Chapter 13: Functional Autoregressive Models

Jakub Černý

Department of Probability and Mathematical Statistics

Stochastic Modelling in Economics and Finance

December 9, 2013
Contents

1 Introduction

2 Existence and uniqueness of the solution

3 Estimation of the operator
Causal model

Definition

Let \( \{\varepsilon_n, n \in \mathbb{Z}\} \) be a white noise and let \( \{c_j, j \in \mathbb{N}_0\} \) be a sequence of constants such that \( \sum_{j=0}^{\infty} |c_j| < \infty \). A random sequence \( \{X_n, n \in \mathbb{Z}\} \) defined as

\[
X_n = \sum_{j=0}^{\infty} c_j \varepsilon_{n-j}, \quad n \in \mathbb{Z}
\]

is called causal linear process.

Above definition briefly says that causal linear processes depend only on the present and past errors, not the future ones.
AR model

Definition

Random sequence \( \{X_n, n \in \mathbb{Z}\} \) is called an autoregressive model of order \( p, \text{AR}(p) \), if

\[
X_n = \sum_{i=1}^{p} \phi_i X_{n-i} + \varepsilon_n
\]

where \( \phi_1, \ldots, \phi_p \) are real constants, \( \phi_p \neq 0 \) and \( \{\varepsilon_n, n \in \mathbb{Z}\} \) is a white noise.

In general, the definition of the AR model is based on the white noise, but [Horváth, Kokoszka (2011)] assume that \( \{\varepsilon_n, n \in \mathbb{Z}\} \) is a sequence of iid mean zero random variables which is a narrower class of processes.
If \( c_j = \psi^j \) and \(|\psi| < 1\), AR(1) model \( \{X_n, n \in \mathbb{Z}\} \) is causal and has a unique solution in form

\[
X_n = \sum_{j=0}^{\infty} \psi^j \varepsilon_{n-j}
\]

Our goal is to use AR(1) model condition \(|\psi| < 1\) and establish analogous condition for the functional model.
To lighten the notation we set $\| \cdot \|_\mathcal{L} = \| \cdot \|$.

**Definition**

We say that a sequence $\{X_n, n \in \mathbb{Z}\}$ of mean zero elements of $L^2$ follows a *functional AR(1) (FAR(1)) model* if

$$X_n = \Psi(X_{n-1}) + \varepsilon_n \quad (1.1)$$

where $\Psi \in \mathcal{L}$ and $\{\varepsilon_n, n \in \mathbb{Z}\}$ is a sequence of iid mean zero errors in $L^2$ satisfying $E\|\varepsilon_n\|^2 < \infty$.

Following theory is developed under the assumption of iid errors. General case can be found in [Bosq (2000)].
Existence - Lemma

**Lemma**

*For any $\Psi \in \mathcal{L}$, the following two conditions are equivalent:*

- **C1** There exists an integer $j_0 \geq 0$ such that $\|\Psi^{j_0}\| < 1$.
- **C2** There exists $a > 0$ and $0 < b < 1$ such that for every $j \geq 0$, $\|\Psi^j\| \leq ab^j$.

**Proof.** The condition C2 clearly implies C1. So we must show only that C1 implies C2.

We set $j = qj_0 + r$ where $q \geq 0$ and $0 \leq r < j_0$. Therefore,

$$\|\Psi^j\| = \|\Psi^{qj_0+r}\| \leq \|\Psi^{j_0}\|^q \|\Psi^r\|.$$  

If $\|\Psi^{j_0}\| = 0$, then C2 holds for any $a > 0$ and $0 < b < 1$, so we assume in the following that $\|\Psi^{j_0}\| > 0$. 


