General-order observation-driven models

François Roueff
LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay

Oaxaca, June, 18, 2018

Outline

Introduction

Statistical inference

Back to consistency

Conclusion

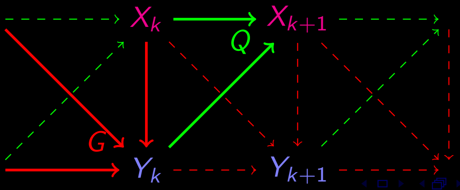
Partially observed Markov models

Definition (Partially observed Markov model (POMM))

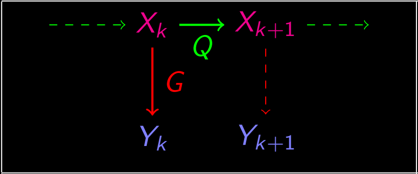
A partially observed Markov model with latent space $(\mathsf{X}, \mathcal{X})$ and observation space $(\mathsf{Y}, \mathcal{Y})$ is a pairwise homogeneous Markov chain $((Y_n, X_n), \mathcal{F}_n)_{n \geqslant 0}$ with kernel $\mathbf{K}^\theta$, $\theta \in \Theta$, generally described as

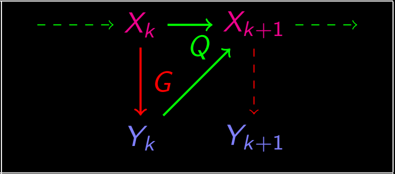
$$\begin{aligned} Y_k | \mathcal{F}_{k-1}, X_k &\sim G^\theta(X_{k-1}, Y_{k-1}, X_k; \cdot) \, , \\ X_{k+1} | \mathcal{F}_k &\sim Q^\theta(X_k, Y_k; \cdot) \, , \end{aligned} \tag{1}$$

and such that only the $\{Y_k\}'$s are observed.



Two important examples





Hidden Markov models (HMM)	Obs.-Driv. models (ODM)

- In both cases, $\{X_k\}$ is a Markov chain.
- An ODM moreover requires that

$$Q^\theta(X_k, Y_k; \cdot) = \delta_{\psi^\theta_{Y_k}(X_k)} ,$$

where δ_x denotes the Dirac mass at point x and, for all $y \in \mathsf{Y}$,

$$\begin{aligned} \psi^\theta : \mathsf{Y} \times \mathsf{X} &\to \mathsf{X} \\ (y, x) &\mapsto \psi^\theta_y(x) . \end{aligned}$$

General-order ODM

For $m \leqslant n$, denote $Z_{m:n} := (Z_m, Z_{m+1}, \ldots, Z_n)$.

General-order ODM

For $m \leqslant n$, denote $Z_{m:n} := (Z_m, Z_{m+1}, \ldots, Z_n)$.

Definition (ODM(p, q))

Let $p, q \geqslant 1$ and, for all $\theta \in \Theta$,

$$\begin{aligned} \psi^\theta \ : \ \mathsf{Y}^p \times \mathsf{X}^q &\to \mathsf{X} \\ (\mathbf{y}, \mathbf{x}) &\mapsto \psi^\theta_{\mathbf{y}}(\mathbf{x}) \ . \end{aligned}$$

An ODM of order (p, q) with link function ψ^θ and observation kernel G^θ satisfies, for all $k \in \mathbb{N}$,

$$\begin{aligned} X_{k+1} &= \psi^\theta_{Y_{(-p+1+k):k}}\left(X_{(-q+1+k):k}\right) \ , \\ Y_{k+1} | \mathcal{F}_k, X_{k+1} &\sim G^\theta\left(X_{k+1}; \cdot\right) \ . \end{aligned} \tag{2}$$

where $\mathcal{F}_k = \sigma(Y_{-p+1}, \ldots, Y_k, X_{-q+1}, \ldots, X_k)$.

Linear ODM

Definition (Linear ODM (LODM))

A linear ODM (LODM) is an ODM

- with parameters $\theta = (\vartheta, \varphi)$ with $\vartheta = (\omega, a_{1:p}, b_{1:q}) \in \mathbb{R}^{1+p+q}$,
- with $\mathsf{X} \subseteq \mathbb{R}$ and link functions of the form

$$\psi^{\theta} : \mathsf{Y}^p \times \mathsf{X}^q \to \mathsf{X}$$

$$(y_{1:p}, x_{1:q}) \mapsto \psi^{\theta}_{\mathbf{y}}(\mathbf{x}) = \omega + \sum_{k=1}^{p} a_k \upsilon(y_k) + \sum_{k=1}^{q} b_k x_k \ , \quad (3)$$

where $\upsilon : \mathsf{Y} \to \mathbb{R}$.

Definition (Linear ODM (LODM))

A linear ODM (LODM) is an ODM

- with parameters $\theta = (\vartheta, \varphi)$ with $\vartheta = (\omega, a_{1:p}, b_{1:q}) \in \mathbb{R}^{1+p+q}$,
- with $\mathsf{X} \subseteq \mathbb{R}$ and link functions of the form

$$\psi^{\theta} : \mathsf{Y}^{p} \times \mathsf{X}^{q} \to \mathsf{X}$$

$$(y_{1:p}, x_{1:q}) \mapsto \psi^{\theta}_{\mathbf{y}}(\mathbf{x}) = \omega + \sum_{k=1}^{p} a_k \upsilon(y_k) + \sum_{k=1}^{q} b_k x_k \,, \quad (3)$$

where $\upsilon : \mathsf{Y} \to \mathbb{R}$.

