Modeling stationary data by classes of generalized Ornstein-Uhlenbeck processes.

Alejandra Cabaña
joint work with Argimiro Arratia and Enrique M. Cabaña

Universitat Autònoma de Barcelona,
Universitat Politècnica de Catalunya and
Universidad de la República

7èmes Journées Statistiques du Sud
Barcelona, 9-11 June, 2014
The link between discrete ARMA processes and stationary processes with continuous time has been of interest for many years and has been studied, among others, by

Durbin, J. (1961) Efficient fitting of linear models for continuous stationary time series from discrete data

and there is a recent upsurge of interest in non Gaussian processes mainly due to the fact that jumps play an important role in realistic modeling in finance and other fields of application.
Example
Series A (Box, Jenkins & Reinsel) consists of 197 lectures of concentration in a certain chemical process, taken every 2 hours.
It can be treated as a series of equally spaced observations of a process developing in continuous time, with possible long term dependence.
The classical model would be an ARMA$(p, q)$, an ARMA$(1,1)$ or subsets of AR(7) have been proposed.

Empirical covariances and covariances of the adjusted (ML) ARMA(11) and AR(7) models for Series A.
From AR(1) to Ornstein Uhlenbeck processes

The simplest ARMA model is an AR(1)

\[ X_t = \phi X_{t-1} + \sigma \epsilon_t \]

that can be written as

\[ (1 - \phi B)X_t = \sigma \epsilon_t \]

where \( \epsilon_t, \ t \in \mathbb{Z} \) is a white noise, \( B \) is the back-shift operator that maps \( X_t \) onto \( BX_t = X_{t-1} \). If \( |\phi| < 1 \), the process \( X_t \) is stationary.

Equivalently, \( X_t \) can be written as

\[ X_t = \sigma \mathcal{MA}(1/\rho)\epsilon_t \]

where \( \mathcal{MA}(1/\rho) \) is the moving average that maps \( \epsilon_t \) onto \( \mathcal{MA}(1/\rho)\epsilon_t, = \sum_{j=0}^{\infty} \frac{1}{\rho^j} \epsilon_{t-j} \).

The covariances of \( X_t \) are

\[ \gamma_h = \mathbf{E}X_tX_{t+h} = \frac{\gamma_0}{\rho^h} \quad \text{where} \quad \gamma_0 = \frac{\sigma^2}{1 - \frac{1}{\rho^2}} \]
Two ways of defining a continuous time analogue $x_t$, $t \in \mathbb{R}$ of AR(1) processes are

- by establishing that $\gamma(h) = \mathbb{E}x(t)x(t + h)$ be $\gamma_0 e^{-\kappa|h|}$

- by replacing the measure $W$ concentrated on the integers defined by $W(A) = \sum_{t \in A} \epsilon_t$, that allows writing

$$X_t = \int_{-\infty}^{t^+} \frac{1}{\rho^{t-s}} dW(s),$$

by a measure $\Lambda$ on $\mathbb{R}$, with stationary, i.i.d. increments and defining (with $\rho = e^\kappa$)

$$x(t) = \int_{-\infty}^{t} e^{-\kappa(t-s)} d\Lambda(s) \quad \Re(\kappa) > 0$$

Both ways lead to the same result: Ornstein-Uhlenbeck type processes.
Lévy driven continuous time ARMA processes

A Lévy process $\Lambda(t)$ is a càdlàg function, with independent and stationary increments, that vanishes in $t = 0$.

As a consequence, $\Lambda(t)$ is, for each $t$, a random variable with an infinitely divisible law.

The characteristic function of $\Lambda(t)$ is $\mathbb{E}e^{iu\Lambda(t)} = (\mathbb{E}e^{iu\Lambda(1)})^t$, and is usually written as $\mathbb{E}e^{iu\Lambda(1)} = e^{\psi_\Lambda(iu)}$.