Since $q > j/j_0 - 1$ and $\|\psi_j\| < 1$, we get

$$\|\psi^j\| \leq \|\psi^j_0\|^{j/j_0 - 1}\|\psi^r\| \leq \left(\|\psi^j_0\|^{1/j_0}\right)^j \|\psi^j_0\|^{-1} \max_{0 \leq r < j_0} \|\psi^r\|,$$

so C2 holds with $a = \|\psi^j_0\|^{-1} \max_{0 \leq r < j_0} \|\psi^r\|$, $b = \|\psi^j_0\|^{1/j_0}$.

\[\square\]

Note that condition C1 is weaker than the condition $\|\psi\| < 1$. Nevertheless, C2 is a sufficiently strong condition to ensure the convergence of the series $\sum_j \psi^j(\varepsilon_{n-j})$ and the existence of a stationary causal solution to FAR(1) equations, as stated in the following theorem.
If condition C1 holds, then there is a unique strictly stationary causal solution to 1.1. This solution is given by

$$X_n = \sum_{j=0}^{\infty} \Psi^j(\varepsilon_{n-j})$$  \hspace{1cm} (2.1)

The series converges almost surely, and in the $L^2$ norm, i.e

$$\mathbb{E} \left\| X_n - \sum_{j=0}^{m} \Psi^j(\varepsilon_{n-j}) \right\|^2 \to 0, \quad m \to \infty.$$
Existence and uniqueness - Proof

Proof. Let us remind that we work with $L^2(\Omega, L^2([0,1]))$ which is a Hilbert space with the inner product $E\langle X, Y \rangle$, $X, Y \in L^2([0,1])$. To show that the sequence $X_n^{(m)} = \sum_{j=0}^{m} \psi^j(\varepsilon_{n-j})$ has a limit in $L^2(\Omega, L^2([0,1]))$ it suffices to check that it is a Cauchy sequence in $m$ for every fixed $n$.

$$E \left\| \sum_{j=m+1}^{m'} \psi^j(\varepsilon_{n-j}) \right\|^2 = \sum_{j,k=m+1}^{m'} E \langle \psi^j(\varepsilon_{n-j}), \psi^k(\varepsilon_{n-k}) \rangle$$

$$= \sum_{j=m+1}^{m'} E \left\| \psi^j(\varepsilon_{n-j}) \right\|^2.$$
Existence and uniqueness - Proof cont.

Therefore by the previous lemma

\[ E \left( \sum_{j=m+1}^{m'} \psi_j(\varepsilon_{n-j}) \right)^2 \leq \left( \sum_{j=m+1}^{m'} \|\psi_j\|^2 \right) E\|\varepsilon_0\|^2 \leq E\|\varepsilon_0\|^2 a^2 \sum_{j=m+1}^{m'} b^{2j}. \]

Thus \( X_n^{(m)} \) converges in \( L^2(\Omega, L^2([0, 1])) \), it is enough to verify that

\[ \sum_{j=0}^{\infty} \|\psi_j(\varepsilon_{n-j})\| < \infty \text{ a.s.} \]

By the condition C2

\[ E \left( \sum_{j=0}^{\infty} \|\psi_j\|\|\varepsilon_{n-j}\| \right)^2 \leq \sum_{j,k=0}^{\infty} \|\psi_j\|\|\psi_k\| E\|\varepsilon_0\|^2 \leq E\|\varepsilon_0\|^2 \left( \sum_{j=0}^{\infty} ab^j \right)^2 \]

which is finite and so \( \sum_{j=0}^{\infty} \|\psi_j\|\|\varepsilon_{n-j}\| < \infty \text{ a.s.} \)
The series $X_n$ is strictly stationary, and it satisfies (1.1). Suppose \{\{X'_n\}\} is another strictly stationary casual sequence satisfying (1.1). Then, iterating 1.1, we obtain, for any $m \geq 1$,

$$X'_n = \sum_{j=1}^{m} \psi^j(\varepsilon_{n-j}) + \psi^{m+1}(X'_{n-m+1}).$$

Therefore

$$E\|X'_n - X_n^{(m)}\| \leq \|\psi^{m+1}\| E\|X'_{n-m+1}\| \leq E\|X_0\| ab^{m+1}.$$ 

Hence $X'_n$ is equal a.s. to the limit of $X_n^{(m)}$, i.e. to $X_n$. 

