Sofia Martinho de Almeida Costa

Master in Mathematics and Applications

Cross-sectional modeling of bank deposits

Dissertação para obtenção do Grau de Mestre em

Matemática e Aplicações

Orientador: Pedro Maria Corte Real Alarcão Júdice,
Diretor, Banco Montepio,
Investigador Associado, ISCTE

Co-orientador: Pedro José dos Santos Palhinhas Mota,
Professor Auxiliar, Faculdade de Ciências e
Tecnologia da Universidade Nova de Lisboa

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

September, 2019

Cross-sectional modeling of bank deposits

Este documento foi gerado utilizando o processador (pdf)LaTeX, com base no template "novathesis" [1] desenvolvido no Dep. Informática da FCT-NOVA [2].
[1] https://github.com/joaomlourenco/novathesis [2] http://www.di.fct.unl.pt

To my family.

Acknowledgements

In the end of this journey, I would like to thank the people that helped to make this possible. They were the major structure that guided me throughout my doubts and aspirations.

I would like to express my deepest gratitude to my advisers, Professor Pedro Júdice and Professor Pedro Mota, for their limitless support and patience, as well as for their incredible availability to answer to my questions and to give feedback.

I would also like to profoundly appreciate Professor Marta Faias for her incredible support since I began my Master's Degree, even when I was still an external student in this School, and for all her amazing availability.

My gratitude also goes to all my Professors that helped me throughout my academical journey, without whom I wouldn't have been able to develop my mathematical skills so well.

To my colleagues that shared their time and work with me while exchanging ideas and tips for studying and researching.

My sincere appreciation to my family, that supported me in all circumstances and struggled for me to have no obstacles in this process.

To my parents, for their major support in my whole life and for doing the possible and the impossible for me to succeed.

To Inês, for being a great companion and for helping me to reach my goals.

To Joana, for being the main inspiration in my academic path by showing me what it is to strive and be determined while helping others.

To Ana Teresa and Hugo Pirraça, for their amazing support, patience and friendship, as also for sharing important moments of our lives together.

To Vítor Augusto, for giving me incredible opportunities and for his genuine friendship.

To the elements of Núcleo de Jogos de Estratégia e Sociedade from Faculdade de Ciências e Tecnologia, from Universidade Nova de Lisboa, for their heartwarming welcoming.

To Teresa Brissos, for her friendship and great advise.

To Diogo Mendes, for his infinite support, friendship and caring.

To Circo Matemático, the most amazing project that I have ever been involved in, for the experiences that it provided and the people that I met.

Abstract

The dynamics of liquidity risk is an important issue in what concerns banks' activity. It can be approached by studying the evolution of banks' clients deposits in order to mitigate the probability of bankruptcy and to efficiently manage banks' resources. A sound liquidity risk model is also an important component of any liquidity stress testing methodology.

In this research[1], we aim to develop a model that can help banks to properly manage their activity, by explaining the evolution of clients deposits throughout time. For this purpose, we considered the momentum, a frequently used tool in finance that helps to clarify observed trends. Therefore, we obtained an AR(2) model that was then used to simulate trajectories, through the use of the *R* software, for possible evolutions of the deposits.

Another feature that we pondered was panel data. By considering different banks in our sample, the simulations would generate varied trajectories, including both good and bad scenarios, which is useful for stress testing purposes. The mostly referred model in the literature is the AR(1) model with only one time series, which often does not generate distress episodes.

In order to validate our model we had to perform several tests, including to the normality and autocorrelation of the residuals of our model. Furthermore, we considered the most used model in the literature for comparison with two different individual banks. We simulated trajectories for all cases and evaluated them through the use of indicators such as the Maximum Drawdown and density plots.

When simulating trajectories for banks' deposits, the panel data model gives more realistic scenarios, including episodes of financial distress, showing much higher drawdowns and density plots that present a wide range of possible values, corresponding to booms and financial crises. Therefore, our methodology is more suitable for planning the management of banks' resources, as well as for conducting liquidity stress tests.

Keywords: Liquidity risk; Non-maturity deposits; Banks' activity management; Autoregressive models; Panel data modeling; Maximum Drawdown; *R* software; Computational simulation.

[1] This research was developed with public data and is independent from Montepio, so it does not reflect the views of this institution.

Resumo

A dinâmica do risco de liquidez é um assunto importante no que diz respeito à atividade bancária. Pode ser abordada através do estudo da evolução dos depósitos bancários dos clientes de forma a mitigar a probabilidade de falência dos bancos e a gerir os recursos dos mesmo de forma eficiente. Um modelo de risco de liquidez sensato é também uma componente importante de qualquer metodologia de *stress testing* de liquidez.

Neste trabalho[2], o nosso objetivo é desenvolver um modelo que possa ajudar os bancos a gerir a sua atividade de forma apropriada, explicando a evolução dos depósitos dos seus clientes ao longo do tempo. Com este propósito, considerámos o *momentum*, uma ferramenta frequentemente utilizada na área financeira que contribui para clarificar tendências observadas. Assim, obtivemos um modelo AR(2) que foi seguidamente utilizado para simular trajetórias, através do uso do *R software*, de possíveis evoluções dos depósitos.

Outra característica que ponderámos foi usar dados em painel. Ao considerar diferentes bancos na nossa amostra, as simulações originariam trajetórias variadas, incluindo tanto cenários bons como cenários maus, o que é útil ao testar cenários de *stress*. O modelo mais referido na literatura é o modelo AR(1) com apenas uma série temporal, o que frequentemente não gera episódios de crise nas simulações.

De forma a validar o nosso modelo, realizámos diversos testes, incluindo à normalidade e à autocorrelação dos resíduos do nosso modelo. Além disso, considerámos o modelo mais utilizado na literatura para fins de comparação aplicando-o a dois bancos de forma individual. Simulámos trajetórias para todos os casos e avaliámo-las através do uso de indicadores tais como o *Maximum Drawdown* e gráficos de densidade de probabilidade.

Ao simular trajetórias para os depósitos bancários, o modelo de dados em painel apresenta cenários mais realistas, incluindo episódios de dificuldade financeira, através de quedas muito mais acentuadas e gráficos de densidade de probabilidade que apresentam uma grande variedade de valores possíveis, correspondendo tanto a períodos de prosperidade como a crises financeiras. Concluindo, a nossa metodologia é mais apropriada para a gestão dos recursos bancários, assim como para executar o *stress testing* de liquidez.

[2] A investigação aqui apresentada é realizada com dados públicos e é independente da atividade no Banco Montepio, não refletindo as visões desta instituição.

