



Universidade de Vigo

Traballo Fin de Mestrado

---

# Modelos cuantís autorregresivos

---

Silvia Novo Díaz

2016–2017

Mestrado en Técnicas Estatísticas



# Proposta de Trabajo Fin de Mestrado

|   |
|---|
| <b>Título en galego:</b> Modelos cuantís autorregresivos  |
| <b>Título en español:</b> Modelos cuantiles autorregresivos   |
| <b>English title:</b> Quantile autoregression models  |
| <b>Modalidade:</b> Modalidad A  |
| <b>Autor/a:</b> Silvia Novo Díaz, Universidade de Santiago de Compostela  |
| <b>Director/a:</b> César A. Sánchez Sello, Universidade de Santiago de Compostela   |
| <b>Breve resumo do traballo:</b><br>El trabajo consiste en revisar los modelos cuantiles autoregresivos, en los cuales se pretende describir la evolución de un cuantil de la variable de interés, teniendo en cuenta los valores precedentes de la propia variable. Además, se ilustrarán los métodos estudiados con una aplicación a los datos de empleo de Estados Unidos. La referencia fundamental de la que partiremos será: Koenker, R. y Xiao, Z. (2006). Quantile autoregression. Journal of the American Statistical Association, 101, pp. 980-990. |
| <b>Outras observacións:</b>   |



Don César A. Sánchez Sello, profesor da Universidade de Santiago de Compostela, informa que o Traballo Fin de Mestrado titulado

**Modelos cuantís autorregresivos**

foi realizado baixo a súa dirección por dona Silvia Novo Díaz para o Mestrado en Técnicas Estatísticas. Estimando que o traballo está rematado, dá a súa conformidade para a súa presentación e defensa ante un tribunal.

En Santiago de Compostela, a 1 de setembro de 2017.

O director/a:

A autora:

Don César A. Sánchez Sello

Dona Silvia Novo Díaz



# Índice xeral

|  |           |
|--|-----------|
| <b>Resumo</b>  | <b>IX</b> |
| <b>Prefacio</b>  | <b>XI</b> |
| <b>1. A regresión cuantil: alternativa á regresión clásica en media</b>  | <b>1</b>  |
| 1.1. Nocións básicas . . . . .   | 1         |
| 1.2. Modelos de regresión en media e de regresión cuantil . . . . .      | 2         |
| 1.3. Función de perda cuantílica . . . . .                               | 4         |
| 1.4. Estimación do modelo . . . . .                                      | 6         |
| 1.5. Inferencia sobre os parámetros . . . . .                            | 8         |
| 1.6. Modelo común para varios cuantís . . . . .                          | 9         |
| <b>2. O modelo autorregresivo clásico</b>                                | <b>11</b> |
| 2.1. Elementos básicos dunha serie de tempo . . . . .                    | 11        |
| 2.2. Definición e características do proceso autorregresivo . . . . .    | 16        |
| 2.2.1. ACF do AR(p) . . . . .  | 16        |
| 2.2.2. PACF do AR(p) . . . . .   | 17        |
| 2.2.3. Proceso autorregresivo en cuantil . . . . .                       | 18        |
| 2.3. Estimación dos coeficientes . . . . .                               | 19        |
| <b>3. Autorregresión cuantil</b>   | <b>21</b> |
| 3.1. O modelo QAR . . . . .  | 21        |
| 3.2. Propiedades do proceso QAR: estacionariedade . . . . .              | 23        |
| 3.3. Estimación do modelo QAR . . . . .                                  | 24        |
| 3.4. Identificación da orde do modelo QAR . . . . .                      | 26        |
| 3.4.1. Correlación simple e parcial en versión cuantil . . . . .         | 26        |
| 3.4.2. Correlación simple e parcial cuantil en versión mostral . . . . . | 28        |
| 3.4.3. Autocorrelacións parciais cuantís e modelo QAR . . . . .          | 29        |
| 3.5. Exemplo con datos simulados . . . . .                               | 33        |
| <b>4. Predición no contexto autorregresivo clásico</b>                   | <b>35</b> |
| 4.1. Intervalos de predición clásicos . . . . .                          | 36        |
| 4.2. Intervalos de predición bootstrap . . . . .                         | 36        |

|   |           |
|---|-----------|
| 4.2.1. Bootstrap de Thombs e Schucany . . . . .           | 37        |
| 4.2.2. Bootstrap de Cao et al. (1997) . . . . .           | 40        |
| 4.2.3. Bootstrap de Pascual, Romo e Ruíz . . . . .        | 41        |
| 4.2.4. Predición mediante a mediana condicional . . . . . | 42        |
| 4.3. Novo método para a predición . . . . .               | 43        |
| 4.4. Estudo de simulación . . . . .                       | 46        |
| <b>5. Aplicación a datos de desemplego</b>                | <b>55</b> |
| 5.1. Modelización en base ao QAR . . . . .                | 56        |
| 5.2. Predición . . . . .                                  | 58        |
| <b>A. Función de autocorrelacións parciais: cálculos</b>  | <b>65</b> |
| <b>B. Código QPACF mostral</b>                            | <b>69</b> |
| <b>Bibliografía</b>                                       | <b>75</b> |



# Resumo

## Resumo en galego

As vantaxes que o enfoque cuantil aporta na regresión, pola análise máis robusta e completa que proporciona, levaron a distintos autores a estender esta metodoloxía ao contexto dos modelos de series temporais, como é o caso dos procesos autorregresivos.

Neste traballo revisamos distintos procedementos para a obtención de intervalos de predición no contexto do modelo autorregresivo clásico, e incorporaremos unha proposta de novo método para a obtención destes intervalos realizando a estimación mediante mínimas desviacións absolutas e mediante bootstrap.

Ademais, para aproveitar as vantaxes que a estimación cuantil aporta baixo heterocedasticidade, estudamos un proceso autorregresivo estacionario máis xeral que permite a dependencia dos coeficientes respecto do cuantil, constituíndo unha alternativa na análise de series temporais con dinámicas asimétricas ou con persistencias locais. Para este modelo estúdase un modo de identificar a súa orde, así como para estimar os seus parámetros.

Finalmente, ilústranse estes procedementos mediante a súa aplicación á análise da serie da taxa de variación trimestral do desemprego en España, cuxo estudo queda incompleto empregando as técnicas clásicas.

## English abstract

The advantages that quantile approach provides in regression, both a more complete and robust analysis, led different authors to extend this methodology to a time series models framework, such as autoregressive processes.

In this project several procedures to obtain prediction intervals in classical autoregressive model are reviewed and a new method to obtain these intervals by least absolute deviation estimates and bootstrap is proposed.

Furthermore, we study a more general autoregressive process to take the advantage of the good behaviour of quantile regression under heteroscedasticity. This process allows coefficients to depend on the quantile order and constitutes an alternative option in time series analysis with asymmetric dynamics or local persistences. For this model, we study a way to identify its order and to estimate its parameters.

Finally, considered procedures are illustrated through an application to real data: the variation of unemployment rate in Spain, whose study is incomplete through classical techniques.



# Prefacio

Na vida real, existen numerosos fenómenos nos que variable actual dun proceso pode ser expresada como unha combinación lineal finita das variables correspondentes a instantes previos do mesmo. Esta estrutura de dependencia adoita ser estudada mediante un modelo de serie temporal autorregresivo. Precisamente, os modelos autorregresivos estacionarios foron moi estudados na literatura de series temporais, non só polo interese que teñen en sí mesmos, senón tamén como piar para a formulación de modelos máis complexos. O enfoque máis explotado no seu estudo céntrase na análise da media condicional do proceso, sendo menos habitual estudar o comportamento do cuantil condicional.

Non obstante, o auxe da regresión cuantil debido ás súas boas propiedades motivou a extensión das súas técnicas a distintos campos, tamén ao das series temporais. As principais vantaxes desta ferramenta son a súa robustez fronte a observacións atípicas e que permite traballar baixo condicións máis xerais, sen necesidade de que se verifiquen as hipóteses de normalidade e homocedasticidade. Precisamente, para aproveitar a versatilidade deste tipo de procedementos Koenker e Xiao (2006) propoñen un proceso autorregresivo máis xeral que o clásico: o modelo cuantil autorregresivo (*quantile autoregression model* en inglés, QAR), onde se permite que os coeficientes autorregresivos do proceso varíen co cuantil, é dicir, que se adapta a condicións de heterocedasticidade. Este modelo permite modelar series que presentan alteracións locais ou asimetrías, mantendo a estacionariedade a grande escala do proceso.

Para traballar con este novo modelo, débese realizar unha extensión das ferramentas empregadas na identificación e na estimación do proceso autorregresivo clásico ao terreo cuantil. Neste contexto, Li et al (2015) definen a función de autocorrelacións parciais cuantís, en base a cal se podería identificar a orde do modelo QAR para cada cuantil, mentres que as técnicas empregadas na estimación se reducirían ás do proceso cuantil autorregresivo homocedástico.

Unha vez proposto un modelo para o proceso, este pode empregarse como medio para a obtención de predicións. O problema redunda en que no contexto dos modelos QAR de Koenker e Xiao (2006) aínda non foron propostos métodos para a obtención de intervalos de predición. Por tal razón, neste traballo tamén nos ocuparemos de revisar distintos procedementos bootstrap para a obtención de intervalos de predición no contexto autorregresivo clásico estimando tamén mediante técnicas cuantís, coa finalidade de nun futuro, abordar a predición no proceso QAR. Ademais disto, incorporamos un novo método bootstrap para a obtención de intervalos de predición no modelo clásico, realizando a predición en base á mediana condicional.

A memoria vaise organizar en seis capítulos:

- Os dous primeiros capítulos inclúense a modo de introdución: no primeiro deles revisaremos as técnicas da regresión cuantil, facendo fincapé na comparación das súas propiedades con respecto ás da regresión clásica en media; no segundo capítulo introducirase o proceso autorregresivo clásico, estudando dous procedementos para estimación dos seus parámetros: mínimos cadrados ordinarios e mínimas desviacións absolutas.
- No Capítulo 3 presentaremos o modelo cuantil autorregresivo proposto por Koenker e Xiao (2006) e estudaremos as súas propiedades en canto a estacionariedade. A dependencia con respecto ao cuantil dos coeficientes do modelo, fai que cobre un especial interese a identificación da orde do proceso autorregresivo para cada cuantil, polo que revisaremos o traballo de Li et al. (2015) onde se propón unha extensión ao terreo cuantil da función de autocorrelacións parciais.

- No Capítulo 4 abordaremos o problema da predición no contexto autorregresivo clásico. Ademais do procedemento Box–Jenkins, revisaremos distintos algoritmos bootstrap para a obtención de intervalos de predición, como son o proposto por Thombs e Schucany (1990), o bootstrap de Cao et al. (1997) ou o de Pascual et al. (2001). A estimación realizarase tanto por mínimos cadrados ordinarios como por mínimas desviacións absolutas (de ser posible). Precisamente, no contexto desta estimación robusta imos propoñer un novo procedemento bootstrap, cuxa consistencia garantiremos teoricamente. Finalmente, compararemos o comportamento práctico dos distintos procedementos presentados por medio dun estudo de simulación.
- A metodoloxía presentada nos Capítulos 3 e 4 será aplicada no Capítulo 5 a un conxunto de datos reais: a serie histórica da taxa de variación intertrimestral do paro en España dende o primeiro trimestre de 1990 ata o segundo de 2016. Iremos estudar o comportamento asimétrico desta serie por medio do modelo QAR e obteremos intervalos de predición para un tramo da mesma aplicando os procedementos estudados no Capítulo 3. Isto enlaza co Capítulo 6, no que se acollen as conclusións que extraemos deste traballo e se citan varias liñas abertas no ámbito da investigación.

# Capítulo 1

## A regresión cuantil: alternativa á regresión clásica en media

A regresión cuantil é unha ferramenta estatística deseñada para estimar e facer inferencia sobre os cuantís dunha variable condicionados a unha ou varias variables explicativas. Ao igual que na regresión lineal clásica, a cal se basea en minimizar a suma de residuos ao cadrado para a estimación da media condicional, no contexto da regresión cuantil deseñáronse procedementos que permiten estimar todo o rango de cuantís condicionais.

Na actualidade a regresión cuantil é un tema de interese na investigación estatística, polo que se están adaptando parte das técnicas da inferencia relacionadas coa regresión en media a este terreo. Isto é debido ás propiedades que confire este tipo de regresión. Por unha parte, os procedementos da regresión cuantil constrúense de maneira flexible, de modo que se adaptan a condicións bastante xerais como a ausencia de normalidade e, xeralmente, tamén sen o requisito de homocedasticidade. Ademais, son robustos fronte a observacións atípicas: ao igual que a mediana é máis robusta que a media para estimar a posición central dunha distribución, tamén a regresión en mediana (ou noutro cuantil) é máis robusta que a regresión en media.

Precisamente, aínda que a regresión cuantil inclúe a mediana como caso particular máis frecuente, tamén permite estudar outros cuantís non centrais, o que aporta unha visión máis completa do efecto que as variables explicativas producen na resposta, é dicir, da relación de dependencia.

Neste primeiro capítulo centrarémonos en desenvolver estas ideas realizando, ao mesmo tempo, unha breve comparación entre os procedementos da regresión cuantil e a metodoloxía da regresión en media no contexto lineal. Na primeira sección presentaremos a definición de cuantil, para nas dúas seguintes adentrarnos xa na regresión lineal e na función de perda cuantílica. Na cuarta sección veremos os dous métodos de estimación do modelo e na quinta incluímos as propiedades que presentan os estimadores dos parámetros. Finalmente, na última sección estudaremos como afecta a hipótese de homocedasticidade na estimación do modelo cuantil. As referencias principais que se empregaron na redacción deste capítulo foron Koenker (2005) e Davino et al. (2014).

### 1.1. Nocións básicas

Para presentar os procedementos da regresión é preciso introducir unhas ideas básicas que nos permitan iniciarnos no tema e fixar notación. Tal é o caso do concepto de función de distribución dunha variable aleatoria e a definición de cuantil.

**Definición 1.1.** Dada unha variable aleatoria  $Y : \Omega \rightarrow \mathbb{R}$ , definida nun espazo de probabilidade  $(\Omega, \mathcal{F}, \mathbb{P})$ , onde  $\Omega$  representa o espazo mostral,  $\mathcal{F}$  é a  $\sigma$ -álgebra de  $\Omega$  e  $\mathbb{P}$  é a medida de probabilidade, o comportamento desta variable aleatoria queda caracterizado pola súa *función de distribución*, a cal se define como:

$$F(y) = \mathbb{P}(Y \leq y), \quad \forall y \in \mathbb{R}.$$

Se a variable aleatoria  $Y$  é discreta, daquela a súa función de distribución sería da forma

$$F(y) = \mathbb{P}(Y \leq y) = \sum_{y_i \leq y} \mathbb{P}(Y = y_i), \quad \forall y \in \mathbb{R},$$

onde  $y_i \leq y$  representa os valores que toma  $Y$  que son inferiores ou iguais a  $y \in \mathbb{R}$ . Por outra parte, se a variable  $Y$  é absolutamente continua tense que

$$F(y) = \mathbb{P}(Y \leq y) = \int_{-\infty}^y f(t)dt, \tag{1.1}$$

onde para calquera  $y \in \mathbb{R}$ ,  $f : \mathbb{R} \rightarrow \mathbb{R}$  é unha función non negativa e integrable tal que  $\int_{-\infty}^{\infty} f(t)dt = 1$ , usualmente denominada *función de densidade*.

**Definición 1.2.** Dada unha variable aleatoria  $Y$  definida nun espazo de probabilidade  $(\Omega, \mathcal{F}, \mathbb{P})$ , o *cuantil de orde*  $\tau$ , para cada  $0 < \tau < 1$ , é o valor  $y \in \mathbb{R}$  tal que verifica que

$$\begin{aligned} \mathbb{P}(Y \leq y) &\geq \tau, \\ \mathbb{P}(Y \geq y) &\geq 1 - \tau. \end{aligned}$$

Como pode observarse, os cuantís son localizacións particulares da distribución dunha variable aleatoria. Un cuantil con nome propio é o de orde  $\tau = 0.5$ , comunmente coñecido como *mediana*.

Notemos que se a función de distribución  $F$  é continua e estrictamente crecente, a relación entre valores da variable e probabilidades é bixectiva, co que  $F^{-1}(\tau)$  é o único número real  $y$  tal que  $F(y) = \tau$ . Se  $F$  é constante nalgún intervalo, calquera punto do mesmo servirá como cuantil  $\tau$ . Por outra banda, se a función de distribución é discreta, nos saltos ocorre que o valor  $y$  da variable é o cuantil  $\tau$  para moitas probabilidades  $\tau$ .

Debido a estes casos nos que a función inversa non estaría ben definida, a función cuantil defínese como a inversa xeneralizada da función de distribución, é dicir,

$$Q_\tau(Y) = F^{-1}(\tau) = \inf\{y : F(y) \geq \tau\} \quad \text{para } \tau \in (0, 1).$$

## 1.2. Modelos de regresión en media e de regresión cuantil

Os modelos de regresión empréganse para representala dependencia entre unha variable aleatoria  $Y$  (coñecida como *variable dependente* ou *resposta*) con respecto a un conxunto de variables  $X_1, X_2, \dots, X_p$ , (*variables independentes* ou *explicativas*) definidas no mesmo espazo de probabilidade  $(\Omega, \mathcal{F}, \mathbb{P})$ .

Para construír o modelo de regresión debe terse en conta o tipo de regresión que interesa levar a cabo (paramétrica ou non paramétrica), a forma da función de regresión (lineal, polinómica ou outras), o tipo de distribución do erro e a súa posible dependencia das variables explicativas e outros aspectos que nos permiten configuralo e determinan a súa natureza e os métodos de estimación que poden ser empregados. Unha vez construído o modelo, este pode usarse tamén para realizar predicións do valor de  $Y$  coñecido o valor que toma o conxunto de variables explicativas, o que o converte nunha ferramenta moi útil na resolución de problemas reais.

Os métodos de regresión clásicos estudan a media condicionada da variable  $Y$  en función do valor que tome o vector aleatorio  $\mathbf{X} = (X_1, X_2, \dots, X_p)$ , é dicir, dado  $\mathbf{x} = (x_1, x_2, \dots, x_p) \in \mathbb{R}^p$  teremos

$$m(\mathbf{x}) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x}).$$

Isto permite descompoñer a variable resposta en función do resultado do vector aleatorio  $\mathbf{X}$  a través da media condicionada, máis unha variable aleatoria non observable que contén a variabilidade da resposta non atribuíble ás variables explicativas, coñecida como *erro*. Deste xeito, obteríamos a seguinte descomposición

$$Y = m(\mathbf{X}) + \varepsilon, \tag{1.2}$$

onde  $\varepsilon$  denota ao erro e verifica  $\mathbb{E}(\varepsilon|\mathbf{X} = \mathbf{x}) = 0$  para todo  $\mathbf{x} \in \mathbb{R}^p$ .

Neste traballo, imos centrarnos en modelos de regresión paramétrica, concretamente lineal, nos cales

$$m(\mathbf{X}) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_p X_p, \tag{1.3}$$

onde  $\theta_0, \theta_1, \dots, \theta_p$  son os coeficientes do modelo, en principio descoñecidos. Así, a variable resposta pode escribirse como combinación lineal das variables explicativas máis un termo de erro,  $\varepsilon$ , é dicir,

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_p X_p + \varepsilon.$$

Como é ben coñecido, na regresión clásica a realización de tarefas de inferencia sobre os parámetros do modelo require asumir que, ademais da linealidade, este verifica unha serie de hipóteses básicas que nos permitan coñecer a distribución dos estimadores dos parámetros. Estas hipóteses a miúdo son as seguintes:

- Homocedasticidade. A varianza do erro é constante, calquera que sexa o valor que tomen as variables explicativas:  $\text{Var}(\varepsilon|\mathbf{X} = \mathbf{x}) = \sigma^2$ .
- Independencia. As variables aleatorias que representan os erros son mutuamente independentes para distintos individuos  $\varepsilon_1, \dots, \varepsilon_n$ , entendendo que vamos obter unha mostra de  $n$  observacións baixo o modelo de regresión.
- Normalidade. Os erros  $\varepsilon_1, \dots, \varepsilon_n$  seguen unha distribución  $N(0, \sigma^2)$ , onde  $\sigma^2$  é un parámetro adicional que require ser estimado.

Como acabamos de ver, os procedementos de regresión en media céntranse exclusivamente nesta localización específica da distribución condicionada de  $Y$ . Neste aspecto, unha das principais vantaxes que presenta regresión cuantil con respecto á regresión en media é, precisamente, a xeneralización deste estudo a diferentes localizacións desa distribución condicionada de  $Y$  con respecto a  $\mathbf{X}$ , proporcionando unha visión global e máis completa da relación de dependencia. Así, fixado  $\tau \in (0, 1)$ , dado  $\mathbf{x} = (x_1, x_2, \dots, x_p) \in \mathbb{R}^p$ , estudaríase o comportamento do cuantil condicional de  $Y$  a cada valor que tome o conxunto de variables  $\mathbf{X}$ :

$$q_\tau(\mathbf{x}) = Q_\tau(Y|\mathbf{X} = \mathbf{x}).$$

Así, analogamente ao exposto para a regresión en media, podemos descompoñer a variable resposta en función do resultado do vector aleatorio  $\mathbf{X}$ , a través do cuantil condicional, máis un termo de *erro*. Deste xeito, escribiríamos que

$$Y = q_\tau(\mathbf{X}) + \varepsilon \quad \text{para } \tau \in (0, 1),$$

onde  $\varepsilon$  denota ao erro e verifica  $Q_\tau(\varepsilon|\mathbf{X} = \mathbf{x}) = 0$  para todo  $\mathbf{x} \in \mathbb{R}^p$ . Concretamente, no caso da regresión lineal teríamos que

$$q_\tau(\mathbf{X}) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_p X_p,$$

onde  $\theta_0, \theta_1, \dots, \theta_p$  son os coeficientes do modelo, en principio descoñecidos.

Unha gran vantaxe da regresión cuantil é que o traballo con cuantís garante a robustez dos seus procedementos fronte a observacións atípicas: como citan Davino et al. (2014) “A regresión cuantil é á clásica o que os cuantís son á media en termos de descrición das localizacións dunha distribución”.

Outro gran beneficio que aporta, é que permite o traballo baixo hipóteses máis xerais que as que se adoitan esixir nos métodos de regresión en media: non se supón a normalidade da distribución do erro, nin tampouco é precisa a homocedasticidade.

**Exemplo 1.3.** A librería `quantreg` do programa R Core Team (2015) contén unha base de datos denominada `engel`, os cales proceden dun estudo acerca da relación entre o gasto en alimentación e os ingresos nos fogares de Bélxica. Estes datos son empregados como ilustración en diversas seccións de Koenker (2005).

Na Figura 1.1 amosamos o diagrama de dispersión xunto coas rectas de regresión axustadas (mediante a función `rq` do paquete `quantreg`) para varios cuantís (entre eles a mediana) e a recta de mínimos cadrados (axustada mediante `lm`). Como se pode ver, a maiores ingresos maior é a variabilidade do gasto en comida, polo que se pode apreciar unha heterocedasticidade clara nas observacións. Concretamente, hai dous fogares cunha inversión en alimentación moi baixa respecto aos seus ingresos. Estes provocan un desprazamento cara abaixo da recta de regresión en media, mentres que a recta en mediana apenas se ve afectada pola presenza desas observacións atípicas.

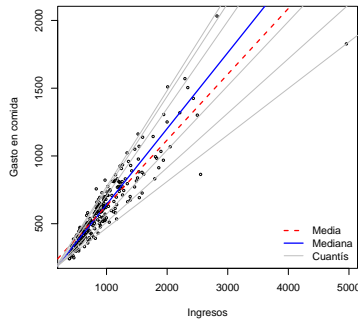


Figura 1.1: Diagrama de dispersión do gasto en alimentación fronte aos ingresos. A recta azul correspóndese co axuste en mediana, a vermella de puntos é a recta de regresión en media, mentres que as rectas grises, de abaixo a arriba da imaxe, correspóndense cos cuantís  $\tau = 0.05, 0.1, 0.25, 0.75, 0.9, 0.95$ .

Por outra parte, a vantaxe que comporta a regresión cuantil é que permite cuantificar o efecto dos ingresos nos fogares de maior (cuantís superiores) e menor consumo (cuantís inferiores), dando unha visión máis pormenorizada da relación de dependencia entre as variables.

### 1.3. Función de perda cuantílica

Para poder empregar na práctica os modelos de regresión é preciso definir o problema de minimización que proporciona como solucións os seus coeficientes ou parámetros. Dada unha variable aleatoria  $Y$ , a súa media é o punto que minimiza o promedio dos cadrados das desviacións, é dicir:

$$\mathbb{E}(Y) = \arg \min_{c \in \mathbb{R}} \mathbb{E} \left[ (Y - c)^2 \right]. \tag{1.4}$$

Esta propiedade, tamén satisfeita pola media condicionada, é a base da regresión lineal clásica (así mesmo, chamada de *mínimos cadrados*) onde os parámetros do modelo (1.3) son a solución dun problema de optimización análogo a (1.4). Isto é, se denotamos por <sup>1</sup>  $\mathbf{x}^1 = (1, x_1, \dots, x_p)^\top \in \mathbb{R}^{p+1}$  e por  $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_p)^\top$ ,

<sup>1</sup>Diferenciamos mediante expoñente un expoñente o feito de se un vector (aleatorio ou non) inclúe ou non o 1 como primeira compoñente ademais das restantes correspondentes a valores de variables. É dicir,  $\mathbf{x}$  denota a  $(x_1, \dots, x_p)$  e  $\mathbf{x}^1$  denota a  $(1, x_1, \dots, x_p)$ .



tense que  $m(\mathbf{x}) = \boldsymbol{\theta}_0^\top \mathbf{x}^1$  onde  $\boldsymbol{\theta}_0$  é a solución de:

$$\boldsymbol{\theta}_0 = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \mathbb{E} \left[ (Y - \boldsymbol{\theta}^\top \mathbf{x}^1)^2 \right].$$

Pola súa parte, entre os cuantís, a mediana satisfai unha propiedade análoga á da media, posto que é a posición que fai mínima a media das desviación absolutas, é dicir:

$$Q_{0.5}(Y) = \arg \min_{c \in \mathbb{R}} \mathbb{E} [|Y - c|].$$

Esta propiedade da mediana resulta interesante posto que é posible estendela aos demais cuantís por medio da chamada *función de perda cuantílica*, a cal se define como:

$$\rho_\tau(u) = \begin{cases} \tau u & \text{si } u \geq 0, \\ -(1 - \tau)u & \text{si } u < 0, \end{cases}$$

con  $u \in \mathbb{R}$ . Mediante esta función, os cuantís poden presentarse como centros da distribución, pois son os puntos que minimizan as desviacións absolutas ponderadas por medio da función de perda cuantílica, é dicir:

$$Q_\tau(Y) = \arg \min_{c \in \mathbb{R}} \mathbb{E} [\rho_\tau(Y - c)]. \quad (1.5)$$

Na Figura 1.2 inclúense tres gráficos nos que se representa a función de perda cuantílica para  $\tau = 0.25, 0.5$  e  $0.75$ . Observemos que no gráfico do medio de dita figura, que se corresponde co da mediana ( $\tau = 0.5$ ), a función cuantil ten igual pendente á dereita e á esquerda do cero, polo que de contar cunha mostra aleatoria simple de  $Y$  de tamaño  $n$ ,  $Y_1, \dots, Y_n$ , para estimar a mediana penalízanse de igual xeito as observacións superiores e inferiores a esa posición central. Se  $\tau = 0.25$ , a función cuantil ten maior pendente á esquerda do cero de maneira que se dará maior peso ás observacións máis pequenas que a posición central, arrastrando así a estimación cara os valores inferiores de  $Y$ . Finalmente, se  $\tau = 0.75$  a función cuantil ten máis pendente á dereita do cero, o que implica que se lles dá maior peso as observacións por riba da posición central.

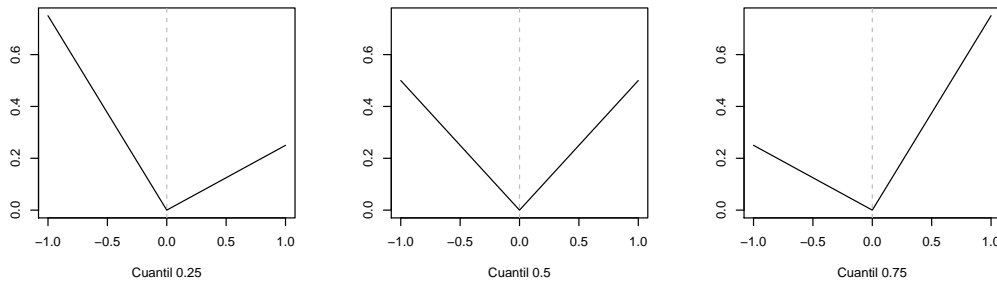


Figura 1.2: Funcións de perda cuantílica para  $\tau = 0.25, 0.5$  e  $0.75$ .

Ao igual que ocurría no caso da media, é posible trasladar á regresión estas ideas. Dado un conxunto de variables explicativas  $X_1, X_2, \dots, X_p$  e unha variable resposta  $Y$ , os coeficientes do modelo de regresión cuantil lineal para un cuantil  $\tau$  xenérico poden verse como a solución ao problema de minimización da perda cuantílica esperada. Para cada  $\mathbf{x}^1 = (1, x_1, \dots, x_p)^\top \in \mathbb{R}^{p+1}$ , se  $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_p)^\top$  tense que  $q_\tau(\mathbf{x}) = \boldsymbol{\theta}_0^\top \mathbf{x}^1$ , onde  $\boldsymbol{\theta}_0$  é a solución do problema de optimización seguinte:

$$\boldsymbol{\theta}_0(\tau) = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \mathbb{E} [\rho_\tau(Y - \boldsymbol{\theta}^\top \mathbf{x}^1)].$$

## 1.4. Estimación do modelo

Para poder levar á práctica o procedemento de estimación dos parámetros do modelo de regresión é necesario trasladar ao terreo mostral esta metodoloxía. No contexto estatístico, é posible distinguir dous tipos de deseño experimental:

- Deseño fixo: os valores da variable explicativa non son aleatorios, senón que están fixados polo experimentador. Só é estocástico o erro. Polo tanto, a mostra resultante dun deseño fixo sería do tipo:  $(\mathbf{x}_1, Y_1), \dots, (\mathbf{x}_n, Y_n)$ , sendo  $n$  o número de individuos que a conforman e  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  con  $i = 1, \dots, n$  os valores correspondentes ao individuo  $i$ -ésimo da mostra.
- Deseño aleatorio: tanto a variable explicativa como a resposta son aleatorias, o que derivaría nunha mostra do tipo  $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ , onde  $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})$  para  $i = 1, 2, \dots, n$ , son as variables correspondentes ao individuo  $i$ -ésimo da mostra,

O usual na literatura é presentar a metodoloxía da regresión asumindo deseño fixo. Por tal razón, nos imos seguir tamén este enfoque, tendo en conta que o paso de deseño aleatorio a fixo se pode realizar condicionando á mostra observada.

Por outra parte, o habitual na práctica é que se descoñeza a distribución da variable resposta. Se dispoñemos dunha mostra aleatoria simple  $Y_1, \dots, Y_n$  de  $Y$ , un estimador natural para  $F$  sería a *distribución empírica* que se define como

$$F_n(y) = \frac{1}{n} \sum_{i=1}^n I(Y_i \leq y),$$

onde  $I(A) = 1$  se  $A$  é certo e  $I(A) = 0$  noutro caso, é dicir, é a función indicador.

En consecuencia, dada unha mostra aleatoria simple do conxunto de variables explicativas  $\mathbf{X} = (X_1, \dots, X_p)$  que denotaremos mediante  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , poderíamos escribir o modelo de regresión lineal múltiple en forma matricial como segue:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{n,p} \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Substituíndo cada vector ou matriz por un símbolo, poderíamos chegar á forma matricial abreviada seguinte:

$$\mathbf{Y} = \vec{\mathbf{X}}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad (1.6)$$

onde  $\mathbf{Y}$  é o vector resposta,  $\vec{\mathbf{X}}$  é unha matriz que contén  $n \times (p+1)$  números onde cada fila representa un individuo e cada columna unha característica,  $\boldsymbol{\theta}$  é o vector de parámetros ou coeficientes do modelo e  $\boldsymbol{\varepsilon}$  contén os erros. Tamén sería posible escribir, na notación que empregaremos usualmente ao longo deste traballo:

$$Y_i = \boldsymbol{\theta}^\top \mathbf{x}_i^1 + \varepsilon_i \quad i = 1, \dots, n. \quad (1.7)$$

Agora, a estimación do modelo de regresión lineal por mínimos cadrados converteríase nun problema do seguinte tipo:

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \int (y - \boldsymbol{\theta}^\top \mathbf{x}_i^1)^2 dF_n(y) = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \frac{1}{n} \sum_{i=1}^n (Y_i - \boldsymbol{\theta}^\top \mathbf{x}_i^1)^2. \quad (1.8)$$

En notación matricial, o problema de minimización podería expresarse de maneira equivalente así:

$$\min_{\boldsymbol{\theta}} (\mathbf{Y} - \vec{\mathbf{X}}\boldsymbol{\theta}) (\mathbf{Y} - \vec{\mathbf{X}}\boldsymbol{\theta}),$$

co que derivando a función obxectivo con respecto a  $\boldsymbol{\theta}$  (o que significa que se deriva a expresión (1.8) con respecto a cada  $\theta_j$  con  $j = 0, \dots, p$ ) e igualando a 0, obtense o que se coñece como *ecuacións normais da regresión*,

$$\vec{\mathbf{X}}^\top \vec{\mathbf{X}} \boldsymbol{\theta} = \vec{\mathbf{X}}^\top \mathbf{Y},$$

cuxa solución é o estimador de  $\boldsymbol{\theta}$  obtido polo método de *mínimos cadrados ordinarios* (MCO),

$$\hat{\boldsymbol{\theta}} = \left( \vec{\mathbf{X}}^\top \vec{\mathbf{X}} \right)^{-1} \vec{\mathbf{X}}^\top \mathbf{Y}. \tag{1.9}$$

Este estimador estará ben definido cando a matriz  $\vec{\mathbf{X}}^\top \vec{\mathbf{X}}$  sexa non singular, dado que só así é posible definir  $\left( \vec{\mathbf{X}}^\top \vec{\mathbf{X}} \right)^{-1}$ . Non obstante, notemos, por unha parte, que  $\vec{\mathbf{X}}^\top \vec{\mathbf{X}}$  é unha matriz cadrada de orde  $(p+1) \times (p+1)$ , simétrica e semidefinida positiva que está relacionada coa estrutura de covarianza das variables. Por outra parte, o seu rango coincide coa dimensión do espazo onde se atopan os vectores  $\mathbf{x}_i^1$  sempre e cando sexa posible obter unha mostra de máis de  $(p+1)$  individuos,  $n \geq (p+1)$ , algo natural nos problemas de regresión clásicos.