If $\mathsf{X} = \mathbb{R}_+$, set $(\omega, a_{1:p}, b_{1:q}) \in \mathbb{R}_+^{1+p+q}$ and $\upsilon : \mathsf{Y} \to \mathbb{R}_+$.

Examples of LODM

- GARCH(p, q), Bollerslev [1986]:

$$\mathsf{X} = \mathbb{R}_+ \,, \;\; G^\theta(x; \cdot) = \mathcal{N}(0, x) \quad \text{and} \quad \upsilon(y) = y^2 \,.$$

Examples of LODM

- GARCH(p,q), Bollerslev [1986]:

$$\mathsf{X} = \mathbb{R}_+\,,\;\; G^\theta(x;\cdot) = \mathcal{N}(0,x) \quad \text{and} \quad \upsilon(y) = y^2\;.$$

- In-GARCH, Davis, Dunsmuir, and Streett [2003]:

$$\mathsf{X} = \mathbb{R}_+\,,\;\; G^\theta(x;\cdot) = \mathrm{Poi}(x) \quad \text{and} \quad \upsilon(y) = y\;.$$

Examples of LODM

- GARCH(p,q), Bollerslev [1986]:

$$\mathsf{X}=\mathbb{R}_+\,,\;\; G^{\theta}(x;\cdot)=\mathcal{N}(0,x) \quad \text{and} \quad \upsilon(y)=y^2\,.$$

- In-GARCH, Davis, Dunsmuir, and Streett [2003]:

$$\mathsf{X}=\mathbb{R}_+\,,\;\; G^{\theta}(x;\cdot)=\mathrm{Poi}(x) \quad \text{and} \quad \upsilon(y)=y\,.$$

- NBINGARCH(p,q), Zhu [2011]:

$$\mathsf{X}=\mathbb{R}_+\,,\;\; G^{\theta}(x;\cdot)=\mathcal{NB}\left(r,\frac{1}{1+x}\right) \quad \text{and} \quad \upsilon(y)=y\,.$$

Examples of LODM

- GARCH(p, q), Bollerslev [1986]:
$$\mathsf{X} = \mathbb{R}_+\,,\ G^\theta(x;\cdot) = \mathcal{N}(0, x) \quad \text{and} \quad \upsilon(y) = y^2\,.$$
- In-GARCH, Davis, Dunsmuir, and Streett [2003]:
$$\mathsf{X} = \mathbb{R}_+\,,\ G^\theta(x;\cdot) = \mathrm{Poi}(x) \quad \text{and} \quad \upsilon(y) = y\,.$$
- NBINGARCH(p, q), Zhu [2011]:
$$\mathsf{X} = \mathbb{R}_+\,,\ G^\theta(x;\cdot) = \mathcal{NB}\left(r, \frac{1}{1+x}\right) \quad \text{and} \quad \upsilon(y) = y\,.$$
- Log-In-GARCH, Fokianos and Tjøstheim [2011]:
$$\mathsf{X} = \mathbb{R}\,,\ G^\theta(x;\cdot) = \mathrm{Poi}(\exp x) \quad \text{and} \quad \upsilon(y) = \ln(1+y)\,.$$

Examples of LODM

- GARCH(p, q), Bollerslev [1986]:
$$\mathsf{X} = \mathbb{R}_+ \,, \; G^\theta(x; \cdot) = \mathcal{N}(0, x) \quad \text{and} \quad \upsilon(y) = y^2 \,.$$

- In-GARCH, Davis, Dunsmuir, and Streett [2003]:
$$\mathsf{X} = \mathbb{R}_+ \,, \; G^\theta(x; \cdot) = \mathrm{Poi}(x) \quad \text{and} \quad \upsilon(y) = y \,.$$

- NBINGARCH(p, q), Zhu [2011]:
$$\mathsf{X} = \mathbb{R}_+ \,, \; G^\theta(x; \cdot) = \mathcal{NB}\left(r, \frac{1}{1+x}\right) \quad \text{and} \quad \upsilon(y) = y \,.$$

- Log-In-GARCH, Fokianos and Tjøstheim [2011]:
$$\mathsf{X} = \mathbb{R} \,, \; G^\theta(x; \cdot) = \mathrm{Poi}(\exp x) \quad \text{and} \quad \upsilon(y) = \ln(1+y) \,.$$

Note that, transforming the observations, we can take $\upsilon(y) = y$, but it modifies the definition of G^θ.

Domination

- In contrast to HMMs, ODMs are not fully dominated: $\mathbf{K}^\theta(x, \cdot)$ is not dominated by $(\mu \otimes \nu)$ for σ-finite measures μ and ν on X and Y, resp.
- We always assume that the ODM is partially dominated: there is a σ-finite measure ν on Y such that, for all $\theta \in \Theta$ and $x \in \mathsf{X}$, $G^\theta(x, \cdot)$ is dominated by ν, and, moreover, the density $g^\theta(x; \cdot) = \mathrm{d}G^\theta(x, \cdot)/\mathrm{d}\nu$ satisfies, for all $y \in \mathsf{Y}$,

$$g^\theta(x; y) > 0 ,$$

- To avoid a trivial case, ν is supposed to be non-degenerate, that is, its support contains at least two points.
- For LODMs, we assume the push measure $\nu \circ \upsilon^{-1}$ to be non-degenerate.