Palavras-chave: Risco de liquidez; Depósitos sem maturidade; Gestão da atividade bancária; Modelos autorregressivos; Modelação de dados em painel; *Maximum Drawdown*; *R software*; Simulação Computacional.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The study of the dynamics of deposit volumes is a very important exercise for two main reasons: first, it is a critical tool for liquidity stress testing purposes; second, it should be used in the development of asset-liability optimization frameworks such as described in Birge and Júdice (2013). Thus, it is necessary to study the liquidity risk, which can be represented by the non-maturity deposits held by clients. Such non-maturity accounts have already been researched by many authors, namely with an AR(1) model with normal residuals, and even including some exogenous variables, such as the market interest rate. However, given that the models are usually only calibrated to a single bank, the models do not incorporate episodes of stress and suffer from survivorship bias.

In a context of bank balance sheet management, both Hałaj (2016) and Lipton (2015) consider an AR(1) model but, when it comes to its estimation, they only give examples of values of the parameters, since they don't have a formal sample for the calibration.

While studying the arbitrage-free valuation and theoretical hedging of deposits, Jarrow and van Deventer (1998) considered an AR(1) model, which was put in practice by Janosi, Jarrow, and Zullo (1999). They tested data of four different types of accounts from the Federal Reserve Bulletin, which lead them to some remarkable values of the R^2, but with at least one non-significant variable in each account, either the trend or the market interest rate. This model was also tested by Benbachir and Hamzi (2016), while comparing it with ARMA models, treating separately deposits from individuals and enterprises held at a Moroccan commercial bank. The AIC (Akaike Information Criterion) suggested an ARMA$(2,2)$ model for individuals and an ARMA$(1,2)$ model for enterprises.

O'Brien (2000) also considered an AR(1) model, including exogenous variables such as the GDP and the market interest rate, and even used data of two types of accounts from

99 different banks. The predictions showed that the deposits in the NOW (Negotiable Order of Withdrawal) accounts would rise with time, while the MMDA's (Money Market Deposit Accounts) would fall. Even though the sample is comprised of many banks, the model only calibrates one bank deposit series. Consequently, such a model cannot take into account episodes of surviving banks and failed banks simultaneously.

Fu and Feng (1985) used an MA(1) model in their research, applied to the second differences of $log(X_t)$, with X_t representing the savings deposits of urban and rural residents. Their prediction was that the deposits would rise with time, which meets the trend of their sample.

Our research has lead us to an AR(2) panel data model, which, unlike previous research, is simultaneously calibrated to a sample of 9 banks, with different number of observations. Moreover, the two lags in the AR(2) model allow us to consider the presence of the momentum, a frequently used indicator in finance. Actually, the trend of the dependent variable revealed to be a relevant indicator of the evolution of clients deposits.

The 9 banks were selected according to the criterion of having an average of at least 10 billion euros in deposits throughout their time series. We proceeded to its estimation, in the *R* software, by using the OLS (Ordinary Least Squares) method, and simulated different possible paths for the deposits in the following 30 years. The results show that there is an increasing trend in every case, with the particularity of simulating some "bank runs", which is due to the panel data sample.

By considering 9 banks simultaneously, liquidity risk isn't underestimated. That is, the model won't discard the possibility of financial crises, as it happens in samples with only one time series. For example, if one bank follows a model that considered a sample of only one time series, not including episodes of stress, it might invest in only a few liquid assets. This situation would be acceptable according to that model, because it wouldn't be able to simulate any crisis, but there's a chance of something going wrong. In that case, that bank would most likely face financial distress because it wouldn't be prepared for such an occurrence, ending up without enough resources to face that situation.

In order to compare our results, we chose two individual banks from our sample. We estimated an AR(1) and an AR(2) models for those and we excluded the lattest because its second coefficient was not relevant at a level of significance of 5% for both banks. Therefore, by considering the AR(1) model for the comparison, we tested both banks for normality and autocorrelation in the residuals. We verified the same results for the two cases: the residuals seemed to follow a Normal distribution and not to be autocorrelated.

Thus, we simulated trajectories for the chosen banks with the referred assumptions and we only obtained optimistic results, since the worst case scenario would be an almost plain path after a very significant growth. Furthermore, the obtained maximum drawdowns were significantly different from the ones that resulted from the panel data model. The latter would be much greater, indicating severe falls of the deposits, representing periods of crisis. This also represents the unrealistic optimism in the individual banks simulations, derived from the use of only a single time series for estimation purposes.

Since we use panel data, our model avoids survivorship bias, as it takes into account events of bankruptcy and acquisitions. Thus, we obtain more realistic simulations because these include bad scenarios as well. Therefore, our model is also more adequate for stress testing.

We now present the structure of this work: Chapter 2 describes the Autoregressive and Panel Data models, with an introduction regarding the Multiple Linear Regression models, followed by their estimators and respective properties. The residuals of the models as well as the tests performed to validate our model are also studied in this chapter. Chapter 3 presents an overall description of the data used and how it was considered in our research. Furthermore, it is in this chapter that the calibration of the model is developed. Chapter 4 discusses the computational simulations of the deposits paths obtained with our model while comparing them to the most frequently considered model. In Chapter 5 some final remarks are presented, concluding our work.

Chapter 2

Autoregressive Processes and Panel Data Modeling

In this chapter we will introduce the models, their parameters' estimators and their properties that were useful to our research. We will also explain different types of data because the goal of our research, of building a model for the deposits evolution by using data from several banks, uses panel data and needs some understanding of what precedes this organization of the collected sample.

2.1 Multiple Linear Regression

2.1.1 Model

A multiple linear regression consists in a model in which there is a dependent variable that is explained by a linear combination of some independent variables (also called explanatory variables or regressors). The dependent variable is what is intended to be studied while the independent variables are the characteristics of the phenomenon we consider that better reach our goal. That is, in order to explain and simulate the dependent variable, we collect data on the explanatory variables.

As Ribeiro (2014) shows, a multiple linear regression model may be presented in equation form:

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \varepsilon_t \tag{2.1}$$

where: y_t represents the t^{th} observation of the dependent variable y; x_{tj} represents the t^{th} observation of the independent variables x_j, with $t = 1, ..., T$ denoting the different observations of each variable in the sample and $j = 1, ..., k$ the number of explanatory variables;

β_j are the coefficients of the corresponding independent variables, the parameters to be estimated; ε_t are the error components. Denoting the error component, ε_t represents the difference between y_t and the sum of $x_{tj}\beta_j$, in each time period, where β_j is not observable. These variables, when estimated, are called the residuals of the model. If we want an independent term in the model we just have to consider $x_{t1} = 1$, obtaining:

$$y_t = \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \varepsilon_t. \tag{2.2}$$

This model can also be represented in matrix form:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & \dots & x_{1k} \\ 1 & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{T2} & \dots & x_{Tk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

as indicated by the following representation:

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon} \tag{2.3}$$

where: $\mathbf{Y}$ denotes the $T \times 1$ vector of the dependent variables' observations; $\mathbf{X}$ denotes the $T \times k$ independent variables' observations; β represents the $k \times 1$ vector of the independent variables' coefficients; $\boldsymbol{\varepsilon}$ represents the $T \times 1$ vector of the residual variables.