Análogamente, o proceso de estimación na regresión cuantil, para un cuantil  $\tau$  dado, pasará pola resolución do problema de minimización<sup>2</sup>

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \int \rho_\tau(y - \boldsymbol{\theta}^\top \mathbf{x}_i^1) dF_n(y) = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \frac{1}{n} \sum_{i=1}^n \rho_\tau(Y_i - \boldsymbol{\theta}^\top \mathbf{x}_i^1). \tag{1.10}$$

O problema é que aquí a función obxectivo non é derivable no punto no que se acada o mínimo. Entón, para achala solución do problema de minimización (1.10) recórrese a ferramentas de programación lineal. Para aplicalas é preciso reformular a función obxectivo mediante a descomposición dos residuos  $Y_i - \boldsymbol{\theta}^\top \mathbf{x}_i^1$  na súa parte positiva,  $u_i$ , e negativa,  $v_i$  con  $i = 1, \dots, n$ . Así, denotando por  $\mathbf{u} = (u_1, \dots, u_n)^\top$  e por  $\mathbf{v} = (v_1, \dots, v_n)^\top$ , isto permitiría reformular o problema de optimización na súa parte positiva e a súa parte negativa

$$\min_{(\boldsymbol{\theta}, \mathbf{u}_i, \mathbf{v}_i) \in \mathbb{R}^{p+1} \times \mathbb{R}_+^{2n}} \sum_{i=1}^n [\tau u_i + (1 - \tau) v_i], \tag{1.11}$$

suxeito a

$$\begin{aligned} \mathbf{Y} - \vec{\mathbf{X}} \boldsymbol{\theta} &= \mathbf{u} - \mathbf{v}, \\ \boldsymbol{\theta} &\in \mathbb{R}^{p+1} \\ u_i \geq 0, v_i &\geq 0, \quad i = 1, 2, \dots, n. \end{aligned} \tag{1.12}$$

onde  $\mathbb{R}_+^{2n}$  denota o espazo dos vectores de dimensión  $2n$  de compoñentes reais positivas. Como se poderá observar, no problema de optimización incluímos  $2n$  variables artificiais  $u_i$  e  $v_i$ , pero tamén están presentes como variables os  $p+1$  coeficientes do vector  $\boldsymbol{\theta}$ . Dado que só temos  $n$  restricións, de existir solución, esta estará formada por  $n$  *variables básicas* ou non nulas: os  $p+1$  coeficientes do modelo máis  $n-p-1$  residuos do modelo.

En consecuencia, teremos exactamente  $p+1$  individuos con residuo cero, polo que a regresión cuantil pasa por  $p+1$  datos da mostra. Isto é coherente co feito de que o propio cuantil dunha variable é un dato da mostra. No caso particular da regresión simple, teríamos que a recta axustada por este procedemento pasaría por dous puntos da mostra e, así, sucesivamente para modelos máis complexos.

A diferenza do que ocorre no método de mínimos cadrados, na regresión cuantil non imos dispoñer dunha expresión explícita para o vector de parámetros  $\boldsymbol{\theta}$ . O algoritmo máis empregado para resolver problemas lineais de regresión cuantil é o de Barrodale e Roberts (1973) que é unha versión simplificada do método Síplex.

<sup>2</sup>Debemos notar que estamos a cometer un abuso da notación ao denotar do mesmo xeito aos estimadores de mínimos cadrados e aos procedentes da minimización da perda cuantílica,  $\hat{\boldsymbol{\theta}}$ , pero, en cada caso, faremos fincapé aclarando a que estimadores en concreto nos referimos.

Outra particularidade deste método de estimación é que se denotamos por  $P$  o número de residuos positivos (individuos que quedan por riba da recta de regresión),  $N$  o número de residuos negativos e  $Z$  o número de residuos que son 0, tense que  $Z = p + 1$  e ademais a proporción de residuos negativos é aproximadamente  $\tau$ , en concreto,

$$\frac{N}{n} \leq \tau \leq \frac{N + p + 1}{n}.$$

mentres que a proporción de residuos positivos é aproximadamente  $1 - \tau$ :

$$\frac{P}{n} \geq 1 - \tau \geq \frac{P + p + 1}{n}.$$

Cabe mencionar finalmente, que no caso particular da mediana, o proceso de estimación adoita chamárselle de *mínimas desviacións absolutas* (MDA) pola propiedade que esta localización da distribución verifica, véxase a Sección 1.3.

## 1.5. Inferencia sobre os parámetros

Como é ben coñecido, cando a estimación do modelo de regresión se realiza por mínimos cadrados, dado que  $\mathbb{E}(\varepsilon) = 0$ , verifícase que  $\hat{\boldsymbol{\theta}}$  é un estimador non nesgado de  $\boldsymbol{\theta}$ . Ademais, o seguinte teorema danos a distribución do estimador:

**Teorema 1.4.** *Supoñamos que  $\varepsilon_1, \dots, \varepsilon_n$ , son variables aleatorias independentes con distribución común  $N(0, \sigma^2)$  e que  $\vec{\mathbf{X}}$  é unha matriz  $n \times (p + 1)$ . Sexa  $\hat{\boldsymbol{\theta}}$  o estimador de MCO de  $\boldsymbol{\theta}$ . Entón:*

$$\hat{\boldsymbol{\theta}} \in N_{p+1} \left( \boldsymbol{\theta}, \sigma^2 \left( \vec{\mathbf{X}}^\top \vec{\mathbf{X}} \right)^{-1} \right).$$

Como citamos con anterioridade, unha das principais vantaxes da regresión cuantil é que non require ningunha suposición restritiva sobre a distribución da mostra, máis aló de que o modelo de regresión sexa lineal e de que os erros sexan independentes.

Un dos posibles contrapuntos dela sería que os estimadores dos parámetros deste modelo non teñen unha expresión explícita a diferenza dos da regresión en media, nin sequera asumindo unha distribución concreta para o erro. Isto non impide, non obstante, o establecemento dunha distribución asintótica para eles, para o que é preciso realizar as seguintes asuncións.

**A1.** Sexan  $Y_1, \dots, Y_n$  variables aleatorias independentes con funcións de distribución  $F_1, \dots, F_n$ . Supoñamos que a función cuantil  $\tau$  condicionada é lineal no vector de covariables  $\mathbf{x}_i$ :

$$q_\tau(\mathbf{x}_i^1) = \mathcal{Q}_\tau(Y_i | \mathbf{X}_i = \mathbf{x}_i) = \boldsymbol{\theta}^\top \mathbf{x}_i^1.$$

As funcións de distribución  $F_i$  son absolutamente continuas, con densidades continuas  $f_i$ , sendo os valores  $f_i(q_\tau(\mathbf{x}_i^1))$  uniformemente limitados fora de cero e de infinito, nos puntos  $q_\tau(\mathbf{x}_i^1)$  cando  $i = 1, \dots, n$ .

**A2.** Existen matrices simétricas e definidas positivas  $\vec{\boldsymbol{\Omega}}_1(\tau)$  e  $\vec{\boldsymbol{\Omega}}_0$  tales que:

1.  $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{x}_i^1 \mathbf{x}_i^{1\top} = \vec{\boldsymbol{\Omega}}_0$ .
2.  $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n f_i(q_\tau(\mathbf{x}_i^1)) \mathbf{x}_i^1 \mathbf{x}_i^{1\top} = \vec{\boldsymbol{\Omega}}_1(\tau)$ .
3.  $\max_{i=1, \dots, n} \|\mathbf{x}_i^1\| / \sqrt{n} \rightarrow 0$ , onde  $\|\cdot\|$  denota a norma euclídea do vector.

**Teorema 1.5.** *Baixo as condicións **A1** e **A2**, supoñamos que o vector de parámetros  $\boldsymbol{\theta}$  se estima mediante:*

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \sum_{i=1}^n \rho_{\tau}(Y_i - \boldsymbol{\theta}^{\top} \mathbf{x}_i^1),$$

*entón a distribución asintótica de  $\hat{\boldsymbol{\theta}}$  vén determinada como segue:*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \rightarrow N_{p+1}\left(\mathbf{0}, \tau(1-\tau) \vec{\boldsymbol{\Omega}}_1^{-1}(\tau) \vec{\boldsymbol{\Omega}}_0 \vec{\boldsymbol{\Omega}}_1^{-1}(\tau)\right).$$

*En particular, no caso de que os erros teñan a mesma distribución, é dicir,  $F_1 = \dots = F_n$ , tense que*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \rightarrow N_{p+1}\left(\mathbf{0}, \tau(1-\tau) \frac{1}{f_1^2(q_{\tau}(\mathbf{x}_1^1))} \vec{\boldsymbol{\Omega}}_0^{-1}\right).$$

*Demostración.* Aínda que non imos desenvolver neste traballo a demostración deste teorema, a proba do mesmo pode consultarse na páxina 121 de Koenker (2005).  $\square$

En canto á distribución asintótica dos estimadores dos parámetros do modelo cuantil, podería chamar a atención a súa varianza, pero cabe notar que o termo  $\tau(1-\tau)1/f_1^2(q_{\tau}(\mathbf{x}_1^1))$  aparece xa na distribución asintótica do cuantil mostral dunha variable, sen estar condicionado a variables explicativas.

Ademais, o termo  $\tau(1-\tau)$  correspóndese simplemente coa varianza dunha distribución de Bernoulli de parámetro  $\tau$ , e deriva do feito de que

$$\mathbb{P}(Y \leq \mathcal{Q}_{\tau}(Y)) = \tau.$$

Por outra parte, a densidade  $f_1$  indica se hai moitos datos próximos ao cuantil. Iso implica que se  $f_1^2(q_{\tau}(\mathbf{x}_1^1))$  é grande, ao estar no denominador, a varianza do estimador é menor, é dicir, estimarase mellor os cuantís que teñen moita densidade no seu carón.

Na varianza do estimador atopamos o produto de matrices  $\vec{\boldsymbol{\Omega}}_1^{-1}(\tau) \vec{\boldsymbol{\Omega}}_0 \vec{\boldsymbol{\Omega}}_1^{-1}(\tau)$ , que se reduce a  $\vec{\boldsymbol{\Omega}}_0^{-1}$  cando os erros teñen a mesma distribución.  $\vec{\boldsymbol{\Omega}}_0$  exerce o papel de matriz de covarianzas das variables explicativas de xeito similar á da matriz  $\vec{\mathbf{X}}^{\top} \vec{\mathbf{X}}$  dos modelos de regresión en media lineais.

A matriz  $\vec{\boldsymbol{\Omega}}_1$  tamén actúa coma unha matriz de covarianzas coa diferenza de que as aportacións de cada dato mostral se ponderan polas densidades  $f_i(q_{\tau}(\mathbf{x}_i^1))$  para cada  $i = 1, \dots, n$ . Se a densidade é grande o dato pesa máis na estimación do cuantil porque aporta máis información. Esta é a maneira na que se manifesta a heterocedasticidade na regresión cuantil.

O problema que se presenta, non obstante, procede da inversa da densidade, tamén coñecida como *sparsity*, na expresión da varianza asintótica, que constitúe unha dificultade importante á hora de obter o erro típico dos estimadores. Para solventar isto propuxéronse distintos métodos de estimación *sparsity*, problema que non abordaremos neste traballo.

## 1.6. Modelo común para varios cuantís

Como vimos ao longo deste capítulo, en particular, no **Exemplo 1.3**, é posible axustar un modelo de regresión para cada cuantil  $\tau$  entre 0 e 1. Non obstante, en vista das propiedades dos estimadores do modelo de regresión cuxa varianza é menor cando se estiman posicións con máis densidade, habitualmente as máis centrais, cabería preguntarse se é posible axustar un mesmo modelo de regresión para todos os cuantís ou para algúns deles.

Por exemplo, no caso dunha única variable explicativa podería ocorrer que todas as rectas tivesen a mesma pendente (ou no caso de varias variables explicativas podería pasar que tivéssemos hiperplanos coa mesma dirección). Isto sería o que pasaría baixo homocedasticidade, onde as rectas (ou hiperplanos) para os

distintos cuantís só se diferenciarían na súa ordenada na orixe (os hiperplanos atoparíanse desprazados por unha constante). Na Figura 1.3 ilustramos este comportamento. As gráficas situadas á esquerda e no centro mostran os diagramas de dispersión xerados dende o modelo:

$$Y_i = 0.75x_i + \varepsilon_i, \quad i = 1, \dots, 1000, \quad (1.13)$$

onde  $\varepsilon_i$  son i.i.d a unha distribución normal estándar,  $N(0,1)$ , no primeiro caso e, no segundo, a unha exponencial de parámetro  $\lambda = 1$ ,  $\text{Exp}(1)$ , centrada no 0. En ambos modelos, as rectas de regresión son practicamente paralelas, polo que sería asumible<sup>3</sup> a igualdade de pendentes. No caso de distribucións do erro simétricas, as rectas correspondentes a cada cuantil sitúanse equiespaciadas a cada lado da recta mediana, cousa que non ocorre cando a distribución do erro é asimétrica. Notemos, ademais, que cando a distribución do erro é distinta da normal, xa non estamos nas hipóteses da regresión clásica.

A terceira gráfica da Figura 1.3 amosa un diagrama de dispersión simulado dende o modelo (1.13), pero agora  $\varepsilon_i$  segue unha distribución normal de media 0 e desviación típica dada por  $0.4 + 0.8x_i$ ,  $N(0, (0.4 + 0.8x_i)^2)$ . Isto significa que a varianza do erro depende do valor da variable explicativa, polo que estamos ante un modelo heterocedástico. Como se pode observar, neste caso cada unha das rectas de regresión presenta un valor da pendente distinto, polo que xa non son paralelas.

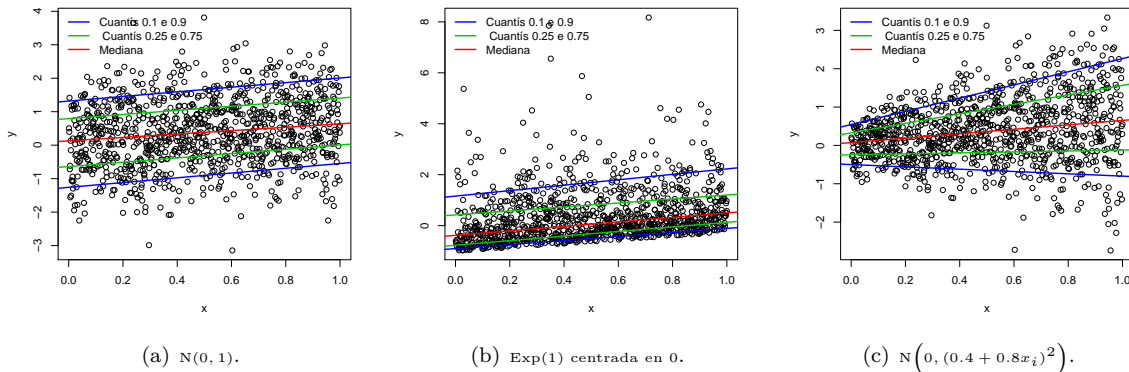


Figura 1.3: Diagramas de dispersión xerados dende o modelo (1.13), para tres distribucións do erro distintas. En cada un deles amósanse as rectas de regresión axustadas para cuantís  $\tau = 0.1, 0.25, 0.5, 0.75$  e  $0.9$ .

Precisamente, a violación da hipótese de homocedasticidade é, quizais, a situación na que a regresión cuantil adquire un papel máis destacado, dado que é precisa unha descrición da relación de dependencia máis complexa que a que proporciona a regresión lineal clásica.

Vistos os beneficios que achega o enfoque cuantil á regresión, nos capítulos vindeiros estudarase a extensión deste enfoque a modelos máis complexos, como son os modelos de series temporais autorregresivos estacionarios. No Capítulo 2 centrarémonos nos procesos autorregresivos usuais, mentres que no Capítulo 3 presentaremos a extensión deste tipo de procesos nunha perspectiva cuantil e heterocedástica, tal como proponen Koenker e Xiao (2006).

<sup>3</sup>É posible a realización dun contraste formal para saber se a diferenza entre as pendentes é estatisticamente significativa, véxase Koenker (2005) páxinas 75–77, pero neste traballo non imos abordar este tema.

## Capítulo 2

# O modelo autorregresivo clásico

Na vida real en moitas ámbitos as observacións que se poden recoller dun proceso estocástico non son independentes, senón que presentan unha estrutura de dependencia temporal, polo que o análise de regresión clásico pode non ser axeitado. Para adaptarse a esas circunstancias propuxéronse os modelos de series de tempo, os cales constitúen unha ferramenta moi estudada na estatística.

A limitación que supón contar cun número finito de observacións, unida á comodidade que proporciona o traballo con modelos paramétricos explica o usual que é considerar técnicas paramétricas no tratamento de series de tempo. Concretamente, os modelos lineais autorregresivos (AR) foron moi estudados na literatura xa que, ademais do interese que poden ter en sí mesmos en áreas como a Economía ou o Control Estatístico da Calidade, constitúen unha peza fundamental no estudo de procesos máis complexos (como son, por exemplo, os modelos autorregresivos integrados de medias móbiles, ARIMA, ou os modelos autorregresivos con heterocedasticidade condicional, ARCH).

Os procesos autorregresivos presentan a peculiaridade de que poden ser estimados usando as ferramentas da regresión, dado que se pode pensar o problema como o de estimación da media do proceso condicionada aos seus valores en instantes anteriores. Isto fai que sexa natural, dende un enfoque menos explotado na literatura, estudalo tamén dende o punto de vista da regresión cuantil, analizando o comportamento do cuantil condicional.

O Capítulo 2, ao igual que o 1, asenta as bases para poder presentar o modelo cuantil autorregresivo introducido por Koenker e Xiao (2006). Por outra parte, dado que un dos propósitos deste traballo é abordar o tema da predición nos modelos autorregresivos (Capítulo 4), é fundamental un bo coñecemento deste proceso para a aplicación de procedementos consistentes.

Neste capítulo, nunha primeira sección introduciremos conceptos fundamentais relacionados coas series temporais, como son as funcións de autocorrelacións simples e parciais. Na Sección 2, revisamos o proceso autorregresivo clásico e presentamos unha primeira xeneralización do mesmo derivada do enfoque cuantil, mentres que na Sección 3 presentaremos dous procedementos para estimar os seus coeficientes, mínimos cadrados ordinarios e mínimas desviacións absolutas, o que enlaza directamente co Capítulo 1 deste traballo. As referencias principais que empregaremos na redacción serán Brockwell e Davis (1990), Wei (2006) e Koenker (2005).

### 2.1. Elementos básicos dunha serie de tempo

Antes de presentar o modelo de serie de tempo autorregresivo, imos revisar unha serie de conceptos fundamentais no contexto dos procesos estocásticos en xeral que son precisos para a correcta comprensión destes modelos. Ademais, algunhas destas ideas introdutorias, como son as funcións de correlacións simple e parcial do proceso, estenderanse no Capítulo 3 ao modelo proposto por Koenker e Xiao (2006), polo que esta sección terá tamén importancia no capítulo seguinte.

**Definición 2.1.** Un *proceso estocástico* é unha colección de variables aleatorias  $\{Y_t\}_{t \in \Gamma}$  definidas no mesmo espazo de probabilidade  $(\Omega, \mathcal{F}, \mathbb{P})$  e indexadas polo parámetro  $t$ , o cal se move no espazo  $\Gamma$ . As variables aleatorias  $Y_t$  toman valores no espazo de estados  $S$ .

Os procesos estocásticos adoitan clasificarse en función dos espazos cos que é posible identificar a  $\Gamma$  e  $S$ . Se  $\Gamma \subset \mathbb{Z}$ , fálase de proceso estocástico en *tempo discreto*, mentres que cando  $\Gamma$  é un intervalo de  $\mathbb{R}$  dirase que é en *tempo continuo*. Se  $S \subset \mathbb{Z}$ , trátase dun proceso estocástico con *espazo de estados discreto*, mentres que cando  $S$  é un intervalo de  $\mathbb{R}$  terá *espazo de estados continuo*. Por outra parte, cada realización do proceso estocástico  $\{Y_t(\omega)\}_{t \in \Gamma}$  con  $\omega \in \Omega$  denomínase *traxectoria*.

Neste traballo imos considerar procesos estocásticos en tempo discreto e con espazo de estados continuo,  $\{Y_t\}_{t \in \mathbb{Z}}$ , onde  $Y_t$  é unha variable aleatoria continua, dos cales suporemos que somos capaces de observar parte da súa traxectoria  $\{Y_t\}_{t=1}^n$ . Á realización particular deste tramo finito da traxectoria (datos)  $\{y_t\}_{t=1}^n$  adoita chamárselle *serie temporal*. A continuación, imos introducir varios conceptos esenciais que caracterizan este tipo de procesos.

Para un proceso estocástico  $\{Y_t\}_{t \in \mathbb{Z}}$ , con espazo de estados continuo, é posible definir a *función de medias* (chamámoslle función xa que pode depender do tempo) como segue:

$$\mu_t = \mathbb{E}(Y_t).$$

A partir dela, se  $\mu_t < \infty$  para cada  $t \in \mathbb{Z}$ , a *función de varianzas* do proceso viría dada por

$$\text{Var}(Y_t) = \mathbb{E}(Y_t - \mu_t)^2.$$

Cando se trata cun número finito de variables aleatorias, adoita ser útil calcular a matriz de covarianzas das mesmas <sup>1</sup>, para cuantificar a relación de dependencia existente entre elas. No caso das series temporais, é preciso estender o concepto de matriz de covarianzas para tratar con coleccións infinitas de variables. A función de autocovarianzas do proceso permite realizar tal extensión.

**Definición 2.2.** Se  $\{Y_t\}_{t \in \mathbb{Z}}$  é un proceso estocástico tal que  $\text{Var}(Y_t) < \infty$  para cada  $t \in \mathbb{Z}$ , entón a función de autocovarianzas do mesmo defínese como:

$$\gamma(r, s) = \text{Cov}(Y_r, Y_s) = \mathbb{E}[(Y_r - \mu_r)(Y_s - \mu_s)], \quad r, s \in \mathbb{Z}.$$

Un concepto directamente relacionado cos momentos do proceso e que resulta moi importante na análise de series temporais é o de estacionaridade.

<sup>1</sup>Dado un vector aleatorio  $(X, Y)$  no que as dúas variables aleatorias que o constitúen,  $X$  e  $Y$ , teñen momentos de segunda orde finitos, a *covarianza* entre elas pódese definir como:

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)],$$

onde  $\mu_X = \mathbb{E}(X)$  e  $\mu_Y = \mathbb{E}(Y)$ . A covarianza é unha medida da variación conxunta das dúas variables aleatorias que indica se existe relación lineal entre elas: cando os valores elevados dunha se corresponden cos valores altos da outra e ocorre o mesmo para os valores baixos, díse que a covarianza é positiva; cando os valores elevados dunha coinciden cos pequenos da outra, será negativa. A covarianza satisfai a propiedade de simetría:  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ .

Por outra parte, a matriz de varianzas-covarianzas dun vector aleatorio bidimensional  $(X, Y)$  (se fose multidimensional a definición é análoga) defínese como

$$\Sigma = \begin{pmatrix} \text{Var}(X) & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & \text{Var}(Y) \end{pmatrix}. \tag{2.1}$$



**Definición 2.3.** O proceso  $\{Y_t\}_{t \in \mathbb{Z}}$  dise que é *estacionario* se verifica as tres condicións seguintes:

1.  $\mathbb{E}|Y_t|^2 < \infty$  para todo  $t \in \mathbb{Z}$ .
2.  $\mu_t = \mu$  para todo  $t \in \mathbb{Z}$ .
3.  $\gamma(r, s) = \gamma(r + t, s + t)$  para todo  $t, r, s \in \mathbb{Z}$ .

A esta definición de estacionariedade na literatura tamén é usual chamarlle estacionariedade débil, covarianza estacionaria ou estacionariedade de segunda orde en sentido amplo. A condición 2 da **Definición 2.3** vén a dicir que a media do proceso é constante no tempo, mentres que a condición 1 implica que o proceso tamén ha de ter varianza finita para ser estacionario. Finalmente, que a función de covarianzas dependa unicamente da separación temporal entre variables, en particular, implica que a varianza do proceso ha de ser constante no tempo.

Un dos procesos estacionarios por excelencia é o coñecido como *ruído branco*, usualmente denotado na literatura por  $\{a_t\}_{t \in \mathbb{Z}}$ , unha colección de variables aleatorias independentes e idénticamente distribuídas, con distribución común,  $F$ , de media 0 e varianza finita  $\sigma_a^2$ .

Notemos tamén que se  $\{Y_t\}_{t \in \mathbb{Z}}$  é un proceso estacionario, daquela  $\gamma(r, s) = \gamma(r - s, 0)$  para todo  $r, s \in \mathbb{Z}$ , polo que é posible reparametrizar a autocovarianza en función dun único argumento:

$$\gamma_k = \gamma(k, 0) = \text{Cov}(Y_{t+k}, Y_t), \quad t, k \in \mathbb{Z}. \quad (2.2)$$

A función  $\gamma_k$  será denominada función de autocovarianzas de  $\{Y_t\}_{t \in \mathbb{Z}}$ . Analogamente, para un proceso estacionario podemos definir a *función de autocorrelacións simples*<sup>2</sup> do proceso (en inglés *autocorrelation function*, ACF) para retardo  $k$  como:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \text{Cor}(Y_{t+k}, Y_t). \quad (2.4)$$

dato que  $\gamma_0 = \text{Var}(Y_t) = \text{Var}(Y_{t+k})$  pola asunción de estacionariedade.

Debemos ter en conta, ademais, que as funcións de autocovarianzas e de autocorrelacións dun proceso estacionario satisfán as seguintes propiedades:

1. Como  $\gamma_0 = \text{Var}(Y_t)$ , tense que  $\rho_0 = 1$ .
2.  $|\gamma_k| \leq \gamma_0$ , polo que  $|\rho_k| \leq 1$ .
3.  $\gamma_k = \gamma_{-k}$  e  $\rho_k = \rho_{-k}$  para calquera  $k \in \mathbb{Z}$ , dado que  $\gamma_k$  e  $\rho_k$  son simétricas con respecto á orixe de tempos  $k = 0$ . Isto é consecuencia inmediata de que a diferenza de tempos entre  $Y_t$  e  $Y_{t+k}$  é a mesma que entre  $Y_t$  e  $Y_{t-k}$ .
4.  $\gamma_k$  e  $\rho_k$  son semidefinidas positivas, é dicir,

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma_{|t_i - t_j|} \geq 0 \quad \text{e} \quad \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho_{|t_i - t_j|} \geq 0,$$

<sup>2</sup>A partir da covarianza entre dúas variables é posible definir o *coeficiente de correlación de Pearson* que nos proporciona unha medida adimensional da relación lineal existente entre elas. Dito coeficiente de correlación é o cociente da covarianza entre as variables dividida polo produto das desviacións típicas, é dicir:

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Sempre se verifica que  $-1 \leq \text{Cor}(X, Y) \leq 1$ , onde os valores 1 e  $-1$  indican unha relación lineal perfecta positiva e negativa, respectivamente, entre as dúas variables, mentres que o valor 0 indica que non hai relación lineal. Por outra parte a matriz de correlacións dun vector aleatorio bidimensional  $(X, Y)$  (se fose multidimensional a definición é análoga) vén dada por:

$$\Sigma = \begin{pmatrix} 1 & \text{Cor}(X, Y) \\ \text{Cor}(X, Y) & 1 \end{pmatrix}. \quad (2.3)$$

para calquera conxunto de tempos  $t_1, t_2, \dots, t_n \in \mathbb{Z}$  e calquera  $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R}$ .

No contexto das series de tempo, ademais de coñecer a autocorrelación entre dúas variables  $Y_t$  e  $Y_{t+k}$  do proceso estacionario  $\{Y_t\}_{t \in \mathbb{Z}}$ , pode ser preciso estudar a relación lineal que hai entre ambas pero tras eliminar delas a influencia das variables intermedias  $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$ . Esta autocorrelación como función de  $k$  (para un proceso estacionario depende unicamente da separación entre as variables), coñécese como *función de autocorrelacións parciais* (do inglés, *partial autocorrelation function*, PACF).

**Definición 2.4.** A *función de autocorrelacións parciais* dun proceso estacionario defínese como

$$\begin{aligned} \phi_{11} &= \text{Cor}(Y_{t+1}, Y_t) = \rho_1, \\ \phi_{kk} &= \text{Cor}(Y_{t+k} - \hat{Y}_{t+k}, Y_t - \hat{Y}_t), \quad k \geq 2, \end{aligned} \tag{2.5}$$

onde  $\hat{Y}_{t+k}$  e  $\hat{Y}_t$  son os mellores predictores lineais de  $Y_{t+k}$  e  $Y_t$ , respectivamente, construídos en base a  $Y_{t+k-1}, \dots, Y_{t+1}$ .

Notemos que a PACF é unha extensión ao contexto das series temporais do coeficiente de correlación parcial da regresión ordinaria <sup>3</sup>. No Apéndice A inclúense os cálculos que permiten chegar a unha expresión explícita para a PACF, en función das correlacións simples do proceso.

### Estimación das funcións de autocorrelacións

De contarmos con varias realizacións do proceso estacionario independentes, estimariamos mediante elas a función de medias e a de varianzas do proceso (ao contar con varias realizacións de cada variable poderíamos obter a media mostral de cada unha delas). Non obstante, na práctica é usual contar cunha única realización do proceso, polo que dispoñemos dunha única observación de cada variable. Neste punto é onde cobra importancia a asunción de estacionariedade, porque permite estimar media e varianza do proceso mediante a media e varianza mostral no tempo.

Entón, coñecido un tramo da traxectoria de tamaño  $n$  dun proceso estacionario,  $\{Y_t\}_{t=1}^n$ , usariamos como estimador da media do proceso  $\mu = \mathbb{E}(Y_t)$  a media mostral no tempo, é dicir,

$$\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t, \tag{2.6}$$

a cal é un estimador non nesgado e consistente (véxase, por exemplo Wei 1990).

Analogamente, contando cunha única realización parcial do proceso estacionario  $\{Y_t\}_{t=1}^n$ , poderíamos

<sup>3</sup>Supoñamos que temos un modelo de regresión lineal con variable resposta  $Y$ , un conxunto de variables explicativas  $\mathbf{Z} = (Z_1, \dots, Z_p)$  de dimensión  $p \times 1$  e, a maiores, unha variable adicional  $X$ . O coeficiente de correlación parcial entre  $Y$  e  $X$  dado o conxunto  $\mathbf{Z}$  coincide co coeficiente de correlación lineal entre os erros da regresión lineal de  $Y$  sobre o conxunto de variables  $\mathbf{Z}$  e os erros da regresión lineal de  $X$  sobre o conxunto de variables  $\mathbf{Z}$ , é dicir,

$$\text{pcor}(X, Y|\mathbf{Z}) = \text{Cor}(Y - \boldsymbol{\theta}_1^\top \mathbf{Z}^1, X - \boldsymbol{\theta}_2^\top \mathbf{Z}^1) = \frac{\text{Cov}(Y - \boldsymbol{\theta}_1^\top \mathbf{Z}^1, X - \boldsymbol{\theta}_2^\top \mathbf{Z}^1)}{\sqrt{\text{Var}(Y - \boldsymbol{\theta}_1^\top \mathbf{Z}^1) \text{Var}(X - \boldsymbol{\theta}_2^\top \mathbf{Z}^1)}},$$

onde  $\mathbf{Z}^1 = (1, Z_1, \dots, Z_p)$ ,  $\boldsymbol{\theta}_1$  e  $\boldsymbol{\theta}_2$  denotan os parámetros da regresión lineal de  $Y$  respecto de  $\mathbf{Z}$  e de  $X$  respecto de  $\mathbf{Z}$ , respectivamente:

$$\boldsymbol{\theta}_1^\top = \arg \min_{\boldsymbol{\theta}} \mathbb{E}(Y - \boldsymbol{\theta}^\top \mathbf{Z}^1)^2, \quad \boldsymbol{\theta}_2^\top = \arg \min_{\boldsymbol{\theta}} \mathbb{E}(X - \boldsymbol{\theta}^\top \mathbf{Z}^1)^2.$$

Como é coñecido, o coeficiente de correlación parcial pode verse como unha medida da relación lineal entre as variables  $Y$  e  $X$  despois de extraer destas a influencia do conxunto de variables  $\mathbf{Z}$ .

empregar o seguinte estimador para a función de autocovarianzas

$$\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (Y_t - \bar{Y}) (Y_{t+k} - \bar{Y}) \quad \text{con } k = 0, 1, \dots, K, \quad K < n, \quad (2.7)$$

aínda que, en principio, podería parecer máis razoable o emprego do estimador

$$\hat{\hat{\gamma}}_k = \frac{1}{n-k} \sum_{t=1}^{n-k} (Y_t - \bar{Y}) (Y_{t+k} - \bar{Y}) \quad \text{con } k = 0, 1, \dots, K, \quad K < n. \quad (2.8)$$

Tanto  $\hat{\gamma}_k$  como  $\hat{\hat{\gamma}}_k$  son estimadores nesgados de  $\gamma_k$ . Como pode verse en Wei (1990), en xeral,  $\hat{\gamma}_k$  ten un nesgo maior que  $\hat{\hat{\gamma}}_k$ , especialmente cando  $k$  é grande con respecto a  $n$ . Por esta razón, adoita recomendarse considerar un valor de  $K$  como moito  $n/4$ . Non obstante, baixo a condición de que a dependencia entre variables tenda a cero cando o retardo considerado tende a infinito, e dado que a varianza da media mostral do proceso tende a cero co aumento do tamaño mostral, ambos estimadores son asintoticamente non nesgados.

Consecuentemente, para comparalos é conveniente empregar o erro cadrático medio. Neste caso, é inferior o de  $\hat{\gamma}_k$  e, ademais,  $\hat{\gamma}_k$  é sempre semidefinido positivo propiedade moi desexable, dado que tamén a verifica  $\gamma_k$ , pero  $\hat{\hat{\gamma}}_k$  non ten por que selo. Por tal razón, empregaremos como estimador de  $\gamma_k$  a  $\hat{\gamma}_k$ .

Analogamente, para a función de autocorrelacións podemos definir o estimador

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y}) (Y_{t+k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}, \quad (2.9)$$

usualmente coñecido como función de autocorrelacións mostrais (ACF mostral).

Finalmente, a PACF mostral,  $\hat{\phi}_{kk}$ , obtense mediante a substitución en (A.7) de  $\rho_i$  polas súas estimacións  $\hat{\rho}_i$ . Para iso, en lugar do cálculo de determinantes adóitase recorrer ao método recursivo dado por Durbin, véxase Wei (2006):

$$\begin{aligned} \hat{\phi}_{11} &= \hat{\rho}_1, \\ \hat{\phi}_{k+1,k+1} &= \frac{\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_j}, \\ \hat{\phi}_{k+1,j} &= \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1} \hat{\phi}_{k,k+1-j}, \quad j = 1, \dots, k. \end{aligned}$$

Cando o tamaño da serie temporal é grande, se esta foi xerada por un proceso de ruído branco verificase que tanto as estimacións das autocorrelacións simples como as estimacións das autocorrelacións parciais, para cada retardo  $k$ , seguen asintoticamente unha distribución normal de media 0 e desviación típica aproximadamente igual a  $1/\sqrt{n}$ :

$$\begin{aligned} \hat{\rho}_k &\approx N\left(0, \frac{1}{n}\right), \\ \hat{\phi}_{kk} &\approx N\left(0, \frac{1}{n}\right). \end{aligned}$$

Daquela, para  $n$  grande, se a serie foi xerada por un proceso de ruído branco, con un nivel de significación individual  $\alpha$  para cada  $k$ , tanto  $\hat{\phi}_{kk}$  como  $\hat{\rho}_k$  deben pertencer ao intervalo

$$\left(-\frac{z_{\alpha/2}}{\sqrt{n}}, \frac{z_{\alpha/2}}{\sqrt{n}}\right), \quad (2.10)$$

onde  $z_{\alpha/2}$  representa o cuantil  $(1 - \alpha/2)$  da distribución normal estándar.

## 2.2. Definición e características do proceso autorregresivo

Un modelo de serie temporal moi empregado na práctica é o autorregresivo, no que un valor do proceso é expresado como unha combinación lineal finita de valores previos do mesmo máis un erro aleatorio.

**Definición 2.5.** Sexa  $\{Y_t\}_{t \in \mathbb{Z}}$  un proceso estacionario, e sexa  $\{a_t\}_{t \in \mathbb{Z}}$  ruído branco. Se o proceso admite unha representación do tipo

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t, \quad (2.11)$$

onde  $\phi_0, \phi_1, \dots, \phi_p$  son constantes, dise que segue un *modelo autorregresivo* de orde  $p$ ,  $AR(p)$ .