The function $\psi_\Lambda$ is called characteristic exponent and has the form

$$
\psi_\Lambda(iu) = aiu - \frac{\sigma^2}{2}u^2 + \int_{|x|<1} (e^{iux} - 1 - iux)dv(x) + \int_{|x|\geq1} (e^{iux} - 1)dv(x)
$$

where $\nu(\{0\}) = 0$, $\int_{|x|<1} x^2dv(x) < \infty$, $\int_{|x|\geq1} dv(x) < \infty$. 
Wiener process $w$ satisfies these properties, and, moreover, is the unique continuous Lévy process.

The compound Poisson process with rate $\lambda$ and i.i.d. jumps $Y_j$ with $\mathbb{E}Y_j = 0, \text{Var}(Y_j) = \eta < \infty$ is also a Lévy process.
Lévy-driven Ornstein-Uhlenbeck processes of higher order

In the same manner that higher order autoregressive processes are used for modeling stationary sequences, higher order Ornstein-Uhlenbeck processes can be used for modeling stationary continuous time processes.

Econometric or physical models apply frequently linear combinations (superpositions) of OU processes driven by either uncorrelated or correlated noise.

\[
\sum_{j=1}^{p} a_j \int_{-\infty}^{t} e^{-\kappa_j(t-s)} d\Lambda_j(s)
\]

Eliazar, I, and Klafter, J. (2009), From Ornstein-Uhlenbeck dynamics to long-memory processes and fractional Brownian motion, PHYSICAL REVIEW E, 79
Levy-driven continuous time ARMA

or models that replace the finite linear combination by a continuous version

$$\int_{s=-\infty}^{t} \int_{\mathbb{R}(\kappa)>0} e^{-\kappa(t-s)} d\Lambda(s, \kappa)$$

Barndorff-Nielsen, O. and Shephard, N. (2001), Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics JRSS, 63.

Bergstrom, A. R. (1984), Continuous time stochastic models and issues of aggregation over time, in Handbook of Econometrics, Volume ll, Edited by Z. Griliches and M. D. Intriligator, Elsevier Science Publishers BV.


Chambers, M. J. and Thornton, M.A. (2012), Discrete time representation of continuous time ARMA processes
Iterated Levy-driven Ornstein-Uhlenbeck processes

We propose to use a parsimonious model, with few parameters, that is able to adjust slowly decaying covariances, obtained by a procedure that resembles the one that allows to build an AR($p$) from an AR(1).

The AR($p$) process

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \sigma \epsilon_t \quad \text{or} \quad \phi(B)X_t = \sigma \epsilon_t,$$

where $\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j = \prod_{j=1}^{p} (1 - z/\rho_j)$ has roots $\rho_j = e^{\kappa_j}$

is obtained by applying the composition of the moving averages $\mathcal{MA}(1/\rho_j)$ to the noise, that is:

$$X_t = \sigma \prod_{j=1}^{p} \mathcal{MA}(1/\rho_j) \epsilon_t$$
Let us denote $\mathcal{M}A_{\kappa} = \mathcal{M}A(e^{-\kappa})$. A continuous version of the operator $\mathcal{M}A_{\kappa}$ that maps $\epsilon_t$ onto

$$\mathcal{M}A_{\kappa}\epsilon_t = \sum_{l \leq t, \text{integer}} e^{-\kappa(t-l)}\epsilon_l.$$ 

is $\mathcal{O}U_{\kappa}$ that maps $y(t)$ onto

$$\mathcal{O}U_{\kappa}y(t) = \int_{-\infty}^{t} e^{-\kappa(t-s)} dy(s)$$

and this suggests the use of the model $\mathcal{O}U(p)$:

$$x_{\kappa,\sigma}(t) = \sigma \prod_{j=1}^{p} \mathcal{O}U_{\kappa_j} \Lambda(t)$$

with parameters $\kappa = (\kappa_1, \ldots, \kappa_p), \sigma$. 