There are some contexts in which we only want to study the evolution of a specific phenomenon or feature over time, so we will only consider its lagged values as explanatory variables. This case will be explained in Section 2.2.

According to Hansen (2000), the following standard assumptions are those that allow estimators with good properties:

1. The sample is i.i.d. (independent and identically distributed);
2. $\mathbb{E}[\varepsilon_i | x_i] = 0$;
3. $\mathbb{E}[\varepsilon_i^2 | x_i] = \sigma^2$;
4. $Q_{xx} = \mathbb{E}[x'_{tj} x_{tj}] > 0$ is invertible;
5. $Cov(\varepsilon_t, \varepsilon_s | \mathbf{X}) = 0$, with $t, s = 1, ..., T$, $t \neq s$.

These assumptions mean the following:

1. Independency means that, in the same sample, the actions of each individual or firm (that is being the target of our research) don't have an impact on the other individuals or firms' actions. Also, if the sample is identically distributed, the obtained observations from one individual or firm follow the same distribution as the observations from the other individuals or firms;

2. There is strict exogeneity, which means that the explanatory variables aren't correlated with the residuals of the model;

3. The model is homoskedastic, which implies that the conditioned variances don't depend on the explanatory variables; that is, they are identical across the individual observations;

4. There is no exact multicollinearity allowed and so the regressors can't be perfectly correlated with each other. This restricts them not to be linear combinations of each other;

5. There is an absence of autocorrelation in the residuals. This condition is particularly important in order to use the Ordinary Least Squares (OLS) method to estimate the coefficients of the variables of the model.

As it is possible to notice, only one out of the five assumptions doesn't include the variable ε_t. This variable, the error component, is very important because it influences the consistency of the estimates of the explanatory variables' coefficients.

2.1.2 Ordinary Least Squares

2.1.2.1 Estimator

When estimating the parameters of a model, we want the residuals to be as small as possible because it means that the estimates of the coefficients are as close as possible to the observed values. Therefore, a way of achieving this goal is to minimize the sum of the squared residuals, which brings us to the OLS (Ordinary Least Squares) estimator.

This is the usual estimator of the parameters that, given the previous assumptions for the multiple linear regression model, is intended to minimize the following expression:

$$\psi(\beta) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \tag{2.4}$$

with respect to $\boldsymbol{\beta}$. This expression has the following components: $\mathbf{Y}$ as the vector of the dependent variables; $\mathbf{X}$ as the matrix of the explanatory variables; $\boldsymbol{\beta}$ as the vector of the coefficients; $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'$ as the transposed matrix of $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$. Through differential calculus, we may obtain the OLS estimator, $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}). \tag{2.5}$$

2.1.2.2 Properties of the Estimator

Given the previous assumptions, Hansen (2000) states that the OLS estimator has the following properties:

1. $\mathbb{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] = \boldsymbol{\beta}$
2. $Var(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$

which mean that:

1. The OLS estimator is unbiased when it is conditioned by the regressors;
2. The estimator's conditional variance-covariance matrix depends on the variance of the regression error, which, considering homoskedasticity, is constant and represented by σ^2. Here, an absence of autocorrelation among the residuals is also evident.

2.1.2.3 Residuals

While studying linear regression models it is important to take the residuals into consideration. For example, the fact that they follow some distribution or not may alter the properties of the estimators of the regression coefficients. Thus, in this context, Hansen (2000) refers that the residuals must have the following properties:

1. $\mathbb{E}[\hat{\varepsilon}_t|\mathbf{X}] = 0$
2. $Var[\hat{\varepsilon}_t|\mathbf{X}] = \mathbf{M}\sigma^2$

1. We can get to this conclusion by using the error's orthogonal projection, the $\mathbf{M}$ matrix from Property 2: $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, because $\hat{\varepsilon}_t = \mathbf{M}\varepsilon_t$;
2. In addition to using the error's orthogonal projection, we have to consider the variance of the regression error, as in the second property of the OLS estimator, σ^2. This result is due to homoskedasticity.

When calculating the residuals, as referred before, we need to have estimates of the parameters of the model. Hence, when computing these quantities, we are also recording the adjusted values of the dependent variable, which allow us to obtain the coefficient of determination.

2.1.2.4 Coefficient of Determination

According to Ribeiro (2014), the coefficient of determination of a model, represented by R^2, is a quotient calculated with two different resources: the adjusted and the observed values of the dependent variable of the model. By considering the squared sum of both of these quantities, the R^2 measures the ability of the independent variables to explain the dependent variable. This may be used as a tool to evaluate the quality of our model and to thus help us understand whether it is reasonable to simulate future values of the variable in study with the adjusted model.

However, this instrument may guide us to misleading conclusions since its value increases every time a new explanatory variable is added. A possible solution for this issue is the adjusted coefficient of determination, $\bar{R}^2$, which considers the residuals of the model, again using their squared sum, and adapts the quantities by considering the number of observations and the number of coefficients. Consequently, there is a compensation if the sum of squared residuals gets smaller when adding a new independent variable.

2.1.3 Normality Tests

Since the normality among the residuals would allow us to make more accurate simulations (see Chapter 4) and to better characterize the estimators' distributions, we needed to test these components. Thus, we will briefly introduce the Shapiro-Wilk and the Jarque-Bera tests for normality.

2.1.3.1 Shapiro-Wilk

This test has been created because distributional assumptions are important when working with statistics since they help to perform inference, even though these might not always be reached. Therefore, according to Shapiro and Wilk (1965), it tests whether a sample follows a Normal distribution or not, under the composite null hypothesis of normality, by comparing its order statistics with the ones from a Normal distribution.

Therefore, to start with, we must consider that a normal sample w_i may be represented as follows:

$$w_i = \mu + \sigma z_i \tag{2.6}$$

with: μ as the mean of the sample and σ as its standard deviation; z_i representing a random variable that follows a Normal distribution with mean equal to 0 and variance equal to 1; $i = 1, ..., n$ representing the number of observations from the sample. Both μ and σ are unknown, but for our case we just need the estimator for σ, $\hat{\sigma}$, which, as cited by Shapiro and Wilk (1965), is the following:

$$\hat{\sigma} = \frac{\mathbf{m}'\mathbf{V}^{-1}(\mathbf{m}\mathbf{1}' - \mathbf{1}\mathbf{m}')\mathbf{V}^{-1}\mathbf{w}}{\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}\mathbf{m}'\mathbf{V}^{-1}\mathbf{m} - (\mathbf{1}'\mathbf{V}^{-1}\mathbf{m})^2} \tag{2.7}$$

with:

- $\mathbf{1}'$ as a row vector of ones;
- $\mathbf{m}' = (m_1, m_2, ..., m_n)$ representing a vector with each element given by $\mathbb{E}[z]_i = m_i$;
- $\mathbf{V}$ representing the variance-covariance matrix of the vector m'.