Para estudar este tipo de modelos, como é usual na literatura, asumiremos que o ruído branco non depende do proceso en instantes previos ou polo menos que está incorrelacionado con eles (nótese que non se está pedindo normalidade), é dicir, que

$$\mathbb{E}(a_t Y_{t-k}) = 0 \quad \text{para } k > 0. \quad (2.12)$$

Por outra parte, é habitual ver o proceso autorregresivo escrito de forma compacta como:

$$\Phi_p(B) Y_t = \phi_0 + a_t,$$

onde  $\Phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ , sendo  $B$  o operador retardo definido como  $BY_t = Y_{t-1}$ . Isto permite expresar a condición de estacionaridade do proceso dicindo que para que se verifique as raíces do polinomio  $\Phi_p(B)$  deben ter módulo maior cá unidade, ou o que é o mesmo, as raíces de  $x^p - \phi_1 x^{p-1} - \phi_2 x^{p-2} - \dots - \phi_p$  deben ter módulo menor que 1. Por exemplo, para modelos  $AR(1)$ :  $Y_t = \phi_0 + \phi_1 Y_{t-1} + a_t$ , esta condición redúcese a pedir que  $|\phi_1| < 1$ .

Un resultado importante que se vai empregar na construción de intervalos de predición no Capítulo 4 é que a asunción de estacionaridade sumada á ausencia de compoñentes deterministas (tendencia), permite representar os procesos  $AR(p)$  seguindo a descomposición de Wold, como

$$Y_t = \psi_0 a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \quad \text{con } \psi_0 = 1 \quad \text{e} \quad \sum_{i=0}^{\infty} \psi_i^2 < \infty, \quad (2.13)$$

onde  $\{a_t\}_{t \in \mathbb{Z}}$  é un proceso de ruído branco.

A continuación, imos particularizar a forma das funcións de autocorrelacións simples e parciais no caso dos procesos  $AR(p)$ , as cales resultan determinantes á hora de identificar este tipo de procesos.

### 2.2.1. ACF do $AR(p)$

Supoñamos, sen perda de xeneralidade, que  $\mu = \mathbb{E}(Y_t) = 0$ . Isto tradúcese en que na ecuación (2.11) a constante sería nula,  $\phi_0 = 0$ . Multiplicando ambos membros de (2.11) con  $\phi_0 = 0$  por  $Y_{t-k}$  con  $k > 0$  obteríamos

$$Y_{t-k} Y_t = \phi_1 Y_{t-k} Y_{t-1} + \phi_2 Y_{t-k} Y_{t-2} + \dots + \phi_p Y_{t-k} Y_{t-p} + Y_{t-k} a_t. \quad (2.14)$$

Se agora tomamos esperanzas en ambos membros da ecuación anterior e tendo en conta a independencia de  $Y_{t-k}$  con respecto a  $a_t$  para  $k > 0$ , que fai que  $\mathbb{E}(Y_{t-k} a_t) = 0$ , obtemos

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p}.$$

Agora, dividindo entre  $\gamma_0$  temos a igualdade

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p}. \quad (2.15)$$

Como vemos, as autocorrelacións satisfán unha igualdade case análoga á ecuación de dependencia que modela o proceso autorregresivo (de non ser porque agora o lado dereito da ecuación é cero),

$$\Phi_p(B) \rho_k = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \rho_k = 0,$$

para  $k > 0$ . Isto permítenos caracterizar a ACF do proceso en base ao polinomio característico, sendo precisamente as raíces de dito polinomio as que determinan a súa forma. En calquera dos casos, esta función de autocorrelacións decae a cero progresivamente no caso dos modelos  $AR(p)$ , sendo unha suma de decrecementos de tipo exponencial: cada raíz real aporta unha compoñente da función de autocorrelacións que decae exponencialmente de xeito amortiguado, mentres que cada par de raíces complexas conxugadas aporta una oscilación en forma de onda, amortiguada exponencialmente. Vexamos isto mediante un exemplo.

**Exemplo 2.6.** Consideremos unha serie temporal de tamaño  $n = 1000$  xerada dende os procesos  $AR(2)$  seguintes (fixando semente 123456789 en R Core Team (2015)), ambos estacionarios:

**Caso 1.**  $Y_t = 0.75Y_{t-1} - 0.5Y_{t-2}$ .

**Caso 2.**  $Y_t = 0.2Y_{t-1} + 0.5Y_{t-2}$ .

No caso 1, o polinomio característico do proceso será  $\Phi_2(x) = 1 - 0.75x + 0.5x^2$ , polo que mediante un sinxelo cálculo podemos ver que as súas raíces son  $0.75 \pm \sqrt{23}/4i$ , é dicir, complexas con parte real positiva. A representación gráfica da ACF mostral para a serie xerada baixo este proceso móstrase na esquerda da Figura 2.1. Esta foi obtida mediante a función `acf` do programa R. No eixo horizontal sitúase o retardo  $k$  para o que se obtén a autocorrelación, mentres que cada barra vertical representa o valor da autocorrelación mostral (con signo) para cada  $k$ . As dúas liñas horizontais punteadas en azul son as bandas de confianza para as autocorrelacións a un nivel de confianza do 95 %, situadas a  $\pm 1.96/\sqrt{n}$ . Como se poderá observar, a ACF mostral do proceso decae progresivamente a cero, ata quedar dentro das bandas que permiten asumir que é nula. Este decaemento faíno oscilando en forma de onda.

No caso 2, o polinomio característico do proceso será  $\Phi_2(x) = 1 - 0.2x - 0.5x^2$  e ten como raíces  $1/5 \pm \sqrt{51}/5$ , é dicir, posúe dúas raíces reais, unha positiva e a outra negativa. A representación da ACF mostral amósase na Figura 2.1 á dereita. Como se poderá observar agora en lugar de termos ondas, a ACF mostral vai decaendo a cero exponencialmente ata quedar comprendida dentro das bandas de confianza, onde non se terían evidencias para rexeitar que é nula.

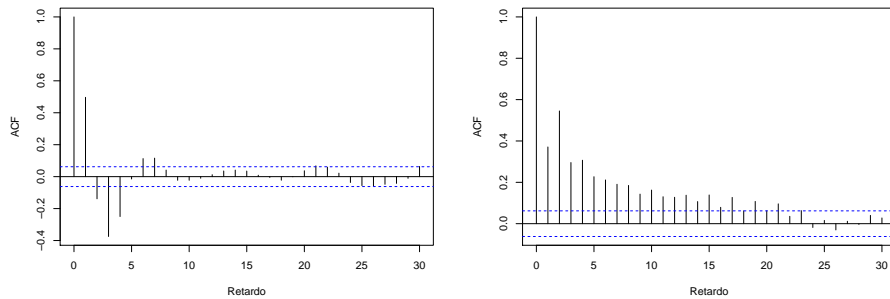


Figura 2.1: Representación da ACF mostral para as series xeradas baixo os procesos dos casos 1, na esquerda e 2, na dereita.

### 2.2.2. PACF do $AR(p)$

En vista da relación verificada polas autocorrelacións simples dada por (2.15), e se temos en conta a expresión de  $\phi_{kk}$  dada por (A.7), vemos que cando  $k > p$  a última columna do determinante do numerador vai ser combinación lineal das restantes columnas, o que implica que  $\phi_{kk} = 0$ . Entón, a PACF dun proceso

autorregresivo de orde  $p$  será nula cando  $k > p$ , feito que resulta útil para identificar este tipo de procesos e para determinar a súa orde na práctica.

Baixo a hipótese de que o proceso é un  $AR(p)$ , a estimación das autocorrelacións parciais para retardo  $p+1$  e maiores seguen asintoticamente unha distribución normal de media 0 e desviación típica aproximadamente  $\frac{1}{\sqrt{n}}$ .

Daquela, cando  $n$  é grande, se a serie foi xerada por un proceso  $AR(p)$  debe verificarse que para cada  $k > p$ , con un nivel de significación individual  $\alpha$ ,  $\hat{\phi}_{kk}$  pertence ao intervalo

$$\left( -\frac{z_{\alpha/2}}{\sqrt{n}}, \frac{z_{\alpha/2}}{\sqrt{n}} \right), \quad (2.16)$$

onde  $z_{\alpha/2}$  representa o cuantil  $(1 - \alpha/2)$  da distribución normal estándar.

**Exemplo 2.7.** Para as series procedentes dos dous procesos autorregresivos presentados no **Exemplo 2.6**, a representación da PACF mostrada ata retardo  $k = 30$  recóllese na Figura 2.2. Para analizar estes gráficos debemos ter en conta que, debido ao nivel de confianza fixado, o 95 %, unha de cada 20 autocorrelacións podería saírse das bandas por azar (pero nese caso, tería un valor moi próximo ao límite do intervalo). Seguindo esta regra, podemos ver que en ambos casos sería asumible que a última autocorrelación parcial non nula é a segunda, o que sumado aos gráficos da Figura 2.1 permitiría identificar o proceso xerador das series coma un  $AR(2)$ .

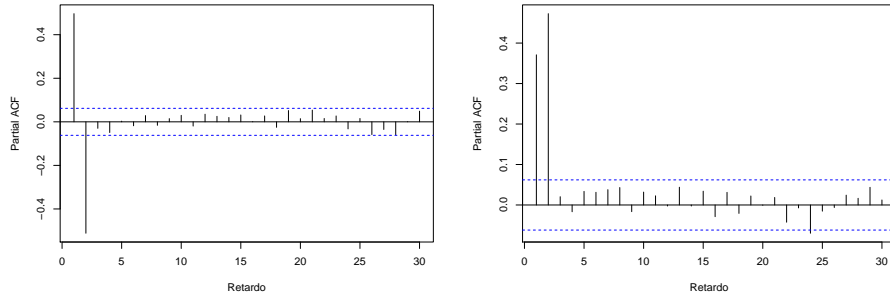


Figura 2.2: Representación da PACF mostrada para as series xeradas baixo os procesos dos casos 1, na esquerda e 2, na dereita.

### 2.2.3. Proceso autorregresivo en cuantil

A estimación do modelo autorregresivo no terreo cuantil, involucra a consideración dun proceso autorregresivo paralelo ao dado na **Definición 2.5**, onde é preciso que a innovación satisfaga outras hipóteses lixeiramente máis suaves, xa que non se pide a existencia do momento de orde dous da innovación do proceso. Seguindo o desenvolvemento de Bloomfield e Steiger (1983) para a autorregresión en mediana, podemos establecer a seguinte definición.

**Definición 2.8.** Sexa  $\{a_t\}_{t \in \mathbb{Z}}$  unha sucesión de variables aleatorias independentes, con distribución común non reticular (“non lattice” en inglés) que denotaremos por  $F_a$ , as cales satisfán tres propiedades:

- $\mathbb{E}(\log_+(a_t)) < \infty$ , onde  $\log_+(a_t) = \max\{0, \log(a_t)\}$ .
- $\mathcal{Q}_\tau(a_t) = 0$  fixado  $\tau$  tal que  $0 < \tau < 1$ .
- $\mathbb{E}(|a_t|) < \infty$ .

Sexa  $\{Y_t\}_{t \in \mathbb{Z}}$  un proceso estacionario, que admite unha representación do tipo:

$$Y_t = \phi_{0\tau} + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t, \quad (2.17)$$

onde  $\phi_{0\tau}, \phi_1, \dots, \phi_p$  son constantes e  $\tau \in (0, 1)$  é o cuantil fixado. Nestas condicións diremos que  $\{Y_t\}_{t \in \mathbb{Z}}$  segue un *modelo cuantil autorregresivo homocedástico* de orde  $p$ ,  $QAR(p)$  homocedástico.

Nótese que o carácter identicamente distribuído de  $\{a_t\}_{t \in \mathbb{Z}}$ , que se identifica coa homocedasticidade, fai que un proceso  $AR(p)$  dado na **Definición 2.5**, en particular, sexa un proceso  $QAR(p)$  para cada valor de  $\tau$  con  $0 < \tau < 1$ . A notación  $\phi_{0\tau}$  fai referencia a que a constante do modelo depende do cuantil  $\tau$  fixado, mentres que os demais coeficientes, baixo a suposición da homocedasticidade do modelo, serían os mesmos para todo  $0 < \tau < 1$ . Isto é consecuencia do explicado na Sección 1.6 deste traballo.

Tamén é posible escribir o proceso autorregresivo cuantil de forma compacta como:

$$\Phi_p(B) Y_t = \phi_{0\tau} + a_t,$$

onde  $\Phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ , sendo  $B$  o operador retardo definido como  $BY_t = Y_{t-1}$ . Como ocurría no proceso autorregresivo en media, a condición de estacionaridade tradúcese en que as raíces do polinomio  $\Phi_p(B)$  deben ser maiores cá unidade.

## 2.3. Estimación dos coeficientes

Aínda que na literatura existen diversos procedementos para estimar os coeficientes dun modelo de serie temporal, neste traballo centrarémonos en describir dous métodos distintos para a estimación do modelo autorregresivo: por unha parte, mínimos cadrados ordinarios (MCO) e, por outra, mínimas desviacións absolutas (MDA). Nas seguintes liñas presentaremos as ideas nas que se basean ámbolos dous procedementos.

Dado un tramo de traxectoria  $\{Y_t\}_{t=1}^n$  que foi xerada por un proceso  $AR(p)$ , os parámetros  $\phi_0, \phi_1, \dots, \phi_p$ , e  $\sigma_a^2$  son descoñecidos. Se denotamos por  $\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p$  as estimacións dos coeficientes, os residuos serían as diferenzas entre os valores observados e os axustados:

$$\hat{a}_t = Y_t - \hat{\phi}_0 - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2} - \dots - \hat{\phi}_p Y_{t-p}, \quad t = p+1, p+2, \dots, n. \quad (2.18)$$

Notemos que para as  $p$  primeiras observacións non seremos capaces de obter o residuo correspondente, dado que para iso teríamos que contar con observacións das variables  $Y_0, Y_{-1}, \dots, Y_{-p+1}$ .

Así, para obter os estimadores dos coeficientes poderíase resolver o problema de optimización

$$\arg \min_{\phi_0, \phi_1, \dots, \phi_p} \sum_{t=p+1}^n (Y_t - \phi_0 - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_p Y_{t-p})^2 = \arg \min_{\phi \in \mathbb{R}^{p+1}} \sum_{t=p+1}^n (Y_t - \phi^\top \mathbf{Z}_{t,p}^1)^2, \quad (2.19)$$

onde  $\mathbf{Z}_{t,p}^1 = (1, Y_{t-1}, \dots, Y_{t-p})^\top$  e  $\phi = (\phi_0, \phi_1, \phi_2, \dots, \phi_p)^\top$ . É dicir, buscaríamos os valores dos coeficientes para os que a suma de residuos ao cadrado é mínima. Este é un problema totalmente análogo ao da regresión ordinaria, polo que poderíamos reproducir o esquema de mínimos cadrados ordinarios presentado no Capítulo 1. Así, chegaríamos a estimadores da forma dos presentados daquela, cunha expresión análoga a (1.9).

Pola súa parte, Bloomfield e Steiger (1983) consideran o proceso dado na **Definición 2.8**, e estudan as propiedades dos estimadores de mínimas desviacións absolutas ( $\tau = 0.5$ ), é dicir, o resultado do problema de minimización

$$\arg \min_{\phi_0, \phi_1, \dots, \phi_p} \frac{1}{n} \sum_{t=p+1}^n |Y_t - \phi_0 - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_p Y_{t-p}| = \arg \min_{\phi \in \mathbb{R}^{p+1}} \sum_{t=p+1}^n |Y_t - \phi^\top \mathbf{Z}_{t,p}^1|, \quad (2.20)$$

onde  $\mathbf{Z}_{t,p}^1 = (1, Y_{t-1}, \dots, Y_{t-p})^\top$  e  $\phi = (\phi_0, \phi_1, \dots, \phi_p)^\top$ . Estes autores proban a consistencia destes estimadores, que novamente, son resultado dun procedemento de estimación análogo ao exposto no Capítulo 1 no contexto da regresión cuantil para o caso particular da mediana. Notemos que nese caso a función de perda cuantílica coincide coa función valor absoluto  $\rho_{0.5}(u) = |u|$ .

Como cita Koenker (2005), cando  $a_t$  ten densidade  $f(0) > 0$  e  $\mathbb{E}(a_t^2) < \infty$  é posible establecer a normalidade asintótica do estimador de mínimas desviacións absolutas. Ademais, no caso particular do proceso

$AR(1)$ , é posible comprobar que o estimador de mínimas desviacións absolutas presenta maior eficiencia asintótica relativa que o de mínimos cadrados ordinarios cando a distribución de  $a_t$  presenta colas pesadas. Martin (1982) proporciona unha xeneralización desta análise da eficiencia ao caso dos procesos  $AR(p)$ , no contexto xeral dos  $M$ -estimadores.

Tamén cabe comentar que Bloomfield e Steiger (1983) mostran vía simulación que nos procesos autorregresivos nos que  $a_t$  ten varianza infinita (estes xa non verificarían a **Definición 2.5**), os estimadores de MDA presentan unha taxa de converxencia mellor que os de MCO.

Realizada esta breve introdución sobre os modelos autorregresivos, podemos observar que o proceso descrito na **Definición 2.8**, aínda non aproveita todas as vantaxes que o cuantil proporciona no contexto da regresión. Por tal razón, Koenker e Xiao (2006) estudan un proceso máis xeral, que permite que os coeficientes autorregresivos do modelo varíen co cuantil, é dicir, que permite certo grao de heterocedasticidade ou comportamentos temporais de tipo raíz unitaria (é dicir, o polinomio característico tal e como o definimos presenta raíces dentro do círculo unidade) violando localmente a hipótese de estacionariedade do proceso.



## Capítulo 3

# Autorregresión cuantil

Os modelos de series temporais de coeficientes constantes foron unha ferramenta moi explotada na literatura. Precisamente, no Capítulo 2 revisamos o modelo autorregresivo clásico e vimos, entre outros feitos, que tamén pode ser estimado empregando as técnicas da regresión cuantil. Non obstante, nese proceso consideramos que os coeficientes autorregresivos non dependen do cuantil que esteamos estimando, é dicir, que son os mesmos para todo o rango de cuantís. Isto supón unha limitación aos beneficios que o enfoque cuantil proporciona no contexto da regresión, dado que os seus procedementos permiten considerar efectos distintos en cuantís diferentes.

Por tal razón, Koenker e Xiao (2006) propoñen un modelo cuantil autorregresivo, QAR, máis flexible que permite relaxar a hipótese de independencia dos coeficientes respecto ao cuantil a estimar. Este modelo pretende captar a influencia das variables correspondentes a instantes previos do proceso non só na localización e na escala da distribución condicional da resposta, senón tamén na forma, polo que proporciona unha alternativa na análise de series temporais con dinámicas asimétricas ou con persistencias locais.

O cambio dos coeficientes autorregresivos co cuantil fai que a fase de identificación da orde do modelo autorregresivo cobre unha importancia especial, polo que neste contexto, Li et al. (2015) introducen a extensión ao terreo cuantil da función de autocorrelacións parciais descrita no Capítulo 2. Esta nova formulación que propoñen caracteriza a estrutura de correlación parcial do proceso, permitindo identificar a orde do modelo para cada cuantil, podéndose obter así unha visión máis ampla da relación de dependencia, ao igual que ocorría na regresión.

Neste Capítulo presentamos o modelo de Koenker e Xiao (2006), enunciando resultados que aseguran a súa estacionariedade, e ocuparémonos da estimación dos seus parámetros. Acto seguido, estudaremos a definición de correlación simple e parcial cuantil dada por Li et al. (2015) e a partir de aí, obteremos as funcións de autocorrelacións simples e parciais do proceso QAR, o que nos permitirá identificar a orde do modelo. Finalmente, incluímos un exemplo con datos simulados que ilustra como se levaría a cabo este proceso de identificación da orde do modelo QAR.

### 3.1. O modelo QAR

Como nos modelos autorregresivos clásicos, no proposto por Koenker e Xiao (2006) o cuantil condicional da resposta  $Y_t$  con  $t \in \mathbb{Z}$  pode ser expresado coma unha combinación lineal finita de valores previos do proceso. A diferenza é que agora os coeficientes son función do cuantil  $\tau$ . O desenvolvemento seguinte recolle de xeito máis preciso en que consiste este proceso.

Sexa  $\{U_t\}_{t \in \mathbb{Z}}$  unha sucesión de variables aleatorias i.i.d. con distribución uniforme en  $[0, 1]$ . Consideremos o proceso dado por

$$Y_t = \phi_0(U_t) + \phi_1(U_t)Y_{t-1} + \dots + \phi_p(U_t)Y_{t-p}, \quad (3.1)$$

onde  $\phi_j(\cdot)$  con  $j = 1, \dots, p$  son funcións tales que  $\phi_j : [0, 1] \rightarrow \mathbb{R}$ . Se agora pedimos que na expresión (3.1)

o lado dereito da igualdade sexa monótono crecente con  $U_t$ , podemos afirmar que a función cuantil  $\tau$  do proceso vén dada por

$$\mathcal{Q}_\tau(Y_t|Y_{t-1}, \dots, Y_{t-p}) = \phi_0(\tau) + \phi_1(\tau)Y_{t-1} + \dots + \phi_p(\tau)Y_{t-p}. \quad (3.2)$$

O paso da expresión (3.1) á expresión (3.2) é consecuencia de dous feitos:

- Dada unha función  $g$  estritamente monótona crecente e unha aleatoria  $X$ , tense que  $F_{g(X)}(g(x)) = \mathbb{P}(g(X) \leq g(x)) = \mathbb{P}(X \leq x) = F_X(x)$  para todo  $x \in \mathbb{R}$ , do que se segue que  $\mathcal{Q}_\tau(g(X)) = g(\mathcal{Q}_\tau(X))$ .
- Ademais, para unha variable aleatoria  $U$  con distribución uniforme en  $[0, 1]$ ,  $\mathcal{Q}_\tau(U) = \tau$ .

Con isto temos que

$$\mathcal{Q}_\tau(g(U)) = g(\mathcal{Q}_\tau(U)) = g(\tau).$$

Por outra parte, notemos que a expresión (3.2) pode ser escrita de xeito compacto como

$$\mathcal{Q}_\tau(Y_t|\mathcal{F}_{t-1}) = \boldsymbol{\phi}(\tau)^\top \mathbf{Z}_{t,p}^1, \quad (3.3)$$

onde  $\mathbf{Z}_{t,p}^1 = (1, Y_{t-1}, \dots, Y_{t-p})^\top$ ,  $\boldsymbol{\phi}(\tau) = (\phi_0(\tau), \phi_1(\tau), \dots, \phi_p(\tau))^\top$  e  $\mathcal{F}_{t-1}$  é a sigma álgebra xerada por  $\{Y_s, s \leq t\}$ .

A este modelo dado pola expresión (3.2) (ou pola (3.1) pedindo a monotónía do lado dereito da ecuación) Koenker e Xiao danlle o nome de *modelo cuantil autorregresivo*, QAR. Neste modelo, como se pode observar, os coeficientes varían co cuantil, polo que os valores correspondentes a instantes precedentes do proceso non só afectan á localización do valor actual do mesmo, senón tamén á súa escala e á súa forma. Para ilustrar a flexibilidade destes modelos Koenker e Xiao (2006) inclúen o seguinte exemplo.

**Exemplo 3.1.** Consideremos o modelo QAR(1) seguinte:

$$\mathcal{Q}_\tau(Y_t|\mathcal{F}_{t-1}) = \phi_0(\tau) + \phi_1(\tau)Y_{t-1}, \quad (3.4)$$

onde  $\phi_0(\tau) = \sigma\Phi^{-1}(\tau)$ , sendo  $\Phi$  a función de distribución dunha normal estándar, e  $\phi_1(\tau) = \min\{\gamma_0 + \gamma_1\tau, 1\}$  para  $\gamma_0 \in (0, 1)$  e  $\gamma_1 > 0$ . Cando  $U_t > (1 - \gamma_0)/\gamma_1$  o modelo xera  $Y_t$  como se fora un proceso autorregresivo gaussiano cunha raíz unitaria dado que  $\phi_1(U_t) = 1$ , mentres que cando  $U_t < (1 - \gamma_0)/\gamma_1$ , o coeficiente autorregresivo aumenta linealmente co cuantil manténdose menor que 1.

Se tomamos, en particular,  $\gamma_0 = 0.5$ ,  $\gamma_1 = 0.8$  e  $\sigma = 1$ , o gráfico secuencial dunha realización particular do proceso de tamaño  $n = 100$  sería o dado na Figura 3.1. Como se poderá observar, neste hai pequenas tendencias locais máis acusadas que nos modelos  $AR(1)$  clásicos. Precisamente, o proceso  $AR(1)$  gaussiano clásico obteríase ao fixar  $\phi_1(\tau) = \phi_1$ , é dicir, a unha constante.

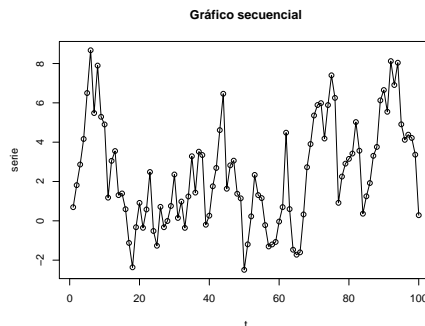


Figura 3.1: Gráfico secuencial dunha serie de tamaño  $n = 100$  do proceso para o cal  $\mathcal{Q}_\tau(Y_t|\mathcal{F}_{t-1}) = \Phi^{-1}(\tau) + \min\{0.5 + 0.8\tau, 1\} Y_{t-1}$ .

O modelo QAR, na súa formulación dada por (3.1), tamén pode ser visto como un tipo especial de modelo autorregresivo de coeficientes aleatorios como o presentado por Nicholls e Quinn (1982). Non obstante, no modelo QAR os coeficientes son dependentes entre sí debido a que todos son función dunha única variable aleatoria  $U_t$ , mentres que nos modelos de coeficientes aleatorios asúmese que estes son independentes.

### 3.2. Propiedades do proceso QAR: estacionariedade

Para enunciarse as propiedades que presenta o proceso QAR, imos reformular a súa expresión de forma máis conveniente, na notación usual dos modelos de coeficientes aleatorios:

$$Y_t = \mu_0 + \phi_{1,t}Y_{t-1} + \dots + \phi_{p,t}Y_{t-p} + u_t, \quad (3.5)$$

onde  $\mu_0 = \mathbb{E}(\phi_0(U_t))$  e  $u_t = \phi_0(U_t) - \mu_0$ , mentres que  $\phi_{j,t} = \phi_j(U_t)$  para  $j = 1, \dots, p$ . Así,  $\{u_t\}_{t \in \mathbb{Z}}$  é unha sucesión de variables aleatorias i.i.d. con función de distribución  $F(x) = \phi_0^{-1}(x + \mu_0)$  e  $\phi_{j,t}$ , con  $j = 1, \dots, p$ , son os coeficientes do modelo que dependen da innovación.

Para establecer as propiedades asintóticas do proceso QAR é preciso introducir tres hipóteses:

**B1.** A sucesión de variables i.i.d.  $\{u_t\}_{t \in \mathbb{Z}}$  teñen media 0 e varianza finita  $\sigma^2$ . Ademais, a súa función de distribución  $F(\cdot)$  ten densidade continua  $f(\cdot)$  verificando que  $f(u) > 0$  para cada  $u$  en  $\mathcal{U} = \{u : 0 < F(u) < 1\}$ .

**B2.** Comecemos expresando (3.5) como un proceso autorregresivo vectorial  $p$ -dimensional de orde 1:

$$\mathbf{Y}_t = \Gamma + \vec{\mathbf{A}}_t \mathbf{Y}_{t-1} + \mathbf{V}_t$$

sendo

$$\Gamma = \begin{pmatrix} \mu_0 \\ \mathbf{0}_{p-1} \end{pmatrix}, \quad \vec{\mathbf{A}}_t = \begin{pmatrix} \mathbf{A}_{p-1,t} & \phi_{p,t} \\ \mathbf{I}_{p-1} & \mathbf{0}_{p-1} \end{pmatrix}, \quad \mathbf{V}_t = \begin{pmatrix} u_t \\ \mathbf{0}_{p-1} \end{pmatrix},$$

onde  $\mathbf{A}_{p-1,t} = (\phi_{1,t}, \dots, \phi_{p-1,t})$ ,  $\mathbf{Y}_t = (Y_t, \dots, Y_{t-p+1})^\top$  y  $\mathbf{0}_{p-1}$  é o vector de dimensión  $p-1$  de ceros. É preciso pedir que se  $\mathbb{E}(\vec{\mathbf{A}}_t \otimes \vec{\mathbf{A}}_t) = \vec{\boldsymbol{\Omega}}_A$ , onde  $\otimes$  representa o produto tensorial de matrices, os autovalores desta matriz  $\vec{\boldsymbol{\Omega}}_A$  teñan módulo inferior á unidade.

**B3.** A distribución condicional do proceso  $\mathbb{P}(Y_t | \mathcal{F}_{t-1})$ , denotada por  $F_{t-1}(\cdot)$  é derivable, e a súa derivada  $f_{t-1}(\cdot)$  é uniformemente integrable en  $\mathcal{U}$ .

Koenker e Xiao (2006) proban a estacionariedade do proceso, que queda enunciada mediante o seguinte resultado:

**Teorema 3.2.** *Baixo as asuncións B1 e B2, o proceso dado por (3.5) é estacionario (usando a definición de estacionariedade vista no Capítulo 2). Ademais, dito proceso satisfai o teorema central do límite, é dicir,*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n (Y_t - \mu_Y) \xrightarrow{d} N(0, \omega_Y^2) \quad (3.6)$$

onde  $\mu_Y = \mu_0 / (1 - \sum_{j=1}^p \mu_j)$  con  $\mu_0 = \mathbb{E}(\phi_0(U_t))$  e  $\mu_j = \mathbb{E}(\phi_{j,t})$  con  $j = 1, \dots, p$ ,  $\omega_Y^2 = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[ \left( \sum_{t=1}^n (Y_t - \mu_Y) \right)^2 \right]$ .

*Demostración.* A demostración deste resultado pode consultarse nas páxinas 988-989 de Koenker e Xiao (2006).  $\square$

Este teorema permítenos afirmar que os dous primeiros momentos do proceso son constantes no tempo, e que a media mostral no tempo da serie observada segue gozando de boas propiedades. Para ver máis concretamente a aplicabilidade deste teorema podemos enunciarnos unha consecuencia do mesmo para procesos  $QAR(1)$ .

**Corolario 3.3.** *Consideremos o proceso  $QAR(1)$  dado por*

$$Y_t = \phi_{1,t} Y_{t-1} + u_t \tag{3.7}$$

onde,  $\phi_{1,t} = \phi_1(U_t)$  e  $u_t = \phi_0(U_t)$  (asumimos  $\mu_0 = 0$ ). Se  $\mathbb{E}(\phi_{1,t}^2) < 1$ , e baixo a hipótese **B1**, o proceso (3.7) é estacionario e satisfai o teorema central do límite, é dicir,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n Y_t \xrightarrow{d} N(0, \omega_Y^2)$$

onde  $\omega_Y^2 = \sigma^2 (1 + \mu_1) / [(1 - \mu_1)(1 - \mathbb{E}(\phi_{1,t}^2))]$  con  $\mu_1 = \mathbb{E}(\phi_{1,t})$ .

Mediante a aplicación do corolario anterior podemos concluir que o proceso dado no **Exemplo 3.1** é estacionario. Notemos que  $\phi_{1,t} = \min\{\gamma_0 + \gamma_1 \tau, 1\}$  vale como moito 1, e tense que a probabilidade de que non valla 1 é positiva:  $\mathbb{P}(|\phi_t| < 1) > 0$ ; así, podemos afirmar que  $\mathbb{E}(\phi_{1,t}^2) < 1$ . Non obstante, aínda que ese proceso é globalmente estacionario, ten localmente un comportamento tipo raíz unitaria. O **Corolario 3.3** asegura que incluso se  $\phi_{1,t} > 1$  para certo rango de cuantís (é dicir, se ten un comportamento do tipo raíz unitaria), mentres que se verifique que  $\mathbb{E}(\phi_{1,t}^2) < 1$  o proceso a longo prazo é estacionario. Noutras palabras, o modelo QAR permite que o proceso teña comportamentos explosivos transitorios mantendo a estacionaridade a grande escala.

Como citan Koenker e Xiao (2006), se comparamos o proceso  $QAR(1)$ , con variable  $t$ -ésima denotada por  $Y_t$ , co  $AR(1)$  con coeficiente autorregresivo  $\mu_1 = \mathbb{E}(\phi_{1,t})$  e variable  $t$ -ésima que denotamos por  $Y_{-t}^1$ , entón a varianza a longo prazo do proceso  $QAR(1)$  é maior que a do proceso  $AR(1)$ , como se podía agardar. A varianza adicional do QAR,  $\omega_Y^2$ , vén dada pola variación de  $\phi_{1,t}$  co cuantil. De feito, tense que a variabilidade do  $QAR(1)$  pode ser descomposta en dúas partes:

- a) A varianza a longo prazo do proceso  $AR(1)$ , que denotamos por  $\omega_{-Y}^2$ .
- b) A variabilidade determinada pola varianza de  $\phi_{1,t}$ .

Entón,

$$\omega_Y^2 = \omega_{-Y}^2 + \frac{\sigma^2}{(1 - \mu_1)^2 (1 - \mathbb{E}(\phi_{1,t}^2))} \text{Var}(\phi_{1,t}), \tag{3.8}$$

onde  $\omega_{-Y}^2 = \sigma^2 / (1 - \mu_1)^2$  é a varianza a longo prazo de  $Y_{-t}$ .

### 3.3. Estimación do modelo QAR

O mecanismo para a estimación dos parámetros que propoñen Koenker e Xiao (2006) coincide co que xa describimos nos Capítulos 1 e 2: para cada cuantil fixado  $\tau$  (usando a notación de (3.2)), resolveríase o .....

<sup>1</sup>Aquí o subíndice temporal co menos diante é simplemente unha notación para distinguir o proceso  $AR(1)$  do proceso  $QAR(1)$ , así como as respectivas características de ámbolos dous.

problema de minimización da perda cuantílica

$$\hat{\phi} = \arg \min_{\phi \in \mathbb{R}^{p+1}} \sum_{t=p+1}^n \rho_{\tau}(Y_t - \phi^{\top} \mathbf{Z}_{t,p}^1), \quad (3.9)$$

onde  $\mathbf{Z}_{t,p}^1 = (1, Y_{t-1}, \dots, Y_{t-p})^{\top}$ ,  $\phi = (\phi_0, \phi_1, \dots, \phi_p)^{\top}$  e  $\hat{\phi} = (\hat{\phi}_0, \hat{\phi}_1, \dots, \hat{\phi}_p)^{\top}$ . Na práctica, non é posible realizar a estimación para cada  $\tau$  en  $(0, 1)$  como sería preciso para obter unha estimación das funcións  $\phi_j(\tau)$  con  $j = 1, \dots, p$ , que denotaremos por  $\hat{\phi}_j(\tau)$ . Por iso, unha posible solución sería estimar nun certo número de cuantís equiespaciados, aínda que a elección óptima dos cuantís é un problema aínda aberto.

Para poder enunciar as propiedades asintóticas destes estimadores comecemos por definir as seguintes cantidades, análogas ás das asuncións **A1** e **A2** previas ao **Teorema 1.5**:

1.  $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{Z}_{i,p}^1 \mathbf{Z}_{i,p}^{1\top} = \vec{\Omega}_0$ .
2.  $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n f_{t-1}(F_{t-1}^{-1}(\tau)) \mathbf{Z}_{i,p}^1 \mathbf{Z}_{i,p}^{1\top} = \vec{\Omega}_1(\tau)$ .

Ademais, denotemos por  $\vec{\Sigma} = \vec{\Omega}_1^{-1}(\tau) \vec{\Omega}_0 \vec{\Omega}_1(\tau)^{-1}$ .

**Teorema 3.4.** *Baixo as asuncións B1, B2 e B3, tense que*

$$\vec{\Sigma}^{-1/2} \sqrt{n} (\hat{\phi}(\tau) - \phi(\tau)) \xrightarrow{d} \mathbf{B}_{p+1}(\tau), \quad (3.10)$$

onde  $\mathbf{B}_{p+1}(\tau)$  é unha ponte browniana  $(p+1)$ -dimensional.