Embedding in an ODM(1,1)

Define $\mathsf{Z} = \mathsf{Y}^{p-1} \times \mathsf{X}^q$ and for all $k \in \mathbb{N}$,

$$Z_k = \left(Y_{(k-p+1):(k-1)}, X_{(k-q+1):k}\right) \in \mathsf{Z} \ .$$

- Then $(Y_k, Z_k)_{k \geqslant 0}$ is an ODM(1,1) with link function

$$\begin{aligned} \Psi^\theta &: \mathsf{Y} \times \mathsf{Z} \to \mathsf{Z} \\ &(y, z) \mapsto \Psi^\theta_y(z) = \psi^\theta_{(\mathbf{y}, y)}(\mathbf{x}) \quad \text{where} \quad z = (\mathbf{y}, \mathbf{x}) \ . \end{aligned}$$

- Given an initial distribution η on Z, we denote by $\mathbb{P}^\theta_\eta$ the distribution of $\{X_k, k > -q, \ Y_\ell, \ell > -p\}$ when

$$\left(Y_{(-p+1):-1}, X_{(-q+1):0}\right) \sim \eta \ .$$

Likelihood

- For any $m \in \mathbb{N}$ and $y_{0:m-1} \in \mathsf{Y}^m$, define

$$\Psi^\theta\langle y_{0:m-1}\rangle = \Psi^\theta_{y_{m-1}} \circ \cdots \circ \Psi^\theta_{y_0} \qquad : \mathsf{Z} \to \mathsf{Z}\,, \qquad (4)$$

$$\psi^\theta\langle y_{0:m-1}\rangle = \Pi_{p+q-1} \circ \Psi^\theta\langle y_{0:m-1}\rangle \qquad : \mathsf{Z} \to \mathsf{X}\,, \qquad (5)$$

where $\Pi_j : \mathsf{Z} \to \mathsf{Y}$ or X denotes the projection over the j-th coordinate.

Likelihood

- For any $m \in \mathbb{N}$ and $y_{0:m-1} \in \mathsf{Y}^m$, define

$$\Psi^\theta\langle y_{0:m-1}\rangle = \Psi^\theta_{y_{m-1}} \circ \cdots \circ \Psi^\theta_{y_0} \qquad : \mathsf{Z} \to \mathsf{Z}\,, \tag{4}$$

$$\psi^\theta\langle y_{0:m-1}\rangle = \Pi_{p+q-1} \circ \Psi^\theta\langle y_{0:m-1}\rangle \qquad : \mathsf{Z} \to \mathsf{X}\,, \tag{5}$$

  where $\Pi_j : \mathsf{Z} \to \mathsf{Y}$ or X denotes the projection over the j-th coordinate.

- Then, for any arbitrary $z^{(\mathrm{i})} \in \mathsf{Z}$ and observations $Y_{0:n}$, the (conditional) log-likelihood (given $Z_0 = z^{(\mathrm{i})}$) reads

$$\mathsf{L}^\theta_{z^{(\mathrm{i})},n} := \sum_{k=0}^{n} \ln g^\theta\left(\psi^\theta\langle Y_{0:(k-1)}\rangle(z^{(\mathrm{i})});\, Y_k\right)\,. \tag{6}$$

Likelihood

- For any $m \in \mathbb{N}$ and $y_{0:m-1} \in \mathsf{Y}^m$, define

$$\Psi^\theta\langle y_{0:m-1}\rangle = \Psi^\theta_{y_{m-1}} \circ \cdots \circ \Psi^\theta_{y_0} \qquad : \mathsf{Z} \to \mathsf{Z}\,, \qquad (4)$$

$$\psi^\theta\langle y_{0:m-1}\rangle = \Pi_{p+q-1} \circ \Psi^\theta\langle y_{0:m-1}\rangle \qquad : \mathsf{Z} \to \mathsf{X}\,, \qquad (5)$$

  where $\Pi_j : \mathsf{Z} \to \mathsf{Y}$ or X denotes the projection over the j-th coordinate.

- Then, for any arbitrary $z^{(\mathrm{i})} \in \mathsf{Z}$ and observations $Y_{0:n}$, the (conditional) log-likelihood (given $Z_0 = z^{(\mathrm{i})}$) reads

$$\mathsf{L}^\theta_{z^{(\mathrm{i})},n} := \sum_{k=0}^{n} \ln g^\theta\left(\psi^\theta\langle Y_{0:(k-1)}\rangle(z^{(\mathrm{i})}); Y_k\right)\,. \qquad (6)$$

- Hence the log-likelihood, as well as its derivatives, can easily be computed using $O(n)$ operations.

Stationary distribution

- The existence and uniqueness of stationary solutions for the standard GARCH(p, q) equation can be treated in the framework of stochastic linear equation, see Bougerol and Picard [1992].

Stationary distribution

- The existence and uniqueness of stationary solutions for the standard GARCH(p, q) equation can be treated in the framework of stochastic linear equation, see Bougerol and Picard [1992].
- For general ODM's, the ergodicity of the Markov chain (Y_k, Z_k) is equivalent to that of the Markov chain (Z_k).