Hence, we have the following test statistic:

$$W = \frac{R^4\hat{\sigma}^2}{C^2S^2} \tag{2.8}$$

where:

- $R^2 = \mathbf{m}'\mathbf{V}^{-1}\mathbf{m}$;
- $C^2 = \mathbf{m}'\mathbf{V}^{-1}\mathbf{V}^{-1}\mathbf{m}$;
- $S^2 = \sum_{i=1}^{n}(w_i - \bar{w})^2$.

and the null-hypothesis should be rejected if the p-value is smaller than the chosen significance level.

This test was later adapted to bigger samples by Royston (1982) and it is easily implemented in the *R-software* with a function that comes in the initial packages, *shapiro.test()*. This version has a simpler test statistic, defined as follows:

$$W'_n = \frac{\left(\sum_{i=1}^{n} b_i w_i\right)^2}{\sum_{i=1}^{n}(w_i - \bar{w})^2} \tag{2.9}$$

with

$$\mathbf{b}^T = (b_1, b_2, ..., b_n) = \frac{\mathbf{m}^T}{(\mathbf{m}^T\mathbf{m})^{\frac{1}{2}}}, \tag{2.10}$$

which makes the test statistic simpler because it doesn't include the variance-covariance matrix anymore. However, it can be applied to the sample because it has similar power compared to the original version of the test.

2.1.3.2 Jarque-Bera

This test, by Jarque and Bera (1987), specializes in the verification of the normality of the residuals and uses the Lagrange multiplier method to test the null hypothesis of normality. Also, it compares the skewness and kurtosis of the sample data to the ones of the Normal distribution. Finally, this method originates an asymptotically efficient test that is computationally feasible because it only needs the first four sample moments of the ordinary least-squares residuals.

Since this test compares the values of the skewness and kurtosis, these are present in the test statistic. Thus, if the respective values are too far from the ones of the Normal distribution, the calculations reject the null hypothesis of normality.

In order to apply this test, we need to consider the regression model presented in equation (2.1) with some assumptions about the error component, as it has to:

- be i.i.d;
- be homoskedastic;
- have zero mean.

Thus, we have that the test statistic, JB_n, follows a χ^2 distribution with 2 degrees of freedom and is given by:

$$JB_n = n\left(\frac{b_1^2}{6} + \frac{(b_2-3)^2}{24}\right) \tag{2.11}$$

where

$$b_1 = \frac{\hat{\mu}_3}{(\hat{\mu}_2)^{\frac{3}{2}}} \text{ and } b_2 = \frac{\hat{\mu}_4}{(\hat{\mu}_2)^2} \tag{2.12}$$

estimate the sample's skewness and kurtosis using the sample moments of the ordinary least-squares residuals, calculated as follows:

$$\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^j,\ j = 2,3,4. \tag{2.13}$$

This test is also easily implemented in the *R-software* with a single function, *jb.norm.test()*, even though a specific package is required.

2.2 Autoregressive Processes

The data must follow a chronological order so that we can study autoregressive models. Here, the dependent variable is in the time period that we want to explain and we need information from previous time periods to make simulations. Thus, stochastic processes have an important role here. Since, according to J. E. P. Box, Jenkins, and Reinsel (1970),

these are sequences of random variables that evolve according to probability laws, we may consider a realization of them: time series.

2.2.1 Time Series

According to Hansen (1994) and Ribeiro (2014), we can say that we are in the presence of a time series when our data refer to the same cross-section unit (an individual, an entity, etc.) for many periods of time, with its observations associated to a chronological order. That is, we get a sequence of observations regarding the same cross-section unit. This feature allows us to make simulations based on events that happened in the past.

In summary, in time series we have:

- One cross-section unit;
- A sequence of time periods.

2.2.2 Model

As referred in Section 2.1.1, an AR(k) (Autoregressive of order k) model implies that the dependent variable's behaviour is justified by its own observations in the previous k time periods. Including an independent term and an error component we have:

$$y_t = \alpha + \beta_1 y_{t-1} + ... + \beta_k y_{t-k} + \varepsilon_t, \tag{2.14}$$

where: y_t represents the dependent variable; α represents the independent term, a constant; $y_{t-1},...,y_{t-k}$ represent the dependent variable's previous k observations; β_j represent the variables' coefficients, the parameters to be estimated; ε_t represents the error component; $j = 1,...,k$ denotes the number of considered previous time periods in the model; $t = 1,...,T$ denotes the time period whose observed values we want to explain, which implies that $t > k$.

There's one more factor to ponder in this context: the error component must be a white noise, which, according to Ribeiro (2014), implies that its:

- expected value is zero;
- variance doesn't depend on time;
- covariance only depends on the considered lags, and not on time itself.

The model's autocorrelation and partial autocorrelation functions, ACF and PACF respectively, are useful to identify the most appropriate order of an AR model. These will be approached in the following section.

2.2.3 Autocorrelation

2.2.3.1 Autocorrelation Function

In order to obtain the ACF, J. E. P. Box et al. (1970) start with the definition of autocovariance at lag k:

$$\gamma_k = cov[z_t, z_{t+k}] = \mathbb{E}[(z_t - \mu)(z_{t+k} - \mu)] \tag{2.15}$$

where: z_t and z_{t+k} are observed values, separated by k time periods, from a time series; μ is the mean of the stochastic process $\{z_t\}$ (assumed to have a constant mean). With this information, the authors explain that we need to consider the autocorrelation at lag k, which is given by:

$$\rho_k = \frac{\gamma_k}{\sqrt{\mathbb{E}[(z_t - \mu)^2]\,\mathbb{E}[(z_{t+k} - \mu)^2]}} \tag{2.16}$$

where the variables and parameters mean the same as in equation (2.15). Here, there's an important concept highlighted by Hansen (1994) that can have an impact on this formula: it is considered that a stochastic process is strictly stationary if the joint distributions of any two elements z_t and z_{t+k}, with k being the number of time periods that separates the observations, doesn't depend on t but on k. Plus, if a process that satisfies this condition has finite second moments, then it is also considered weakly stationary, which means that their mean and autocovariances don't depend on the time period t. That is, a stochastic process is weakly stationary if its mean and the covariance between any two of its components only depend on their relative positions in the succession and not on their absolute positions. Thus, in the presence of a stationary process, its variance is the same in every time period, including both times t and $t+k$. Therefore, we obtain the following expression for the autocorrelation at lag k:

$$\rho_k = \frac{\gamma_k}{\sigma_z^2} = \frac{\gamma_k}{\gamma_0} \tag{2.17}$$

where σ_z^2 equals the variance of the process $\{z_t\}$.

Once having these data, we just need to plot ρ_k against k. By considering only the positive lags, we get the desired ACF, which is the same as analyzing the upper triangular part of the autocorrelation matrix.

The Autocorrelation Function of an AR process should geometrically converge to zero as the lag variable grows. For an autoregressive process to be stationary, some restrictions on the coefficients of the model must be met. We'll limit our research to two different cases:

- The AR(1) model: $y_t = \alpha + \beta_1 y_{t-1} + \varepsilon_t$, where $|\beta_1| < 1$;

- The AR(2) model: $y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t$, where:
  - $\beta_1 + \beta_2 < 1$;
  - $\beta_2 - \beta_1 < 1$;
  - $-1 < \beta_2 < 1$.