---

**The OU$_p$**
OU($p$) as a superposition of OU(1)

The *Ornstein-Uhlenbeck process* with parameters $\kappa = (\kappa_1, \ldots, \kappa_p)$ and $\sigma$, 

$$x_{\kappa,\sigma} = \prod_{j=1}^{p} \text{OU}_{\kappa_j} (\sigma \Lambda)$$

can be written as a linear combination of $p$ processes of order 1 when the components of $\kappa$ are pairwise different:

$$x_{\kappa,\sigma} = \sum_{j=1}^{p} K_j(\kappa) \xi_{\kappa_j}, \quad \xi_{\kappa_j}(t) = \int_{-\infty}^{t} e^{-\kappa_j(t-s)} d(\sigma \Lambda(s)).$$

The coefficients are:

$$K_j(\kappa) = \frac{1}{\prod_{\kappa_l \neq \kappa_j} (1 - \kappa_l / \kappa_j)}$$
When $\kappa$ has componentes $\kappa_h$ repeated $p_h$ times ($h = 1, 2, \ldots, q$, $\sum_{h=1}^{q} p_h = 1$) the linear combination is:

$$x_{\kappa, \sigma} = \sum_{h=1}^{q} K_h(\kappa) \sum_{j=0}^{p_h-1} \binom{p_h-1}{j} \xi^{(j)}_{\kappa_h}$$

where

$$\xi^{(j)}_{\kappa_h}(t) = \int_{-\infty}^{t} e^{-\kappa_h(t-s)} \frac{(-\kappa_h(t-s))^j}{j!} d(\sigma \Lambda(s))$$
The autocovariances of $x_{\kappa, \sigma}$ are

$$\gamma_{\kappa, \sigma}(t) = \sum_{h'=1}^{q} \sum_{i'=0}^{p_{h'-1}} \sum_{h''=1}^{q} \sum_{i''=0}^{p_{h''-1}} K_{h'}(\kappa) \overline{K}_{h''}(\kappa) \left( \frac{p_{h'-1}}{i'} \right) \left( \frac{p_{h''-1}}{i''} \right) \gamma_{\kappa_{h'}, \kappa_{h''}, \sigma}(t)$$

$$= \sigma^2 (-\kappa_1)^{i_1} (-\bar{\kappa}_2)^{i_2} \int_{-\infty}^{0} e^{-\kappa_1(t-s)} \frac{(t-s)^{i_1}}{i_1!} e^{-\bar{\kappa}_2(s)} \frac{(-s)^{i_2}}{i_2!} ds,$$

and when the components of $\kappa$ are pairwise different, the covariances can be written as

$$\gamma_{\kappa, \sigma}(t) = \sum_{h'=1}^{p} \sum_{h''=1}^{p} K_{h'}(\kappa) \overline{K}_{h''}(\kappa) \gamma_{\kappa_{h'}, \kappa_{h''}, \sigma}(t).$$
A state space representation of the OU($p$) process

The decomposition of the OU($p$) process

$$x_{\kappa,\sigma}(t)$$

as a linear combination of simpler processes of order 1, leads to an expression of the process by means of a state space model.

State space modeling provides us with

- a unified approach for computing the likelihood of $x_{\kappa,\sigma}(t)$ through a Kalman filter.
- a tool to show that the covariances of $x_{\kappa,\sigma}(t)$ coincide with those of an ARMA($p, p - 1$) whose coefficients can be computed from $\kappa$. 
In order to ease notation, we consider that the components of $\kappa$ are all different.
The decomposition of $x_{\kappa,\sigma}(t) = \sum_{j=1}^{p} K_j \xi_{\kappa_j}(t)$ as a linear combination of the OU(1) processes

$$
\xi_{\kappa_j}(t) = \int_{-\infty}^{t} e^{-\kappa_j(t-s)} d(\sigma \Lambda(s)) = e^{-\kappa_j} \xi_{\kappa_j}(t-1) + \int_{t-1}^{t} e^{-\kappa_j(t-s)} d(\sigma \Lambda(s))
$$

with innovations $\eta_{\kappa_j}$ with components

$$
\eta_{\kappa_j}(t) = \int_{t-1}^{t} e^{-\kappa_j(t-s)} d\Lambda(s)
$$

provides a representation of the OU($p$) process in the space of states $\xi_{\kappa} = (\xi_{\kappa_1}, \ldots, \xi_{\kappa_p})^{tr}$. 
The transitions in the state space are

\[ \xi_{\kappa}(t) = \text{diag}(e^{-\kappa_1}, \ldots, e^{-\kappa_p}) \xi_{\kappa}(t-1) + \eta_{\kappa}(t), \]

and

\[ x(t) = K^{tr}(\kappa) \xi(t) \]