□
Consider an integral Hilbert-Schmidt operator on $L^2$ defined by

$$\Psi(x)(t) = \int \psi(t, s)x(s)ds, \ x \in L^2,$$

which satisfies

$$\int \int \psi^2(t, s)dtds < 1. \quad (2.2)$$

From the introduction ([Horváth, Kokoszka (2011), chapter 2.2]) we know that the left-hand side of above inequality is equal to $\|\Psi\|_S^2$. Since $\|\psi\| \leq \|\Psi\|_S$, we see that (2.2) implies condition C1 with $j_0 = 1$. 

First, we will focus on the AR(1) model

\[ X_n = \psi X_{n-1} + \varepsilon_n. \]

We assume that \(|\psi| < 1\) so there is a stationary solution. Then, multiplying previous equation by \(X_{n-1}\) and taking the expected value, we get \(\gamma_1 = \psi \gamma_0\), where \(\gamma_k = \text{Cov}(X_n, X_{n-k}) = E(X_nX_{n-k})\). The usual estimator of \(\psi\) is \(\hat{\psi} = \hat{\gamma}_1/\hat{\gamma}_0\) where \(\hat{\gamma}_k\) are sample autocovariances. This approach is known as the Yule-Walker estimation.
Estimation of the operator $\Psi$

We want to apply this technique to the functional model. By 1.1 and under condition C1

$$E \left[ \langle X_n, x \rangle X_{n-1} \right] = E \left[ \langle \psi(X_{n-1}), x \rangle X_{n-1} \right], \quad x \in L^2.$$  

The lag-1 autocovariance operator is defined by

$$C_1(x) = E \left[ \langle X_n, x \rangle X_{n+1} \right].$$

and denote $\bullet^T$ the adjoint operator, then $C_1^T = C \psi^T$, i.e.

$$C_1 = \psi C$$

The above identity is analogous to the AR(1) process, so we would like to obtain an estimate of $\psi$ by using a finite sample version of the following relation

$$\psi = C_1 C^{-1}.$$
Estimation of the operator $\Psi$ - operator $C$

However, the operator $C$ does not have a bounded inverse on the whole $H$. Recall that $C^{-1}(C(x)) = x$, where

$$
C^{-1}(y) = \sum_{j=1}^{\infty} \lambda_j^{-1} \langle y, v_j \rangle v_j.
$$

The operator $C^{-1}$ is defined if all $\lambda_j$ are positive. Since $\|C^{-1}(v_n)\| = \lambda_n^{-1} \to \infty$, as $n \to \infty$ it is unbounded. This makes the estimation of the bounded operator $\Psi$ using the relation $\Psi = C_1 C^{-1}$ difficult.

Note that if $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > \lambda_{p+k} = 0$ for all $k = 1, 2, \ldots$ then $\{X_n\}$ is in the space spanned by $\{v_1, \ldots, v_p\}$. On this subspace we define $C^{-1}(y) = \sum_{j=1}^{p} \lambda_j^{-1} \langle y, v_j \rangle v_j$. 
A practical solution is to use only $p$ most important EFPC’s $\hat{v}_j$, and to define

$$\hat{IC}_p(x) = \sum_{j=1}^{p} \hat{\lambda}_j^{-1} \langle x, \hat{v}_j \rangle \hat{v}_j$$

The operator $\hat{IC}_p$ is defined on the whole of $L^2$, and it is bounded if $\hat{\lambda}_j > 0$ for $j \leq p$.

By reasonable choice of $p$ we can find a balance between retaining the relevant information in the sample, and the danger of working with the small eigenvalues.
To derive a computable estimator of $\Psi$, we will use an empirical version of $C_1 = \Psi C$. $C_1$ is estimated by

$$\hat{C}_1(x) = \frac{1}{N-1} \sum_{k=1}^{N-1} \langle X_k, x \rangle X_{k+1},$$

we get, for any $x \in L^2$,

$$\hat{C}_1 \hat{C}_p(x) = \hat{C}_1 \left( \sum_{j=1}^{p} \hat{\lambda}_j^{-1} \langle x, \hat{v}_j \rangle \hat{v}_j \right)$$

$$= \frac{1}{N-1} \sum_{k=1}^{N-1} \left( \sum_{j=1}^{p} \hat{\lambda}_j^{-1} \langle x, \hat{v}_j \rangle \hat{v}_j \right) X_{k+1}$$