Depending on the values of the coefficients, we get two possible scenarios for the ACF function, for instance in the AR(1) case:

- If β_1 is positive, the ACF graphic decays to zero;
- If β_1 is negative, the ACF graphic only decays to zero in modulus; that is, the graphic will be alternating between positive and negative values on each turn, each time being smaller to the previous one in modulus.

2.2.3.2 Partial Autocorrelation Function

As shown by J. E. P. Box et al. (1970), the Partial Autocorrelation Function is the last coefficient in an autoregressive representation of order k. That is, in the models:

$$y_j = \beta_{k1} y_{j-1} + ... + \beta_{k(k-1)} y_{j-k+1} + \beta_{kk} y_{j-k},\ j = 1, 2, ..., k \tag{2.18}$$

where each β_{kj} represents β_j in the respective AR(k) model, the Partial Autocorrelation Function is represented by β_{kk}.

For an AR(k) process, the PACF cuts the effect of the correlation between variables that are separated by a difference of at least k time periods. This turns the values of the function into zero for all the higher lags and characterizes the most appropriate lag order for the considered model.

2.2.3.3 Ljung-Box Test for Autocorrelation

This test was developed by G. Box and Ljung (1978) as an improved version of the Box-Pierce test for Autocorrelation (see G. Box and Pierce (1970)) and it contributes to the diagnostic checking of a model, originally proposed by J. E. P. Box et al. (1970). It aims to evaluate the goodness of fit of a stationary AR(k) model by assessing its sample's residuals' autocorrelations, verifying whether these are equal to zero or not. Therefore, its null hypothesis corresponds to the absence of autocorrelation among the residuals, which means that these are i.i.d..

The authors begin by considering a stationary process, $\{w_t\}$, defined by an AR(k) model, as well as its sequence of errors, $\{a_j\}$ which are i.i.d.. These follow a Normal distribution with zero mean and constant variance equal to σ^2.

The estimated autocorrelations, since the values of the residuals are used, are calculated by:

$$\hat{r}_p = \frac{\sum_{t=p+1}^{n} \hat{a}_t \hat{a}_{t-p}}{\sum_{t=1}^{n} \hat{a}_t^2}, \; p = 1, 2, ... \tag{2.19}$$

where: $t = 1, ..., n$ is the number of observations of the time series; $p = 1, 2, ...$ is the considered lag separating the variables.

With these data, the authors suggest the following test statistic:

$$\tilde{Q}(\hat{r}) = n(n+2) \sum_{p=1}^{m} (n-p)^{-1} \hat{r}_p^2 \tag{2.20}$$

with m representing the first sample autocorrelations and being small in comparison to n. This quantity approximates a χ^2 distribution with $m - k$ degrees of freedom, only exceeding its variance. The test statistic shown in equation (2.20) has better results compared to the one originally proposed by G. Box and Pierce (1970) because the latter is an approximation of 2.20, which biases its distribution.

This test has another advantage: there's statistical evidence that the non-normality of the residuals doesn't affect the results. The errors only need to have finite variance in order to ensure the referred results.

2.3 Panel Data

Since our research focuses on panel data, there are two more types of data that we must introduce and are clarified in Ribeiro (2014): cross-sectional data and panel data.

2.3.1 Cross-sectional Data

Cross-sectional data differ in two aspects when compared to time series: cross-sectional data do not consider only one cross-section unit; in different periods of time there may be observations for several cross-section units. That is, in cross-sectional data the data collection isn't necessarily loyal to the previously stored observations. Therefore, while in time series we know that we are always referring to the same individual or the same firm, here that's not likely to happen. In fact, in different time periods we may have different cross-section units. Hence, we have observations from many cross-section units and also a chronological order.

In summary, in cross-sectional data we have:

- Many cross-section units;

- A sequence of time periods;
- The hypothesis of having different cross-section units in different periods of time.

2.3.2 Panel Data

The goal of our research is to study the evolution of clients deposits for several banks simultaneously and to be able to simulate those deposits evolution through the use of the momentum. Since that can be done by means of panel data, we must develop this concept before clarifying the respective model. We first described the concepts of time series and of cross-sectional data because panel data is a combination of both, as explained below.

Panel data consists on a data structure that includes both cross-section units (that can be individuals, firms, etc.) and sequenced time. Essentially, the aim of a study with a sample of this nature is to observe the evolution of certain attributes of many entities over time in order to make better simulations in comparison to only one time series. For instance, in the case of a small sample, the amount of observations in time is compensated for having more individuals or firms with recorded observations. Hence, in this context, the cross-section units must remain the same (there can't be a swap, as an opposition to cross-sectional data) and the observations for each one of them must be chronologically organized. Moreover, there may be the case in which the number of observations of at least one cross-section unit is different from the remaining. When this happens, we can say that we have an unbalanced panel.

In summary, in panel data we have:

- Many cross-section units;
- A sequence of time periods;
- The same cross-section units over time.

2.3.3 Model

We will work with the basic linear panel data model used in econometrics that Wooldridge (2002) describes by the following equation, for each $i = 1,...,n$:

$$y_{it} = \beta_{1i}x_{1it} + ... + \beta_{ki}x_{kit} + \varepsilon_{it}, \quad t = 1,...,T \tag{2.21}$$

where: y_{it} is the dependent variable; x_{jit}, with $j = 1,2,...,k$ and $t = 1,2,...,T$, are the independent variables; β_{ji}, with $j = 1,2,...,k$, are the linear model coefficients; $j = 1,...,k$ is the number of independent variables; $i = 1,...,n$ is the individual index; $t = 1,...,T$ is the time index; ε_{it}, with $t = 1,2,...,T$, is the error component, a random disturbance term. We can also represent this model with a different notation, where the model in the population is:

$$y_t = \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \tag{2.22}$$

where we observe the same variables, for all the considered T time periods, but for each cross-section unit i. Here: $\mathbf{x}_t$ is a $1 \times k$ vector, for all t; $\boldsymbol{\beta}$ is a $k \times 1$ vector; both y_t and ε_t are scalars. $\mathbf{x}_t$ can be represented as shown below:

$$\mathbf{x}_t = [x_{t1}\ x_{t2}\ \ldots\ x_{tk}]. \tag{2.23}$$

If we want to refer to an equation that specifically targets a cross-section unit i, during a concrete time period t, we should represent equation (2.22) with the i subscript as well:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \varepsilon_{it} \tag{2.24}$$

with

$$\mathbf{x}_{it} = [x_{1it}\ x_{2it}\ \ldots\ x_{kit}]. \tag{2.25}$$

Finally, this model also has a matrix representation:

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} x_{1i1} & x_{2i1} & \ldots & x_{ki1} \\ x_{1i2} & x_{2i2} & \ldots & x_{ki2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1iT} & x_{2iT} & \ldots & x_{kiT} \end{bmatrix} \begin{bmatrix} \beta_{1i} \\ \beta_{2i} \\ \vdots \\ \beta_{ki} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}$$

defined by the following equation:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \tag{2.26}$$

where: $\mathbf{Y}_i$ is a $T \times 1$ vector; $\mathbf{X}_i$ is the $T \times k$ matrix $\mathbf{X}_i = (\mathbf{x}'_{i1}, \mathbf{x}'_{i2}, \ldots, \mathbf{x}'_{iT})'$; $\boldsymbol{\beta}$ is a $k \times 1$ vector; $\boldsymbol{\varepsilon_i}$ is a $T \times 1$ vector that, as it will be explained in Section 2.3.5, may have its own components:

$$\boldsymbol{\varepsilon}_i = c_i\mathbf{A} + \mathbf{u}_i \tag{2.27}$$

that are: c_i, the unobserved effect, as a constant for each cross-section unit i; $\mathbf{A}$ as a column vector of dimension T of ones; $\mathbf{u}_i$ as a $T \times 1$ vector that contains the idiosyncratic errors.

2.3.4 Ordinary Least Squares in a Panel Data Context

For the OLS estimates of $\boldsymbol{\beta}$ to be consistent in a Panel Data context, Wooldridge (2002) describes the assumptions that had to be verified:

1. $E[\mathbf{x}'_t\varepsilon_t] = \mathbf{0}$
2. $rank(\sum_{t=1}^{T} E[\mathbf{x}'_t\mathbf{x}_t]) = k$

3. a) $E[\varepsilon_t^2 \mathbf{x}_t' \mathbf{x}_t] = \sigma^2 E[\mathbf{x}_t' \mathbf{x}_t]$ where $E[\varepsilon_t^2] = \sigma^2$

   b) $E[\varepsilon_t \varepsilon_s \mathbf{x}_t' \mathbf{x}_s] = \mathbf{0}$, $s \neq t$(with $\mathbf{0}$ as a null vector)

considering that:

1. For the orthogonality condition to hold, it's only necessary to check if the explanatory variables are correlated or not to the residual variable. In case they are not, the condition is satisfied because:

$$E[\mathbf{x}_t' \varepsilon_t] = E[E[\mathbf{x}_t' \varepsilon_t | \mathbf{x}_t']] = E[\mathbf{x}_t' E[\varepsilon_t | \mathbf{x}_t']] = 0.$$

   Here, it is still important to notice that both the explanatory and the residual variables are referring to the same time period, so we have no information about the relation between $\mathbf{x}_t$ and ε_s, with $s \neq t$.

2. No variable is a linear combination of any other variable. Here, k represents the number of explanatory variables.

3. This condition includes two topics:

   a) Homoskedasticity, with an explanation similar to the first property, but regarding the conditional variance:

$$E[\mathbf{x}_t' \mathbf{x}_t \varepsilon_t^2] = E[E[\mathbf{x}_t' \mathbf{x}_t \varepsilon_t^2 | \mathbf{x}_t' \mathbf{x}_t]] = E[\mathbf{x}_t' \mathbf{x}_t E[\varepsilon_t^2 | \mathbf{x}_t' \mathbf{x}_t]] = E[\mathbf{x}_t' \mathbf{x}_t \sigma_t^2] = \sigma^2 E[\mathbf{x}_t' \mathbf{x}_t];$$

   b) Absence of correlation between the errors and the variables in different time periods. For this condition to hold, it is only necessary that $E[\varepsilon_t \varepsilon_s | \mathbf{x}_t \mathbf{x}_s] = 0$.

2.3.5 Unobserved Effects

Considering that there are unobserved effects, Wooldridge (2002) presents the following model:

$$y_{it} = \mathbf{x}_{it} \boldsymbol{\beta} + c_i + u_{it} \tag{2.28}$$

where c_i is the unobserved effects component and u_{it} denotes the idiosyncratic error component. These, as indicated in Section 2.3.3, are the elements that make $\varepsilon_{it} = c_i + u_{it}$ the composite error. Here, the idiosyncratic error must be i.i.d. and have zero as its expected value and σ_u^2 as its variance.

The unobserved effects highlight the specific characteristics of each cross-section unit: they add another reason to justify why panel data models are useful. Indeed, more data is relevant to get more information and avoid problems, such as trends, meaning that if we have a sample of only one individual, then we cannot simulate behaviours that are different from the ones that that individual has. For example, if we are studying

the behaviour of a bank that is really successful and has a time series of its deposits continuously growing, we cannot expect a model using that information to simulate a fall in any point of time.

2.3.6 Estimation Methods

In panel data models it is usual to consider the residual variable as a sum of two components: the idiosyncratic error, that was denoted by u_{it} in equation (2.28), and a more specific component usually referred to as the unobserved effects, denoted by c_i in the same equation. The latter is an unobserved variable that only changes according to each cross-section unit and not according to time. Since it is constant in time, it is commonly stated as a "fixed effect". It used to be considered that the unobserved effects could either be constant (fixed effects) or a random variable (random effects) but, most recently, this approach has turned into the study of existence of correlation between the unobserved effects and the independent variables. Depending on the results, the unobserved effects can either be considered fixed effects or random effects.

2.3.7 Random Effects

While studying random effects, we consider that the unobserved effects are not correlated with the explanatory variables. In this case, there are more restrictions to take into consideration compared to the OLS estimator presented in the basic linear panel data model. For example, we need to add strict exogeneity to the orthogonality condition. Thus, the necessary assumptions are the following:

1. a) $\mathbb{E}[u_{it}|\mathbf{x}_i, c_i] = 0, t = 1, ..., T$
   b) $\mathbb{E}[c_i|\mathbf{x}_i] = \mathbb{E}[c_i] = 0$
2. $rank(\mathbb{E}[\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i]) = k$ with $\mathbf{\Omega} = \mathbb{E}[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i']$
3. a) $\mathbb{E}[\mathbf{u}_i\mathbf{u}_i'|\mathbf{x}_i, c_i] = \sigma_u^2\mathbf{I}_T$
   b) $\mathbb{E}[c_i^2|\mathbf{x}_i] = \sigma_c^2$

which mean that:

1. a) The independent variables are uncorrelated with the idiosyncratic error in any considered time period due to strict exogeneity;
   b) There is orthogonality between the independent variables and the unobserved effects. This specific condition is useful to obtain the asymptotic variance of the random effects estimator in a GLS (Generalized Least Squares) context;
2. A rank condition implies that the estimator is consistent in the GLS method. Here, it is assumed that the variance-covariance matrix of the composite error conditional on the independent variables, $\mathbf{\Omega}$, is constant;

3. a) The idiosyncratic errors are uncorrelated with each other over time and have constant variances;

   b) The unobserved effects are homoskedastic.