The assumption \( \mathbb{E} \Lambda(1)^2 = 1 \) implies that the innovations have variance

\[ \text{Var}(\eta_{\kappa,\tau}(t)) = ((v_{j,l})) \]

\[ v_{j,l} = \mathbb{E} \int_{t-1}^{t} e^{-(\kappa_j + \bar{\kappa}_l)(t-s)} ds = \frac{1-e^{-(\kappa_j + \bar{\kappa}_l)}}{\kappa_j + \bar{\kappa}_l}. \]
OU(\(p\)) as an ARMA\((p, p - 1)\)

Now apply the AR operator \(\prod_{j=1}^{p} (1 - e^{-\kappa_j B})\) to \(x_\kappa\) and obtain

\[
\prod_{j=1}^{p} (1 - e^{-\kappa_j B}) x_\kappa(t) = \sum_{j=1}^{p} K_j G_j(B) \eta_{\kappa_j}(t) =: \zeta(t),
\]

with \(G_j(z) = \prod_{l \neq j} (1 - e^{-\kappa_l z}) := 1 - \sum_{l=1}^{p-1} g_{j,l} z^l\).

This process has the same second-order moments as the ARMA\((p, p - 1)\) \(^1\)

\[
\prod_{j=1}^{p} (1 - e^{-\kappa_j B}) x_\kappa(t) = \sum_{j=0}^{p-1} \theta_j \epsilon(t - j) =: \zeta'(t) \quad (\epsilon \text{ is a white noise})
\]

when the covariances \(c_j = \mathbf{E}\zeta(t)\bar{\zeta}(t - j)\) and \(c'_j = \mathbf{E}\zeta'(t)\bar{\zeta}'(t - j)\) coincide.

\(^1\)When \(\Lambda\) is a Wiener process, it is in fact an ARMA\((p, p - 1)\)
The covariances $c'_j$ and $c_j$ are given respectively by the generating functions

$$
\left( \sum_{h=0}^{p-1} \theta_h z^h \right) \left( \sum_{k=0}^{p-1} \bar{\theta}_k z^{-h} \right) = \sum_{l=-p+1}^{p-1} c'_l z^l
$$

and

$$
J(z) := \sum_{j=1}^{p} \sum_{l=1}^{p} K_j \bar{K}_l G_j(z) \bar{G}_l(1/z) v_{j,l} = \sum_{l=-p+1}^{p-1} c_l z^l.
$$

Since $J(z)$ can be computed once $\kappa$ is known, the coefficients $\theta = (\theta_0, \theta_1, \ldots, \theta_{p-1})$ are obtained by identifying the coefficients of the polynomials $z^{p-1} J(z)$ and $z^{p-1} \left( \sum_{h=0}^{p-1} \theta_h z^h \right) \left( \sum_{k=0}^{p-1} \bar{\theta}_k z^{-h} \right)$. 
OU($p$) as an ARMA($p, p-1$)

A state space representation and its implications on the covariances of the OU process in the general case are slightly more complicated.

\[
\begin{align*}
\xi(t) &= A\xi(t-1) + \eta(t) \\
\mathbf{x}(t) &= \mathbf{K}^{\text{tr}}\xi(t)
\end{align*}
\]

When $\kappa_1, \ldots, \kappa_q$ are all different, $p_1, \ldots, p_q$ are positive integers, $\sum_{h=1}^{q} p_h = p$ and $\mathbf{\kappa}$ is a $p$-vector with $p_h$ repeated components equal to $\kappa_h$, the OU($p$) process $\mathbf{x}_{\kappa}$ is a linear function of the state space vector

\[
\left( \xi_{\kappa_1}^{(0)}, \xi_{\kappa_1}^{(1)}, \ldots, \xi_{\kappa_1}^{(p_1-1)}, \ldots, \xi_{\kappa_q}^{(0)}, \xi_{\kappa_q}^{(1)}, \ldots, \xi_{\kappa_q}^{(p_q-1)} \right)
\]

and the transition equation is no longer expressed by a diagonal matrix.
Reparameterization

Our purpose is to insert the expression of the covariance in a numeric optimization procedure in order to compute the maximum likelihood estimates of the parameters.