Pola definición de movemento browniano (que pode ser consultada, por exemplo, en Billingsley 1995), para cada  $\tau$  fixo, tense que  $\vec{\Sigma}^{-1/2} \sqrt{n} (\hat{\phi} - \phi)$  converxe en distribución a unha normal  $N(\mathbf{0}, \tau(1-\tau)\mathbf{I}_{p+1})$ .

O **Teorema 3.4** ten unha consecuencia que enlaza co caso de coeficientes constantes, é dicir, cando o proceso é un  $QAR(p)$  homocedástico como o estudado no Capítulo 2. No caso de coeficientes autorregresivos constantes en todo o rango de cuantís  $\vec{\Omega}_1 = f(F^{-1}(\tau)) \vec{\Omega}_0$ , onde  $f(\cdot)$  e  $F(\cdot)$  son as funcións de densidade e de distribución de  $u_t$ . O corolario que imos enunciar agora dános a distribución asintótica dos estimadores dos coeficientes do proceso obtidos mediante a minimización da perda cuantílica no contexto da autorregresión clásica, en particular, dos de mínimas desviacións absolutas (MDA), os cales foron vistos no Capítulo 2 (lembremos que naquel contexto, só a constante do modelo dependía do cuantil para o que estabamos a estimar).

**Corolario 3.5.** *Baixo as asuncións B1, B2 e B3, tense que*

$$f(F^{-1}(\tau)) \vec{\Omega}_0^{1/2} \sqrt{n} (\hat{\phi}(\tau) - \phi(\tau)) \xrightarrow{d} \mathbf{B}_{p+1}(\tau). \quad (3.11)$$

Outro punto importante a destacar é que a función cuantil condicional debe ser monótona en  $\tau$ , como xa comentáramos, e isto impón tamén certas restricións sobre as funcións  $\phi_j(\tau)$  con  $j = 1, \dots, p$ . O problema do cruce de cuantís tamén aparecía na regresión ordinaria, pero aquí, onde o deseño experimental é necesariamente aleatorio, podería ser algo máis crítico nas aplicacións prácticas.

Koenker e Xiao (2006) apuntan que os cruces de cuantís van existir salvo no caso no que as funcións cuantís sexan estritamente paralelas, o cal sería volver ao modelo de localización–escala típico. Non obstante, o modelo proposto segue sendo interesante dado que xeralmente o que interesa é poder garantir que estas funcións proporcionan unha aproximación local válida sobre unha zona de interese.

### 3.4. Identificación da orde do modelo QAR

Na práctica, dado un conxunto de datos con estrutura de serie temporal, o primeiro paso é identificar un modelo tentativo para os mesmos. No caso dos modelos QAR, a idoneidade da estrutura autorregresiva pode verse empregando a función de autocorrelacións simples como no modelo AR (vendo se as autocorrelacións decaen a 0 exponencialmente) <sup>2</sup>.

Non obstante, a proposta da orde do modelo QAR de Koenker e Xiao (2006) require unha extensión da función de autocorrelacións parciais ao terreo cuantil. Na presente sección presentaremos unha serie de conceptos e resultados que nos permitirán levar a cabo esta extensión e, así, poder identificar na práctica a orde do modelo QAR.

#### 3.4.1. Correlación simple e parcial en versión cuantil

Para estender a correlación ao ámbito da regresión cuantil, imos comezar presentando o concepto de covarianza cuantil dado por Li et al. (2015). Para iso, usaremos o seguinte resultado:

**Lema 3.6.** *Sexan  $X$  e  $Y$  dúas variables aleatorias e sexa  $\tau$  tal que  $0 < \tau < 1$ . Se denotamos por  $\mathcal{Q}_\tau(Y)$  a función cuantil  $\tau$  incondicionada para  $Y$ , e por  $\mathcal{Q}_\tau(Y|X)$  a función cuantil  $\tau$  de  $Y$  condicionada a  $X$ , cúmprese que  $\mathcal{Q}_\tau(Y|X)$  é independente de  $X$ , é dicir,  $\mathcal{Q}_\tau(Y|X) = \mathcal{Q}_\tau(Y)$  con probabilidade 1 se, e só se, as variables aleatorias  $I(Y - \mathcal{Q}_\tau(Y) > 0)$  e  $X$  son independentes, onde  $I(\cdot)$  denota a función indicador.*

Este resultado permítenos establecer a seguinte definición.

**Definición 3.7.** Dadas dúas variables aleatorias  $X$  e  $Y$  definidas sobre o mesmo espazo de probabilidade, tales que  $\mathbb{E}(X^2) < \infty$  e dado  $0 < \tau < 1$ , a *covarianza cuantil* de  $Y$  e  $X$ , que denotaremos mediante  $\text{qcov}_\tau(Y, X)$ , defínese como

$$\begin{aligned} \text{qcov}_\tau(Y, X) &= \text{Cov}[I(Y - \mathcal{Q}_\tau(Y) > 0), X] \\ &= \mathbb{E}[[I(Y - \mathcal{Q}_\tau(Y) > 0) - \mathbb{E}[I(Y - \mathcal{Q}_\tau(Y) > 0)]](X - \mathbb{E}(X))]. \end{aligned}$$

Nótese que a covarianza cuantil non posúe a propiedade de simetría da covarianza clásica, é dicir,  $\text{qcov}_\tau(Y, X) \neq \text{qcov}_\tau(X, Y)$ . Isto é debido a que o primeiro argumento da covarianza fai referencia ao cuantil  $\tau$ , mentres que o segundo é o mesmo que o da covarianza usual.

A expresión da covarianza cuantil dada na **Definición 3.7** pode ser simplificada se temos en conta que

$$\mathbb{E}[I(Y - \mathcal{Q}_\tau(Y) > 0)] = \mathbb{P}(Y - \mathcal{Q}_\tau(Y) > 0) = \mathbb{P}(Y > \mathcal{Q}_\tau(Y)) = 1 - \mathbb{P}(Y \leq \mathcal{Q}_\tau(Y)) = 1 - \tau.$$

Isto permitiríanos escribir, finalmente, que

$$\begin{aligned} \text{qcov}_\tau(Y, X) &= \mathbb{E}[[I(Y - \mathcal{Q}_\tau(Y) > 0) - (1 - \tau)](X - \mathbb{E}(X))] \\ &= \mathbb{E}[[\tau - I(Y - \mathcal{Q}_\tau(Y) < 0)](X - \mathbb{E}(X))] \\ &= \mathbb{E}[\psi_\tau(Y - \mathcal{Q}_\tau(Y))(X - \mathbb{E}(X))], \end{aligned} \tag{3.12}$$

onde  $\psi_\tau(w) = \tau - I(w < 0)$ .

A partir da covarianza cuantil podemos definir a *correlación cuantil* como extensión do coeficiente de correlación de Pearson ao plano cuantil. Esta será denotada por  $\text{qcor}_\tau(Y, X)$ , e virá dada pola expresión .....

<sup>2</sup>Keith Knight fai un comentario ao artigo de Koenker e Xiao (2006) probando que a estrutura de autocovarianzas do proceso  $QAR(p)$  é a mesma que a dun proceso  $AR(p)$ .

seguinte:

$$\text{qcor}_\tau(Y, X) = \frac{\text{qcov}_\tau(Y, X)}{\sqrt{\text{Var}[\psi_\tau(Y - \mathcal{Q}_\tau(Y))] \text{Var}(X)}}. \quad (3.13)$$

Como ocorre co coeficiente de correlación usual  $-1 \leq \text{qcor}_\tau(Y, X) \leq 1$ . Notemos, ademais, que dado que  $\text{qcov}_\tau(Y, X) \neq \text{qcov}_\tau(X, Y)$ , terase que  $\text{qcor}_\tau(Y, X) \neq \text{qcor}_\tau(X, Y)$ .

Facendo operacións pode obterse unha simplificación da expresión da correlación cuantil dada por (3.13). Para iso debemos ter en conta que

$$\text{Var}[\psi_\tau(Y - \mathcal{Q}_\tau(Y))] = \text{Var}[\tau - I(Y - \mathcal{Q}_\tau(Y) < 0)] = \text{Var}[I(Y - \mathcal{Q}_\tau(Y) < 0)] = \tau(1 - \tau) = \tau - \tau^2. \quad (3.14)$$

Entón, substituíndo en (3.13) as expresións (3.12) e (3.14) e se denotamos por  $\sigma_X^2 = \text{Var}(X)$  chegamos á seguinte expresión simplificada para a correlación cuantil:

$$\text{qcor}_\tau(Y, X) = \frac{\mathbb{E}[\psi_\tau(Y - \mathcal{Q}_\tau(Y))(X - \mathbb{E}(X))]}{\sqrt{(\tau - \tau^2)\sigma_X^2}}. \quad (3.15)$$

Agora, a correlación parcial, que foi estudada no contexto clásico na Sección 2.1 do Capítulo 2, pode estenderse ao terreo cuantil como detallamos a continuación. Sexa:

$$(\alpha_1, \beta_1^\top) = \arg \min_{\alpha, \beta} \mathbb{E}(X - \alpha - \beta^\top \mathbf{Z})^2$$

onde  $(\alpha, \beta^\top)^\top$  é o vector de parámetros descoñecidos. Deste xeito,  $\alpha_1 + \beta_1^\top \mathbf{Z}$  é o efecto do vector de variables  $\mathbf{Z}$ , de dimensión  $p$ , sobre a variable  $X$ .

Para agora extraer o efecto de  $\mathbf{Z}$  sobre o cuantil  $\tau$  de  $Y$ , consideremos

$$(\alpha_2, \beta_2^\top) = \arg \min_{\alpha, \beta} \mathbb{E}(\rho_\tau(Y - \alpha - \beta^\top \mathbf{Z})).$$

En consecuencia,  $\alpha_2 + \beta_2^\top \mathbf{Z}$  é o efecto lineal de  $\mathbf{Z}$  sobre o cuantil  $\tau$  da variable  $Y$ . Ademais,  $\alpha_1, \beta_1, \alpha_2$  e  $\beta_2$  son únicos se o vector aleatorio  $(Y, X, \mathbf{Z}^\top)^\top$  ten densidade conxunta con  $\mathbb{E}(X^2) < \infty$  e  $\mathbb{E}\|\mathbf{Z}\|^2 < \infty$ .

A partir destes feitos podemos definir a correlación parcial cuantil como:

$$\begin{aligned} \text{qqcor}_\tau(Y, X|\mathbf{Z}) &= \frac{\text{Cov}(\psi_\tau(Y - \alpha_2 - \beta_2^\top \mathbf{Z}), X - \alpha_1 - \beta_1^\top \mathbf{Z})}{\sqrt{\text{Var}(\psi_\tau(Y - \alpha_2 - \beta_2^\top \mathbf{Z})) \text{Var}(X - \alpha_1 - \beta_1^\top \mathbf{Z})}} \\ &= \frac{\mathbb{E}[[\psi_\tau(Y - \alpha_2 - \beta_2^\top \mathbf{Z}) - \mathbb{E}(\psi_\tau(Y - \alpha_2 - \beta_2^\top \mathbf{Z}))](X - \alpha_1 - \beta_1^\top \mathbf{Z} - \mathbb{E}(X - \alpha_1 - \beta_1^\top \mathbf{Z}))]}{\sqrt{\text{Var}(\psi_\tau(Y - \alpha_2 - \beta_2^\top \mathbf{Z})) \text{Var}(X - \alpha_1 - \beta_1^\top \mathbf{Z})}}. \end{aligned} \quad (3.16)$$

Non obstante, debemos ter en conta que  $\mathbb{E}(X - \alpha_1 - \beta_1^\top \mathbf{Z}) = 0$  e  $\mathbb{E}(\psi_\tau(Y - \alpha_2 - \beta_2^\top \mathbf{Z})) = 0$  por construción dos modelos. Por outra parte,  $\mathbb{E}(\mathbf{Z}\psi_\tau(Y - \alpha_2 - \beta_2^\top \mathbf{Z})) = \mathbf{0}$  pola independencia de  $\mathbf{Z}$  respecto do cuantil  $\tau$  de  $Y$  unha vez foi extraído o efecto deste conxunto de variables  $\mathbf{Z}$  do mesmo, onde  $\mathbf{0}$  é un vector de ceros de dimensión  $p$ . Isto permítenos obter a simplificación da expresión (3.16) seguinte:

$$\text{qqcor}_\tau(Y, X|\mathbf{Z}) = \frac{\mathbb{E}(\psi_\tau(Y - \alpha_2 - \beta_2^\top \mathbf{Z})(X - \alpha_1 - \beta_1^\top \mathbf{Z}))}{\sqrt{(\tau - \tau^2)\mathbb{E}(X - \alpha_1 - \beta_1^\top \mathbf{Z})^2}}.$$

Agora, realizando o produto e aplicando novamente a independencia de  $\mathbf{Z}$  respecto de  $\psi_\tau(Y - \alpha_2 - \beta_2^\top \mathbf{Z})$  chegamos a que:

$$\text{qpcor}_\tau(Y, X|\mathbf{Z}) = \frac{\mathbb{E}(\psi_\tau(Y - \alpha_2 - \beta_2^\top \mathbf{Z}) X)}{\sqrt{(\tau - \tau^2)\sigma_{X|\mathbf{Z}}^2}}, \quad (3.17)$$

onde  $\sigma_{X|\mathbf{Z}}^2 = \mathbb{E}(X - \alpha_1 - \beta_1^\top \mathbf{Z})^2$ .

### 3.4.2. Correlación simple e parcial cuantil en versión mostral

Para poder empregar na práctica as correlacións cuantís é preciso dispoñer dun modo de estimalas dada unha mostra. Non obstante, a soa estimación é pouco informativa e, a maiores, resulta preciso coñecer o comportamento asintótico das mesmas, o que permitirá realizar tarefas de inferencia. Para iso, na presente subsección imos encargarnos da estimación destas correlacións e da obtención da súa distribución asintótica seguindo, novamente, a formulación de Li et al. (2015).

Supoñamos que temos unha mostra de vectores aleatorios  $\{(Y_i, X_i, \mathbf{Z}_i^\top)^\top, i = 1, \dots, n\}$  que son independentes e identicamente distribuídos, xerados da distribución de  $(Y, X, \mathbf{Z}^\top)^\top$ . Podemos definir a correlación cuantil mostral entre as variables  $X$  e  $Y$  a partir da expresión (3.15) do xeito seguinte:

$$\widehat{\text{qcor}}_\tau(Y, X) = \frac{1}{\sqrt{(\tau - \tau^2)\hat{\sigma}_X^2}} \frac{1}{n} \sum_{i=1}^n \psi_\tau(Y_i - \hat{Q}_\tau(Y)) (X_i - \bar{X}),$$

onde  $\bar{X} = n^{-1} \sum_{i=1}^n X_i$  e  $\hat{\sigma}_X^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$  e  $\hat{Q}_\tau(Y) = \inf\{y : F_n(y) \geq \tau\}$  é o cuantil mostral de  $Y$ .

A distribución asintótica da correlación cuantil mostral queda resumida no seguinte teorema.

**Teorema 3.8.** *Denotemos por  $f_Y(\cdot)$  e por  $f_{Y|X}(\cdot)$  a densidade de  $Y$  e a densidade de  $Y$  condicionada a  $X$ , respectivamente. Supoñamos que  $\mathbb{E}(X^4) < \infty$  e que existe un  $\delta > 0$  tal que a densidade condicional  $f_{Y|X}(\cdot)$  é uniformemente integrable en  $[Q_\tau(Y) + \delta, Q_\tau(Y) - \delta]$ . Ademais, asumamos que a densidade  $f_Y(\cdot)$  é continua e positiva. Entón*

$$\sqrt{n}(\widehat{\text{qcor}}_\tau(Y, X) - \text{qcor}_\tau(Y, X)) \xrightarrow{d} N(0, \Omega_1),$$

onde a expresión de  $\Omega_1$  se obtén a partir das seguintes:

$$\begin{aligned} \mu_X &= \mathbb{E}(X), \\ \sigma_X^2 &= \text{Var}(X), \\ \mu_{X|Y} &= \frac{\mathbb{E}[f_{Y|X}(Q_\tau(Y)) X]}{f_Y(Q_\tau(Y))}, \\ \Sigma_{11} &= \mathbb{E}(X - \mu_X)^4 - \sigma_X^4, \\ \Sigma_{12} &= \mathbb{E}[\psi_\tau(Y - Q_\tau(Y)) (X - \mu_{X|Y})]^2 - [\text{qcov}_\tau(Y, X)]^2, \\ \Sigma_{13} &= \mathbb{E}[\psi_\tau(Y - Q_\tau(Y)) (X - \mu_{X|Y}) (X - \mu_X)^2] - \sigma_X^2 \cdot \text{qcov}_\tau(Y, X). \end{aligned}$$

Finalmente,

$$\Omega_1 = \frac{1}{\tau - \tau^2} \left[ \frac{\Sigma_{11} (\text{qcov}_\tau(Y, X))^2}{4\sigma_X^6} - \frac{\Sigma_{13} \cdot \text{qcov}_\tau(Y, X)}{\sigma_X^4} + \frac{\Sigma_{12}}{\sigma_X^2} \right].$$

*Demostración.* A proba deste teorema pode ser consultada no Apéndice de probas (páxs. 259–260) de Li et al. (2015).  $\square$



O **Teorema 3.8** permítenos coñecer a distribución asintótica da correlación cuantil, pero para a súa aplicación é preciso estimar a varianza asintótica  $\Omega_1$ . Isto involucra a estimación de  $\mu_{X|Y}$ . Para isto, dado que pode probarse que  $\mu_{X|Y} = \mathbb{E}(X|Y = \mathcal{Q}_\tau(Y))$ , Li et al. (2015) recorren a técnicas non paramétricas de estimación da función de regresión. A súa proposta baséase en estimar  $m(y) = \mathbb{E}(X|Y = y)$  empregando o estimador de tipo Nadaraya–Watson. Dita estimación denotárase mediante  $\hat{m}(y)$ . A partir disto, o valor buscado pode estimarse mediante  $\hat{\mu}_{X|Y} = \hat{m}(\hat{\mathcal{Q}}_\tau(Y))$ , onde  $\hat{\mathcal{Q}}_\tau(Y)$  é o cuantil  $\tau$  mostral de  $Y$ .

As demais cantidades,  $\mu_X$ ,  $\sigma_X^2$ ,  $\text{qcov}_\tau(X, Y)$ ,  $\Sigma_{11}$ ,  $\Sigma_{12}$  e  $\Sigma_{13}$  poden ser estimadas mediante os seus análogos mostrais, respectivamente:

$$\begin{aligned}\hat{\mu}_X &= \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \\ \hat{\sigma}_X^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu}_X)^2, \\ \widehat{\text{qcov}}_\tau(Y, X) &= \frac{1}{n} \sum_{i=1}^n \psi_\tau(Y_i - \hat{\mathcal{Q}}_\tau(Y)) (X_i - \hat{\mu}_X), \\ \hat{\Sigma}_{11} &= \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu}_X)^4 - \hat{\sigma}_X^4, \\ \hat{\Sigma}_{12} &= \frac{1}{n} \sum_{i=1}^n \left[ \psi_\tau(Y_i - \hat{\mathcal{Q}}_\tau(Y)) (X_i - \hat{\mu}_{X|Y}) \right]^2 - (\widehat{\text{qcov}}_\tau(Y, X))^2, \\ \hat{\Sigma}_{13} &= \frac{1}{n} \sum_{i=1}^n \left[ \psi_\tau(Y_i - \hat{\mathcal{Q}}_\tau(Y)) (X_i - \hat{\mu}_{X|Y}) (X_i - \hat{\mu}_X)^2 \right] - \hat{\sigma}_X^2 \cdot \widehat{\text{qcov}}_\tau(Y, X).\end{aligned}$$

Isto permitíranos obter unha estimación de  $\Omega_1$  que denotaremos mediante  $\hat{\Omega}_1$ .

De xeito análogo a como se fixo para a correlación cuantil, é posible obter unha estimación da correlación parcial cuantil e tamén se pode obter a distribución asintótica para a mesma. Dado que esta última non vai ser empregada de xeito directo, simplemente indicamos que para ela temos un resultado análogo ao **Teorema 3.8**, que pode consultarse en Li et al. (2015).

En vista da expresión (3.17), para obter a expresión da correlación parcial cuantil mostral sexan

$$\begin{aligned}(\hat{\alpha}_1, \hat{\beta}_1^\top) &= \arg \min_{\alpha, \beta} \frac{1}{n} \sum_{i=1}^n (X_i - \alpha - \beta^\top \mathbf{Z}_i)^2, \\ (\hat{\alpha}_2, \hat{\beta}_2^\top) &= \arg \min_{\alpha, \beta} \frac{1}{n} \sum_{i=1}^n \rho_\tau(Y_i - \alpha - \beta^\top \mathbf{Z}_i).\end{aligned}\tag{3.18}$$

Entón a correlación parcial cuantil mostral viría dada por:

$$\widehat{\text{cor}}_\tau(Y, X|\mathbf{Z}) = \frac{1}{\sqrt{(\tau - \tau^2)\hat{\sigma}_{X|\mathbf{Z}}^2}} \frac{1}{n} \sum_{i=1}^n \psi_\tau(Y_i - \hat{\alpha}_2 - \hat{\beta}_2^\top \mathbf{Z}_i) X_i,\tag{3.19}$$

onde  $\hat{\sigma}_{X|\mathbf{Z}}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\alpha}_1 - \hat{\beta}_1^\top \mathbf{Z}_i)^2$ .

### 3.4.3. Autocorrelacións parciais cuantís e modelo QAR

Nas subseccións previas definíronse os conceptos de correlación cuantil e correlación parcial cuantil, o que nos servirá de base para estender ao ámbito cuantil a clásica función de autocorrelacións parciais das series temporais seguindo como referencia o artigo de Li et al. (2015). O obxectivo desta subsección será concretar o

valor de  $p$  no modelo QAR de Koenker e Xiao (2006) dado pola expresión (3.2), é dicir, determinar o número de instantes anteriores dos que depende o cuantil  $\tau$  da variable correspondente ao instante  $t$ -ésimo  $Y_t$ .

Nos modelos autorregresivos en media, as autocorrelacións simples e parciais xogaban un papel fundamental na identificación do modelo. Como xa se viu, para eles as autocorrelacións simples han de tender a cero exponencialmente, mentres que a última autocorrelación parcial non nula indica a orde autorregresiva. Nos modelos cuantís autorregresivos as autocorrelacións cuantís xogarán un papel totalmente análogo, como se verá ao longo desta subsección.

A partir da expresión da correlación parcial cuantil (3.17) e tendo en conta a expresión da función de autocorrelacións parciais dos dun proceso estocástico estacionario dada na **Definición 2.4**, imos obter a expresión da función de autocorrelacións parciais cuantís. Para iso, comecemos fixando a notación.

Sexa  $k$  un enteiro positivo, denotamos por  $\mathbf{Z}_{t,k-1} = (Y_{t-1}, \dots, Y_{t-k+1})^\top$ , e

$$\begin{aligned} (\alpha_1, \boldsymbol{\beta}_1^\top) &= \arg \min_{\alpha, \boldsymbol{\beta}} \mathbb{E} (Y_{t-k} - \alpha - \boldsymbol{\beta}^\top \mathbf{Z}_{t,k-1})^2, \\ (\alpha_{2,\tau}, \boldsymbol{\beta}_{2,\tau}^\top) &= \arg \min_{\alpha, \boldsymbol{\beta}} \mathbb{E} (\rho_\tau (Y_t - \alpha - \boldsymbol{\beta}^\top \mathbf{Z}_{t,k-1})). \end{aligned} \quad (3.20)$$

A correlación parcial cuantil entre  $Y_t$  e  $Y_{t-k}$ , obtérase extraendo de ambas variables o efecto das variables aleatorias en instantes intermedios recollidas en  $\mathbf{Z}_{t,k-1}$ , é dicir,

$$\phi_{kk,\tau} = \text{qpcor}_\tau (Y_t, Y_{t-k} | \mathbf{Z}_{t,k-1}) = \frac{\mathbb{E} (\psi_\tau (Y_t - \alpha_{2,\tau} - \boldsymbol{\beta}_{2,\tau}^\top \mathbf{Z}_{t,k-1}) Y_{t-k})}{\sqrt{(\tau - \tau^2) \mathbb{E} (Y_{t-k} - \alpha_1 - \boldsymbol{\beta}_1^\top \mathbf{Z}_{t,k-1})^2}}, \quad k \geq 2. \quad (3.21)$$

Debemos de notar que para  $k = 1$  tense que  $\phi_{11,\tau} = \text{qcor}_\tau (Y_t, Y_{t-1})$ . Nese caso, empregárase o desenvolvemento teórico visto na subsección anterior.

Debido á hipótese de estacionaridade de  $\{Y_t\}_{t \in \mathbb{Z}}$ ,  $\phi_{kk,\tau}$  é independente do índice  $t$ , polo que esta función variará unicamente co cuantil  $\tau$  e co retardo  $k$ . Seguindo a notación usual de series temporais, denotáremola por QPACF (siglas en inglés de *quantile partial autocorrelation function*).

A QPACF cumpre a mesma función que a PACF nos modelos autorregresivos clásicos no tocante á identificación da orde do modelo autorregresivo. Esta utilidade resúmese lema seguinte.

**Lema 3.9.** *Supoñamos que  $Y_t$  ten densidade condicional na  $\sigma$ -álgebra  $\mathcal{F}_{t-1}$  e que  $\mathbb{E} (Y_t^2) < \infty$ . Se o proceso pode ser descrito mediante un modelo QAR dado pola expresión (3.2) se  $\phi_p(\tau)$  é o último coeficiente que sabemos non nulo do proceso, cando  $\phi_p(\tau) \neq 0$  con  $p > 0$ , daquela  $\phi_{pp,\tau} \neq 0$  e  $\phi_{kk,\tau} = 0$  para  $k > p$ .*

Isto vén a dicir que o retardo co que se corresponde a última autocorrelación parcial cuantil non nula dáanos a orde do proceso QAR, é dicir, determina o último coeficiente autorregresivo non nulo do proceso.

Para poder aplicar na práctica o **Lema 3.9**, é preciso obter unha estimación de  $\phi_{kk,\tau}$  e coñecer a distribución asintótica da mesma. Para iso, necesitamos estimar os parámetros dos modelos presentados en (3.20). Sexan

$$\begin{aligned} (\hat{\alpha}_1, \hat{\boldsymbol{\beta}}_1^\top) &= \arg \min_{\alpha, \boldsymbol{\beta}} \sum_{t=k+1}^n (Y_{t-k} - \alpha - \boldsymbol{\beta}^\top \mathbf{Z}_{t,k-1})^2, \\ (\hat{\alpha}_{2,\tau}, \hat{\boldsymbol{\beta}}_{2,\tau}^\top) &= \arg \min_{\alpha, \boldsymbol{\beta}} \sum_{t=k+1}^n \rho_\tau (Y_t - \alpha - \boldsymbol{\beta}^\top \mathbf{Z}_{t,k-1}). \end{aligned}$$

e  $\hat{\sigma}_{Y|\mathbf{Z}}^2 = n^{-1} \sum_{t=k+1}^n (Y_{t-k} - \hat{\alpha}_1 - \hat{\boldsymbol{\beta}}_1^\top \mathbf{Z}_{t,k-1})^2$ . Así, a estimación de  $\phi_{kk,\tau}$ , tendo en conta a expresión

(3.19), viría dada por:

$$\hat{\phi}_{kk,\tau} = \frac{1}{\sqrt{(\tau - \tau^2) \hat{\sigma}_{Y|\mathbf{Z}}^2}} \frac{1}{n} \sum_{t=k+1}^n \psi_{\tau} \left( Y_t - \hat{\alpha}_{2,\tau} - \hat{\boldsymbol{\beta}}_{2,\tau}^{\top} \mathbf{Z}_{t,k-1} \right) Y_{t-k}, \quad (3.22)$$

e referirémonos a ela mediante QPACF mostral. Como vemos, na expresión de  $\hat{\phi}_{kk,\tau}$  aparece no denominador  $n$  en lugar de  $n - k$ . Isto novamente é debido ao comentado na Subsección 2.1 (o erro cadrático medio do estimador é menor con  $n$  no denominador e, ademais, mantense carácter semidefinido positivo que ten a función de autocorrelacións parciais cuantís teórica do proceso).

O erro asociado ao modelo para o cuantil condicional  $\tau$  sería o dado por:

$$e_{t,\tau} = Y_t - \phi_0(\tau) - \phi_1(\tau)Y_{t-1} - \dots - \phi_p(\tau)Y_{t-p}.$$

A variable aleatoria  $I(e_{t,\tau} > 0)$  é independente de  $Y_{t-k}$  para calquera  $k > 0$  (dado que  $e_{t,\tau}$  é o resultado de extraer de  $Y_t$ , precisamente, o efecto das variables  $Y_{t-k}$ , con  $k > 0$ , das que depende) e  $(\alpha_{2,\tau}, \boldsymbol{\beta}_{2,\tau}^{\top}) = (\phi_0(\tau), \phi_1(\tau), \dots, \phi_p(\tau), 0, \dots, 0)$  para  $k > p$ .

O estudo da distribución asintótica da QPACF mostral require formular unha condición similar á hipótese **B3** introducida por Koenker e Xiao (2006). É a seguinte:

**Condición 3.10.** *Imos supoñer que  $\mathbb{E}(Y_t^2) < \infty$ ,  $\mathbb{E}(Y_t - \mathbb{E}(Y_t|\mathcal{F}_{t-1}))^2 > 0$ , e que  $f_{t-1}(\cdot)$  é uniformemente integrable en  $U = \{u : 0 < F(u) < 1\}$ , sendo  $F(\cdot)$  a densidade marxinal de  $e_{t,\tau}$  e  $f_{t-1}(\cdot)$ , a densidade condicionada de  $e_{t,\tau}$  na  $\sigma$ -álgebra  $\mathcal{F}_{t-1}$ .*

Esta condición permitíranos resumir o comportamento asintótico da QPACF nun teorema.

**Teorema 3.11.** *Sexa  $\mathbf{Z}_{t,k-1}^1 = \left(1, \mathbf{Z}_{t,k-1}^{\top}\right)^{\top} = (1, Y_{t-1}, \dots, Y_{t-k+1})^{\top}$ . Dados  $\tau_1, \tau_2 \in \mathcal{I}$ , onde  $\mathcal{I} \subset (0, 1)$  é un intervalo pechado definimos:*

$$\begin{aligned} A_0 &= \mathbb{E} \left[ Y_{t-k} \mathbf{Z}_{t,k-1}^1 \right], \\ A_1(\tau) &= \mathbb{E} \left[ f_{t-1}(0) Y_{t-k} \mathbf{Z}_{t,k-1}^1 \right], \\ \vec{\Sigma}_{30} &= \mathbb{E} \left[ \mathbf{Z}_{t,k-1}^1 \mathbf{Z}_{t,k-1}^{1\top} \right], \\ \vec{\Sigma}_{31}(\tau) &= \mathbb{E} \left[ f_{t-1}(0) \mathbf{Z}_{t,k-1}^1 \mathbf{Z}_{t,k-1}^{1\top} \right], \\ \Sigma_{32}(\tau_1, \tau_2) &= \mathbb{E} \left( Y_t^2 \right) - A_1^{\top}(\tau_1) \vec{\Sigma}_{31}^{-1}(\tau_1) A_0 - A_1^{\top}(\tau_2) \vec{\Sigma}_{31}^{-1}(\tau_2) A_0 + A_1^{\top}(\tau_1) \vec{\Sigma}_{31}^{-1}(\tau_1) \vec{\Sigma}_{30} \vec{\Sigma}_{31}^{-1}(\tau_2) A_1(\tau_2), \end{aligned}$$

co que, finalmente, obtemos a expresión

$$\Omega_3(\tau_1, \tau_2) = \frac{\mathbb{E} \left[ \psi_{\tau_1}(e_{t,\tau_1}) \psi_{\tau_2}(e_{t,\tau_2}) \right] \Sigma_{32}(\tau_1, \tau_2)}{\sqrt{(\tau_1 - \tau_1^2)(\tau_2 - \tau_2^2)} \mathbb{E} \left( Y_{t-k} - \alpha_1 - \boldsymbol{\beta}_1^{\top} \mathbf{Z}_{t,k-1} \right)^2}. \quad (3.23)$$

Supoñamos que para cada  $\tau \in \mathcal{I}$ ,  $A_1(\tau)$  e  $\vec{\Sigma}_{31}(\tau)$  son finitas e  $\vec{\Sigma}_{31}(\tau)$  é unha matriz definida positiva. Se se satisfai a **Condición 3.10** e  $k > p$ , entón

$$\sqrt{n} \hat{\phi}_{kk,\tau} \xrightarrow{d} B_1(\tau),$$

para cada  $\tau \in \mathcal{I}$ , onde  $B_1(\tau)$  é un proceso gaussiano con media cero e covarianza dada por

$$\Omega_3(\tau_1, \tau_2) = \mathbb{E} [B_1(\tau_1) B_1(\tau_2)],$$

para  $\tau_1, \tau_2 \in \mathcal{I}$ .

Debemos ter en conta que cando os erros  $\{e_{t,\tau}\}_{t \in \mathbb{Z}}$  asociados ao cuantil condicional son i.i.d., a variable aleatoria  $e_{t,\tau} - \mathbb{E}(e_{t,\tau})$  é independente de  $\tau$ , e podemos escribir:  $e_t = e_{t,\tau} - \mathbb{E}(e_{t,\tau})$ . Notemos que  $e_{t,\tau} = e_t - \mathcal{Q}_\tau(e_{t,\tau})$  con  $\mathbb{E}(e_t) = 0$ . Entón  $f_{t-1}(0) = f(\mathcal{Q}_\tau(e_{t,\tau}))$ , onde  $f(\cdot)$  é a función de densidade de  $e_t$  e, así,  $\vec{\Sigma}_{31}(\tau) = \mathbb{E} \left[ f_{t-1}(0) \mathbf{Z}_{t,k-1}^1 \mathbf{Z}_{t,k-1}^{1\top} \right] = f(\mathcal{Q}_{e_{t,\tau}}(\tau)) \mathbb{E} \left[ \mathbf{Z}_{t,k-1}^1 \mathbf{Z}_{t,k-1}^{1\top} \right]$ . Dado que  $\vec{\Sigma}_{31}(\tau)$  é definida positiva, tense que  $f(\mathcal{Q}_\tau(e_{t,\tau})) > 0$  para cada  $\tau \in \mathcal{I}$ , e dado que é unha matriz finita tense que  $f(\mathcal{Q}_\tau(e_{t,\tau})) < \infty$  para cada  $\tau \in \mathcal{I}$ . Así,  $0 < f_{t-1}(0) < \infty$ .

O **Teorema 3.11** indícanos que para un  $\tau \in \mathcal{I}$ , tense que  $\sqrt{n} \hat{\phi}_{kk} \xrightarrow{d} N(0, \Omega_3(\tau, \tau))$ . Polo tanto, a estimación de  $\Omega_3(\tau, \tau)$  permitiríanos ter totalmente determinada a distribución das autocorrelacións parciais cuantís, e así, contrastar a nulidade das mesmas para determinar a orde do autorregresivo (é dicir, o valor de  $p$ ).