Stationary distribution

- The existence and uniqueness of stationary solutions for the standard GARCH(p, q) equation can be treated in the framework of stochastic linear equation, see Bougerol and Picard [1992].
- For general ODM's, the ergodicity of the Markov chain (Y_k, Z_k) is equivalent to that of the Markov chain (Z_k).
- Integer valued ODM's are not ϕ-irreducible and require particular attention.

Stationary distribution

- The existence and uniqueness of stationary solutions for the standard GARCH(p, q) equation can be treated in the framework of stochastic linear equation, see Bougerol and Picard [1992].
- For general ODM's, the ergodicity of the Markov chain (Y_k, Z_k) is equivalent to that of the Markov chain (Z_k).
- Integer valued ODM's are not ϕ-irreducible and require particular attention.
  - Existence: Weak Feller + geometric Drift condition, see Tweedie [1988].
  - Uniqueness: coupling argument based on technical assumptions on G^θ, see Douc, Doukhan, and Moulines [2013].

Stationary distribution

- The existence and uniqueness of stationary solutions for the standard GARCH(p, q) equation can be treated in the framework of stochastic linear equation, see Bougerol and Picard [1992].
- For general ODM's, the ergodicity of the Markov chain (Y_k, Z_k) is equivalent to that of the Markov chain (Z_k).
- Integer valued ODM's are not ϕ-irreducible and require particular attention.
  - Existence: Weak Feller + geometric Drift condition, see Tweedie [1988].
  - Uniqueness: coupling argument based on technical assumptions on G^θ, see Douc, Doukhan, and Moulines [2013].
- A similar coupling argument yields β mixing properties for (Y_k), see Doukhan and Neumann [2017].

Notation

Recall that, for all $\theta \in \Theta$ and initial distribution η on $(\mathsf{Z}, \mathcal{Z})$, $\mathbb{P}_\eta^\theta$ denotes the distribution of $\{Y_k, X_\ell : k > -p, \ell > -q\}$ for $Z_0 \sim \eta$.

Notation

Recall that, for all $\theta \in \Theta$ and initial distribution η on $(\mathsf{Z}, \mathcal{Z})$, $\mathbb{P}_\eta^\theta$ denotes the distribution of $\{Y_k, X_\ell : k > -p, \ell > -q\}$ for $Z_0 \sim \eta$.

If the model is ergodic, i.e. there exists a unique probability measure η such that $\mathbb{P}_\eta^\theta$ is shift-invariant, we denote

- the stationary distribution extended on $\left((\mathsf{Y} \times \mathsf{X})^\mathbb{Z}, (\mathcal{X} \otimes \mathcal{Y})^{\otimes \mathbb{Z}}\right)$ by $\mathbb{P}^\theta$,
- the marginalization of $\mathbb{P}^\theta$ on $\left(\mathsf{Y}^\mathbb{Z}, \mathcal{Y}^{\otimes \mathbb{Z}}\right)$ by $\tilde{\mathbb{P}}^\theta$.

Notation

Recall that, for all $\theta \in \Theta$ and initial distribution η on $(\mathsf{Z}, \mathcal{Z})$, $\mathbb{P}_\eta^\theta$ denotes the distribution of $\{Y_k, X_\ell \,:\, k > -p, \ell > -q\}$ for $Z_0 \sim \eta$.

If the model is ergodic, i.e. there exists a unique probability measure η such that $\mathbb{P}_\eta^\theta$ is shift-invariant, we denote

- the stationary distribution extended on $\left((\mathsf{Y} \times \mathsf{X})^{\mathbb{Z}}, (\mathcal{X} \otimes \mathcal{Y})^{\otimes \mathbb{Z}}\right)$ by $\mathbb{P}^\theta$,
- the marginalization of $\mathbb{P}^\theta$ on $\left(\mathsf{Y}^{\mathbb{Z}}, \mathcal{Y}^{\otimes \mathbb{Z}}\right)$ by $\tilde{\mathbb{P}}^\theta$.

Remark : when establishing ergodicity through a drift condition, we obtain some minimal finite moment condition, referred to as (M) in the following, for X_0 (and Y_0) under the stationary distribution.

Statistical inference

- The Maximum Likelihood Estimator (MLE) $\widehat{\theta}_{z^{(i)},n}$ is defined as

$$\widehat{\theta}_{z^{(i)},n} \in \operatorname{argmax}\left\{\mathsf{L}^{\theta}_{z^{(i)},n} \;:\; \theta \in \Theta\right\} \tag{7}$$

for some arbitrary initial point $z^{(i)} \in \mathsf{Z}$.

- In well-specified models, a standard consistency result consists in showing that

$$\lim_{n\to\infty} \widehat{\theta}_{\eta,n} = \theta_\star\,, \quad \mathbb{P}^{\theta_\star}\text{-a.s.} \tag{8}$$

- and asymptotic normality consists in showing that

$$\sqrt{n}(\widehat{\theta}_{z^{(i)},n} - \theta_\star) \overset{\mathbb{P}^{\theta_\star}}{\Longrightarrow} \mathcal{N}(0, \mathcal{J}^{-1}(\theta_\star)) \tag{9}$$

where $\mathcal{J}(\theta_\star)$ is a nonsingular $d \times d$-matrix.

Asymptotic behavior of the likelihood

To fulfill this program, an essential ingredient is to study the asymptotic behavior of the likelihood (6).

Asymptotic behavior of the likelihood

To fulfill this program, an essential ingredient is to study the asymptotic behavior of the likelihood (6).

