Themodel doesn't require to prespecify any marginal cost proxy driving the inflation. Hence, it facilitates computing the most plausible driving process given the estimated parameter values
- Consequently, inflation persistence appears to be intrinsic as opposed to being inherited from a persistent driving process

- Andrews, B., Calder, M., & Davis, R. A. (2009). Maximum likelihood estimation for α-stable autoregressive processes. The Annals of Statistics, 37(4), 1946-1982.
- Cline, D. (1983). Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data. PhD thesis of Colorado State University.
- Cline, D. B., & Brockwell, P. J. (1985). Linear prediction of ARMA processes with infinite variance. Stochastic Processes and their applications, 19(2), 281-296.
- Davis, R., & Resnick, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. The Annals of Statistics, 533-558.
- Davis, R. A., & Zhang, J. (2018). Semi-Parametric Estimation for Non-Gaussian Non-Minimum Phase ARMA Models. Journal of Time Series Analysis, 39(3), 251-272.
- de Truchis, G., Fries, S., & Thomas, A. (2025). Forecasting extreme trajectories using seminorm representations.
- Drost, F. C., & Nijman, T. E. (1993). Temporal aggregation of GARCH processes. Econometrica, 909-927.
- Fries, S., & Zakoian, J. M. (2018). Mixed causal-noncausal ar processes and the modelling of explosive bubbles. Econometric Theory, 35(6), 1234-1270.
- Fries, S. (2022). Conditional moments of noncausal alpha-stable processes and the prediction of bubble crash odds. Journal of Business & Economic Statistics, 40(4), 1596-1616.
- Gourieroux, C., & Jasiak, J. (2016). Filtering, prediction and simulation methods for noncausal processes. Journal of Time Series Analysis, 37(3), 405-430.
- Gouriéroux, C., & Zakoïan, J. M. (2017). Local explosion modelling by non-causal process. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 79(3), 737-756.
- Gourieroux, C., & Jasiak, J. (2023). Generalized covariance estimator. Journal of Business & Economic Statistics, 41(4), 1315-1327.

- Hecq, A., Lieb, L., & Telg, S. (2016). Identification of Mixed Causal-Noncausal Models in Finite Samples. Annals of Economics and Statistics, (123/124), 307-331.
- Samorodnitsky, G., & Taqqu, M. S. (1991). Conditional moments and linear regression for stable random variables. Stochastic processes and their applications, 39(2), 183-199.
- Wu, R. (2011). Least absolute deviation estimation for general ARMA time series models with infinite variance. Statistica Sinica, 779-805.
- Wu, R. (2014). Least absolute deviation estimation for general fractionally integrated autoregressive moving average time series models. Statistics & Probability Letters, 94, 69-76.
- Wu, R., & Davis, R. A. (2010). Least absolute deviation estimation for general autoregressive moving average time-series models. Journal of Time Series Analysis, 31(2), 98-112.