Para estimar  $\Omega_3(\tau, \tau)$ , comecemos por determinar un modo de estimar  $f_{t-1}(0)$ . Para isto Hendriks e Koenker (1992) propoñen o seguinte método:

$$\hat{f}_{t-1}(0) = \frac{2h}{\hat{Q}_{\tau+h}(Y_t | \mathcal{F}_{t-1}) - \hat{Q}_{\tau-h}(Y_t | \mathcal{F}_{t-1})}, \quad (3.24)$$

onde  $\hat{Q}_\tau(Y_t | \mathcal{F}_{t-1}) = \hat{\phi}_0(\tau) + \hat{\phi}_1(\tau)Y_{t-1} + \dots + \hat{\phi}_k(\tau)Y_{t-k}$  é a estimación do cuantil condicionado de  $Y_t$ , e  $h$  é o parámetro ventá, seleccionado mediante un método apropiado. Segundo Koenker e Xiao (2006), poderíamos considerar como ventá a proposta por Hall e Sheather (1988) que se basea nos desenvolvementos de Edgeworth (introducindo de xeito plug-in a distribución normal):

$$h_{HS} = n^{-1/3} z_{\alpha/2}^{2/3} \left[ \frac{1.5\phi^2(\Phi^{-1}(\tau))}{2(\Phi^{-1}(\tau))^2 + 1} \right]^{1/3},$$

onde  $z_{\alpha/2}$  é o cuantil  $1 - \alpha/2$  da normal estándar,  $\Phi(\cdot)$  e  $\phi(\cdot)$  son as funcións de distribución e de densidade tamén da normal estándar e  $\tau$  é o cuantil para o que se fai a estimación. Outra opción que propoñen é considerar a ventá de Bofinger (1975) baseada na minimización do erro cadrático medio do estimador da densidade (introducindo de xeito plug-in a distribución normal):

$$h_B = n^{-1/5} \left[ \frac{4.5\phi^4(\Phi^{-1}(\tau))}{(2(\Phi^{-1}(\tau))^2 + 1)^2} \right]^{1/5}. \quad (3.25)$$

Entón, se ademais estimamos  $\alpha_1$  e  $\beta_1$  do modo usual, pódese obter a versión mostral de  $A_0, A_1(\tau), \vec{\Sigma}_{30}, \vec{\Sigma}_{31}(\tau), \mathbb{E}(Y_t^2)$  e de  $\mathbb{E}(Y_{t-k} - \alpha_1 - \beta_1^\top \mathbf{Z}_{t,k-1})^2$ .

Así, para un  $\tau$  fixado, podemos obter a estimación de  $\Omega_3(\tau, \tau)$ , que denotaremos mediante  $\hat{\Omega}_3$ , o que nos permitirá construír intervalos de confianza para  $\phi_{kk,\tau}$  para un nivel de confianza  $1 - \alpha$  fixado. Para cada retardo  $k$  considerado, estes virían dados por:

$$\left( 0 - z_{\alpha/2} \sqrt{\frac{\hat{\Omega}_3}{n}}, 0 + z_{\alpha/2} \sqrt{\frac{\hat{\Omega}_3}{n}} \right),$$

onde  $z_{\alpha/2}$  representa o cuantil  $1 - \alpha/2$  da normal estándar. Así, poderíamos construír bandas de confianza que nos informan da significación das autocorrelacións parciais cuantís para cada retardo  $k$ , sendo posible a identificación na práctica da orde do modelo cuantil autorregresivo.

### 3.5. Exemplo con datos simulados

Finalmente, mostraremos un exemplo práctico con datos simulados para ilustrar o funcionamento da metodoloxía presentada para a identificación da orde do modelo QAR. Para isto, dado que descoñecemos a existencia de librerías de R que conteñan estes métodos implementados, foi preciso o desenvolvemento dun código propio.

Supoñamos que temos un conxunto de  $n = 500$  observacións  $\{y_1, \dots, y_{500}\}$  xeradas dende o modelo:

$$Y_t = \Phi^{-1}(U_t) + a(U_t)Y_{t-1}$$

onde  $\Phi$  denota a función de distribución dunha normal estándar,  $\{U_t\}_{t \in \mathbb{Z}}$  é unha sucesión i.i.d. a unha uniforme no intervalo  $[0, 1]$  e  $a(x) = \max\{0.8 - 1.6x, 0\}$ . En base ao modelo teórico, para cuantís inferiores a 0.5 (este incluído) a orde autorregresiva teórica sería 1, mentres que para cuantís superiores a 0.5 sería 0, é dicir, que temos ruído branco.

O gráfico secuencial da serie amósase na esquerda da Figura 3.2, mentres que na dereita da mesma figura incluímos o gráfico de dispersión de  $Y_t$  respecto de  $Y_{t-1}$ . Como se pode observar neste último, a variabilidade cambia en función do valor de  $Y_{t-1}$ , polo que un análise clásico da serie podería non ser apropiado.

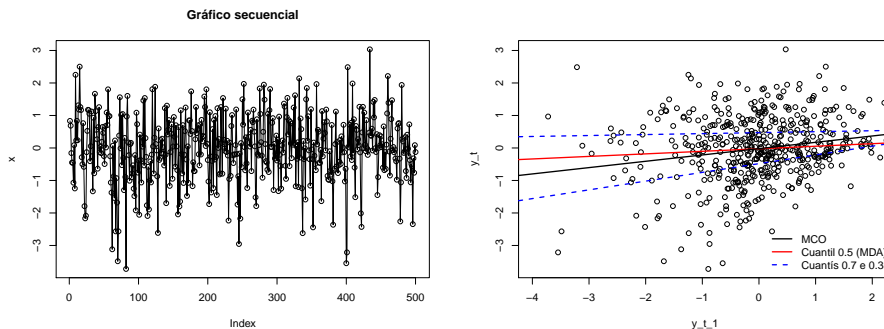


Figura 3.2: Esquerda: gráfico secuencial da serie simulada; dereita: gráfico de dispersión de  $Y_t$  fronte a  $Y_{t-1}$  para a serie simulada.

Como vimos ao longo deste capítulo, podemos realizar un estudo máis completo que o da media condicionada do proceso. Imaxinemos que estamos interesados en coñecer o comportamento do cuantil condicional para  $\tau = 0.1, 0.3, 0.5, 0.7$  e  $0.9$ , é dicir, que desexamos axustar un modelo QAR para cada un deses cuantís. Vexamos se a metodoloxía presentada identifica ben a orde do modelo QAR en cada caso.

A Figura 3.3 amosa a representación da QPACF mostral para cada  $\tau$  considerado. Para a obtención destes gráficos empregouse a función `qpacf(x, tau, max_lag)` incluída no Apéndice B, onde `x` é a serie, `tau` é o cuantil e `max_lag` é o maior retardo para o que se obtén a autocorrelación parcial cuantil. Como se poderá observar, representáronse as autocorrelacións parciais ata retardo  $k = 20$ . As liñas de puntos en azul son bandas de confianza para as autocorrelacións a nivel 0.95. Para obter estas bandas, na estimación de  $f_{t-1}(0)$  empregouse como parámetro de suavizado o proposto por Bofinger (1975), cuxa expresión viña dada en (3.25).

Dado que representamos 20, para analizar estes gráficos debemos ter en conta que, debido ao nivel de confianza fixado, unha delas podería saírse das bandas por azar (pero nese caso, tería un valor moi próximo ao límite do intervalo). Como sempre interesa quedarse con modelos sinxelos (con menos parámetros), en principio, parece que a orde  $p = 1$  é axeitada para  $\tau = 0.1, 0.3$  e  $0.5$ . Non obstante, para  $\tau = 0.7, 0.9$  podería asumirse que  $p = 0$ . Polo cal, mediante a metodoloxía presentada identificaríase ben a orde do modelo simulado, constituíndo esta unha ferramenta útil na práctica.

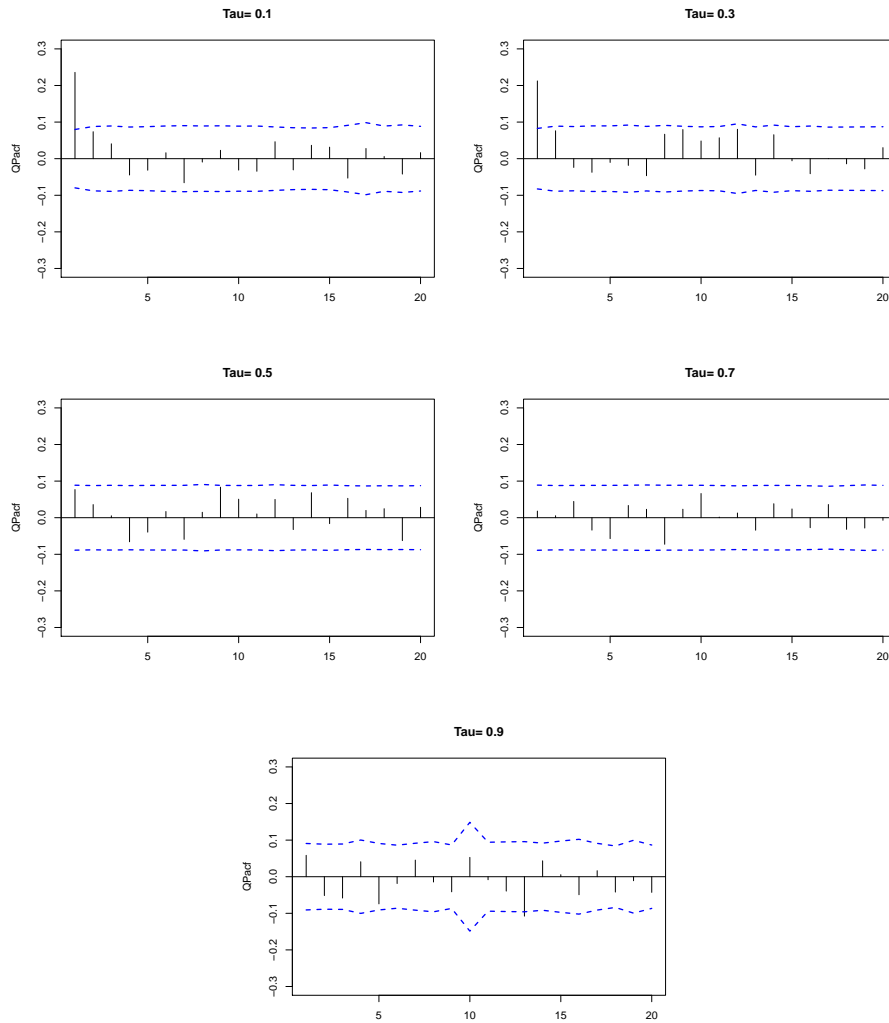


Figura 3.3: QPACF mostral para  $\tau = 0.1, 0.3, 0.5, 0.7$  e  $0.9$ . As liñas de puntos en azul son as bandas de confianza para un nivel do 0.95.

Unha vez identificada a orde do proceso QAR, procederíase á estimación dos seus parámetros. Li et al (2015) tamén abordan o tema da validación do modelo dende a análise dos residuos, aínda que neste traballo non imos tratar este tema. Unha vez proposto un modelo válido para unha serie temporal, este podería ser empregado para a realización de predicións. Precisamente, a obtención de intervalos de predición no caso dos modelos QAR é un problema aínda aberto e descoñecemos a existencia de artigos que o tratasen. Por tal razón, neste traballo imos abordar primeiro a predición no contexto autorregresivo clásico dende a estimación baseada en métodos cuantís, para nun futuro poder adaptar tales procedementos ao terreo dos modelos QAR de Koenker e Xiao (2006).

## Capítulo 4

# Predicción no contexto autorregresivo clásico

Un dos problemas máis importantes da análise de series de tempo e, en particular, dos modelos autorregresivos é o da predicción: dado o tramo de traxectoria observada dun proceso ata o instante  $n$ , queremos obter información do que pode ocorrer nun tempo  $n + k$ . Neste contexto, as predicións puntuais adoitan acompañarse dun intervalo de predicción, un conxunto aleatorio que contén os posibles valores da futura observación  $Y_{n+k}$  cunha elevada probabilidade.

Os métodos clásicos para obtención de intervalos de predicción para modelos  $AR(p)$ , como o proposto por Box–Jenkins (1976), baséanse na asunción dunha distribución específica para o erro, a normal, e poden proporcionar resultados pouco axeitados cando a hipótese de normalidade non se cumpre. Por tal razón, varios autores propuxeron procedementos alternativos baseados no bootstrap para a obtención destes intervalos seguindo, principalmente, dous enfoques distintos.

Por unha parte, Stine (1987) propón un método bootstrap para estimar o erro cadrático medio de predicción do mellor predictor lineal estimado no contexto dos  $AR(p)$ . Non obstante, o seu plan de remostraxe non é apropiado para a predicción, dado que non se centra en replicar a distribución condicional de  $Y_{n+k}$  dados os valores da serie, senón en replicar o mecanismo aleatorio de estimación dos parámetros.

Por outra parte, varios autores céntranse en replicar directamente a distribución de  $Y_{n+k}$  condicionada á serie observada. Thombs e Schucany (1990) propoñen un método bootstrap para a obtención de intervalos de predicción que involucra a representación cara atrás dos modelos  $AR(p)$ . O seu plan de remostraxe incorpora a variabilidade debida ao termo residual, así como a ocasionada pola estimación dos parámetros do modelo. Cao et al. (1997) estudan un método bootstrap alternativo ao de Thombs e Schucany que é computacionalmente máis rápido e tamén consistente. Este procedemento evita a representación cara atrás dos  $AR(p)$ , pero non incorpora a variabilidade debida á estimación dos parámetros do modelo. Por tal razón, Pascual et al. (2001) propoñen unha variante dese método, que tampouco involucra a representación cara atrás e incorpora a variabilidade que ocasiona a estimación dos parámetros da serie. Este método presenta a vantaxe de que pode estenderse a outros modelos máis complexos no ámbito das series temporais.

Neste Capítulo, realizaremos unha breve revisión destes procedementos propostos para a obtención dos intervalos de predicción no contexto dos modelos  $AR(p)$ , analizando o seu comportamento práctico mediante un estudo de simulación, cando se estima o modelo mediante mínimos cadrados. Tamén se estudará o funcionamento dos algoritmos bootstrap propostos se a estimación se realiza mediante mínimas desviacións absolutas (considerando o modelo  $QAR(p)$  de erros i.i.d.), comparando as coberturas medias obtidas en cada caso coas da estimación por mínimos cadrados.

Finalmente, no contexto da estimación da mediana condicional, proporemos un novo método bootstrap para a obtención de intervalos de predicción baseado nas ideas de Rao e Zhao (1992). Ademais de probar a consistencia desta proposta, analizaremos o seu comportamento práctico en comparación cos demais procedementos revisados neste capítulo no estudo de simulación.

A información organizarase en catro seccións. Na primeira revisamos os intervalos de predición clásicos; na seguinte inclúense catro procedementos bootstrap propostos na literatura e na terceira presentamos o novo método que propoñemos para a predición. Finalmente, a cuarta sección acolle o estudo de simulación comparativo dos métodos presentados, incluíndo a nova proposta.

## 4.1. Intervalos de predición clásicos

No Capítulo 2 víronse as características que presentaban os modelos autorregresivos clásicos, así como dúas metodoloxías distintas para a estimación dos seus parámetros. Como se adiantou na introdución, dispoñendo dunha serie de lonxitude  $n$  xerada por un proceso  $AR(p)$  (dado pola **Definición 2.5**), imos centrar a nosa atención no que ocorre en tempo  $n+k$ , concretamente, na obtención de intervalos de predición para a variable  $Y_{n+k}$ . Para iso, imos comezar presentando o procedemento proposto por Box–Jenkins (1976), un dos máis empregados na literatura e tamén un dos máis sinxelos de realizar. As ideas fundamentais deste método son as que se presentan a continuación.

Se denotamos por  $\hat{Y}_{n+k}$  a predición realizada en base ao modelo axustado á serie por mínimos cadrados (é dicir,  $\hat{Y}_{n+k} = \hat{\phi}_0 + \sum_{j=1}^p \hat{\phi}_j \hat{Y}_{n+k-j}$ ), os erros de predición virían dados pola diferenza  $Y_{n+k} - \hat{Y}_{n+k}$  que, cando o ruído branco é gaussiano, asintoticamente seguen unha distribución normal de media cero e varianza dada por:

$$\sigma_a^2 (1 + \psi_1^2 + \dots + \psi_{k-1}^2),$$

onde  $\psi_i$  con  $i = 0, \dots, k-1$ , son os primeiros  $k$  coeficientes da expresión (2.13). Así, poderían obterse os intervalos para a predición ao  $100(1 - \alpha)\%$  que serían da forma:

$$\left( \hat{Y}_{n+k} - z_{\alpha/2} \left[ \hat{\sigma}_a^2 \sum_{j=0}^{k-1} \hat{\psi}_j^2 \right]^{1/2}, \hat{Y}_{n+k} + z_{\alpha/2} \left[ \hat{\sigma}_a^2 \sum_{j=0}^{k-1} \hat{\psi}_j^2 \right]^{1/2} \right),$$

onde  $z_{\alpha/2}$  representa o cuantil  $(1 - \alpha/2)$  da distribución normal estándar,  $\hat{\sigma}_a^2$  é a estimación usual de  $\sigma_a^2$  e  $\hat{\psi}_j$ , con  $j = 1, \dots, k-1$  son os coeficientes do polinomio  $\hat{\Psi}(B)$  calculados dende a fórmula  $\hat{\Psi}(B) = \hat{\Phi}^{-1}(B)$ .

## 4.2. Intervalos de predición bootstrap

Como se puido observar, a construción dos intervalos de predición Box–Jenkins descansa na asunción de normalidade do proceso do erro. Cando esta condición falla, os resultados que proporciona poderían ser pouco axeitados. Isto levou á proposta de diversos procedementos baseados no bootstrap para a obtención de intervalos de predición cunha cobertura asintótica apropiada, sen facer ningunha asunción paramétrica acerca da distribución do erro. Como é coñecido, as diferenzas entre os posibles métodos bootstrap presentados atopáranse no plan de remostraxe deseñado.

As ideas de Stine (1987) baséanse no feito de que se denotamos por  $H_k$  á distribución incondicional da variable aleatoria  $Y_{n+k} - \hat{Y}_{n+k}$ , os intervalos de predición teóricos para tempo  $n+k$  serían os dados por:

$$\left( \hat{Y}_{n+k} + H_k^{-1} \left( \frac{\alpha}{2} \right), \hat{Y}_{n+k} + H_k^{-1} \left( 1 - \frac{\alpha}{2} \right) \right),$$

onde  $H_k^{-1}(b)$  é o cuantil de orde  $b$  de  $H_k$ . Dado que a distribución  $H_k$  é descoñecida, podería estimarse mediante bootstrap e posteriormente introducir na fórmula do intervalo o valor axeitado para o cuantil. Non obstante, este enfoque non é apropiado para a predición. Como mostran Pascual et al. (2001) vía simulación, centrar o intervalo de predición no predictor lineal produce infracobertura cando o erro presenta distribución asimétrica.

Como se citou na introdución, alternativamente a este enfoque xurdiron outros na literatura baseados en replicar directamente a distribución de  $Y_{n+k}$  condicionada á serie. Será neste tipo de intervalos de predición nos que nos centraremos neste traballo.



Unha característica que posúen os métodos que imos presentar nas seguintes subseccións, é que involucran a remostraxe dos residuos. Definamos o  $t$ -ésimo residuo como:

$$\hat{a}_t = Y_t - \hat{\phi}_0 - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2} - \dots - \hat{\phi}_p Y_{t-p}, \quad t = p+1, p+2, \dots, n. \quad (4.1)$$

onde  $\hat{\phi}_1, \dots, \hat{\phi}_p$  denotan os estimadores de mínimos cadrados ordinarios dos parámetros do modelo.

Dado que  $\hat{a}_t$  é o estimador do erro verdadeiro  $a_t$ , na obtención de réplicas bootstrap resulta moi importante proporcionar un estimador axeitado para a súa función de distribución  $F_a$ . Cando se emprega o procedemento de estimación de mínimos cadrados, os residuos do modelo presentarán a propiedade de ter media 0, pero de empregarse outro procedemento (como o de máxima verosimilitude) sería preciso centrar os residuos na orixe. Stine (1987) tamén propón reescalalos para a aplicación das técnicas bootstrap. Deste xeito, definiríamos os residuos reescalados centrados como:

$$\hat{a}'_t = \sqrt{\frac{n-p}{n-2p}} \left( \hat{a}_t - \frac{1}{n-p} \sum_{j=1}^{n-p} \hat{a}_j \right). \quad (4.2)$$

Entón un estimador apropiado para  $F_a$ , sería precisamente a distribución empírica dos residuos reescalados e centrados, que denotaremos por  $\hat{F}_n^{\hat{a}'}$ :

$$F_n^{\hat{a}'}(x) = \frac{1}{n-p} \sum_{t=p+1}^n 1_{\{\hat{a}'_t \leq x\}}.$$

Este estimador será o empregado polos métodos que se presentarán nas seguintes subseccións.

### 4.2.1. Bootstrap de Thombs e Schucany

No contexto dos modelos  $AR(p)$ , Thombs e Schucany (1990) propoñen un método bootstrap primeiro cara atrás e logo cara adiante para a obtención de intervalos de predición para horizonte  $k$ . Este será o procedemento que expoñeremos a continuación, explicando previamente as ideas que o fundamentan para, finalmente, presentar o algoritmo que o resume.

Como é sabido, para os modelos  $AR(p)$  a distribución de  $Y_{n+k}$  condicionada a todos os valores pasados, é a mesma que a distribución condicional de  $Y_{n+k}$  dados os últimos  $p$  valores da serie. Daquela para simular a distribución condicional de  $Y_{n+k}$ , debería propoñerse un plan de remostraxe no que as réplicas bootstrap xeradas cumprisen dúas condicións:

1. Reproducir a estrutura de correlación da serie para a que se realizan as tarefas de predición.
2. Os últimos  $p$  valores de cada réplica deberían coincidir cos da serie:  $Y_t^* = Y_t$  para  $t = n-p+1, \dots, n$ .

Para cumprir este cometido, Thombs e Schucany (1990) recorren á chamada *representación cara atrás* dos modelos  $AR(p)$  estacionarios, nos que a variable aleatoria  $Y_t$  é expresada coma unha combinación lineal dos valores futuros máis un termo de erro:

$$Y_t = \phi_0 + \phi_1 Y_{t+1} + \phi_2 Y_{t+2} + \dots + \phi_p Y_{t+p} + e_t. \quad (4.3)$$

É importante destacar que a estrutura de correlación do proceso xerada por (2.11) e a xerada por (4.3) é a mesma (razón pola cal chamamos do mesmo xeito aos seus coeficientes). Isto está relacionado cos cálculos que incluímos no Apéndice A para a función de autocorrelacións parciais, onde se pode ver que os coeficientes da regresión de mínimos cadrados de  $Y_t$  sobre  $Y_{t+1}, \dots, Y_{t+k-1}$  son os mesmos que os de  $Y_{t+k}$  sobre  $Y_{t+1}, \dots, Y_{t+k-1}$ . Dado que a ecuación (4.3) require a xeración de variables  $Y_t$  cara atrás no tempo, unha elección razoable para os primeiros  $p$  valores da remostraxe sería tomar os últimos  $p$  valores da serie observada. Así, estas réplicas bootstrap cara atrás reproducirían a estrutura de correlación da serie e tamén mimetizarían a estrutura condicional aos últimos  $p$  valores da mesma.

Para describir de xeito explícito como obteríamos esas remostras precisamos un estimador dos erros da representación cara atrás  $e_t$ , que sería o dado polos residuos:

$$\hat{e}_t = Y_t - \hat{\phi}_0 - \hat{\phi}_1 Y_{t+1} - \hat{\phi}_2 Y_{t+2} - \dots - \hat{\phi}_p Y_{t+p}, \quad t = 1, 2, \dots, n-p, \quad (4.4)$$

onde  $\hat{\phi}_0, \hat{\phi}_1, \dots, \hat{\phi}_p$  denotan os estimadores de mínimos cadrados ordinarios dos parámetros do modelo. Do mesmo xeito que cos residuos cara adiante, imos denotar por  $\hat{F}_n^{\hat{e}'}$  a distribución empírica dos residuos cara atrás centrados e reescalados, os cales teñen unha expresión análoga a (4.2).

Un feito importante que destacan Thombs e Schucany (1990) é que  $F_a$  e  $F_e$  son a mesma cando os erros son normais. Cando isto non ocorre, non teñen por que. De feito, no caso non normal os erros cara atrás non serán independentes, soamente incorrelados. Isto poñería en dúbida que a remostraxe dende a empírica dos residuos cara atrás fose o máis indicado, dado que só se adoita recomendar a remostraxe i.i.d. cando os erros son i.i.d.. Non obstante, os autores apuntan que ese procedemento segue a proporcionar unha boa aproximación, polo que será o empregado.

Expostas as ideas xerais nas que se basea o procedemento para a obtención de intervalos de predición de Thombs e Schucany (TS), imos describir brevemente o algoritmo que se seguiría para obter ditos intervalos:

**Paso 1.** Construimos os residuos cara atrás,  $\hat{e}_t$  con  $t = 1, 2, \dots, n-p$ , segundo a expresión (4.4), e calculamos a súa versión corrixida  $\hat{e}'_t$ .

**Paso 2.** Lanzamos erros bootstrap cara atrás,  $\hat{e}_t^*$  dende a distribución empírica dos residuos centrados e reescalados,  $\hat{F}_n^{\hat{e}'}$ .

**Paso 3.** Definimos as réplicas bootstrap cara atrás:

$$Y_t^* = \hat{\phi}_0 + \hat{\phi}_1 Y_{t+1}^* + \hat{\phi}_2 Y_{t+2}^* + \dots + \hat{\phi}_p Y_{t+p}^* + \hat{e}_t^*, \quad t = n-p, n-p-1, \dots, 1.$$

con  $Y_t^* = Y_t$  cando  $t = n-p+1, n-p+2, \dots, n$ . Como vemos, os valores inicio da remostraxe cara atrás fixémoslos ao valor das  $p$  últimas observacións da serie.

**Paso 4.** Obtemos as versións bootstrap dos estimadores  $\hat{\phi}_1^*, \hat{\phi}_2^*, \dots, \hat{\phi}_p^*$ .

**Paso 5.** Construimos os residuos cara adiante,  $\hat{a}_t$  con  $t = p+1, p+2, \dots, n$ , seguindo a expresión (4.1) e obtemos a súa versión corrixida,  $\hat{a}'_t$ , dada por (4.2).

**Paso 6.** Lanzamos erros bootstrap cara adiante,  $\hat{a}_t^*$ , dende a función de distribución empírica dos residuos cara adiante corrixidos,  $\hat{F}_n^{\hat{a}'}$ .

**Paso 7.** Definimos as réplicas bootstrap cara adiante:

$$Y_{n+j}^* = \hat{\phi}_0^* + \hat{\phi}_1^* Y_{n+j-1}^* + \hat{\phi}_2^* Y_{n+j-2}^* + \dots + \hat{\phi}_p^* Y_{n+j-p}^* + \hat{a}_{n+j}^*, \quad j = 1, 2, \dots, k.$$

Dada a natureza condicional das remostras, cuxos valores dende  $n-p+1$  ata o  $n$  coincidirán cos  $p$  últimos valores da serie, as observacións bootstrap  $Y_t^*$  cando  $t = n-p+1, n-p+2, \dots, n$ , non contribuirán á variabilidade de  $Y_{n+j}^*$  con  $j = 1, \dots, k$ . Por exemplo, cando  $k = 1$ , a variabilidade de  $Y_{n+1}^*$  vén determinada pola debida á estimación dos parámetros  $\hat{\phi}_1^*, \hat{\phi}_2^*, \dots, \hat{\phi}_p^*$  e polos valores de  $\hat{a}_{n+1}^*$ , que varían con cada réplica bootstrap. Isto é coherente co feito de que o predictor de  $Y_{n+k}$  ten asociado un erro de predición cuxa varianza se pode descompoñer en dúas partes:

- A debida ao termo residual.
- A debida á estimación dos parámetros.

Para  $k = 2$ , antes de obter  $Y_{n+2}^*$  debemos obter  $Y_{n+1}^*$ . Como  $Y_{n+1}^*$  é un valor futuro, variará con cada réplica bootstrap incrementando a variabilidade de  $Y_{n+2}^*$ . Isto é razoable, pois a precisión da predición diminúe ao aumentar o valor do retardo  $k$ .

Para a obtención dos intervalos de predición, debemos repetir os pasos do 1–7 un número elevado de veces,  $B$ . Así, obteriamos un conxunto de futuros valores  $(Y_{n+k}^{*(1)}, Y_{n+k}^{*(2)}, \dots, Y_{n+k}^{*(B)})$ , onde o superíndice denota a remostras coa que se corresponden. Entón, os límites de predición virían dados polos cuantís da estimación por Monte Carlo da distribución bootstrap de  $Y_{n+k}^*$ . Concretamente, se definimos a distribución bootstrap de  $Y_{n+k}^*$  como:

$$\hat{G}(x) = \mathbb{P}^*(Y_{n+k}^* \leq x),$$

onde  $\mathbb{P}^*$  denota a probabilidade condicionada á serie observada, a súa estimación por Monte Carlo sería

$$\hat{G}_B(y) = \frac{1}{B} \sum_{b=1}^B I(Y_{n+k}^{*(b)} \leq y),$$

e para unha cobertura nominal  $\beta$ , o intervalo de predición para  $Y_{n+k}$  ao  $100\beta\%$  viría dado por

$$(I_B^*(y), S_B^*(y)) = \left( \hat{Q}_B \left( \frac{1-\beta}{2} \right), \hat{Q}_B \left( \frac{1+\beta}{2} \right) \right), \quad (4.5)$$

onde  $\hat{Q}_B = \hat{G}_B^{-1}$ .

Thombs e Schucany (1990) proban a validez asintótica deste método que queda resumida no seguinte teorema.

**Teorema 4.1.** *Sexa  $\{Y_t\}_{t \in \mathbb{Z}}$  un proceso autorregresivo estacionario verificando que  $\mathbb{E}(a_t) = 0$  e  $\mathbb{E}(|a_t^\alpha|) < \infty$  con  $t \in \mathbb{Z}$  para algún  $\alpha > 2$ . Sexa  $(y_1, \dots, y_n)$  unha realización particular de  $\{Y_t\}_{t \in \mathbb{Z}}$  ata instante  $n$ , e consideremos os estimadores de MCO dos parámetros do proceso:  $\hat{\phi}_i$  con  $i = 0, \dots, p$ . Entón, para case todas as secuencias mostrais, cando  $n$  tende a infinito tense que:*

- (a) *As estimacións bootstrap dos parámetros converxen ao valor verdadeiro dos mesmos en probabilidade condicionada.*
- (b)  *$Y_{n+k}^* \rightarrow Y_{n+k}$  en distribución condicionada.*

*Demostración.* A continuación inclúese o esquema da proba deste teorema desenvolvido por Thombs e Schucany (1990), páxina 489.

- (a) Thombs e Schucany (1990) refiren a primeira parte da demostración deste teorema ao traballo de Freedman (1985).
- (b) A proba da segunda parte realízana por indución.

**Caso  $k = 1$ .** Sabemos que

$$Y_{n+1}^* = \hat{\phi}_0^* + \hat{\phi}_1^* y_n + \dots + \hat{\phi}_p^* y_{n-p+1} + \hat{a}_{n+1}^*. \quad (4.6)$$

Polo resultado dado no apartado (a), tense que  $\hat{\phi}_0^*$  converxe en probabilidade condicionada a  $\phi_0$ . Ademais, na expresión (4.6) os valores  $(y_{n-p+1}, \dots, y_n)$  poden ser vistos como constantes por seren parte da mostra. Entón, tense que a suma de termos

$$\hat{\phi}_1^* y_n + \dots + \hat{\phi}_p^* y_{n-p+1} \xrightarrow{P} \phi_1 y_n + \dots + \phi_p y_{n-p+1}, \quad (4.7)$$

e  $\phi_1 y_n + \dots + \phi_p y_{n-p+1}$  é unha constante.

Daquela  $Y_{n+1}^*$  está conformada por unha suma cuxos primeiros termos converxen en probabilidade a unha constante:

$$\hat{\phi}_0^* + \hat{\phi}_1^* y_n + \dots + \hat{\phi}_p^* y_{n-p+1} \xrightarrow{P} \phi_0 + \phi_1 y_n + \dots + \phi_p y_{n-p+1}. \quad (4.8)$$

Por outra parte, o último termo de (4.6),  $\hat{a}_{n+1}^*$ , foi xerado con función de distribución a empírica dos residuos centrados,  $F_n^{\hat{a}}$ , e en Freedman (1985) próbase que  $|F_n^{\hat{a}'} - F_a| \xrightarrow{d} 0$ , polo que  $\hat{a}_t^* \xrightarrow{d} a_t$  para  $t \in \mathbb{Z}$  (en particular,  $\hat{a}_{n+1}^*$  converge en distribución a  $a_{n+1}$  ao ser  $\{a_t\}_{t \in \mathbb{Z}}$  unha sucesión de variables i.i.d.).

Finalmente, a aplicación do Lema de Slutsky permite concluír que  $Y_{n+1}^* \xrightarrow{d} Y_{n+1}$ .

**Caso  $k = 2$ .**

$$Y_{n+2}^* = \hat{\phi}_0^* + \hat{\phi}_1^* Y_{n+1}^* + \hat{\phi}_2^* y_n + \dots + \hat{\phi}_p^* y_{n-p+2} + \hat{a}_{n+2}^*. \quad (4.9)$$

Polo **Caso  $k = 1$**  e volvendo aplicar o Lema de Slutsky, tense que  $\hat{\phi}_1^* Y_{n+1}^* \xrightarrow{d} \phi_1 Y_{n+1}$ . De novo, temos garantido que  $\hat{a}_{n+2}^*$  converge en distribución a  $a_{n+2}$ .

Dado que  $a_{n+2}$  é independente de  $Y_{n+1}$  (polas asuncións do proceso autorregresivo), e  $a_{n+2}^*$  é independente de  $Y_{n+1}^*$  (condicionalmente á mostra), é posible concluír que  $Y_{n+2}^* \xrightarrow{d} Y_{n+2}$ .

**Caso  $k = m$ .** Supoñamos que (b) é certo para un enteiro  $k = m - 1$ . Vexamos, entón, que é certo tamén para  $k = m$ .

$$Y_{n+m}^* = \hat{\phi}_0^* + \hat{\phi}_1^* Y_{n+m-1}^* + \dots + \hat{\phi}_p^* Y_{n+m-p}^* + \hat{a}_{n+m}^*. \quad (4.10)$$

Dado que  $\hat{\phi}_j^* Y_{n+m-j}^* \xrightarrow{d} \phi_j Y_{n+m-j}$  (pola asunción e o Lema de Slutsky) para  $j = 1, \dots, p$ , é posible volver a aplicar o teorema de Slutsky a cada termo para  $j = 1, \dots, p$ . Polo que, finalmente, polos mesmos argumentos dados anteriormente é posible concluír que

$$Y_{n+m}^* \xrightarrow{d} Y_{n+m},$$

en distribución condicionada.

□

Entón, posto que se ten a converxencia débil da distribución bootstrap de  $Y_{n+k}$  (que denotamos por  $\hat{G}$  e que é obtida fixando as últimas  $p$  observacións de cada réplica) á verdadeira distribución condicional, tense a converxencia puntal dos cuantís:  $\hat{Q}(u) \rightarrow Q(u)$  para cada  $0 < u < 1$  onde  $\hat{Q} = \hat{G}^{-1}$  é a función cuantil bootstrap teórica. Dado que o número de remostras considerado,  $B$ , vai ser finito, emprégase o teorema de Glivenko–Cantelli para concluír que  $\hat{Q}_B(u)$  converge en probabilidade a  $\hat{Q}(u)$  cando  $B \rightarrow \infty$ . Estes resultados conclúen a validez asintótica do intervalo de predición proporcionado polo método, é dicir:

$$\lim_{n \rightarrow \infty} \left( \lim_{B \rightarrow \infty} \mathbb{P}^* (I_B^*(y) \leq Y_{n+k} \leq S_B^*(y)) \right) = \beta.$$

### 4.2.2. Bootstrap de Cao et al. (1997)

O maior inconveniente do procedemento presentado por Thombs e Schucany é que o seu algoritmo é relativamente lento computacionalmente, debido, especialmente, ao feito de que en cada remostra cara atrás é preciso estimar de novo os parámetros do modelo  $AR(p)$ . Nalgúns casos, pode precisarse dun método máis rápido para obter intervalos de predición. Ademais ao involucrar a representación cara atrás dos  $AR(p)$ , supón unha dificultade á hora de estender este procedemento a outros modelos de series temporais.