Though $\gamma(t)$ depends continuously on $\kappa$, the same does not happen with each term in the expression for the covariance, because of the lack of boundedness of the coefficients of the linear combination when two different values of the components of $\kappa$ approach each other.

Since we wish to consider real processes $x$ and the process itself and its covariance $\gamma(t)$ depend only of the unordered set of the components of $\kappa$, we shall reparameterize the process.
With the notation $K_{j,i} = \frac{1}{(-\kappa_j)^i} \prod_{l \neq j} (1 - \kappa_l / \kappa_j)$ (in particular, $K_{j,0}$ is the same as $K_j$)

- The processes $x_i(t) = \sum_{j=1}^{p} K_{j,i} \xi_j(t)$ and the coefficients $\phi = (\phi_1, \ldots, \phi_p)$ of the polynomial

$$g(z) = \prod_{j=1}^{p} (1 + \kappa_j z) = 1 - \sum_{j=1}^{p} \phi_j z^j$$

satisfy

$$\sum_{i=1}^{p} \phi_i x_i(t) = x(t).$$

Therefore, the new parameter $\phi = (\phi_1, \ldots, \phi_p) \in \mathbb{R}^p$ shall be adopted.
ML estimation of the parameters of OU($p$) in the Gaussian case

From the observations $\{\mu + x(i) : i = 0, 1, \ldots, n\}$, obtain the likelihood $L$ of the vector $x = (x(1)), \ldots, x(n))$:

$$\log L(x; \phi, \sigma) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(\det(V(\phi, \sigma))) - \frac{1}{2} x^\text{tr}(V(\phi, \sigma))^{-1} x$$

with $V(\phi, \sigma)$ equal to the $n \times n$ matrix with components

$$V_{h,i} = \gamma(|h - i|)$$

that reduce to $\gamma(0)$ at the diagonal, $\gamma(1)$ at the 1st sub and super diagonals, ...

Obtain via numerical optimisation the MLE $\hat{\phi}$ of $\phi$ and $\hat{\sigma}^2$ of $\sigma^2$.

The estimations $\hat{\kappa}$ follow by solving $\prod_{j=1}^p (1 + \hat{\kappa}_j z) = 1 - \sum_{j=1}^p \hat{\phi}_j z^j$. 
Matching correlations estimation

From the closed formula for the covariance $\gamma$ and the relationship between $\kappa$ and $\phi$, we have a mapping $(\phi, \sigma^2) \mapsto \gamma(t)$, for each $t$. Since $\rho^{(T)} := (\rho(1), \ldots, \rho(T))^\text{tr} = (\gamma(1), \ldots, \gamma(T))^\text{tr} / \gamma(0)$ does not depend on $\sigma^2$, these equations determine a map

$$\mathcal{C} : (\phi, T) \mapsto \rho^{(T)} = \mathcal{C}(\phi, T)$$

for each $T$.

After choosing a value of $T$ and obtaining an estimate $\rho_e^{(T)}$ of $\rho^{(T)}$ based on $x$, we propose as a first estimate of $\phi$, the vector $\tilde{\phi}_T$ such that all the components of the corresponding $\kappa$ have positive real parts, and such that the euclidean norm $\|\rho_e^{(T)} - \mathcal{C}(\tilde{\phi}_T, T)\|$ reaches its minimum, that is, a procedure that resembles the method of moments.
The Gaussian case: examples

When $\Lambda$ is a Wiener process, the OU process of order $p$ belongs to a subclass with $p+1$ parameters of the classical family of the $2p$-parameters Gaussian ARMA($p$, $p - 1$)

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \theta_0 \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_{p-1} \epsilon_{t-p+1}$$

where $\phi_1, \ldots, \phi_p$ i $\theta_0, \ldots, \theta_q$ are parameters and $\epsilon_t$ is a Gaussian noise with variance 1.
Examples

Gaussian case

Maximum likelihood estimation of $\kappa$ and $\sigma$

The parameters $\kappa, \sigma$ determine the Gaussian likelihood of $OU_{\kappa,\sigma}w$, and are estimated by the values $\hat{\kappa}$ and $\hat{\sigma}$ that maximize that likelihood.