$$= \frac{1}{N-1} \sum_{k=1}^{N-1} \sum_{j=1}^{p} \hat{\lambda}_j^{-1} \langle x, \hat{v}_j \rangle \langle X_k, \hat{v}_j \rangle X_{k+1}.$$
The estimator $\hat{C}_1 \hat{IC}_p$ can be used in principle, but typically an additional smoothing step is introduced by using the approximation

$$X_{k+1} \approx \sum_{i=1}^{p} \langle X_{k+1}, \hat{v}_i \rangle \hat{v}_i.$$  

From this follows

$$\hat{\Psi}_p(x) = \frac{1}{N-1} \sum_{k=1}^{N-1} \sum_{j=1}^{p} \sum_{i=1}^{p} \hat{\lambda}_j^{-1} \langle x, \hat{v}_j \rangle \langle X_k, \hat{v}_j \rangle \langle X_{k+1}, \hat{v}_i \rangle \hat{v}_i.$$  

The estimator is consistent if $p = p_N$ is a function of the sample size $N$. Sufficient conditions for $\|\hat{\Psi}_p - \Psi\| \to 0$ can be found in [Bosq (2000)].
The estimator is a kernel operator with the kernel $\hat{\psi}_p$ such that

$$\hat{\Psi}_p(x)(t) = \int \hat{\psi}_p(t, s)x(s)ds$$

from which follows that

$$\hat{\psi}_p(t, s) = \frac{1}{N - 1} \sum_{k=1}^{N-1} \sum_{j=1}^{p} \sum_{i=1}^{p} \lambda_j^{-1} \langle X_k, \hat{v}_j \rangle \langle X_{k+1}, \hat{v}_i \rangle \hat{v}_j(s)\hat{v}_i(t).$$

The kernel estimation can be obtained using R function `pca.fd` which is contained in a `fda` package (Functional Data Analysis). More about this package can be found at http://cran.r-project.org/web/packages/fda/fda.pdf.
Estimation of the operator $\Psi$ - Example

Figure: The kernel surface $\psi(t, s)$ (top left) and its estimates $\hat{\psi}(t, s)$ for $p = 2, 3, 4$ (Source: [Horváth, Kokoszka (2011), p. 241])
Estimation of the operator $\Psi$ - Example cont.

In previous figure are illustrated Gaussian kernel

$$\psi(t, s) = \alpha \exp\{- (t^2 + s^2)/2\}$$

with $\alpha$ chosen so that the Hilbert-Schmidt norm is $1/2$, and three estimates (series length $N = 100$) for $p = 2, 3, 4$ and $\varepsilon_n$ were generated as Brownian bridges. We may see very large discrepancies in magnitude and in shape (which were observed for other kernels and innovations).

Moreover, by any reasonable distance measure between two surfaces, the distance between $\psi$ and $\hat{\psi}_p$ increases as $p$ increases. This is counterintuitive because by using more EFPC’s $\hat{v}_j$, we would expect the approximation to improve.
The sums $\sum_{j=1}^{p} \hat{\lambda}_j$ explain 74, 83 and 87% of the variance for $p = 2, 3, 4$, but the absolute deviation distances between $\psi$ and $\hat{\psi}_p$ are 0.4, 0.44 and 0.55. The same pattern was obtained for the RMSE distance $\|\hat{\psi} - \psi\|_S$. As $N$ increases, these distances decrease, but their tendency to increase with $p$ remains.

This problem is partially due to the fact that for many FAR(1) models, the estimated eigenvalues $\hat{\lambda}_j$ are very small, except $\hat{\lambda}_1$ and $\hat{\lambda}_2$, and small error in their estimation means large error in the rest $\hat{\lambda}_j^{-1}$. Partial solution to this problem can be found in [Kokoszka, Zhang (2010)] by adding a positive baseline to the $\hat{\lambda}_j$.

However, as we shall further see, precise estimation of the kernel $\psi$ is not necessary to obtain satisfactory predictions.
References


Thank you for your attention!

Jakub Černý