Neste contexto, Cao et al. (1997) propoñen un procedemento bootstrap alternativo deseñado para aforrar tempo computacional. Nel, non se constrúen as remostras cara atrás, polo que non se incorpora a variabilidade procedente da estimación dos parámetros. Tal razón lévaos a darlle o nome de *bootstrap condicional* (CB).

A súa proposta baséase en que, en vista da expresión (2.11), pódese entender  $Y_{n+k}$  coma unha suma de termos, uns que só dependen da serie observada e dos verdadeiros valores dos parámetros e outros que son combinacións lineais dos erros futuros  $a_t$  con  $t = n + 1, \dots, n + k$ , cuxos coeficientes dependen dos parámetros descoñecidos. Entón, a distribución condicional de  $Y_{n+k}$  dada a serie observada, será a convolución de  $k$  distribucións obtidas ao reescalar a distribución do erro  $F_a$ . Se tanto esta como os parámetros fosen coñecidos,

tamén o sería a distribución condicional de  $Y_{n+k}$ , pero como isto non ocorre na práctica, unha opción razoable é substituír eses valores teóricos polos estimadores de mínimos cadrados e por un estimador da distribución do erro. Para valores moi pequenos de  $k$  poderíase ter de xeito exacto a distribución bootstrap pero, en xeral, adoita recorrerse a Monte Carlo.

Así, no bootstrap condicional só se obteñen réplicas dos valores futuros  $Y_{n+1}^*, \dots, Y_{n+k}^*$  da serie. O seu plan de remostraxe é o seguinte:

**Paso 1.** Estímense os parámetros do modelo, obtéñense os residuos segundo a expresión (4.1), céntranse e reescálanse (véxase (4.2)) seguindo a recomendación de Stine (1987).

**Paso 2.** Constrúese a función empírica dos residuos cara adiante centrados reescalados,  $\hat{F}_n^{\hat{a}'}$  e xéranse a partir dela erros bootstrap con reempazamento:  $\hat{a}_t^*$  con  $t = n + 1, \dots, n + k$ .

**Paso 3.** Definimos  $Y_t^* = Y_t$  para cada  $t = n - p + 1, n - p + 2, \dots, n$  e construímos observacións bootstrap futuras usando a ecuación:

$$Y_{n+j}^* = \hat{\phi}_0 + \hat{\phi}_1 Y_{n+j-1}^* + \hat{\phi}_2 Y_{n+j-2}^* + \dots + \hat{\phi}_p Y_{n+j-p}^* + \hat{a}_{n+j}^* \quad j = 1, 2, \dots, k.$$

Este procedemento repetiríase un número elevado de veces,  $B$ , e a distribución bootstrap de  $Y_{n+k}^*$  empregárase para aproximar a distribución de  $Y_{n+k}$  dada a serie observable. Chegados a este punto, obteríamos os intervalos de predición cun razoamento análogo ao que se describiu para o método de Thombs e Schucany.

Como se puido observar, o mecanismo de remostraxe non foi empregado para construír réplicas bootstrap da serie orixinal, senón para obter posibles valores futuros da serie, os cales non foron observados cando se tomou a mostra. Este procedemento vai ser consistente no mesmo sentido que o de Thombs e Schucany, e a súa validez quedaría recollida nun resultado completamente análogo ao **Teorema 4.1**, segundo consta no traballo de Cao et al. (1997).

### 4.2.3. Bootstrap de Pascual, Romo e Ruíz

Seguindo a liña de Cao et al. (1997), Pascual et al. (2001) propoñen un método bootstrap para a obtención de intervalos de predición que non involucra a representación cara atrás do modelo  $AR(p)$ . Non obstante, na súa proposta obtéñense as versións bootstrap dos parámetros para realizar a predición, polo que tamén se incorpora a variabilidade debida á súa estimación, a diferenza do que ocorría co bootstrap de Cao et al. (1997).

O procedemento que describen estes autores (PRR) podería verse como un algoritmo en dúas etapas. Nunha primeira etapa o mecanismo de remostraxe emprégase para construír réplicas bootstrap da serie orixinal, o que permite obter as versións bootstrap dos parámetros. Nunha segunda etapa, fíxanse as últimas observacións da remostraxe ás da serie (para conservar o carácter condicional da predición), e empréganse para obter valores futuros bootstrap da serie, os cales non foron observados cando se tomou a mostra.

O algoritmo que resume este procedemento é o seguinte.

**Paso 1.** Estímense os parámetros do modelo, obtéñense os residuos segundo a expresión (4.1), céntranse e reescálanse (véxase (4.2)) seguindo a recomendación de Stine (1987).

**Paso 2.** Constrúese a distribución empírica dos residuos cara adiante centrados reescalados,  $\hat{F}_n^{\hat{a}'}$  e xéranse a partir dela erros bootstrap con reempazamento:  $\hat{a}_t^*$  con  $t = p + 1, \dots, n + k$ .

**Paso 3.** Definimos as réplicas bootstrap da serie, fixando os  $p$  primeiros valores da remostraxe aos  $p$  primeiros valores da serie orixinal (para ter valores de inicio):

$$Y_t^* = \hat{\phi}_0 + \hat{\phi}_1 Y_{t+1}^* + \hat{\phi}_2 Y_{t+2}^* + \dots + \hat{\phi}_p Y_{t+p}^* + \hat{a}_t^*, \quad t = p + 1, p + 2, \dots, n,$$

con  $Y_t^* = Y_t$  cando  $t = 1, \dots, p$ .

**Paso 4.** A partir da remostraxe, calculamos as versións bootstrap dos estimadores dos parámetros:  $\hat{\phi}_1^*, \hat{\phi}_2^*, \dots, \hat{\phi}_p^*$ .

**Paso 5.** Fixamos os  $p$  últimos valores da remostrá aos da serie orixinal (para reproducir o carácter condicional da predición), e obtemos valores futuros da serie empregando as versións bootstrap dos estimadores dos parámetros:

$$Y_{n+j}^* = \hat{\phi}_0^* + \hat{\phi}_1^* Y_{n+j-1}^* + \hat{\phi}_2^* Y_{n+j-2}^* + \dots + \hat{\phi}_p^* Y_{n+j-p}^* + \hat{a}_{n+j}^*, \quad j = 1, 2, \dots, k.$$

con  $Y_t^* = Y_t$  para  $t = n - p + 1, n - p + 2, \dots, n$ .

Repítense os pasos do 2 ao 5 un número elevado de veces,  $B$ , procédese como nos dous mecanismos bootstrap que vimos anteriormente para a obtención dos intervalos de predición. Como xa se citou na introdución do capítulo, a principal vantaxe que comporta este mecanismo bootstrap con respecto ao de Thombs e Schucany é que pode ser estendido a modelos máis complexos no seo das series temporais. De feito, Pascual et al. (2001), modifican este procedemento para adaptalo a modelos non estacionarios, como é o caso dos modelos  $ARIMA(p, d, q)$ . Ademais, estes autores proban a consistencia do seu método para calquera tipo de M-estimador <sup>1</sup>. O teorema que enuncián é o seguinte:

**Teorema 4.2.** *Sexa  $\{Y_t\}_{t \in \mathbb{Z}}$  un proceso  $ARIMA(p, d, q)$  verificando que  $\mathbb{E}(a_t) = 0$  e  $\mathbb{E}(a_t^4) < \infty$  e as raíces do polinomio autorregresivo satisfacendo a condición de estacionaridade usual. Sexa  $(y_1, \dots, y_n)$  unha realización particular de  $\{Y_t\}_{t \in \mathbb{Z}}$  ata instante  $n$  e consideremos  $\hat{\phi}_i$  con  $i = 0, \dots, p$ , e  $\hat{\theta}_j$  con  $j = 1, \dots, q$ , M-estimadores calquera dos parámetros do modelo  $ARIMA(p, d, q)$ . Daquela, dada a serie observada,  $Y_{n+k}^*$  converxe en distribución condicionada a  $Y_{n+k}$ , en probabilidade respecto á serie observada cando  $n$  tende a infinito.*

*Demostración.* A demostración deste teorema pode ser consultada en Pascual et al. (2004), páxinas 455–456. □

#### 4.2.4. Predición mediante a mediana condicional

O algoritmo bootstrap proposto por Thombs e Schucany precisa da modificación do plan de remostraxe dos residuos cando a estimación se realiza mediante MDA para distribucións do erro asimétricas. Isto é debido a que a estrutura de correlación do proceso cara adiante e cara atrás xa non son a mesma.

Na Figura 4.1 amosamos o gráfico de dispersión de  $Y_t$  respecto de  $Y_{t-1}$  dunha serie de tamaño  $n = 1000$  do modelo autorregresivo  $Y_t = 0.75Y_{t-1} + a_t$ , onde  $a_t$  segue unha  $\mathcal{X}^2$  con 1 grao de liberdade.

<sup>1</sup>Como é coñecido, na regresión lineal clásica introducida no Capítulo 1,

$$Y_i = \boldsymbol{\theta}^\top \mathbf{x}_i^1 + \varepsilon_i, \quad i = 1, \dots, n$$

onde  $\varepsilon_i$  son erros aleatorios *i.i.d.*, é posible definir estimadores dos parámetros do modelo máis xerais que os baseados na perda cuantílica ou en mínimos cadrados. Estes son un caso particular dos chamados M-estimadores, que denotaremos mediante  $\hat{\boldsymbol{\theta}}_M$ , definidos como segue:

$$\hat{\boldsymbol{\theta}}_M = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \frac{1}{n} \sum_{i=1}^n \rho(Y_i - \boldsymbol{\theta}^\top \mathbf{x}_i^1), \quad (4.11)$$

onde  $\rho(\cdot)$  é unha función convexa definida en  $\mathbb{R}$  tal que

$$\min \mathbb{E}(\rho(\varepsilon_1 + u) - \rho(\varepsilon_1)) = 0. \quad (4.12)$$

No caso da estimación mediante mínimos cadrados a función  $\rho(\cdot)$  é simplemente o cadrado  $\rho(x) = x^2$ , mentres que para a estimación cuantil  $\rho(\cdot)$  é a función de perda cuantílica definida en (1.3).

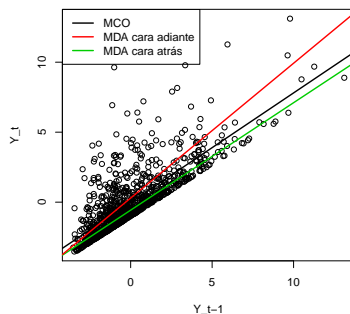


Figura 4.1: Representación gráfica da nube de puntos de  $Y_t$  respecto a  $Y_{t-1}$  xunto coas rectas de regresión axustadas. En negro a recta axustada por MCO, en vermello a estimada mediante MDA de  $Y_t$  sobre  $Y_{t-1}$  e, en verde, a estimada por MDA de  $Y_{t-1}$  respecto a  $Y_t$ .

Cando se estima por MCO, a recta de regresión axustada de  $Y_t$  sobre  $Y_{t-1}$  é a mesma que a de  $Y_{t-1}$  sobre  $Y_t$  mentres que, como se pode observar, cando se estima mediante MDA esa propiedade non se verifica. Breidt et al. (1995) propoñen unha modificación do plan de remostraxe bastante complexa que permite realizar a estimación mediante MDA, involucrando tamén a representación cara atrás do  $AR(p)$ . Non obstante, neste traballo non imos abordala.

Pola contra, o bootstrap condicional de Cao et al. (1997) é aplicable directamente no contexto da predición mediante a mediana condicional ao non involucrar a representación cara atrás do  $AR(p)$ , e como vimos no Capítulo 2, os estimadores de mínimas desviacións absolutas son consistentes no contexto da autorregresión.

Neste caso, os residuos do  $QAR(p)$  homocedástico (que é un  $AR(p)$  cando se lle pide varianza finita para á innovación) virían dados por

$$\hat{a}_t = Y_t - \hat{\phi}_0 - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2} - \dots - \hat{\phi}_p Y_{t-p}, \quad t = p+1, p+2, \dots, n. \quad (4.13)$$

onde  $\hat{\phi}_0, \hat{\phi}_1, \dots, \hat{\phi}_p$  denotan os estimadores de MDA dos parámetros do modelo. O algoritmo deseñado por Cao et al. (1997) presentado na Subsección 4.2.2 seguiríase igualmente, pero agora realizando sempre a estimación mediante MDA.

Tamén é posible empregar o algoritmo de Pascual et al. (2001), pois polo **Teorema 4.2**, este será consistente para calquera tipo de M-estimadores, polo que, en particular, serao para os de mínimas desviacións absolutas. Novamente, agora os residuos virían dados por (4.13) e realizaríanse as estimacións dos parámetros mediante MDA.

### 4.3. Novo método para a predición

Neste Capítulo, ata o de agora, revisamos distintos procedementos da literatura para a obtención de intervalos de predición no contexto dos modelos autorregresivos. Neste apartado, pola contra, propoñemos un procedemento bootstrap para a predición nese tipo de modelos, que pretende mellorar a cobertura práctica dos intervalos de predición obtidos mediante os métodos existentes, especialmente cando contamos con tamaño de mostra reducido.

Este novo procedemento está baseado no bootstrap de pesos aleatorios de Rao e Zhao (1992) no contexto dos M-estimadores para a regresión. Ao igual que o bootstrap de Cao et al. (1997), non require a obtención de réplicas bootstrap da serie orixinal pero, a diferenza deste último, inclúe a variabilidade debida á estimación dos parámetros mediante a inclusión de pesos aleatorios na función de perda cuantílica.

Para presentar este procedemento, o algoritmo que o resume e os resultados que aseguran a súa consistencia, imos comezar primeiro revisando as ideas de Rao e Zhao (1992).

### Aproximación da distribución dun M-estimador mediante bootstrap de pesos aleatorios

Rao e Zhao (1992) aproximan a distribución dun M-estimador, véxase a expresión (4.11) empregando o bootstrap de pesos aleatorios. Este é un procedemento no que a versión bootstrap  $\hat{\boldsymbol{\theta}}_M^*$  do M-estimador de  $\boldsymbol{\theta}$  se define mediante:

$$\hat{\boldsymbol{\theta}}_M^* = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \frac{1}{n} \sum_{i=1}^n w_i \rho(Y_i - \boldsymbol{\theta}^\top \mathbf{x}_i^1) = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \frac{1}{n} \sum_{i=1}^n w_i \rho(\varepsilon_i), \quad (4.14)$$

onde  $\{w_i\}_{i=1}^n$  son os pesos aleatorios: variables i.i.d. sempre positivas, verificando que  $\mathbb{E}(w_1) = 1$ ,  $\text{Var}(w_1) = 1$  e independentes de  $\{Y_t\}_{t \in \mathbb{Z}}$ . Así, aproximaríase a distribución na mostraxe do estadístico  $\vec{\mathcal{S}}_n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$  mediante a distribución empírica condicional de  $\vec{\mathcal{S}}_n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}})$ , onde  $\vec{\mathcal{S}}_n = \sum_{i=1}^n \mathbf{x}_i^1 \mathbf{x}_i^{1\top}$ .

Para efectuar a demostración da consistencia deste enfoque, Rao e Zhao (1992) introducen as 5 hipóteses seguintes:

- H1.**  $\rho(\cdot)$  é unha función convexa en  $\mathbb{R}$  con derivadas pola dereita e pola esquerda,  $\psi_+(\cdot)$  e  $\psi_-(\cdot)$  respectivamente, e  $\psi(\cdot)$  é unha función definida en  $\mathbb{R}$ , tal que  $\psi_-(u) \leq \psi(u) \leq \psi_+(u)$  para todo  $u \in \mathbb{R}$ .
- H2.**  $\mathbb{E}(\psi(\varepsilon_1)) = 0$ , e  $G(u) = \mathbb{E}(\psi(\varepsilon_1 + u))$  ten derivada  $\lambda$  en  $u = 0$ .
- H3.**  $\sigma^2 = \mathbb{E}(\psi(\varepsilon_1)^2) < \infty$  e  $\mathbb{E}(\psi(\varepsilon_1 + u) - \psi(\varepsilon_1))^2 \rightarrow 0$  cando  $u \rightarrow 0$ .
- H4.** Cúmprese que  $\vec{\mathcal{S}}_n = \sum_{i=1}^n \mathbf{x}_i^1 \mathbf{x}_i^{1\top} > 0$  para  $n$  grande e  $d_n^2 = \max \mathbf{x}_i^{1\top} \vec{\mathcal{S}}_n^{-1} \mathbf{x}_i^1 \rightarrow 0$  cando  $n \rightarrow \infty$ .
- H5.** As variables  $\{w_i\}_{i=1}^n$  son variables i.i.d. tales que  $\mathbb{P}(w_1 > 0) = 1$ , verificando que  $\mathbb{E}(w_1) = 1$  e  $\mathbb{E}(w_1^2) = 2$ . Ademais  $\{w_i\}_{i=1}^n$  e  $\{Y_t\}_{t \in \mathbb{Z}}$  son independentes. Un exemplo de distribución que verifica isto é a exponencial de parámetro  $\lambda = 1$ .

A hipótese **H3** é equivalente a que  $\mathbb{E}(\psi(\varepsilon_1 + u)^2)$  é finita para  $u$  nun entorno de 0 e continua en  $u = 0$ , mentres que **H4** implica que  $\vec{\mathcal{S}}_n^{-1} \rightarrow 0$  cando  $n \rightarrow \infty$ . Así, a validez do bootstrap de Rao e Zhao (1992) pode resumirse no seguinte teorema.

**Teorema 4.3.** *Baixo o modelo de regresión lineal usual e baixo as hipóteses **H1–H5**, supoñamos que  $\mathbb{E}|\psi(\varepsilon_1)|^\delta < \infty$  para algún  $\delta > 2$ . Sexan  $\hat{\boldsymbol{\theta}}_M$  e  $\hat{\boldsymbol{\theta}}_M^*$  definidos como en (4.11) e (4.12), respectivamente. Daquela cando  $n \rightarrow \infty$ , por unha parte tense que*

$$\vec{\mathcal{S}}_n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_M^* - \hat{\boldsymbol{\theta}}_M) \xrightarrow{d} N_{p+1}\left(\mathbf{0}, \lambda^{-2} \sigma^2 \vec{\mathbf{I}}_{p+1}\right), \quad (4.15)$$

onde  $\vec{\mathbf{I}}_{p+1}$  denota á matriz identidade de orde  $p$ . Por outra,

$$\sup_{\mathbf{u}} \left| \mathbb{P}^* \left( \vec{\mathcal{S}}_n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_M^* - \hat{\boldsymbol{\theta}}_M) \leq \mathbf{u} \right) - \mathbb{P}(\mathbf{N} \leq \mathbf{u}) \right| = \sup_{\mathbf{r}} \left| \mathbb{P}^* \left( \vec{\mathcal{S}}_n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_M^* - \hat{\boldsymbol{\theta}}_M) \leq \mathbf{r} \right) - \mathbb{P}(\mathbf{N} \leq \mathbf{r}) \right| \xrightarrow{P} 0, \quad (4.16)$$

onde  $\mathbf{N}$  é un vector de dimensión  $p+1$  distribuído coma unha  $N_{p+1}\left(\mathbf{0}, \lambda^{-2} \sigma^2 \vec{\mathbf{I}}_{p+1}\right)$ , onde  $\mathbf{u}$  se move sobre todos os vectores de  $p+1$  compoñentes, mentres que  $\mathbf{r}$  móvese só naqueles de  $p+1$  compoñentes racionais, e  $\mathbb{P}^*$  representa a probabilidade condicionada á mostra observada.

*Demostración.* A proba deste teorema pode consultarse en Rao e Zhao (1992), páxinas 326–330. □



Das expresións (4.15) e (4.16) séguese que

$$\sup_{\mathbf{u}} \left| \mathbb{P}^* \left( \vec{\mathcal{S}}_{\frac{1}{n}}^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_M^* - \hat{\boldsymbol{\theta}}_M) \leq \mathbf{u} \right) - \mathbb{P} \left( \vec{\mathcal{S}}_{\frac{1}{n}}^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}) \leq \mathbf{u} \right) \right| \xrightarrow{P} 0, \quad (4.17)$$

o que significa que o teorema anterior garante que a aproximación da distribución de  $\vec{\mathcal{S}}_{\frac{1}{n}}^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta})$  mediante a distribución condicionada á mostra de  $\vec{\mathcal{S}}_{\frac{1}{n}}^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_M^* - \hat{\boldsymbol{\theta}}_M)$  é válida no sentido débil.

### Bootstrap de pesos aleatorios para a predición en mediana en modelos autorregresivos

Dado que os cuantís mostrais son un caso particular de M-estimador, parece razoable pensar que se podería artellar un algoritmo bootstrap para a predición nos modelos autorregresivos homocedásticos baseado neste enfoque.

Intuitivamente, o primeiro problema co que nos atoparíamos é que a función de perda cuantílica non é derivable en 0 (pero si que o é pola dereita e pola esquerda, co que se verifica **H1**). Para que se verifiquen as hipóteses **H2** e **H3**, é suficiente con pedir o erro teña densidade  $f(\cdot)$  positiva en 0. Por outra parte, no contexto da autorregresión trabállase con deseño aleatorio, mentres que os distintos métodos de estimación no Capítulo 1 foron expostos baixo deseño fixo. O paso dun tipo de deseño a outro faise condicionando á mostra no caso aleatorio para pasar a deseño fixo. Ademais, como se viu no Capítulo 2, tal e como definimos o proceso  $QAR(p)$  homocedástico, este verifica que  $\mathbb{E}(Y_{t-j}a_t) = 0$  con  $j = 1, \dots, p$ , o que permite usar na estimación do modelo as ferramentas usuais da regresión clásica.

Nas seguintes liñas resúmese o algoritmo bootstrap que se propón para a predición nos modelos autorregresivos.

**Paso 1.** Axústase o modelo  $AR(p)$  (usando os estimadores MDA), obtéñense os residuos segundo a expresión (4.13) e constrúese a súa distribución empírica.

**Paso 2.** Xerar  $\hat{a}_j^*$  para  $j = 1, \dots, k$  con distribución a empírica dos residuos dados por (4.13).

**Paso 3.** Lanzar unha mostra de variables positivas  $w_t$  con  $t = 1, \dots, n$ , independentes de  $\{Y_t\}_{t \in \mathbb{Z}}$ , tales que  $\mathbb{E}(w_t) = 1$  e  $\mathbb{E}(w_t^2) = 2$ . Por exemplo,  $w_t$  poden ser i.i.d. a unha Exponencial de parámetro  $\lambda = 1$ . Estas variables farán o papel de pesos aleatorios.

**Paso 4.** Obter unha estimación dos coeficientes da serie introducindo na función obxectivo os pesos aleatorios:

$$\hat{\boldsymbol{\phi}}^* = \arg \min_{\boldsymbol{\phi} \in \mathbb{R}^{p+1}} \sum_{t=p+1}^n w_t \rho_{\tau}(Y_t - \boldsymbol{\phi}^{\top} \mathbf{Z}_{t,p}^1), \quad (4.18)$$

onde  $\mathbf{Z}_{t,p}^1 = (1, Y_{t-1}, \dots, Y_{t-p})^{\top}$ ,  $\boldsymbol{\phi} = (\phi_0, \phi_1, \dots, \phi_p)^{\top}$  e, consecuentemente,  $\hat{\boldsymbol{\phi}}^* = (\hat{\phi}_0^*, \hat{\phi}_1^*, \dots, \hat{\phi}_p^*)^{\top}$ .

**Paso 5.** Definimos  $Y_t^* = Y_t$  para cada  $t = n - p + 1, n - p + 2, \dots, n$  e construímos observacións bootstrap futuras usando a ecuación:

$$Y_{n+j}^* = \hat{\phi}_0^* + \hat{\phi}_1^* Y_{n+j-1}^* + \hat{\phi}_2^* Y_{n+j-2}^* + \dots + \hat{\phi}_p^* Y_{n+j-p}^* + \hat{a}_{n+j}^* \quad j = 1, 2, \dots, k.$$

A continuación procederíase como nos métodos descritos nos apartados anteriores, estimaríamos a distribución bootstrap de  $Y_{n+k}^*$  por Monte Carlo, mediante a repetición nun número elevado de veces dos pasos do 2 ao 5, e obteríanse os intervalos de predición do mesmo xeito que nos métodos bootstrap antes descritos.

Outra opción sería seguir a formulación de Stine (1987) e estudar a distribución do erro cuadrático medio de predición empregando as mesmas ideas. Se definimos

$$Y_{n+j}^* = \hat{\phi}_0 + \hat{\phi}_1 Y_{n+j-1}^* + \hat{\phi}_2 Y_{n+j-2}^* + \dots + \hat{\phi}_p Y_{n+j-p}^* + \hat{a}_{n+j}^* \quad j = 1, 2, \dots, k. \quad (4.19)$$

onde  $\hat{\phi}_i$  con  $i = 0, \dots, p$  son os estimadores de MDA dos parámetros, e

$$\hat{Y}_{n+j}^* = \hat{\phi}_0^* + \hat{\phi}_1^* \hat{Y}_{n+j-1}^* + \hat{\phi}_2^* \hat{Y}_{n+j-2}^* + \dots + \hat{\phi}_p^* \hat{Y}_{n+j-p}^* \quad j = 1, 2, \dots, k.$$

o análogo bootstrap do erro cuadrático medio de predición (en inglés *mean squared prediction error*) viría dado por:

$$MSPE^* = \frac{1}{B} \sum_{b=1}^B \left( \hat{Y}_{n+j}^* - Y_{n+j}^* \right)^2, \quad (4.20)$$

sendo  $B$  o número de réplicas bootstrap empregadas para estimar por Monte Carlo a distribución bootstrap.

A vantaxe que ten para nós o emprego de pesos aleatorios na estimación dos parámetros, é que nos permite incluír no noso algoritmo a variabilidade debida á súa estimación sen necesidade de obter réplicas bootstrap da serie orixinal. Isto, especialmente cando se dispón de series autorregresivas de pequeno tamaño, pode presentar certa mellora na taxa de cobertura dos intervalos de predición, dado que a variabilidade conferida mediante a mera replicación da serie pode non ser suficiente, como se poderá observar no estudo de simulación.

A consistencia do método dado resumímola no seguinte teorema:

**Teorema 4.4.** *Sexa  $\{Y_t\}_{t \in \mathbb{Z}}$  un proceso  $AR(p)$  estacionario, onde  $\mathbb{E}(a_t^2) < \infty$ . Supoñamos que  $a_t$  ten densidade continua e positiva no 0. Sexa  $(y_1, \dots, y_n)$  unha realización particular de  $\{Y_t\}_{t \in \mathbb{Z}}$  ata instante  $n$ , e consideremos  $\hat{\phi}_i$  con  $i = 0, \dots, p$  as estimacións MDA dos parámetros. Entón, para case todas as secuencias mostrais cando  $n$  tende a infinito tense que:*

- (a) *As estimacións bootstrap dos parámetros converxen ao valor verdadeiro dos mesmos en probabilidade condicionada.*
- (b)  *$Y_{n+k}^* \rightarrow Y_{n+k}$  en distribución condicionada.*

*Demostración.* A proba deste resultado é análoga á desenvolvida no **Teorema 4.1**, pero agora encadeando resultados referentes á estimación mediante MDA.

- (a) A proba deste apartado é consecuencia directa do Teorema 5 de Li et al. (2015).
- (b) Aquí séguese un esquema totalmente análogo á proba do **Teorema 4.1**. Aplicariamos igualmente o razoamento de indución no horizonte  $k$ , pero agora para probar a converxencia en distribución de  $\hat{a}_t^*$  a  $a_t$  con  $t \in \mathbb{Z}$ , recorreremos ao Teorema 3.1 de Kreiss e Franke (1992) que asegura que  $d(\hat{F}_n^a, F_a) \rightarrow 0$  en probabilidade.

□

## 4.4. Estudio de simulación

Nesta sección imos analizar o comportamento práctico dos procedementos bootstrap presentados para a obtención de intervalos de predición no contexto dos modelos autorregresivos homocedásticos. A estimación dos parámetros do modelo realizarase mediante os seguintes procedementos, en cada caso:

- MCO e MDA para o bootstrap condicional (CB) e o de Pascual, Ruiz e Romo (PRR).
- Só MCO no caso do bootstrap de Thombs e Schucany (TS).
- Só MDA para a nosa proposta (P).

Neste último caso, compararemos tamén os resultados obtidos cando se replica directamente a distribución condicional (P) con cando se estima o erro cuadrático medio de predición ( $P_p$ ) usando intervalos de predición tipo percentil (segundo as ideas de Stine, 1987). Compararemos, ademais, os resultados que ofrecen todos estes métodos cos obtidos mediante o procedemento estándar Box–Jenkins (BJ) estimando por MCO.

Este estudo permitiranos concluir que algoritmos bootstrap proporcionan uns intervalos máis apropiados en termos de cobertura, así como estudar as vantaxes que pode comportar a estimación mediante MDA, menos explotada na literatura de series temporais.

O estudo vai estar baseado no modelo autorregresivo de orde dous seguinte:

$$Y_t = 0.75Y_{t-1} - 0.50Y_{t-2} + a_t, \quad (4.21)$$

considerando tres distribucións distintas para  $a_t$ :

- A  $N(0, 1)$ .
- A exponencial de parámetro 1 desprazada de modo que teña media 0.
- A T de Student con 3 graos de liberdade.
- A mixtura de normais #10 de Marron e Wand (1992), en R Core Team (2015) denotada como modelo `MW.nm10`.

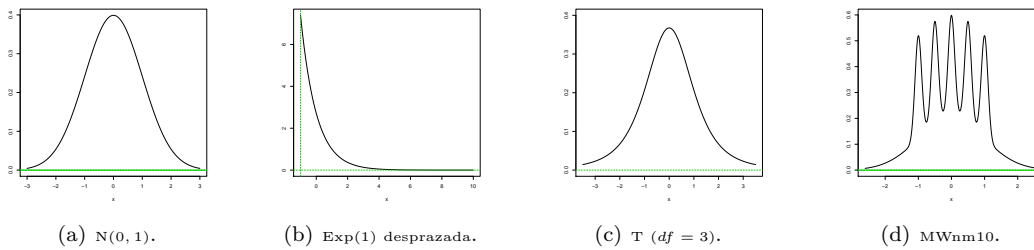


Figura 4.2: Representación gráfica das densidades das distribucións consideradas para o erro.

Por outra parte, consideraremos tamaños de mostra  $n = 25$  e  $50$  e horizontes de predición  $k = 1, 2, 3$ , e traballaremos con coberturas nominais  $\beta = 0.95$  e  $\beta = 0.99$ . A elección dese tamaño de serie e dese horizonte non é arbitraria, senón que nos baseamos nos estudos de simulación xa realizados na literatura, véxase Cao et al. (1997) ou Pascual et al. (2001). Os valores da serie foron obtidos tras un período inicial de estabilización (rexeitando os primeiros 300 valores obtidos). No caso da nosa proposta, os pesos que se empregarán no estudo seguen unha distribución exponencial de parámetro 1.

Para cada combinación de distribución do erro, horizonte de predición e tamaño de mostra, obteremos unha serie de medidas resumo que nos informarán do comportamento práctico de cada método. A estimación da probabilidade de cobertura e da lonxitude dos intervalos de predición, para cada realización de (4.21) farase como detallamos a continuación:

1. Simulamos unha serie baixo o modelo (4.21), cunha determinada distribución para o erro,  $F_a$ , e xeramos por Monte Carlo  $R = 1000$  valores futuros verdadeiros para cada horizonte de predición considerado, empregando o valor real dos parámetros, os últimos  $p$  valores da serie e  $F_a$ .
2. Obtemos os intervalos de predición Box–Jenkins,  $(I_{BJ}, S_{BJ})$ .
3. Seguimos cos algoritmos bootstrap citados, e calculamos o intervalo de predición para cada un deles en base a  $B = 1000$  réplicas bootstrap; por exemplo, para a nosa proposta (P) teriamos un intervalo da forma  $(I_P^*, S_P^*)$ .

4. Estimamos a cobertura condicionada á serie simulada mediante os tres métodos. Para  $P$  teríamos (para os demais algoritmos obteríase analogamente)

$$\hat{\beta}_{i,P}^* = \frac{\sum_{r=1}^R I\left(I_P^* \leq Y_{n+k}^{(r)} \leq S_P^*\right)}{R}, \quad (4.22)$$

onde  $Y_{n+k}^{(r)}$  con  $r = 1, \dots, R$  representan os valores futuros verdadeiros da serie simulada no primeiro paso. Así mesmo, estimamos a lonxitude dos intervalos:

$$\hat{L}_{i,P}^* = S_P^* - I_P^*, \quad (4.23)$$

Repetimos  $M = 500$  veces os pasos do 1 ao 3 (o que significa aplicar os tres procedementos de obtención de intervalos de predición sobre 500 series xeradas a partir de (4.21)). Entón é posible obter a media e o erro estándar para a probabilidade de cobertura e a lonxitude dos intervalos. Por exemplo, para o bootstrap que propoñemos estas medidas virían dadas por:

$$\begin{aligned} \bar{\beta}_P^* &= \frac{1}{M} \sum_{i=1}^M \hat{\beta}_{i,P}^*, \\ SE(\bar{\beta}_P^*) &= \sqrt{\frac{1}{M(M-1)} \sum_{i=1}^M \left(\hat{\beta}_{i,P}^* - \bar{\beta}_P^*\right)^2}, \\ \overline{\text{Lonx}}_P^* &= \frac{1}{M} \sum_{i=1}^M \hat{L}_{i,P}^*, \\ SE(\overline{\text{Lonx}}_P^*) &= \sqrt{\frac{1}{M(M-1)} \sum_{i=1}^M \left(\hat{L}_{i,P}^* - \overline{\text{Lonx}}_P^*\right)^2}. \end{aligned}$$

Para os intervalos de predición mediante os outros métodos obteríanse analogamente estas medidas. Como sinalan Thombs e Schucany (1990), a cobertura  $C(Y)$  é unha variable aleatoria que satisfai  $\mathbb{E}[C(Y)] = \beta$ . En consecuencia,  $\mathbb{P}[C(Y) \geq \beta] = \gamma$ , ha de ser aproximadamente igual a 0.5, é dicir, agárdase que aproximadamente a metade dos intervalos de predición condicionados excedan o valor nominal. Así, pódese obter tamén unha estimación de  $\gamma$  que, por exemplo, para o bootstrap proposto (para os demais mecanismos de predición presentados faríase de xeito análogo) viría dada por:

$$\hat{\gamma}^* = \frac{\sum_{i=1}^M I\left(\hat{\beta}_{i,P}^* \geq \beta\right)}{M},$$

a cal serve coma unha segunda comprobación da adecuación da cobertura proporcionada por cada procedemento.

O estudo de simulación aquí descrito realizouse no programa R Core Team (2015). Cabe mencionar que para levar a cabo a estimación dos parámetros, chámase á función `lm` da librería `stats` e á función `rq` da librería `quantreg`. Por outra parte, para xerar facilmente datos procedentes dunha mixtura de normais empregáronse as funcións `normMix` e `rnorMix` da librería `nor1mix`. Ademais, para garantir a reproducibilidade do estudo, fixouse semente 123456789 coa función `set.seed`.

Os resultados do estudo para tamaño mostral  $n = 25$  amósanse na Táboa 4.1. Na primeira columna amosamos o horizonte de predición ( $k = 1, 2$  e  $3$ ), na seguinte, o método de estimación empregado (MCO ou MDA) e, despois, para cada distribución do erro considerada (normal e exponencial) mostramos a cobertura media dos intervalos de predición, o seu erro estándar, a lonxitude media do intervalo de predición, o seu erro estándar, a lonxitude media dos intervalos, o seu erro estándar e o valor de  $\hat{\gamma}$ .