In particular, we need a stationary approximation of

$$\psi^{\theta}\langle Y_{0:(k-1)}\rangle(z^{(\mathrm{i})}) , \qquad k \geqslant 0 .$$

Asymptotic behavior of the likelihood

To fulfill this program, an essential ingredient is to study the asymptotic behavior of the likelihood (6).

In particular, we need a stationary approximation of

$$\psi^{\theta}\langle Y_{0:(k-1)}\rangle(z^{(i)}) , \qquad k \geqslant 0 .$$

This is done by approximating $Y_{0:(k-1)}$ by $Y_{-\infty:(k-1)}$, defined, in the case $k = 1$ by the backward limit

$$\lim_{j \to -\infty} \psi^{\theta}\langle Y_{j:0}\rangle(z^{(i)}) .$$

To fulfill this program, an essential ingredient is to study the asymptotic behavior of the likelihood (6).

In particular, we need a stationary approximation of

$$\psi^{\theta}\langle Y_{0:(k-1)}\rangle(z^{(\mathrm{i})}) \;, \qquad k \geqslant 0 \;.$$

This is done by approximating $Y_{0:(k-1)}$ by $Y_{-\infty:(k-1)}$, defined, in the case $k = 1$ by the backward limit

$$\lim_{j \to -\infty} \psi^{\theta}\langle Y_{j:0}\rangle(z^{(\mathrm{i})}) \;.$$

Let $(\mathsf{X}, \delta_{\mathsf{X}})$ and $(\mathsf{Z}, \delta_{\mathsf{Z}})$ be complete metric spaces in such a way that $\Pi_{p+q+1} : \mathsf{Z} \to \mathsf{X}$ is 1-Lipschitz.

By (4) and (5), $\delta_{\mathsf{X}}\left(\psi^{\theta}\langle Y_{i:0}\rangle(z^{(\mathrm{i})}), \psi^{\theta}\langle Y_{i+1:0}\rangle(z^{(\mathrm{i})})\right)$ is bounded from above by

$$\left(\sup_{y \in \mathsf{Y}^{i}, z, z'} \frac{\delta_{\mathsf{Z}}\left(\Psi^{\theta}\langle y\rangle(z), \Psi^{\theta}\langle y\rangle(z')\right)}{\delta_{\mathsf{Z}}\left(z, z'\right)}\right) \delta_{\mathsf{Z}}\left(\Psi^{\theta}\langle Y_{i+1}\rangle(z^{(\mathrm{i})}), z^{(\mathrm{i})}\right) .$$

where we used a uniform Lipschitz constant of the i- iterate of the link function.

Uniform Lipschitz condition

By (4) and (5), $\delta_{\mathsf{X}}\left(\psi^\theta\langle Y_{i:0}\rangle(z^{(\mathrm{i})}), \psi^\theta\langle Y_{i+1:0}\rangle(z^{(\mathrm{i})})\right)$ is bounded from above by

$$\left(\sup_{y \in \mathsf{Y}^i, z, z'} \frac{\delta_{\mathsf{Z}}\left(\Psi^\theta\langle y\rangle(z), \Psi^\theta\langle y\rangle(z')\right)}{\delta_{\mathsf{Z}}\left(z, z'\right)}\right) \delta_{\mathsf{Z}}\left(\Psi^\theta\langle Y_{i+1}\rangle(z^{(\mathrm{i})}), z^{(\mathrm{i})}\right) .$$

where we used a uniform Lipschitz constant of the i- iterate of the link function.

Define for all $i \in \mathbb{N}^*$,

$$\mathrm{Lip}_i^\theta = \sup\left\{\frac{\delta_{\mathsf{Z}}\left(\psi^\theta\langle y\rangle(z), \psi^\theta\langle y\rangle(z')\right)}{\delta_{\mathsf{Z}}\left(z, z'\right)} \;:\; y \in \mathsf{Y}^i, z, z' \in \mathsf{Z}\right\} .$$

We use the following condition:

(A-1) For all $\theta \in \Theta$, we have $\mathrm{Lip}_0^\theta < \infty$ and $\mathrm{Lip}_n^\theta \to 0$ as $n \to \infty$,

Asymptotic behavior of the likelihood (cont.)

Under (A-1)+ (M), for all $\theta \in \Theta$, there exists a measurable function $\psi^{\theta}\langle\cdot\rangle : \mathsf{Y}^{\mathbb{Z}_-} \to \mathsf{X}$ such that for all $\theta, \theta_\star \in \Theta$,

$$X_1 = \psi^{\theta}\langle Y_{-\infty:0}\rangle \qquad \mathbb{P}^{\theta}\text{-a.s.}$$

$$\psi^{\theta}\langle Y_{-\infty:0}\rangle = \lim_{n\to\infty} \psi^{\theta}\langle Y_{-n:0}\rangle(z^{(i)}) \qquad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Note that, under $\tilde{\mathbb{P}}^{\theta_\star}$, we have that

- ▷ $y \mapsto \mathrm{p}^{\theta}(y \mid Y_{-\infty:0}) := g^{\theta}\left(\psi^{\theta}\langle Y_{-\infty:0}\rangle; y\right)$ is a density w.r.t. ν
- ▷ $y \mapsto \mathrm{p}^{\theta_\star}(y \mid Y_{-\infty:0})$ is the conditional density of Y_1 given $Y_{-\infty:0}$.