The matching correlations estimators can be used as the starting point of an optimization procedure leading to compute the ML estimators.

We have observed in several examples that the covariances of the process with the maximum likelihood estimators as parameters, follow closely the empirical covariances of the series. We have simulated the sample paths for the Wiener-driven $OU(p)$ for different values of the parameters.
Examples

Gaussian case

\[ n=500 \quad \text{kappa} = 0.04 \quad 0.21 \quad 1.87 \]
Examples

Gaussian case

\[ n=500 \quad \kappa = 0.04 \quad 0.21 \quad 1.87 \]
Examples

Gaussian case

n=500 \( \kappa = 0.04, 0.21, 1.87 \)

Original ○  MCE●  MLE●
Examples

Gaussian case

Original
Examples
Gaussian case

Original ○ Empirical +
Examples

Gaussian case

Original ○ Empirical + MCE●
Examples

- Gaussian case

Original ○ Empirical + MCE ● MLE ●
Examples

Gaussian case

$n = 500$ $\kappa = 0.2 + 0.4i$ $0.2 - 0.4i$ $0.9 + 0i$

Original ○ MCE ○ MLE ○
Examples

Gaussian case

Original ◦ Empirical + MCE ○ MLE ○
A series \((x_i)_{i=0,1,...,n}\) of \(n = 300\) observations of the OU\(\kappa\) process \(x\), \(p = 3\)

<table>
<thead>
<tr>
<th></th>
<th>(\kappa)</th>
<th>(\phi)</th>
<th>(\kappa)</th>
<th>(\sigma^2 = 1)</th>
<th>(\sigma^2 = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>original (\kappa)</td>
<td>0.9</td>
<td>0.2 + 0.4(i)</td>
<td>0.2 − 0.4(i)</td>
<td>0.9001</td>
<td>0.2273 + 0.4582(i)</td>
</tr>
<tr>
<td>original (\phi)</td>
<td>−1.30</td>
<td>−0.56</td>
<td>−0.18</td>
<td>1.6368</td>
<td>0.1439 + 0.4196(i)</td>
</tr>
<tr>
<td>MCE(\phi)(T)</td>
<td>−1.9245</td>
<td>−0.6678</td>
<td>−0.3221</td>
<td>0.9001</td>
<td>0.2273 + 0.4582(i)</td>
</tr>
<tr>
<td>(\hat{\kappa})</td>
<td>0.9001</td>
<td>0.2273 + 0.4582(i)</td>
<td>0.2273 − 0.4582(i)</td>
<td>0.9001</td>
<td>0.2273 + 0.4582(i)</td>
</tr>
<tr>
<td>MLE (\phi)</td>
<td>−1.3546</td>
<td>−0.6707</td>
<td>−0.2355)</td>
<td>(\hat{\sigma}^2 = 0.8958)</td>
<td>(\hat{\sigma}^2 = 0.8958)</td>
</tr>
</tbody>
</table>
Series A (Box, Jenkins & Reinsel) consists of 197 lectures of concentration in a certain chemical process, taken every 2 hours.
The parameters $\phi$ and $\theta$ can be obtained from $\kappa$ and $\sigma$ using the expressions for the covariances of both processes. For Series A, the parameters of the OU(3) fitted by maximum likelihood are

$$\hat{\kappa} = (0.8293, 0.0018 + 0.0330i, 0.0018 - 0.0330i) \quad \text{and} \quad \hat{\sigma} = 0.4401$$

The corresponding ARMA(3,2) is

$$(1 - 2.4316B + 1.8670B^2 - 0.4348B^3)x = 0.4401(1 - 1.9675B + 0.9685B^2) \epsilon$$