En negro destacamos o valor da cobertura media máis próximo ao nivel nominal e en vermello, todos aqueles que non son significativamente distintos deste. En vista dos resultados, cabe recalcar que na maioría dos casos de tamaño de mostra pequeno as taxas de cobertura son significativamente distintas de 0.95 (se realizamos un contraste paramétrico para a media da probabilidade de cobertura, con nivel de significación fixado a  $\alpha = 0.05$ , pode comprobarse que esta media vai ser significativamente distinta do nivel fixado). Ademais, para cada caso considerado, xeralmente, as diferenzas entre cada par de métodos tamén son estatisticamente significativas.

Cando a distribución do erro é normal para tamaño de mostra tan pequeno como  $n = 25$ , dado que o procedemento Box Jenkins se asenta en teoría asintótica, presenta infracobertura. O mesmo ocorre cos demais algoritmos bootstrap presentados; non obstante, o método proposto neste traballo parece que é o que proporciona un resultado máis próximo ao nivel nominal en canto a cobertura e, así mesmo, un erro estándar máis baixo. Esta maior taxa de cobertura vai acompañada tamén de intervalos de predición máis longos en xeral. Por outra parte, neste caso o procedemento P\_p amosa peores resultados que o P, aínda que nalgúns casos non sexan significativas as diferenzas.

Conclusións similares poderíamos extraer cando a distribución do erro é exponencial. Pode sorprender que teñamos mellores coberturas incondicionais neste caso que no da normal, pero isto é o efecto da media de coberturas moi elevadas con coberturas máis pequenas (se estudásemos coberturas condicionais os resultados serían mellores no caso da normal). Isto vese reflectido nun maior erro estándar da cobertura media, aínda que no método que propoñemos neste traballo, P, segue a ser o máis pequeno. De incluír unha columna que recollese o porcentaxe de valores quedan por riba e por baixo do intervalo de predición, veríamos que no caso da exponencial, o procedemento Box–Jenkins proporciona intervalos totalmente descentrados, que deixan moitos valores por riba e poucos por baixo. Agora vemos que o método P devolve resultados moi distintos ao P\_p en termos de cobertura, sendo o primeiro claramente mellor.

Por outra parte, se na Táboa 4.1 comparamos os resultados obtidos polos algoritmos CB e PPR cando se estima mediante MDA cos obtidos por MCO, vemos que MDA ofrece mellores coberturas.

Na Táboa 4.2, cuxa estrutura é análoga á da Táboa 4.1, acollemos os resultados cando a distribución do erro é unha T de Student con 3 graos de liberdade e cando é a mixtura de normais do modelo MW.nm10 da librería `nor1mix`. En ambos casos, especialmente no da T de Student, as coberturas son máis baixas que as da Táboa 4.1, pero podemos observar que o método proposto P ofrece unha cobertura media maior que a dos demais métodos en todos os casos, que ademais é significativamente maior a maioría de veces. Tamén é posible observar, novamente, que a estimación mediante MDA ofrece mellores resultados que a de MCO no caso dos procedementos bootstrap PPR e CB.

Para apreciar o efecto do aumento do tamaño mostral, así como o da consideración de intervalos de predición como o 99%, na Táboa 4.3 recollemos os resultados que se obteñen para tamaño mostral  $n = 50$ , unicamente para horizonte de predición  $k = 3$ , para as catro distribucións do erro que viñemos considerando con cobertura nominal 95% e 99%. Agora a estrutura da táboa é lixeiramente distinta, pois na primeira columna en lugar do horizonte de predición, recóllese a cobertura nominal fixada. Ademais, nunha mesma táboa acollemos os resultados para as catro distribucións do erro.

| Horiz. | Esti. | Método | Normal ( $\mu = 0, \sigma = 1$ ) |           |       |           |       | Exponencial ( $\lambda = 1$ ) |           |       |           |       |
|--------|-------|--------|----------------------------------|-----------|-------|-----------|-------|-------------------------------|-----------|-------|-----------|-------|
|        |       |        | Cober.                           | Erro est. | Lonx. | Erro est. | gamma | Cober.                        | Erro est. | Lonx. | Erro est. | gamma |
| 1      | MCO   | BJ     | 90.344                           | 0.268     | 3.656 | 0.026     | 0.214 | 91.477                        | 0.311     | 3.557 | 0.048     | 0.330 |
|        |       | TS     | 92.148                           | 0.263     | 4.020 | 0.032     | 0.378 | 94.557                        | 0.288     | 4.081 | 0.064     | 0.620 |
|        |       | CB     | 89.961                           | 0.311     | 3.790 | 0.033     | 0.248 | 89.940                        | 0.495     | 3.776 | 0.061     | 0.438 |
|        |       | PRR    | 92.316                           | 0.253     | 4.034 | 0.032     | 0.406 | 94.558                        | 0.259     | 4.061 | 0.064     | 0.614 |
|        | MDA   | CB     | 91.098                           | 0.340     | 4.109 | 0.036     | 0.360 | 90.752                        | 0.491     | 4.029 | 0.062     | 0.492 |
|        |       | PRR    | 93.083                           | 0.276     | 4.337 | 0.035     | 0.466 | 94.467                        | 0.303     | 4.216 | 0.062     | 0.624 |
|        |       | P      | <b>93.736</b>                    | 0.252     | 4.387 | 0.039     | 0.520 | <b>95.047</b>                 | 0.259     | 4.225 | 0.063     | 0.658 |
|        |       | P_p    | 93.471                           | 0.258     | 4.401 | 0.040     | 0.482 | 93.155                        | 0.369     | 4.337 | 0.065     | 0.580 |
| 2      | MCO   | BJ     | 88.905                           | 0.329     | 4.490 | 0.036     | 0.200 | 89.567                        | 0.392     | 4.352 | 0.060     | 0.316 |
|        |       | TS     | 91.735                           | 0.297     | 5.003 | 0.044     | 0.368 | 92.312                        | 0.345     | 4.887 | 0.072     | 0.474 |
|        |       | CB     | 89.413                           | 0.341     | 4.635 | 0.039     | 0.248 | 87.606                        | 0.468     | 4.391 | 0.063     | 0.298 |
|        |       | PRR    | 91.799                           | 0.290     | 4.982 | 0.043     | 0.358 | 92.174                        | 0.342     | 4.853 | 0.072     | 0.456 |
|        | MDA   | CB     | 89.391                           | 0.396     | 4.863 | 0.044     | 0.284 | 88.174                        | 0.469     | 4.574 | 0.065     | 0.322 |
|        |       | PRR    | 92.297                           | 0.317     | 5.347 | 0.050     | 0.456 | 92.298                        | 0.342     | 5.018 | 0.069     | 0.464 |
|        |       | P      | <b>93.818</b>                    | 0.276     | 5.754 | 0.065     | 0.554 | <b>94.062</b>                 | 0.294     | 5.319 | 0.078     | 0.586 |
|        |       | P_p    | 93.109                           | 0.316     | 5.684 | 0.067     | 0.520 | 90.910                        | 0.409     | 5.348 | 0.081     | 0.450 |
| 3      | MCO   | BJ     | 89.640                           | 0.307     | 4.563 | 0.036     | 0.214 | 90.339                        | 0.353     | 4.408 | 0.061     | 0.350 |
|        |       | TS     | 92.197                           | 0.279     | 5.060 | 0.045     | 0.408 | 92.945                        | 0.308     | 4.952 | 0.073     | 0.500 |
|        |       | CB     | 90.205                           | 0.308     | 4.713 | 0.040     | 0.278 | 89.609                        | 0.373     | 4.460 | 0.064     | 0.322 |
|        |       | PRR    | 92.535                           | 0.247     | 5.055 | 0.042     | 0.396 | 93.119                        | 0.302     | 4.911 | 0.072     | 0.512 |
|        | MDA   | CB     | 91.051                           | 0.302     | 4.994 | 0.046     | 0.326 | 90.408                        | 0.371     | 4.689 | 0.066     | 0.368 |
|        |       | PRR    | 93.686                           | 0.239     | 5.532 | 0.053     | 0.518 | 93.468                        | 0.302     | 5.151 | 0.071     | 0.532 |
|        |       | P      | <b>94.796</b>                    | 0.225     | 6.128 | 0.079     | 0.612 | <b>94.688</b>                 | 0.270     | 5.592 | 0.087     | 0.620 |
|        |       | P_p    | 94.288                           | 0.249     | 5.896 | 0.077     | 0.558 | 91.957                        | 0.338     | 5.493 | 0.086     | 0.470 |

Táboa 4.1: Resultados obtidos en base a  $M=500$  series distintas de  $n = 25$  datos, considerando  $R=1000$  valores futuros e  $B=1000$  réplicas bootstrap. Semente: 123456789. Cobertura nominal considerada 95% . Destacado en negro aparece o valor de cobertura media máis próximo ao nivel nominal e en vermello todos aqueles que non son significativamente distintos. Os métodos empregados na obtención dos intervalos de predición foron: Box–Jenkins (BJ), Thombs e Schucany (TS), Bootstrap condicional (CB), Pascual, Romo e Ruíz (PRR) e a nosa proposta (P). A estimación realizouse ben por mínimos cadrados ordinarios (MCO) ou ben mediante mínimas desviacións absolutas (MDA).

| Horiz. | Esti. | Método | T ( $df = 3$ ) |           |       |           |       | MW.nm10 |           |       |           |       |
|--------|-------|--------|----------------|-----------|-------|-----------|-------|---------|-----------|-------|-----------|-------|
|        |       |        | Cober.         | Erro est. | Lonx. | Erro est. | gamma | Cober.  | Erro est. | Lonx. | Erro est. | gamma |
| 1      | MCO   | BJ     | 90.385         | 0.285     | 5.946 | 0.154     | 0.242 | 91.817  | 0.285     | 3.199 | 0.022     | 0.336 |
|        |       | TS     | 91.280         | 0.279     | 7.139 | 0.174     | 0.304 | 92.844  | 0.263     | 3.482 | 0.030     | 0.426 |
|        |       | CB     | 89.600         | 0.342     | 6.888 | 0.170     | 0.248 | 90.189  | 0.340     | 3.279 | 0.032     | 0.298 |
|        |       | PRR    | 91.223         | 0.275     | 7.110 | 0.171     | 0.310 | 93.030  | 0.250     | 3.490 | 0.031     | 0.434 |
|        | MDA   | CB     | 91.231         | 0.313     | 7.596 | 0.226     | 0.330 | 91.260  | 0.357     | 3.560 | 0.035     | 0.380 |
|        |       | PRR    | 92.181         | 0.262     | 7.719 | 0.223     | 0.356 | 93.827  | 0.267     | 3.787 | 0.032     | 0.568 |
|        |       | P      | 92.255         | 0.249     | 7.592 | 0.212     | 0.362 | 94.453  | 0.239     | 3.852 | 0.035     | 0.612 |
|        |       | P_p    | 91.389         | 0.302     | 7.617 | 0.213     | 0.326 | 94.120  | 0.253     | 3.855 | 0.035     | 0.588 |
| 2      | MCO   | BJ     | 88.577         | 0.344     | 7.330 | 0.187     | 0.214 | 89.431  | 0.339     | 3.928 | 0.030     | 0.238 |
|        |       | TS     | 90.041         | 0.307     | 8.329 | 0.180     | 0.264 | 91.964  | 0.303     | 4.348 | 0.038     | 0.398 |
|        |       | CB     | 87.256         | 0.403     | 7.723 | 0.169     | 0.168 | 89.462  | 0.352     | 4.033 | 0.034     | 0.244 |
|        |       | PRR    | 89.730         | 0.335     | 8.252 | 0.177     | 0.252 | 92.137  | 0.286     | 4.347 | 0.037     | 0.398 |
|        | MDA   | CB     | 88.542         | 0.397     | 8.248 | 0.219     | 0.190 | 89.783  | 0.367     | 4.235 | 0.036     | 0.304 |
|        |       | PRR    | 90.444         | 0.325     | 8.756 | 0.219     | 0.284 | 93.016  | 0.286     | 4.682 | 0.039     | 0.500 |
|        |       | P      | 91.291         | 0.287     | 9.052 | 0.213     | 0.336 | 94.633  | 0.245     | 5.101 | 0.053     | 0.616 |
|        |       | P_p    | 89.721         | 0.402     | 8.956 | 0.213     | 0.282 | 93.659  | 0.281     | 5.005 | 0.054     | 0.538 |
| 3      | MCO   | BJ     | 89.129         | 0.324     | 7.425 | 0.187     | 0.236 | 90.250  | 0.308     | 3.995 | 0.031     | 0.264 |
|        |       | TS     | 90.414         | 0.300     | 8.426 | 0.181     | 0.300 | 92.388  | 0.296     | 4.407 | 0.038     | 0.428 |
|        |       | CB     | 88.013         | 0.375     | 7.810 | 0.167     | 0.190 | 90.362  | 0.319     | 4.093 | 0.034     | 0.294 |
|        |       | PRR    | 90.267         | 0.330     | 8.360 | 0.178     | 0.296 | 92.693  | 0.270     | 4.412 | 0.037     | 0.432 |
|        | MDA   | CB     | 89.452         | 0.304     | 8.371 | 0.221     | 0.218 | 91.212  | 0.335     | 4.347 | 0.038     | 0.334 |
|        |       | PRR    | 91.195         | 0.269     | 8.897 | 0.220     | 0.316 | 94.020  | 0.264     | 4.833 | 0.043     | 0.574 |
|        |       | P      | 92.083         | 0.257     | 9.428 | 0.216     | 0.374 | 95.484  | 0.216     | 5.469 | 0.067     | 0.686 |
|        |       | P_p    | 90.369         | 0.370     | 9.162 | 0.218     | 0.296 | 94.982  | 0.220     | 5.198 | 0.063     | 0.616 |

Táboa 4.2: Resultados obtidos en base a  $M=500$  series distintas de  $n = 25$  datos, considerando  $R=1000$  valores futuros e  $B=1000$  réplicas bootstrap. Semente: 123456789. Cobertura nominal considerada 95% . Destacado en negro aparece o valor de cobertura media máis próximo ao nivel nominal e en vermello todos aqueles que non son significativamente distintos. Os métodos empregados na obtención dos intervalos de predición foron: Box–Jenkins (BJ), Thombs e Schucany (TS), Bootstrap condicional (CB), Pascual, Romo e Ruíz (PRR) e a nosa proposta (P). A estimación realizouse ben por mínimos cadrados ordinarios (MCO) ou ben mediante mínimas desviacións absolutas (MDA).

|       |          |        | Normal ( $\mu = 0, \sigma = 1$ ) |           |        |           |       | Exponencial ( $\lambda = 1$ ) |           |       |           |       |
|-------|----------|--------|----------------------------------|-----------|--------|-----------|-------|-------------------------------|-----------|-------|-----------|-------|
| Nivel | Estimad. | Método | Cober.                           | Erro est. | Lonx.  | Erro est. | gamma | Cober.                        | Erro est. | Lonx. | Erro est. | gamma |
| 95    | MCO      | BJ     | 92.646                           | 0.168     | 4.712  | 0.024     | 0.268 | 93.176                        | 0.190     | 4.648 | 0.044     | 0.404 |
|       |          | TS     | 93.753                           | 0.168     | 4.974  | 0.027     | 0.422 | 94.893                        | 0.195     | 4.928 | 0.053     | 0.606 |
|       |          | CB     | 92.893                           | 0.179     | 4.812  | 0.026     | 0.340 | 92.603                        | 0.245     | 4.652 | 0.047     | 0.416 |
|       |          | PRR    | 93.883                           | 0.163     | 4.983  | 0.027     | 0.452 | 94.926                        | 0.192     | 4.901 | 0.051     | 0.608 |
|       | MDA      | CB     | 93.249                           | 0.181     | 4.916  | 0.029     | 0.350 | 92.686                        | 0.263     | 4.763 | 0.048     | 0.404 |
|       |          | PRR    | 94.471                           | 0.162     | 5.187  | 0.032     | 0.506 | 94.565                        | 0.218     | 5.024 | 0.050     | 0.586 |
|       |          | P      | 94.994                           | 0.161     | 5.347  | 0.034     | 0.614 | 95.294                        | 0.177     | 5.084 | 0.052     | 0.630 |
|       |          | P_p    | 94.440                           | 0.165     | 5.214  | 0.033     | 0.526 | 88.423                        | 0.351     | 5.029 | 0.052     | 0.236 |
| 99    | MCO      | BJ     | 97.938                           | 0.083     | 6.193  | 0.031     | 0.288 | 96.314                        | 0.123     | 6.108 | 0.058     | 0.116 |
|       |          | TS     | 98.018                           | 0.097     | 6.440  | 0.038     | 0.388 | 98.088                        | 0.098     | 6.397 | 0.071     | 0.426 |
|       |          | CB     | 97.481                           | 0.113     | 6.161  | 0.035     | 0.256 | 97.125                        | 0.142     | 5.929 | 0.063     | 0.286 |
|       |          | PRR    | 98.080                           | 0.094     | 6.450  | 0.037     | 0.376 | 98.062                        | 0.100     | 6.351 | 0.069     | 0.422 |
|       | MDA      | CB     | 97.679                           | 0.107     | 6.344  | 0.042     | 0.320 | 97.144                        | 0.157     | 6.078 | 0.064     | 0.314 |
|       |          | PRR    | 98.329                           | 0.084     | 6.746  | 0.045     | 0.462 | 98.153                        | 0.103     | 6.509 | 0.068     | 0.442 |
|       |          | P      | 98.695                           | 0.075     | 7.136  | 0.051     | 0.574 | 98.420                        | 0.084     | 6.692 | 0.071     | 0.494 |
|       |          | p_p    | 98.323                           | 0.083     | 6.808  | 0.048     | 0.440 | 96.691                        | 0.180     | 6.606 | 0.072     | 0.348 |
|       |          |        | T ( $df = 3$ )                   |           |        |           |       | MW.nm10                       |           |       |           |       |
| Nivel | Esti.    | Método | Cober.                           | Erro est. | Lonx.  | Erro est. | gamma | Cober.                        | Erro est. | Lonx. | Erro est. | gamma |
| 95    | MCO      | BJ     | 92.156                           | 0.189     | 7.629  | 0.103     | 0.252 | 92.739                        | 0.171     | 4.104 | 0.023     | 0.326 |
|       |          | TS     | 92.937                           | 0.179     | 8.716  | 0.148     | 0.348 | 93.541                        | 0.171     | 4.310 | 0.027     | 0.438 |
|       |          | CB     | 92.157                           | 0.195     | 8.488  | 0.143     | 0.284 | 92.477                        | 0.191     | 4.144 | 0.026     | 0.306 |
|       |          | PRR    | 92.969                           | 0.177     | 8.751  | 0.147     | 0.338 | 93.684                        | 0.167     | 4.312 | 0.027     | 0.426 |
|       | MDA      | CB     | 92.613                           | 0.189     | 8.683  | 0.138     | 0.336 | 93.044                        | 0.193     | 4.296 | 0.028     | 0.376 |
|       |          | PRR    | 93.213                           | 0.172     | 8.896  | 0.143     | 0.378 | 94.516                        | 0.165     | 4.565 | 0.030     | 0.546 |
|       |          | P      | 93.361                           | 0.166     | 8.936  | 0.143     | 0.394 | 95.144                        | 0.156     | 4.712 | 0.033     | 0.608 |
|       |          | P_p    | 92.331                           | 0.193     | 8.789  | 0.141     | 0.286 | 94.586                        | 0.167     | 4.611 | 0.033     | 0.544 |
| 99    | MCO      | BJ     | 96.205                           | 0.110     | 10.141 | 0.164     | 0.090 | 97.967                        | 0.075     | 5.400 | 0.030     | 0.264 |
|       |          | TS     | 96.557                           | 0.118     | 11.923 | 0.236     | 0.120 | 97.892                        | 0.091     | 5.593 | 0.039     | 0.316 |
|       |          | CB     | 96.000                           | 0.134     | 11.487 | 0.221     | 0.094 | 97.173                        | 0.112     | 5.325 | 0.038     | 0.210 |
|       |          | PRR    | 96.529                           | 0.118     | 11.920 | 0.242     | 0.118 | 97.890                        | 0.087     | 5.566 | 0.037     | 0.290 |
|       | MDA      | CB     | 96.270                           | 0.127     | 11.767 | 0.223     | 0.102 | 97.463                        | 0.112     | 5.518 | 0.041     | 0.282 |
|       |          | PRR    | 96.792                           | 0.102     | 12.145 | 0.223     | 0.104 | 98.349                        | 0.082     | 5.957 | 0.043     | 0.440 |
|       |          | P      | 96.934                           | 0.097     | 12.242 | 0.214     | 0.118 | 98.735                        | 0.064     | 6.301 | 0.048     | 0.546 |
|       |          | P_p    | 96.324                           | 0.134     | 11.977 | 0.214     | 0.090 | 98.372                        | 0.080     | 5.991 | 0.044     | 0.456 |

Táboa 4.3: Resultados obtidos para horizonte de predición  $k = 3$ , en base a  $M=500$  series distintas de  $n = 50$  datos, considerando  $R=1000$  valores futuros e  $B=1000$  réplicas bootstrap. Semente: 123456789. Coberturas nominais consideradas 95 % e 99 % .



Os resultados obtidos mostran que ao aumentar o tamaño de mostra, como era de agardar, redúcense as diferenzas entre os métodos presentados e, en xeral, mellora a cobertura media que ofrecen. Con todo, a nosa proposta P segue a proporcionar a mellor taxa de cobertura na maior parte dos casos, a excepción de cando o erro segue distribución exponencial para o 95 %, pero nese caso a cobertura que ofrece non é significativamente distinta da que proporciona o PRR en media, que nesas condicións deu o mellor resultado. Por outra parte, para esa mesma distribución, fixado o nivel do 99 % o noso método P ofrece o mellor resultado e unha cobertura significativamente distinta das ofrecidas polos demais métodos.

Noutra orde de cousas, cabe recalcar que, precisamente neste caso de distribución do erro asimétrica, o método  $P_p$  non é consistente, xa que ofrece resultados claramente peores que para tamaño de mostra  $n = 25$ . Isto coincide co que observaron Pascual, Romo e Ruíz (2004) no seu estudo de simulación: é preferible replicar directamente a distribución condicional da predición.

En resumo, o estudo de simulación realizado parece indicar que o método bootstrap proposto neste traballo (P) presenta resultados máis axeitados que outros procedementos presentados na literatura para a obtención de intervalos de predición en modelos autorregresivos. Esta mellora queda reflectida, especialmente, cando o tamaño da serie é pequeno.

Por outra parte, para este método vimos que replicar directamente a distribución condicional da predición ofrece mellores resultados que replicar o erro de predición, procedemento non consistente cando a distribución do erro é asimétrica. Ademais, a estimación mediante MDA parece comportar beneficios en termos de cobertura e erro estándar respecto da estimación por MCO.

Cabería estudar unha extensión deste procedemento, por exemplo, á estimación mediante mínimos cadrados ou outros tipos de M-estimadores, dado que tamén sería posible a inclusión de pesos aleatorios no problema de minimización que permite a obtención dos parámetros. Por outra parte, podería estudarse se é posible estender esta metodoloxía a modelos de series temporais máis complexos que podan ter unha parte de medias móbiles ou, incluso, que non sexan estacionarios.



## Capítulo 5

# Aplicación a datos de desempleo

Nesta última parte do traballo imos ilustrar o funcionamento da metodoloxía presentada nos Capítulos 3 e 4 mediante un conxunto de datos reais. Como é ben coñecido, na práctica é usual que as series temporais relacionadas co ámbito da Economía presenten dinámicas asimétricas. Concretamente, este tipo de comportamento é habitual cando se analizan as taxas de desemprego, dado que o paro tende a aumentar nos períodos de recesión e a diminuír nas épocas de expansión económica, habendo así alteracións locais con respecto á dinámica xeral da serie. Por tal razón, na análise dese tipo de series temporais é preciso dispoñer de modelos axeitados que incorporen este comportamento, como é o caso do modelo QAR.

Precisamente, neste capítulo imos analizar unha serie económica: a taxa de variación intertrimestral do desemprego en España desestacionalizada, é dicir, a diferenza entre a taxa de desemprego dun trimestre con respecto ao mesmo trimestre do ano anterior, corrixida de efectos estacionais e de calendario. Os datos que manexamos abarcan dende o segundo trimestre de 1990 ata o segundo de 2016 e proceden da Enquisa de Poboación Activa (EPA) realizada polo Instituto Nacional de Estadística (INE). O noso obxectivo é establecer un modelo que permita describir de xeito coherente e preciso a evolución da taxa de variación e, así mesmo, realizar predicións axeitadas acerca da mesma, dada a relevancia que ten a taxa de desemprego no devir da economía.

O gráfico secuencial da serie amósase na Figura 5.1. A simple vista o nivel da serie parece constante, polo que é axeitado asumir que foi xerada por un proceso estacionario. Non obstante, pode apreciarse que no primeiro trimestre do ano 2001 prodúcese unha grande caída da taxa de variación que inmediatamente queda contrarrestada por unha posterior subida, polo que aí teríamos un posible dato atípico. Ademais, coincidindo coa recente crise económica mundial, produciuse un ascenso progresivo que acada o máximo no primeiro trimestre do 2009.

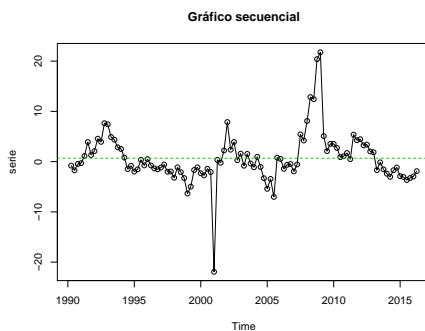


Figura 5.1: Gráfico secuencial da taxa de variación intertrimestral do paro dende o segundo trimestre de 1990 ata o segundo de 2016.

A análise da serie dende o punto de vista clásico, faríanos identificar como modelo tentativo un  $AR(1)$ . A Figura 5.2 amosa a ACF (esquerda) e a PACF (dereita) mostrais da mesma. Como se poderá observar, a ACF mostrais decae a cero exponencialmente, mentres que a PACF pode asumirse que é nula despois do primeiro retardo, características que encaixan cun proceso  $AR(1)$  como xerador da serie:  $Y_t = \phi_0 + \phi_1 Y_{t-1} + a_t$ .

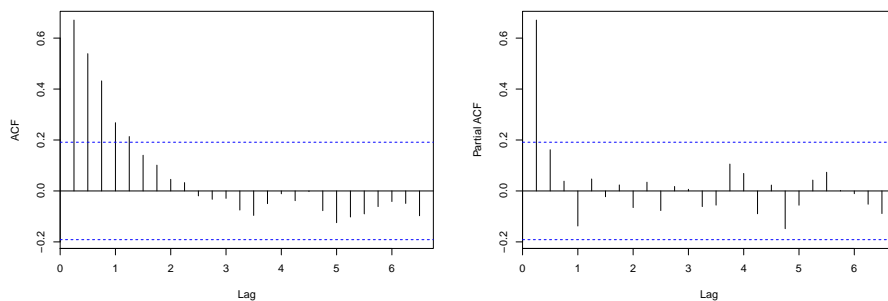


Figura 5.2: Esquerda: gráfico de autocorrelacións simples mostrais; dereita: gráfico de autocorrelacións parciais mostrais para a serie da taxa de variación intertrimestral do paro en España.

Non obstante, se observamos a Figura 5.3, onde amosamos o gráfico de dispersión de  $Y_t$  fronte a  $Y_{t-1}$ , vemos que hai heterocedasticidade: a variabilidade cambia co valor de  $Y_{t-1}$ . Neste gráfico incluíronse as rectas de regresión axustadas por mínimos cadrados e as estimadas para varios cuantís. Como pode observarse as rectas en media e mediana teñan pendentes distintas, como consecuencia da heterocedasticidade e da presenza de datos atípicos. Ademais, tampouco parece asumible a igualdade de pendentes das rectas de regresión correspondentes aos distintos cuantís. Polo tanto, a análise mediante o procedemento clásico non parece o máis razoable.

Unha boa opción para dispoñer dun estudo máis completo da serie parece recorrer ao modelo QAR, dado que este permite que os coeficientes autorregresivos cambien co cuantil.

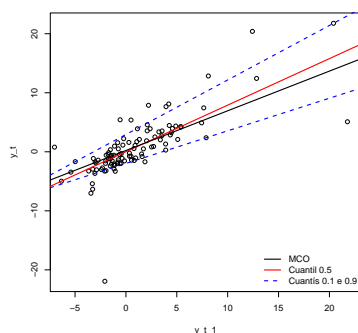


Figura 5.3: Gráfico de dispersión de  $Y_t$  fronte a  $Y_{t-1}$  para a serie de taxa de variación intertrimestral do desemprego en España.

## 5.1. Modelización en base ao QAR

Para estudar o comportamento da serie e ver se efectivamente o valor dos coeficientes autorregresivos depende do cuantil imos centrarnos en cinco deciles: 0.1, 0.3, 0.5, 0.7 e 0.9. Para cada un deles, analizaremos cal é a orde tentativa do modelo QAR aplicando a metodoloxía presentada por Li et al. (2015) baseada na QPACF. Dado que o número de datos é moderado (un total de 105) e como este procedemento está baseado

en teoría asintótica e involucra varias estimacións non paramétricas, imos considerar soamente 10 retardos.

Na Figura 5.4 amosamos os gráficos da QPACF mostral para cada un dos cuantís antes citados, os cales foron obtidos mediante a función `qpacf` incluída no Apéndice B. En vista dos gráficos obtidos, parece que é posible asumir orde 1 para  $\tau = 0.1, 0.3, 0.5, 0.7$ , mentres que para  $\tau = 0.9$  a última autocorrelación parcial cuantil non nula é a cuarta, polo que que o modelo tentativo sería un  $QAR(4)$ . Dada a complexidade dos modelos que en ocasións resultan da análise dos gráficos de autocorrelacións parciais cuantís, Li et al. (2015) recomandan axudarse dalgún procedemento de selección de variables para determinar a orde do modelo. Concretamente, propoñen o uso dos criterios de información AIC (Akaike) e do BIC (Bayes).

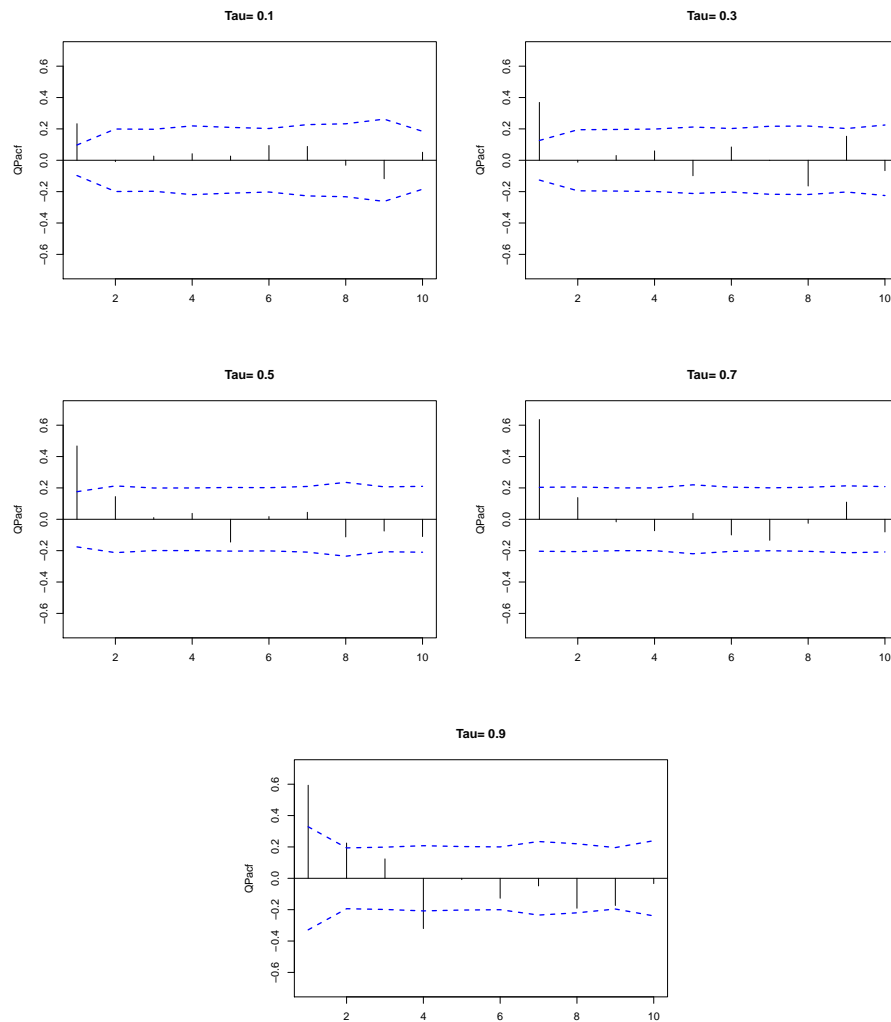


Figura 5.4: Gráficos de autocorrelacións parciais cuantís mostrais para a taxa de variación do desemprego en España.

A Táboa 5.1 amosa os modelos resultantes para os cinco cuantís, logo de realizar a estimación dos coeficientes mediante o procedemento cuantil. Como se poderá observar, ao aumentar  $\tau$  incrementase o valor do coeficiente autorregresivo. Cabe mencionar tamén que no caso de  $\tau = 0.9$  o criterio AIC non nos permite descartar ningunha variable, polo que mantivemos as catro. Polo tanto, vemos que a taxa de variación do paro presenta un comportamento asimétrico, reflectido non só na variación do coeficiente autorregresivo co cuantil, senón tamén co cambio da orde do modelo. Parece que para explicar os cuantís elevados da taxa de variación, aumenta a influencia do trimestre anterior, e esta incluso pode verse lixeiramente afectada por

variacións anteriores (as obtidas con marxe dun ano).

| Cuantil | Orden QPACF | Modelo QAR |                  |                  |                                   |
|---------|-------------|------------|------------------|------------------|-----------------------------------|
| 0.1     | 1           | -1.975     | +0.553 $y_{t-1}$ |                  |                                   |
| 0.3     | 1           | -1.063     | +0.722 $y_{t-1}$ |                  |                                   |
| 0.5     | 1           | 0.018      | +0.788 $y_{t-1}$ |                  |                                   |
| 0.7     | 1           | 1.398      | +0.894 $y_{t-1}$ |                  |                                   |
| 0.9     | 4           | 3.422      | +0.811 $y_{t-1}$ | +0.183 $y_{t-2}$ | +0.065 $y_{t-3}$ -0.160 $y_{t-4}$ |

Táboa 5.1: Orde do modelo segundo os gráficos de autocorrelacións cuantís parciais e modelo axustado para cada cuantil.

## 5.2. Predición

A análise clásica, baseada nas autocorrelacións simples e parciais do proceso, faríanos identificar como modelo tentativo para os nosos datos un  $AR(1)$ . A pesar de que, como vimos, este modelo non parece a elección máis idónea para a serie, imos obter intervalos de predición en base a el, dado que isto nos permite tamén ilustrar o funcionamento da metodoloxía presentada no Capítulo 4, en particular, do novo método bootstrap proposto neste traballo. Por outra parte, dado que descoñecemos a existencia na literatura de métodos de predición en base ao modelo QAR, imos proporcionar un esbozo dun posible método para a obtención de intervalos de predición para dito modelo e ver como funciona nestes datos.

As estimacións do modelo autorregresivo de orde 1 mediante mínimos cadrados ordinarios (MCO) e mínimas desviacións absolutas (MDA) son as seguintes:

$$\hat{y}_{t,MCO} = 0.220 + 0.672y_{t-1},$$

$$\hat{y}_{t,MDA} = 0.018 + 0.788y_{t-1}.$$

En base a estes dous axustes, podemos obter intervalos de predición para horizonte 4 (é dicir, ata o segundo trimestre de 2017, por ser ata onde dispoñemos de datos para realizar a comprobación) mediante todos os procedementos presentados no Capítulo 4.