Moreover, with some continuity conditions, for n large,

$$\mathrm{L}^{\theta}_{z^{(i)},n} \approx \sum_{k=0}^{n} \ln \mathrm{p}^{\theta}(Y_k \mid Y_{-\infty:(k-1)}) \qquad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \tag{10}$$

Asymptotic behavior of the likelihood (cont.)

Under (A-1)+ (M), for all $\theta \in \Theta$, there exists a measurable function $\psi^{\theta}\langle \cdot \rangle : \mathsf{Y}^{\mathbb{Z}_-} \to \mathsf{X}$ such that for all $\theta, \theta_\star \in \Theta$,

$$
\begin{aligned}
X_1 &= \psi^{\theta}\langle Y_{-\infty:0} \rangle && \mathbb{P}^{\theta}\text{-a.s.} \\
\psi^{\theta}\langle Y_{-\infty:0} \rangle &= \lim_{n \to \infty} \psi^{\theta}\langle Y_{-n:0} \rangle (z^{(i)}) && \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}
\end{aligned}
$$

Asymptotic behavior of the likelihood (cont.)

Under (A-1)+ (M), for all $\theta \in \Theta$, there exists a measurable function $\psi^{\theta}\langle\cdot\rangle : \mathsf{Y}^{\mathbb{Z}_-} \to \mathsf{X}$ such that for all $\theta, \theta_\star \in \Theta$,

$$X_1 = \psi^{\theta}\langle Y_{-\infty:0}\rangle \qquad \mathbb{P}^{\theta}\text{-a.s.}$$

$$\psi^{\theta}\langle Y_{-\infty:0}\rangle = \lim_{n\to\infty} \psi^{\theta}\langle Y_{-n:0}\rangle(z^{(i)}) \qquad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Note that, under $\tilde{\mathbb{P}}^{\theta_\star}$, we have that

- $y \mapsto \mathrm{p}^{\theta}(y \mid Y_{-\infty:0}) := g^{\theta}\left(\psi^{\theta}\langle Y_{-\infty:0}\rangle; y\right)$ is a density w.r.t. ν
- $y \mapsto \mathrm{p}^{\theta_\star}(y \mid Y_{-\infty:0})$ is the conditional density of Y_1 given $Y_{-\infty:0}$.

Asymptotic behavior of the likelihood (cont.)

Under (A-1)+ (M), for all $\theta \in \Theta$, there exists a measurable function $\psi^{\theta}\langle\cdot\rangle : \mathsf{Y}^{\mathbb{Z}_-} \to \mathsf{X}$ such that for all $\theta, \theta_\star \in \Theta$,

$$X_1 = \psi^{\theta}\langle Y_{-\infty:0}\rangle \qquad \mathbb{P}^{\theta}\text{-a.s.}$$

$$\psi^{\theta}\langle Y_{-\infty:0}\rangle = \lim_{n\to\infty} \psi^{\theta}\langle Y_{-n:0}\rangle(z^{(\mathrm{i})}) \qquad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Note that, under $\tilde{\mathbb{P}}^{\theta_\star}$, we have that

- $y \mapsto \mathrm{p}^{\theta}(y \mid Y_{-\infty:0}) := g^{\theta}\left(\psi^{\theta}\langle Y_{-\infty:0}\rangle; y\right)$ is a density w.r.t. ν
- $y \mapsto \mathrm{p}^{\theta_\star}(y \mid Y_{-\infty:0})$ is the conditional density of Y_1 given $Y_{-\infty:0}$.

Moreover, with some continuity conditions, for n large,

$$\mathrm{L}^{\theta}_{z^{(\mathrm{i})},n} \approx \sum_{k=0}^{n} \ln \mathrm{p}^{\theta}(Y_k \mid Y_{-\infty:(k-1)}) \qquad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \qquad (10)$$

The steps to the CLT for the MLE

Using ergodicity +(M) :

- ▷ Approximate the likelihood as above;

The steps to the CLT for the MLE

Using ergodicity $+$(M) :

- Approximate the likelihood as above;
- Prove consistency: $\lim_{n\to\infty} \widehat{\theta}_{z^{(i)},n} = \theta_\star$, $\tilde{\mathbb{P}}^{\theta_\star}$-a.s.,

The steps to the CLT for the MLE

Using ergodicity $+$(M) :

- Approximate the likelihood as above;
- Prove consistency: $\lim_{n\to\infty} \widehat{\theta}_{z^{(i)},n} = \theta_\star$, $\tilde{\mathbb{P}}^{\theta_\star}$-a.s.,
- Taylor expansion of the score function

$$\mathbf{S}_n(\theta) = \sum_{k=0}^{n} \partial_\theta \ln \mathrm{p}^\theta(Y_k \mid Y_{-\infty:(k-1)})$$

around $\theta_\star$;

The steps to the CLT for the MLE

Using ergodicity $+$(M) :

- Approximate the likelihood as above;
- Prove consistency: $\lim_{n\to\infty} \widehat{\theta}_{z^{(i)},n} = \theta_\star$, $\tilde{\mathbb{P}}^{\theta_\star}$-a.s.,
- Taylor expansion of the score function

$$\mathbf{S}_n(\theta) = \sum_{k=0}^{n} \partial_\theta \ln \mathrm{p}^\theta(Y_k \mid Y_{-\infty:(k-1)})$$

  around $\theta_\star$;
- CLT for the martingale $\mathbf{S}_n(\theta_\star)$ (under $\tilde{\mathbb{P}}^{\theta_\star}$),
- A.s. convergence for the Hessian.