In order to calculate this estimator, we must focus on Assumption 2., where we consider the following matrix for the variances and covariances of ε_i, which has the structure presented in equation (2.27):

$$\mathbb{E}[\varepsilon_i \varepsilon_i'] = \begin{pmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \dots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & \dots & \sigma_c^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_c^2 & \sigma_c^2 & \dots & \sigma_c^2 + \sigma_u^2 \end{pmatrix}.$$

Therefore, it is possible to notice that $\mathbb{E}[\varepsilon_i \varepsilon_i'] = \sigma_u^2 \mathbf{I}_T + \sigma_c^2 \mathbf{B}$, with $\mathbf{B}$ being a $T \times T$ matrix of ones. So, considering $\hat{\sigma}_u^2$ and $\hat{\sigma}_c^2$ as consistent estimators of σ_u^2 and σ_c^2, respectively, we get the consistent estimator of the whole matrix: $\hat{\mathbf{\Omega}} = \hat{\sigma}_u^2 \mathbf{I}_T + \hat{\sigma}_c^2 \mathbf{B}$. Thus, we obtain the random effects estimator:

$$\hat{\beta} = \left(\sum_{i=1}^{n} \mathbf{X}_i' \hat{\mathbf{\Omega}}^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^{n} \mathbf{X}_i' \hat{\mathbf{\Omega}}^{-1} \mathbf{y}_i \right). \tag{2.29}$$

If the assumptions of this model are verified, this estimator is efficient, as well as asymptoticaly equivalent to the GLS case.

2.3.8 Fixed Effects

Now considering that the unobserved effects, which are constant, are correlated with the explanatory variables, we have simpler assumptions:

1. $\mathbb{E}[u_{it} | \mathbf{x}_i, c_i] = 0$
2. $rank(\sum_{t=1}^{T} \mathbb{E}[\ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it}]) = rank(\mathbb{E}[\ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i]) = k$
3. $\mathbb{E}[\mathbf{u}_i \mathbf{u}_i' | \mathbf{x}_i, c_i] = \sigma_u^2 \mathbf{I}_T$

in which we must consider that:

1. This assumption is equal to Assumption 1.a) from the random effects model. It dismisses Assumption 1.b) because, in the fixed effects context, there is no need to restrict $\mathbb{E}[c_i | x_i]$. Hence, this analysis is more robust than the random effects perspective. However, since the unobserved effects are correlated to the independent variables, the latter must not contain features that are time-constant. That is because it is impossible to distinguish whether the effects of time-constant factors come from the observable or from the unobservable variables.

2. In order to have Assumption 2., we consider that $\ddot{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$, with $\bar{\mathbf{x}}_i = T^{-1} \sum_{t=1}^{T} \mathbf{x}_{it}$, obtaining time-demeaning values that cut out the impact of the unobserved effects. As a result, time-constant elements of the $\mathbf{x}_{it}$ vectors are replaced by zero for all t. Here, is is intended to make sure that the estimator works asymptotically, forbidding the independent variables to have time-constant elements.

3. This assumption has the same meaning as Assumption 3.a from the random effects model and has the purpose to ensure efficiency of the fixed effects estimator.

Once again, we start by Assumption 2. to get to our estimator:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i \right)^{-1} \left(\sum_{i=1}^{n} \ddot{\mathbf{X}}_i' \ddot{\mathbf{y}}_i \right) \tag{2.30}$$

which is unbiased when conditioned by $\mathbf{X}$. Furthermore, Assumption 3. assures that this estimator is efficient.

Chapter 3

Deposits Data and Model

In this chapter, first we will describe the data that was used to estimate the model and then we will present the obtained estimates and observe how they explain our final model choice. Furthermore, we will interpret our models' coefficients and take a closer look at its residuals.

Remark 3.0.1. *Notice that the pictures (graphics and tables) representing values of clients deposits will be in thousands of euros.*

3.1 Deposits Data

3.1.1 The Whole Sample

For this research, we collected information from 51 Portuguese banks, including successful ones, some that went bankrupt and some that were acquired by other banks that are also in the sample. Their deposits were registered annually from the year of 1992 until 2017. All of these banks contained at least 4 observations and had up to 26, summing a total of 713 observations. The fact that each bank disposes of a small amount of records doesn't harm the results because, by using panel data, there is a compensation, since the whole sample is included in the same model. More details about the whole sample can be found in Table 3.1.

Table 3.1: Whole sample's summary statistics

Minimum	Median	Maximum	Mean	Standard Deviation
379	1 002 690	73 426 264	6 310 595	12 350 636

In the table above, we can see that there is a wide range of clients deposits that varies from 379 thousand euros to 73 billion euros. However, the median of the sample is only 1 billion euros, which doesn't mean that there are more small banks, because there are small banks with only 4 observations as well as there are small banks with over 20 observations. Furthermore, the mean of the clients deposits is approximately 6 billion euros, which is relevant, since considering that the median is 1 billion euros, more than one half of the observations does not reach the mean value. This implies that, not only a considerable amount of observations does not attain the 6 billion euros as also the remaining deposits are significantly greater than the median. Finally, there is a fluctuation of approximately 12 billion euros in the overall deposits evidenced by their standard deviation.

However, we didn't consider the 51 banks in our research. We selected a smaller sample of banks according to the criterion of having an average of at least 10 billion euros in clients deposits. This is because the average growth rate of banks with different dimensions are distinct from one another. For instance, if a bank has 100 million euros in clients deposits, it can easily double its value even within one year. However, a bank that has 10 billion euros will take several years to double that amount. By adjusting models to banks by their dimension, it is possible to make better simulations in each context.

3.1.2 Large Banks

In the used sample we consider annual deposits data from 9 Portuguese banks and we present the collected data in Table A.1 and in Table A.2. The clients deposits were collected between 1992 and 2017, summing a total of 129 observations, though only 111 estimated residuals resulted from the estimation. This is due to the use of an AR(2) model in the estimation, since it can't consider the first 2 observations of each bank as it would need previous values to those.

According to our data structure, we are considering panel data: our cross-section units are the 9 Portuguese banks and the time considered starts in 1992 and ends in 2017. However, our panel is unbalanced since we don't have observations for all the banks in all the considered years. This happens because some of the banks that were taken into account have gone bankrupt or because there was only available information from a date later than 1992. The historical series for the banks deposits can be found from Figure 3.1 to Figure 3.9.