Na Figura 5.5 representamos a serie histórica da taxa de variación do desemprego, xunto cos 4 valores futuros reais e os intervalos de predición obtidos en base a 10000 réplicas bootstrap (nos procedementos que requiren remostraxe). Como podemos observar, todos os intervalos conteñen o valor futuro verdadeiro da serie. Ademais, en xeral, os intervalos a curto prazo son máis estreitos, mentres que a longo prazo son máis amplos, dado que aumenta a incerteza sobre o comportamento da serie. Para mostrar máis claramente os resultados, na Táboa 5.2 recolleemos o valor dos seus extremos, e tamén o valor futuro verdadeiro da serie. No caso dos intervalos de predición Box–Jenkins, a propia construción leva a que os extremos dos mesmos a partir dun certo horizonte sexan practicamente iguais, dado que este procedemento fai que a predición a partir dun certo instante sexa a media do proceso. Cabe comentar tamén, que ao aplicar o test de normalidade de Shapiro–Wilk e o de Jarque–Bera sobre os residuos dos axustes, rexeitamos a normalidade da innovación do proceso que xerou a serie e, nese tipo de circunstancias, os intervalos de predición Box–Jenkins poden non ser fiables.

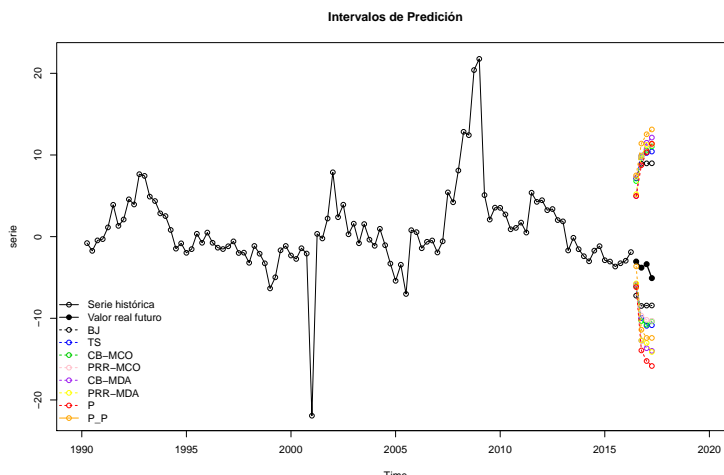


Figura 5.5: Serie histórica de taxa de variación intertrimestral do desemprego en España dende o segundo trimestre de 1990 ata o segundo trimestre de 2016, xunto cos intervalos de predición para os catro anos seguintes mediante os métodos presentados: Box–Jenkins (BJ), Thombs e Schucany (TS), Cao et al. (1997) ou bootstrap condicional mediante MCO (CB–MCO) e MDA (CB–MDA), Pascual, Romo e Ruíz mediante MCO (PRR–MCO) e MDA (PRR–MDA), a nosa proposta (P), e a proposta percentil (P\_P). En trazo negro destácase o valor verdadeiro da serie nesas catro instantes futuros (dende o terceiro trimestre de 2016 ata o segundo de 2017).

Como vimos, a serie que estamos a manexar presenta un comportamento máis complexo do que cabería esperar baixo o modelo autorregresivo clásico, polo que para obter uns resultados máis axeitados sería preciso desenvolver un método de predición baseado no modelo QAR de Koenker e Xiao (2006), o que permitiría que a predición recollese esa información aportada pola serie.

Nestas condicións, dado que descoñecemos a existencia de propostas para a obtención de intervalos de predición para este modelo, suxerimos o seguinte método:

**Paso 1.** Identificamos a orde do proceso *QAR* para unha malla de cuantís apropiada, e axustámolo para cada un deles. Esta podería ser: 0.05, 0.15, 0.25, ..., 0.95. Asumiríamos que nos intervalos comprendidos entre os deciles os coeficientes do modelo *QAR* son constantes e iguais aos do cuantil central do intervalo.

**Paso 2.** Lanzamos unha mostra de pesos aleatorios, verificando as condicións da proposta para a predición do Capítulo 4: variables positivas  $w_t$  con  $t = 1, \dots, n$ , independentes de  $\{Y_t\}_{t \in \mathbb{Z}}$ , tales que  $\mathbb{E}(w_t) = 1$  e  $\mathbb{E}(w_t^2) = 2$ . Por exemplo,  $w_t$  poden ser i.i.d. a unha Exponencial de parámetro  $\lambda = 1$ .

**Paso 3.** Para cada cuantil  $\tau$  da malla, obtemos unha versión bootstrap dos coeficientes da serie introducindo na función obxectivo os pesos aleatorios (os mesmos para todos os cuantís):

$$\hat{\phi}^* = \arg \min_{\phi \in \mathbb{R}^{p+1}} \sum_{t=p+1}^n w_t \rho_{\tau}(Y_t - \phi^{\top} \mathbf{Z}_{t,p}^1), \quad (5.1)$$

onde  $\hat{\phi}^* = (\hat{\phi}_0^*, \hat{\phi}_1^*, \dots, \hat{\phi}_p^*)^{\top}$ ,  $\mathbf{Z}_{t,p}^1 = (1, Y_{t-1}, \dots, Y_{t-p})^{\top}$  e  $\phi = (\phi_0, \phi_1, \dots, \phi_p)^{\top}$ .

**Paso 4.** Baseándonos na expresión (3.1) do modelo de Koenker e Xiao (2006), xeramos unha variable uniforme no intervalo  $(0, 1)$ ,  $U_t$ . O número que resultante é o cuantil  $\tau$ , cuxos coeficientes serven para facer a combinación lineal e obter o valor futuro bootstrap para horizonte 1. Eses coeficientes son os que resultan do paso 3, isto é, a versión bootstrap dos mesmos. É dicir, definimos  $Y_t^* = Y_t$  para cada  $t = n - p + 1, n - p + 2, \dots, n$  e construímos a primeira observación futura usando a ecuación:

$$Y_{n+1}^* = \hat{\phi}_0^*(\tau) + \hat{\phi}_1^*(\tau)Y_n^* + \hat{\phi}_2^*(\tau)Y_{n-1}^* + \dots + \hat{\phi}_p^*(\tau)Y_{n-p+1}^*$$

A partir de  $Y_{n+1}^*$  obtemos  $Y_{n+2}^*$  mediante un procedemento análogo, xerando unha variable uniforme no intervalo  $(0, 1)$  e empregando os coeficientes bootstrap correspondentes ao intervalo ao que pertence o cuantil, así ata o horizonte desexado  $k$ . Deste modo:

$$Y_{n+j}^* = \hat{\phi}_0^*(\tau) + \hat{\phi}_1^*(\tau)Y_{n+j-1}^* + \hat{\phi}_2^*(\tau)Y_{n+j-2}^* + \dots + \hat{\phi}_p^*(\tau)Y_{n+j-p}^* \quad j = 1, 2, \dots, k.$$

Para a obtención dos intervalos de predición, debemos repetir os pasos do 2–4 un número elevado de veces,  $B$ . Así, teriamos un conxunto de futuros valores  $(Y_{n+k}^{*(1)}, Y_{n+k}^{*(2)}, \dots, Y_{n+k}^{*(B)})$ , onde o superíndice denota a remostrá coa que se corresponden, e como fixemos no Capítulo 4, os límites de predición virían dados polos cuantís da estimación por Monte Carlo da distribución bootstrap de  $Y_{n+k}^*$ .

No caso particular dos datos que estamos a manexar, aínda que na sección anterior para  $\tau = 0.9$  a orde que obtiñamos para o  $QAR$  en base á análise da QPACF era 4, parece que un  $QAR(1)$  para todos os cuantís tampouco aporta un mal axuste. De feito, para  $\tau = 0.95$  parece indicado asumir orde  $p = 1$ . Por tal razón, axustamos un  $QAR(1)$  para todos os cuantís da malla.

Empregando este procedemento descrito, obtivéronse os intervalos de predición que presentamos na Figura 5.6 en cor azul cian para horizontes 1, 2, 3 e 4. Como vemos, estes intervalos son máis curtos que os que ofrecen os métodos baseados no modelo  $AR(1)$  e conteñen o valor verdadeiro da predición, polo que este algoritmo sería o que proporciona mellores resultados neste caso. Con todo, debemos interpretar este feito con cautela, xa que se trata dun exemplo concreto e sería preciso probar a validez do procedemento de maneira teórica e vía simulación. Isto podería servir, quizais, para abrir unha liña futura de investigación.

En canto aos resultados obtidos, vemos que a taxa de variación do desemprego actualmente presenta unha dinámica decrecente, o que produce que o valor verdadeiro da predición se sitúe na metade inferior dos intervalos obtidos.

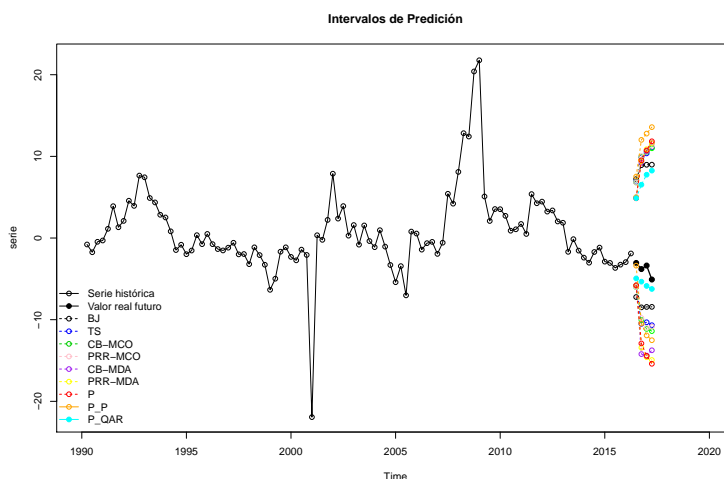


Figura 5.6: Serie histórica de taxa de variación intertrimestral do desemprego en España dende o segundo trimestre de 1990 ata o segundo de 2016, xunto cos intervalos de predición para os catro trimestres seguintes mediante os métodos presentados: Box–Jenkins (BJ), Thombs e Schucany (TS), Cao et al. ou bootstrap condicional mediante MCO (CB–MCO) e MDA (CB–MDA), Pascual, Romo e Ruíz mediante MCO (PRR–MCO) e MDA (PRR–MDA), a nosa proposta (P), e a proposta percentil (P\_P). En color cian inclúense os intervalos de predición obtidos empregando o modelo QAR. En trazo negro destácase o valor verdadeiro da serie nes catro instantes futuros (dende o terceiro trimestre de 2016 ata o segundo trimestre de 2017).



| $k$ | MCO          |                |                |                | MDA            |                |                 |                 |               | Val. real |
|-----|--------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|---------------|-----------|
|     | BJ           | TS             | CB             | PRR            | CB             | PRR            | P               | P_p             | QAR           |           |
| 1   | (-7.23,7.23) | (-6.17,6.87)   | (-6.01,6.81)   | (-5.97, 6.73)  | (-5.82, 5.00)  | (-5.86, 5.13)  | (-5.93,5.02)    | (-3.44 , 7.66)  | (-4.82,4.88)  | -3.05     |
| 2   | (-8.49,8.93) | (-9.91,9.55)   | (-8.94,9.89)   | (-10.21,9.82)  | (-13.55,9.49)  | (-12.87,9.90)  | (-13.53,9.66)   | (-10.62, 12.01) | (-5.30,6.73)  | -3.80     |
| 3   | (-8.44,8.97) | (-10.29,10.15) | (-10.35,10.67) | (-10.97,10.41) | (-14.17,10.77) | (-14.22,10.57) | (-15.07, 10.92) | (-11.95,12.83)  | (-5.85, 7.78) | -3.37     |
| 4   | (-8.43,8.99) | (-10.37,10.64) | (-10.85,10.81) | (-10.64,10.99) | (-14.13,11.67) | (-14.94,10.82) | (-15.75,12.00 ) | (-12.30,13.89)  | (-6.09,8.67)  | -5.08     |

Táboa 5.2: Intervalos de predición para horizontes 1, 2, 3 e 4 (dende o terceiro de 2016 ata o segundo trimestre de 2017) mediante os métodos presentados: Box–Jenkins (BJ), Thombs e Schucany (TS), Cao et al. (1997) ou bootstrap condicional por MCO (CB–MCO) e MDA (CB–MDA), Pascual, Romo e Ruíz mediante MCO (PRR–MCO) e MDA (PRR–MDA), a nosa proposta (P) e a proposta percentil (P\_p). A columna QAR acolle os intervalos resultantes do algoritmo que esbozamos, baseado no modelo de Koenker e Xiao (2006).

Cabe comentar que para que a taxa de desemprego, cuxo gráfico secuencial amosamos na Figura 5.7, volva ao nivel que tiña previamente á crise económica é preciso que se sigan a producir caídas similares na taxa de variación intertrimestral aos incrementos que se produciron.

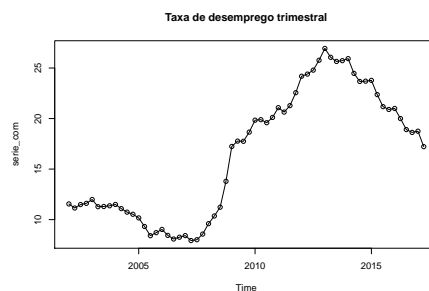


Figura 5.7: Serie histórica de taxa de desemprego en España dende o primeiro trimestre do ano 2002 ata o segundo de 2017.



# Conclusiones e temas abertos

Ao longo deste documento, abordouse a análise de procesos autorregresivos estacionarios dende o punto de vista da regresión cuantil. Esta análise seguiu dúas liñas diferentes:

- Por unha parte, para a autorregresión clásica estudamos o problema da predición, revisamos distintos algoritmos bootstrap propostos na literatura para a obtención de intervalos de predición e vimos as vantaxes que comporta o emprego de estimadores de mínimas desviacións absolutas en canto á obtención de taxas de cobertura apropiadas. O estudo de todos estes procedementos permitiunos incluír neste traballo unha nova proposta de bootstrap neste contexto, do que garantimos a súa consistencia teórica e do que comparamos o seu comportamento práctico con respecto aos outros procedementos presentados mediante un estudo de simulación. Os resultados de dito estudo mostráronnos as melloras que comporta o noso método en canto a cobertura, especialmente en series de tamaño pequeno.
- Por outra parte, revisamos o proceso cuantil autorregresivo (QAR), proposto por Koenker e Xiao (2006), que pretende aproveitar no terreo da autorregresión as vantaxes que a estimación cuantil comporta na regresión baixo heterocedasticidade. Para este modelo vimos as propiedades que presenta en canto a estacionariedade e expuxemos o modo de estimar os seus parámetros na práctica. Ademais, presentamos un modo de identificar a súa orde, proposto por Li et al. (2015), que se basea no estudo da función de autocorrelacións parciais cuantís. O descoñecemento de funcións do programa R Core Team (2015) que permitisen facer esta identificación de xeito automático, levounos a crear unha propia para poder empregar esta metodoloxía na práctica.

Desenvolver un método bootstrap para a obtención de intervalos de predición no modelo QAR é un tema totalmente aberto na actualidade, pois descoñecemos a existencia de métodos para predición neste modelo. A extensión do novo procedemento bootstrap incluído no Capítulo 4 a este contexto máis xeral do modelo QAR, tal e como esbozamos na aplicación a datos, intuímos que podería proporcionar bos resultados; en calquera caso, sería preciso perfilar algúns aspectos (a maneira de replicar valores futuros no proceso, o número de cuantís para os que é conveniente estimar ou para cales en concreto se debería facer) e estudar de xeito teórico e vía simulación o seu funcionamento, o que podería derivar nun traballo futuro de investigación.



## Apéndice A

# Función de autocorrelacións parciais: cálculos

Neste capítulo do apéndice incluímos as contas precisas para obter unha expresión explícita para a función de autocorrelacións parciais (PACF) dun proceso estocástico estacionario. O seu cálculo ten interese por dous motivos:

1. Presentar unha extensión ao terreo cuantil da PACF. Dita extensión foi proposta por Li et al. (2015), e pode aplicarse na identificación da orde do modelo de Koenker e Xiao (2006).
2. Permite xustificar que no algoritmo para a predición de Thombs e Schucany (1990) (véxase Subsección 4.2.1), os coeficientes da representación cara atrás do modelo autorregresivo sexan os mesmos que os da representación cara adiante.

Para presentar a expresión explícita da PACF, supoñamos sen perda de xeneralidade que  $\mathbb{E}(Y_t) = 0$ . Se denotamos por  $\hat{Y}_{t+k}$  o predictor lineal de  $Y_{t+k}$  baseado en  $Y_{t+k-1}, Y_{t+k-2}, \dots, Y_{t-k+1}$ , daquela:

$$\hat{Y}_{t+k} = \alpha_{01}Y_{t+k-1} + \alpha_{02}Y_{t+k-2} + \dots + \alpha_{0k-1}Y_{t+1},$$

onde  $\alpha_{0i}$  con  $1 \leq i \leq k-1$  son a solución do problema de mínimos cadrados, resultado da minimización da seguinte esperanza

$$\mathbb{E} \left[ \left( Y_{t+k} - \hat{Y}_{t+k} \right)^2 \right] = \mathbb{E} \left[ \left( Y_{t+k} - \alpha_1 Y_{t+k-1} - \alpha_2 Y_{t+k-2} - \dots - \alpha_{k-1} Y_{t+1} \right)^2 \right].$$

Para resolver este problema, seguindo o procedemento usual, fariáanse os cálculos pertinentes na expresión anterior, derivaríase con respecto a cada un dos coeficientes e igualaríase a cero a ecuación resultante, obténdose o seguinte sistema de ecuacións:

$$\gamma_i = \alpha_1 \gamma_{i-1} + \alpha_2 \gamma_{i-2} + \dots + \alpha_{k-1} \gamma_{i-k+1}, \quad 1 \leq i \leq k-1.$$

Entón, ao dividir cada ecuación do sistema por  $\gamma_0$ , poderíamos escribir, de xeito equivalente, que

$$\rho_i = \alpha_1 \rho_{i-1} + \alpha_2 \rho_{i-2} + \dots + \alpha_{k-1} \rho_{i-k+1}, \quad 1 \leq i \leq k-1.$$

e, en forma matricial (usando a simetría da función de correlacións con respecto á orixe) teríamos que

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \dots & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{k-1} \end{pmatrix}$$

Igualmente, denotando por  $\hat{Y}_t$  o mellor predictor lineal de  $Y_t$  baseado en  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$  teriamos que

$$\hat{Y}_t = \beta_{01}Y_{t+k-1} + \beta_{02}Y_{t+k-2} + \dots + \beta_{0k-1}Y_{t+1},$$

onde, novamente,  $\beta_{0i}$  con  $1 \leq i \leq k-1$  serían as estimacións de MCO dos parámetros do modelo, que se obterían mediante a minimización da esperanza

$$\mathbb{E} \left[ (Y_t - \hat{Y}_t)^2 \right] = \mathbb{E} \left[ (Y_t - \beta_1 Y_{t+k-1} - \beta_2 Y_{t+k-2} - \dots - \beta_{k-1} Y_{t+1})^2 \right].$$

Aplicando o procedemento antes descrito, chegaríamos ao sistema

$$\rho_i = \beta_1 \rho_{i-1} + \beta_2 \rho_{i-2} + \dots + \beta_{k-1} \rho_{i-k+1}, \quad 1 \leq i \leq k-1.$$

que en forma matricial viría dado por

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \dots & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-1} \end{pmatrix},$$

o que implicaría que  $\alpha_{0i} = \beta_{0i}$  para  $1 \leq i \leq k-1$ , pois ambos conxuntos de coeficientes son solución do mesmo sistema de ecuacións (que é compatible determinado).

Daquela, a autocorrelación parcial entre  $Y_t$  e  $Y_{t+k}$  sería igual á correlación ordinaria entre os residuos da regresión de  $Y_t$  sobre  $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$  e os residuos da regresión de  $Y_{t+k}$  sobre as mesmas variables  $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$ , é dicir, se denotamos por  $\phi_{kk}$  a correlación parcial

$$\phi_{kk} = \frac{\text{cov}[(Y_t - \hat{Y}_t), (Y_{t+k} - \hat{Y}_{t+k})]}{\sqrt{\text{Var}(Y_t - \hat{Y}_t)} \sqrt{\text{Var}(Y_{t+k} - \hat{Y}_{t+k})}}. \quad (\text{A.1})$$

Esta expresión aínda pode concretarse máis, notando que

$$\begin{aligned} \text{Var}(Y_{t+k} - \hat{Y}_{t+k}) &= \mathbb{E} \left[ (Y_{t+k} - \alpha_{01}Y_{t+k-1} - \alpha_{02}Y_{t+k-2} - \dots - \alpha_{0k-1}Y_{t+1})^2 \right] \\ &= \mathbb{E} [Y_{t+k} (Y_{t+k} - \alpha_{01}Y_{t+k-1} - \alpha_{02}Y_{t+k-2} - \dots - \alpha_{0k-1}Y_{t+1})] \\ &\quad - \alpha_1 \mathbb{E} [Y_{t+k-1} (Y_{t+k} - \alpha_{01}Y_{t+k-1} - \alpha_{02}Y_{t+k-2} - \dots - \alpha_{0k-1}Y_{t+1})] \\ &\quad \dots - \alpha_{k-1} \mathbb{E} [Y_{t+1} (Y_{t+k} - \alpha_{01}Y_{t+k-1} - \alpha_{02}Y_{t+k-2} - \dots - \alpha_{0k-1}Y_{t+1})] \\ &= \mathbb{E} [Y_{t+k} (Y_{t+k} - \alpha_{01}Y_{t+k-1} - \alpha_{02}Y_{t+k-2} - \dots - \alpha_{0k-1}Y_{t+1})], \end{aligned}$$

posto que  $\mathbb{E}(Y_t) = 0$  e que as variables  $Y_{t+k-i}$  con  $i = 1, \dots, k-1$  van ser independentes do resultado de eliminar de  $Y_{t+k}$  a influencia lineal delas mesmas. Entón, usando a asunción de estacionaridade

$$\text{Var}(Y_{t+k} - \hat{Y}_{t+k}) = \text{Var}(Y_t - \hat{Y}_t) = \gamma_0 - \alpha_1 \gamma_1 - \dots - \alpha_{k-1} \gamma_{k-1}. \quad (\text{A.2})$$

Agora notando que  $\alpha_i = \beta_i$  para  $1 \leq i \leq k-1$ , usando argumentos similares chegamos a

$$\begin{aligned} \text{Cov} (Y_t - \hat{Y}_t, Y_{t+k} - \hat{Y}_{t+k}) &= \mathbb{E} [(Y_t - \alpha_{01}Y_{t+1} - \alpha_{02}Y_{t+2} - \dots - \alpha_{0k-1}Y_{t+k-1}) (Y_{t+k} - \alpha_{01}Y_{t+k-1} - \alpha_{02}Y_{t+k-2} - \dots - \alpha_{0k-1}Y_{t+1})] \\ &= \mathbb{E} [(Y_t - \alpha_{01}Y_{t+1} - \alpha_{02}Y_{t+2} - \dots - \alpha_{0k-1}Y_{t+k-1}) Y_{t+k}] \\ &= \gamma_k - \alpha_1 \gamma_{k-1} - \dots - \alpha_{k-1} \gamma_1. \end{aligned}$$

Así, temos que

$$\phi_{kk} = \frac{\gamma_k - \alpha_1 \gamma_{k-1} - \dots - \alpha_{k-1} \gamma_1}{\gamma_0 - \alpha_1 \gamma_1 - \dots - \alpha_{k-1} \gamma_{k-1}}, \quad (\text{A.3})$$

ou, de xeito equivalente, dividindo numerador e denominador entre  $\gamma_0$ ,

$$\phi_{kk} = \frac{\rho_k - \alpha_1 \rho_{k-1} - \dots - \alpha_{k-1} \rho_1}{1 - \alpha_1 \rho_1 - \dots - \alpha_{k-1} \rho_{k-1}}. \quad (\text{A.4})$$

Así, pola regra de Cramer é posible obter unha expresión para a PACF que non dependa do valor de cada  $\alpha_i$ , simplemente, substituíndo a columna  $i$ -ésima por  $(\rho_1, \dots, \rho_{k-1})$  como segue

$$\alpha_i = \frac{\begin{vmatrix} 1 & \rho_1 & \dots & \rho_{i-2} & \rho_1 & \rho_i & \dots & \rho_{k-2} \\ \rho_1 & 1 & \dots & \rho_{i-3} & \rho_2 & \rho_{i-1} & \dots & \rho_{k-3} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \dots & \rho_{k-i} & \rho_{k-1} & \rho_{k-i-2} & \dots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \dots & \rho_{i-2} & \rho_{i-1} & \rho_i & \dots & \rho_{k-2} \\ \rho_1 & 1 & \dots & \rho_{i-3} & \rho_{i-2} & \rho_i & \dots & \rho_{k-3} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \dots & \rho_{k-i} & \rho_{k-i-1} & \rho_{k-i-2} & \dots & 1 \end{vmatrix}}. \quad (\text{A.5})$$

Substituíndo os valores de  $\alpha_i$  en (A.4), e multiplicando numerador e denominador de (A.4) polo determinante

$$\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \dots & 1 \end{vmatrix}, \quad (\text{A.6})$$

obtense a seguinte expresión para a PACF:

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & 1 \end{vmatrix}}. \quad (\text{A.7})$$





## Apéndice B

# Código QPACF mostral

Neste Apéndice inclúese a función empregada na identificación da orde do modelo *QAR*, a cal devolve un gráfico coas autocorrelacións parciais cuantís mostrais (QPACF mostral) e bandas de confianza para elas ao 95%. Para empregala deben introducírselle como argumentos:

- A serie observada, *x*.
- O valor do cuantil para o que estamos a identificar a orde do modelo, *tau*.
- O retardo máximo para o que se calculan as autocorrelacións parciais *lag\_max*.

```
qpacf=function(x,tau,lag_max){

#Función auxiliar cálculo psi
psi<-function(tau,w){
a=tau-ifelse(w<0,1,0)
return(a)
}

#Función que coloca as x segundo o valor de k das autocorrelacións parciais
lag<-function(x,k,tau){
n<-length(x)
x_lag=matrix(0,ncol=k-1,nrow=n-k)
if (k>1){
for(i in 1:(n-k)){
x_lag[i,1:(k-1)]=x[(i+1):(i+k-1)]
}
x_lag=x_lag[, (k-1):1]
}
else{
y=x[(k+1):n]
x_lag=rep(sort(y)[floor((n-k)*tau)],n-k)
}
return(x_lag)
}

#Función que obtén o valor de ft_1(0):

spar=function(tau,k,x){
```

---

```

tol=1/(10^6)
n=length(x)

x_lag2=matrix(0,ncol=k,nrow=n-k)
for(i in 1:(n-k)){
x_lag2[i,1:k]=x[i:(i+k-1)]
}
x_k=x_lag2[,k:1]
y=x[(k+1):n]

t=tau
q=qnorm(t)
f4=(1/sqrt(2*pi))^4*exp(-2*(q^2));f4
h_B=0.6*(n)^(-1/5)*((4.5*f4)/(2*q^2+1)^2)^(1/5);h_B

tau1=t+h_B
tau2=t-h_B
mod_qh=rq(y~x_k,tau=tau1);mod_qh
mod_q_h=rq(y~x_k,tau=tau2);mod_q_h

#Cálculo ft_1(0):

qh=mod_qh$fitted.values;qh
q_h=mod_q_h$fitted.values;q_h
f_0=numeric(n-k)
for(i in 1:(n-k)){
f_0[i]=max(0,2*h_B/(qh[i]-q_h[i]-tol))
}
return(f_0)
}

n=length(x)
IC_sup=numeric(lag_max)
IC_inf=numeric(lag_max)
om_3=numeric(lag_max)
phi_kk=numeric(lag_max)

#Para k=1, qpacf=qacf, debemos calculalo a parte.

k=1
q_y=lag(x,k,tau);q_y
q_y=as.numeric(q_y)
X=x[1:(n-k)]
y_t=x[(k+1):n]

muX=mean(X)
ei_tk=X-muX;ei_tk
ei_t=y_t-q_y;ei_t
s_yz_2=1/(n)*sum(ei_tk^2);s_yz_2
ps=psi(tau,ei_t);ps
phi_kk[1]=1/sqrt((tau-tau^2)*s_yz_2)*1/(n)*(ps%*%ei_tk)

```

---

```

h=dpill(y_t,X)
loc=locpoly(y_t,X,bandwidth=h,degree=0)

points<-loc$x
q_tau_y=sort(y_t)[floor((n-k)*tau)]
points_new<-abs(q_tau_y-points)
c_loc<-which(points_new==min(points_new))
c<-points[c_loc]
y_loc<-which(loc$x==c)
m_tau<-loc$y[y_loc]

nX=length(X);nX
sigma2X=1/(n)*sum((X-muX)^2);sigma2X
q_cov=1/(n)*(ps%*%ei_tk);q_cov

si11=1/(n)*sum((X-muX)^4)-sigma2X^2;si11
si12=1/(n)*sum((ps*(X-m_tau))^2)-(q_cov)^2;si12
si13=1/(n)*sum(ps*(X-m_tau)*(X-muX)^2)-sigma2X*q_cov ;si13

om_3[1]=1/(tau-tau^2)*((si11*(q_cov)^2)/(4*sigma2X^6)-(si13*q_cov)/sigma2X^4+(si12)/sigma2X)
z=qnorm(0.975)
IC_sup[1]=z*sqrt(om_3[1]/(n));IC_sup
IC_inf[1]=-z*sqrt(om_3[1]/(n));IC_inf

#Valores das autocorrelaci3ns para distintos valores de k>1

for (l in 2:lag_max){
k=l
z_tk=lag(x,k,tau);z_tk

#Modelo ols para y_t-k

y_tk=x[1:(n-k)]
mod_tk=lm(y_tk~z_tk)
ei_tk=mod_tk$residuals;ei_tk
s_yz_2=1/(n)*sum(ei_tk^2)

#Modelo qar para y_t

y_t=x[(k+1):n];y_t
modt=rq(y_t~z_tk,tau=tau)
ei_t=modt$residuals;ei_t

ps=psi(tau,ei_t);ps

#C3lculo autocorrelaci3ns
phi_kk[k]=1/sqrt((tau-tau^2)*s_yz_2)*1/(n)*(ps%*%y_tk)

```

```

#####
## 2-Obtención bandas de confianza para a QPACF mostral ##
#####

###Obtención varianza:

z_tk_as=cbind(rep(1,n-k),z_tk); z_tk_as
z_tk_as=t(z_tk_as)
A_0=matrix(0,nrow=k,ncol=1)
for(i in 1:(n-k)){
A_0=A_0+(y_tk[i]*t(z_tk_as)[i,])/(n)
}

#Necesitamos estimar a sparsity f_t-1(0)
#para iso debemos escoller h:
#ventás de Koenker e Xiao (2006)

#Nivel de significación fixado
alpha=0.05

f_t1_0=spar(tau,k,x);f_t1_0
A_1=matrix(0,nrow=k,ncol=1)
for(i in 1:(n-k)){
A_1=A_1+(f_t1_0[i]*y_tk[i]*t(z_tk_as)[i,])/(n)
}

sig_30=(z_tk_as%*%t(z_tk_as))/(n)
sig_30

matriz=matrix(0,nrow=k,ncol=k)
for(i in 1:(n-k)){
matriz=matriz+(f_t1_0[i]*outer(z_tk_as[,i],t(z_tk_as)[i,],"*"))/(n)
}
sig_31=matriz

sig_32=mean(y_t^2)-2*t(A_1)%*%solve(sig_31)%*%A_0+
t(A_1)%*%solve(sig_31)%*%sig_30%*%solve(sig_31)%*%A_1

res_tk=mod_tk$residuals
e1=psi(tau,ei_t)
om_3[k]=(mean(e1^2)*sig_32)/(sqrt((tau-tau^2)^2)*mean(res_tk^2))
z=qnorm(1-alpha/2)
IC_sup[k]=z*sqrt(om_3[k]/(n));IC_sup
IC_inf[k]=-z*sqrt(om_3[k]/(n));IC_inf
}

gr=plot(phi_kk,type="h",ylim=c(-0.7,0.7),xlab="",ylab="QPacF",main=paste("Tau=",tau))
l1=abline(h=0)
l2=lines(IC_sup,type="l",lty=2,col=4,lwd=2)
l3=lines(IC_inf,type="l",lty=2,col=4,lwd=2)

return(list(gr,l1,l2,l3,"f_t1_0"=f_t1_0,"phi_kk"=phi_kk,

```

---

```
"omega3"=om_3,"IC_sup"=IC_sup,"IC_inf"=IC_inf))  
}
```



# Bibliografía

- [1] Barrodale, I. and Roberts, F. D. K. (1973). An improved algorithm for discrete l1 linear approximation. *SIAM Journal of Numerical Analysis*, Vol. 10, 839-848.
- [2] Billingsley, P. (1995). *Probability and Measure*. Wiley & Sons, New York.
- [3] Bloomfield, P. and Steiger, W. S. (1983). *Least Absolute Deviations. Theory, Applications and Algorithms*. Birkhäuser.
- [4] Bofinger, E. (1975). Estimation of a density function using order statistics. *Australian Journal of Statistics*, **17**, 1-7.
- [5] Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. San Francisco, Holden-Day.
- [6] Breidt, F. J., Davis, R. A. and Dunsmuir, W. T. M. (1995). Improved bootstrap prediction intervals for autoregressions. *Journal of Time Series Analysis*. Vol 16, **2**, 177-200.
- [7] Brockwell, J. and Davis R. A. (1993). *Time Series: Theory and Methods*. Springer-Verlag, New York.
- [8] Cao, R., Febrero-Bande, M., González-Manteiga, W., Prada-Sánchez, J.M. and García-Jurado, I. (1997). Saving computer time in constructing consistent bootstrap prediction intervals for autoregressive processes. *Communications in Statistics-Simulation and Computation*, **26**, 961-978.
- [9] Davino, C., Furno, M. and Vistocco, D. (2014). *Quantile Regression: Theory and Applications*. Wiley & Sons.
- [10] Freedman, D. A. (1985). On bootstrapping two-stage least-squares estimates in stationary linear models. *The Annals of Statistics*, **9**, 827-842.
- [11] Hall, P. and Sheather, S. (1988). On the distribution of a studentized quantile. *Journal of the Royal Statistical Society. Ser B*, **5**, 381-391.
- [12] Hendricks, W., and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity, *Journal of the American Statistical Association*, **87**, 58-68.
- [13] Koenker R., (2005). *Quantile Regression*. Cambridge University Press.
- [14] Koenker R. e Xiao, (2006). Quantile autoregression. *Journal of the American Statistical Association*, **101**, 980-990.
- [15] Kreiss, J. P. and Franke, J. (1992). Bootstrapping stationary autoregressive moving-average models. *Journal of Time Series Analysis*, Vol. 13, 4, 297-316.
- [16] Mächler, M. (2014). nor1mix: Normal (1-d) Mixture Models (S3 Classes and Methods). R package version 1.2-0.
- [17] Nicholls, D. F. and Quinn, B. G. (1982). *Random Coefficient Autoregressive Models: An Introduction*, Springer-Verlag, Berlin.

- 
- [18] Pascual, L., Romo, J. and Ruiz, E. (2001). Effects of parameter estimation on prediction densities: a bootstrap approach. *International Journal of Forecasting*, **17**, 83–103.
- [19] Pascual, L., Romo, J. and Ruiz, E. (2004). Bootstrap predictive inference for ARIMA processes. *Journal of Time Series Analysis*, 25, 4, 449–465.
- [20] R Core Team (2015), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- [21] Rao, C. R. and Zhao, L. C. (1992). Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap. *Sankhya: The Indian Journal of Statistics, Series A*, Vol. 54, 3, 323–331.
- [22] Stine, R. A. (1987). Estimating properties of autoregressive forecasts. *Journal of the American Statistical Association*, **82**, 1072–1078.
- [23] Thombs, L. A. and Schucany, W. R. (1990). Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association* Vol. **85**, 410, 486–492.
- [24] Wei, W. W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Addison Wesley, Boston.