The steps to the CLT for the MLE

Using ergodicity $+$(M) :

- Approximate the likelihood as above;
- Prove consistency: $\lim_{n\to\infty} \widehat{\theta}_{z^{(i)},n} = \theta_\star$, $\tilde{\mathbb{P}}^{\theta_\star}$-a.s.,
- Taylor expansion of the score function

$$\mathbf{S}_n(\theta) = \sum_{k=0}^{n} \partial_\theta \ln \mathrm{p}^\theta(Y_k \mid Y_{-\infty:(k-1)})$$

  around $\theta_\star$;
- CLT for the martingale $\mathbf{S}_n(\theta_\star)$ (under $\tilde{\mathbb{P}}^{\theta_\star}$),
- A.s. convergence for the Hessian.

All theses steps can be carried out for the previously mentioned models, with some restrictions on the parameter set Θ, sometimes appearing in technical conditions.

Consistency: missing steps

From the approximation in (10), we only get that

$$\lim_{n\to\infty} \Delta\left(\widehat{\theta}_{z^{(i)},n}, \Theta^\star\right) = 0 \qquad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}\ , \tag{11}$$

where Δ is the metric on Θ and the limit maximizing set $\Theta^\star$ is defined for all $\theta_\star$ by

$$\Theta^\star = \operatorname{argmax}\left\{\tilde{\mathbb{E}}^{\theta_\star}[\ln p^\theta(Y_1 \mid Y_{-\infty:0})] \ :\ \theta \in \Theta\right\}$$

To prove consistency, it remains to go through two additional steps:

- we need to show that (see Douc, Roueff, and Sim [2016] for any POMM)

$$\Theta^\star = [\theta_\star] := \left\{\theta \in \Theta \ :\ \tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta_\star}\right\}\ .$$

  Then, with (11), we get equivalence class consistency (as introduced by Leroux [1992]).
- To conclude, find conditions to have that $[\theta_\star] = \{\theta_\star\}$. (So that the model restricted to these $\theta_\star$s is identifiable).

Identifiability: basic assumption on the observation kernel

We suppose that the observation kernel satisfies:

(A-2) We can write $\theta_\star = (\vartheta_\star, \varphi_\star)$ and, for all $\theta = (\vartheta, \varphi)$ in Θ and $x, x' \in \mathsf{X}$,

$$G^{\theta}(x; \cdot) = G^{\theta_\star}(x'; \cdot) \quad \text{if and only if} \quad \varphi = \varphi_\star \quad \text{and} \quad x = x' \; .$$

(i.e. φ is the part of the parameter θ that can be identified directly from the conditional distribution of one observation)

Identifiability: basic assumption on the observation kernel

We suppose that the observation kernel satisfies:

(A-3) We can write $\theta_\star = (\vartheta_\star, \varphi_\star)$ and, for all $\theta = (\vartheta, \varphi)$ in Θ and $x, x' \in \mathsf{X}$,

$$G^{\theta}(x; \cdot) = G^{\theta_\star}(x'; \cdot) \quad \text{if and only if} \quad \varphi = \varphi_\star \quad \text{and} \quad x = x' .$$

(i.e. φ is the part of the parameter θ that can be identified directly from the conditional distribution of one observation)

Denote by $[\theta_\star]$ the equivalence class:

$$[\theta_\star] = \left\{ \theta \in \Theta \, : \, \tilde{\mathbb{P}}^{\theta} = \tilde{\mathbb{P}}^{\theta_\star} \right\}$$

Identifiability: a general result

We have the following result.

Theorem

Consider an ergodic ODM(p, q) satisfying (A-2). Suppose that, for all $\theta \in \Theta$, there exists $\psi^{\theta}\langle\cdot\rangle : \mathsf{Y}^{\mathbb{Z}_-} \to \mathsf{X}$ such that

$$X_1 = \psi^{\theta}\langle Y_{-\infty:0}\rangle \qquad \mathbb{P}^{\theta}\text{-a.s.} \tag{12}$$

Then $[\theta_\star]$ coincides with the set of all $\theta = (\vartheta, \varphi_\star) \in \Theta$ such that

$$\psi^{\theta}\langle Y_{-\infty:0}\rangle = \psi^{\theta_\star}\langle Y_{-\infty:0}\rangle \qquad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s. },$$

$$\psi^{\theta}\langle Y_{-\infty:0}\rangle = \psi^{\theta}_{Y_{(-p+1):0}}\left(\left(\psi^{\theta}\langle Y_{-\infty:j}\rangle\right)_{-q\leqslant j\leqslant -1}\right) \qquad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Identifiability: a general result

We have the following result.

Theorem

Consider an ergodic ODM(p, q) satisfying (A-2). Suppose that, for all $\theta \in \Theta$, there exists $\psi^\theta\langle\cdot\rangle : \mathsf{Y}^{\mathbb{Z}_-} \to \mathsf{X}$ such that

$$X_1 = \psi^\theta \langle Y_{-\infty:0} \rangle \qquad \mathbb{P}^\theta\text{-a.s.} \tag{12}$$

Then $[\theta_\star]$ coincides with the set of all $\theta = (\vartheta, \varphi_\star) \in \Theta$ such that

$$\psi^\theta \langle Y_{-\infty:0} \rangle = \psi^{\theta_\star} \langle Y_{-\infty:0} \rangle \qquad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} ,$$

$$\psi^\theta \langle Y_{-\infty:0} \rangle = \psi^\theta_{Y_{(-p+1):0}} \left(\left(\psi^\theta \langle Y_{-\infty:j} \rangle \right)_{-q \leqslant j \leqslant -1} \right) \qquad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Recall that (12) follows from (A-1)+(M), see ▸ likelihood approx.