On the other hand, the ARMA(3,2) fitted by maximum likelihood is

$$(1 - 0.7945B - 0.3145B^2 + 0.1553B^3)x = 0.3101(1 - 0.4269B - 0.2959B^2) \epsilon.$$
Examples

Gaussian case

ARMA (AIC 110.46) and OU(3) (AIC 109.9) fitted by maximum likelihood
Empirical correlations and adjusted by maximum likelihood
OU (—) (AIC 109.9) and FOU (AIC 106.6)(····)
Non Gaussian case: Estimating the shape of the noise

The estimation of the parameters of $\psi_\Lambda$, two real numbers and a measure (the so called \textit{Lévy-Hinčin triplet}), is difficult and requires a large amount of information.

Jongbloed, van der Meulen and van der Vaart (Bernoulli 11(5), 2005, 759–791) have proposed nonparametric estimation for the Lévy noise driving an Ornstein-Uhlenbeck process.

A simpler setting is to assume that the admissible exponents belong to a parametric class $\Psi = \{\psi_\theta : \theta \in \Theta\}$ where $\Theta \subset \mathbb{R}^d$, and obtain the value of $\theta$ for which a chosen quadratic distance between the exponential of $\psi_\theta(iu)$ and the empirical characteristic function of the residuals is minimum.
Examples

Non Gaussian case: Estimating the shape of the noise

Compound PP of intensity 2 and Gamma(2,2) jumps, and the corresponding OU(1)
Let us denote $\psi_\Lambda(iu)$ the characteristic exponent of the Lévy process $\Lambda$,

$$\psi_\Lambda(iu) = \log \mathbb{E} e^{iu\Lambda(1)}$$

The innovation in each component $\xi_j$ is

$$\eta_j(t) = \int_{t-1}^{t} e^{-\kappa_j(t-s)} d\Lambda(s),$$

so that the innovation of $x_{\kappa,\sigma}$ is

$$\eta(t) = \int_{t-1}^{t} g(t-s) d\Lambda(s),$$

where $g(t) = \sum_{j=1}^{p} K_j e^{-\kappa_j t}$.

Hence,

$$\eta \sim \int_{0}^{1} g(1-s) d\Lambda(s) \sim \int_{0}^{1} g(s) d\Lambda(s)$$

and its characteristic exponent is therefore

$$\psi_\eta = \log \mathbb{E} e^{iu\eta} = \log \mathbb{E} e^{iu \int_{0}^{1} g(s) d\Lambda(s)} = \int_{0}^{1} \psi_\Lambda(iug(s)) ds$$
A simple example: estimation of a noise sum of a Poisson process plus a Gaussian term

Let us assume that the noise is given by

$$\Lambda(t) = \sigma w(t) + c(N(t) - \lambda t)$$

where $w$ is a standard Wiener process and $N$ is a Poisson process with intensity $\lambda$. The family of possible noises depends on the three parameters ($\sigma, \lambda, c$). In this case, the characteristic exponent has a simple form:

$$\psi_{\Lambda(1)}(iu) = -\frac{\sigma^2 u^2}{2} + \lambda(e^{iuc} - iuc - 1)$$

hence

$$\psi_{\eta}(iu) = \int_0^1 \left(-\frac{\sigma^2 u^2 g^2(s)}{2} + \lambda(e^{iug(s)c} - iug(s)c - 1)\right) ds$$
With $g_h = \int_0^1 g^h(s)ds$,

$$
\psi \eta(iu) = -\frac{\sigma^2 u^2 g_2}{2} + \lambda \left( -\frac{u^2 g_2 c_2^2}{2} - i\frac{u^3 g_3 c_3^3}{6} + \frac{u^4 g_4 c_4^4}{24} + \ldots \right)
$$

Then we propose to estimate the parameters by equating the coefficients of $u^2, u^3, u^4$ in $\psi \eta(iu)$ with the corresponding ones in the logarithm of the empirical characteristic function of the residuals. Assuming that the mean of the residuals $r_1, r_2, \ldots, r_n$ is zero, their empirical characteristic function is

$$
\frac{1}{n} \sum_{h=1}^{n} e^{iur_h} = 1 - \frac{1}{2} u^2 R_2 - \frac{1}{6} iu^3 R_3 + \frac{1}{24} u^4 R_4 + \ldots
$$

where $R_m = \frac{1}{n} \sum_{h=1}^{n} r_h^m$. 
Then the logarithm has the expansion
\[ \log \frac{1}{n} \sum_{h=1}^{n} e^{iur_h} = -\frac{1}{2} u^2 R_2 - \frac{1}{6} i u^3 R_3 + \frac{1}{24} u^4 R_4 - \frac{1}{8} u^4 R_2^2 + \ldots \]
Consequently, the estimation equations are
\[ (\sigma^2 + \lambda c^2) g_2 = R_2, \lambda c^3 g_3 = R_3, \lambda c^4 g_4 = R_4 - 3R_2^2 \]
from which the estimators follow:
\[ \tilde{c} = \frac{R_4 - 3R_2^2}{R_3} \frac{g_3}{g_4}, \quad \tilde{\lambda} = \frac{R_3^4}{(R_4 - 3R_2^2)^3} \frac{g_4^3}{g_3^4} \]
\[ \tilde{\sigma}^2 = \frac{R_2^2}{g_2} - \frac{R_3^2}{(R_4 - 3R_2^2)} \frac{g_4}{g_3^2}. \]
Next figures show the empirical c.d.f. of 90 estimators of the parameters obtained from simulated series of 200 terms. The residuals were obtained by applying a Kalman filter to the space state formulation, starting from the actual value of $\kappa$ used at the simulation (red), that in practical situations is unknown, and from the estimators obtained by matching correlations (green) and by maximum likelihood (blue).

Estimation of the parameters of the noise from 90 replications of $\{x_\kappa(t) : t = 0, 1, \ldots, 200\}$, $\kappa = (0.01 \pm 0.1i, 0.2)$, driven by $\Lambda(t) = 0.1w(t) + N_{0.3}(t) - 0.3t$.\(^2\)

The estimators are not sharp at all, but the ones obtained by the same procedure applied directly on the unfiltered noise $\Lambda$ (dashed lines) are equally rough. Larger series (of size 10000 and 1000000) produce sharper estimates, also shown in the figures by dotted lines.

\(^2\)Normality is rejected in 100% of all cases.
Examples
Non Gaussian case: Estimating the shape of the noise

Estimation of the parameters of the noise from 90 replications of \( \{ x_\kappa(t) : t = 0, 1, 2, \ldots, 200 \} \), \( \kappa = (0.0018 + 0.033i, 0.0018 - 0.033i, 0.083) \), driven by \( \Lambda(t) = w(t) + N_1(t) - t \).\(^3\) \( \kappa \) used at the simulation (red), estimated by matching correlations (green) and by maximum likelihood (blue).

\(^3\) Normality is rejected in 30%, 36% and 36% of the cases.
Conclusions

- We have proposed a family of continuous time stationary processes, based on $p$ iterations of the linear operator that maps a Wiener process onto an Ornstein-Uhlenbeck process, or more generally, wrt Levy Processes.
- These operators have some nice properties, such as being commutative, and their $p$-compositions decompose as a linear combination of simple operators of the same kind.
- An OU($p$) process depends on $p + 1$ parameters that can be easily estimated by either maximum likelihood (ML) or matching correlations (MC) procedures. Matching correlation estimators provide a fair estimation of the covariances of the data, even if the model is not well specified.
Conclusions

• When sampled on equally spaced instants, the OU($p$) family can be written as a discrete time state space model; i.e., a VARMA model in a space of dimension $p$. As a consequence, the families of OU($p$) models are a parsimonious subfamily of the ARMA($p, p-1$) processes in the Gaussian case.

• Furthermore, the coefficients of the ARMA can be deduced from those of the corresponding OU($p$).

• We have shown examples for which the ML-estimated OU model is able to capture a long term dependence that the ML-estimated ARMA model does not show. This leads to recommend the inclusion of OU models as candidates to represent stationary series to the users interested in such kind of dependence.