Uniform Lipschitz assumption: the linear case.

For a linear link function (3) with $\upsilon(y) = y$,
Ass. (A-1) is equivalent to

(L-1) For all $\theta = (\vartheta, \varphi) \in \Theta$ with $\vartheta = (\omega, a_{1:p}, b_{1:q})$, we have $b_{1:q} \in \mathcal{S}_q$,

where

$$\mathcal{S}_q := \left\{ c_{1:q} \in \mathbb{R}^q \,:\, \forall z \in \mathbb{C}, |z| \leqslant 1 \text{ implies } 1 - \sum_{k=1}^{q} c_k z^k \neq 0 \right\} .$$

Identifiability: the linear case.

We have the following result.

Theorem

Consider an ergodic ODM(p, q) satisfying (L-1) and (A-2)+(M). Then, for any $\theta_\star = (\omega^\star, a^\star{}_{1:p}, b^\star{}_{1:q}, \varphi^\star)$ in the interior of Θ,

$$[\theta_\star] = \{\theta_\star\} \qquad \text{if and only if}$$

(L-2) The polynomials $\mathrm{P}_p(\cdot; a^\star{}_{1:p})$ and $\mathrm{Q}_q(\cdot; b^\star{}_{1:q})$ have no common complex roots,

where we defined

$$\mathrm{P}_p(z; a_{1:p}) = \sum_{k=0}^{p-1} a_{k+1}\ z^{p-1-k}$$

$$\mathrm{Q}_q(z; b_{1:q}) = z^q - \sum_{k=1}^{q} b_k z^{q-k}\ .$$

Concluding remarks

- Details can be found in the thesis of Sim [2016] and in a forthcoming paper.

Concluding remarks

- Details can be found in the thesis of Sim [2016] and in a forthcoming paper.
- The steps leading to consistency and asymptotic normality are standard for all POMMs, some of which can be treated in a very general fashion.

Concluding remarks

- Details can be found in the thesis of Sim [2016] and in a forthcoming paper.
- The steps leading to consistency and asymptotic normality are standard for all POMMs, some of which can be treated in a very general fashion.
- The initial step is in fact to prove ergodicity of the complete Markov chain, and define Θ accordingly (the integer valued case being of special interest).

Concluding remarks

- Details can be found in the thesis of Sim [2016] and in a forthcoming paper.
- The steps leading to consistency and asymptotic normality are standard for all POMMs, some of which can be treated in a very general fashion.
- The initial step is in fact to prove ergodicity of the complete Markov chain, and define Θ accordingly (the integer valued case being of special interest).
- The same identifiability condition is valid in the general case of LODMs.

Concluding remarks

- Details can be found in the thesis of Sim [2016] and in a forthcoming paper.
- The steps leading to consistency and asymptotic normality are standard for all POMMs, some of which can be treated in a very general fashion.
- The initial step is in fact to prove ergodicity of the complete Markov chain, and define Θ accordingly (the integer valued case being of special interest).
- The same identifiability condition is valid in the general case of LODMs.
- Open question : GARCH(p, q) processes are known to be regularly varying. How can this be extended to integer valued cases ?

References I

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31:307–327, 1986.

P. Bougerol and N. Picard. Stationarity of garch processes and of some nonnegative time series. *J. Econometrics*, 52(1992):115 – 127, 1992. ISSN 0304-4076. doi: 10.1016/0304-4076(92)90067-2.

Richard A Davis, William TM Dunsmuir, and Sarah B Streett. Observation-driven models for poisson counts. *Biometrika*, 90 (4):777–790, December 2003.

R. Douc, P. Doukhan, and E. Moulines. Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stoch. Proc. App.*, 123(7): 2620–2647, 2013.

References II

R. Douc, F. Roueff, and T. Sim. The maximizing set of the asymptotic normalized log-likelihood for partially observed markov chains. *Annals of Applied Probability*, 26(4):2357–2383, 2016.

P. Doukhan and M. H. Neumann. Absolute regularity of semi-contractive GARCH-type processes. *ArXiv e-prints*, November 2017.

Konstantinos Fokianos and Dag Tjøstheim. Log-linear poisson autoregression. *Journal of Multivariate Analysis*, 102(3): 563–578, 2011.

B. G. Leroux. Maximum-likelihood estimation for hidden Markov models. *Stoch. Proc. Appl.*, 40:127–143, 1992.

Tepmony Sim. *Maximum likelihood estimation in partially observed Markov models with applications to time series of counts*. PhD thesis, Télécom ParisTech, 2016.

References III

R. L. Tweedie. Invariant measures for Markov chains with no irreducibility assumptions. *Journal of Applied Probability*, 25: 275–285, 1988.

Fukang Zhu. A negative binomial integer-valued GARCH model. *J. Time Series Anal.*, 32(1):54–67, 2011. ISSN 0143-9782. doi: 10.1111/j.1467-9892.2010.00684.x. URL http://dx.doi.org/10.1111/j.1467-9892.2010.00684.x.