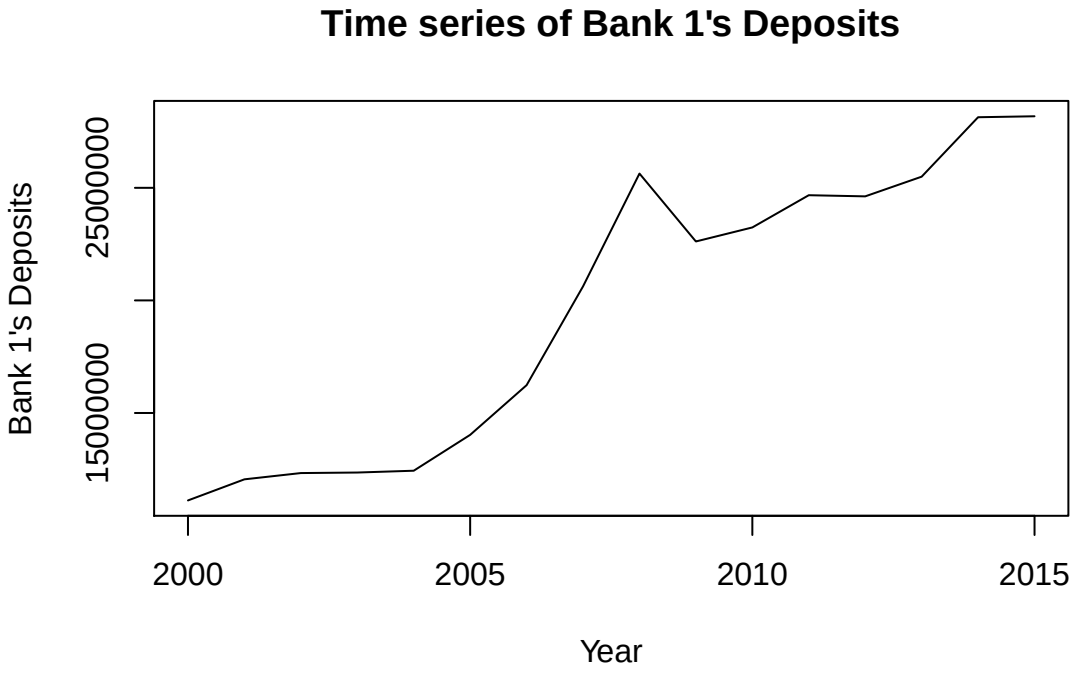


Figure 3.1: Historical time series of Bank 1's deposits

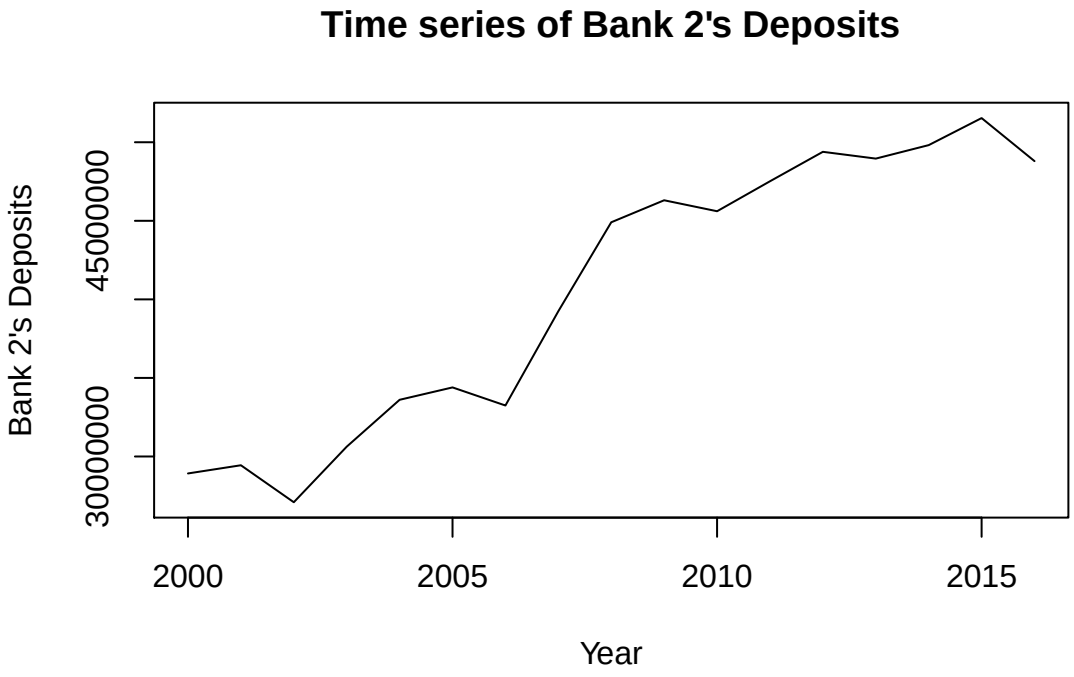


Figure 3.2: Historical time series of Bank 2's deposits

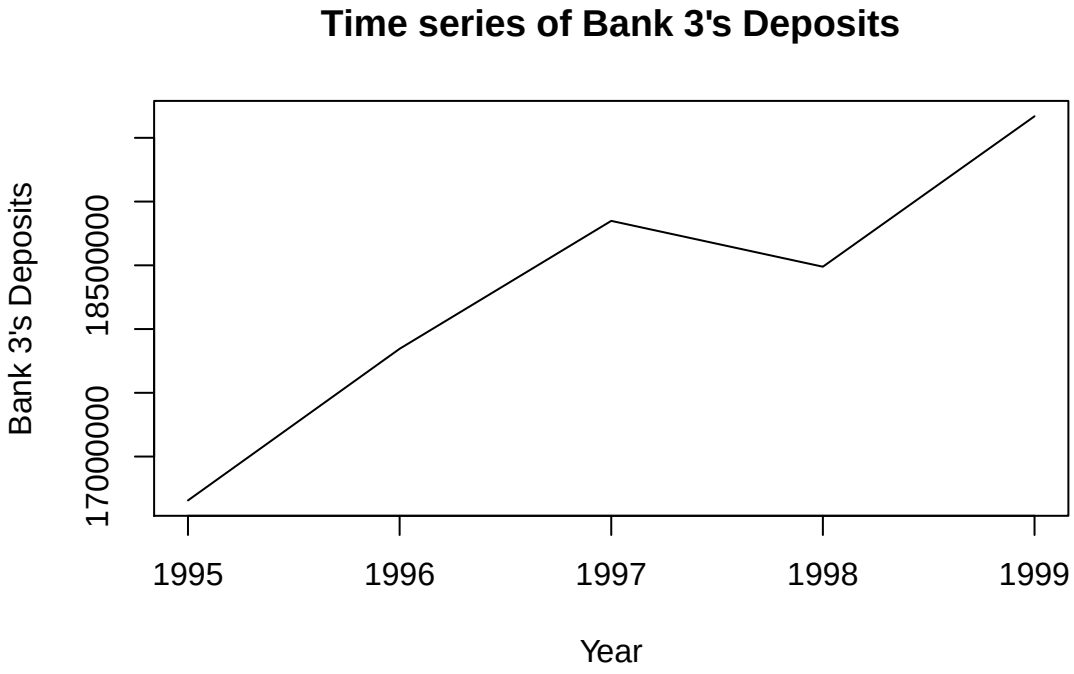


Figure 3.3: Historical time series of Bank 3's deposits

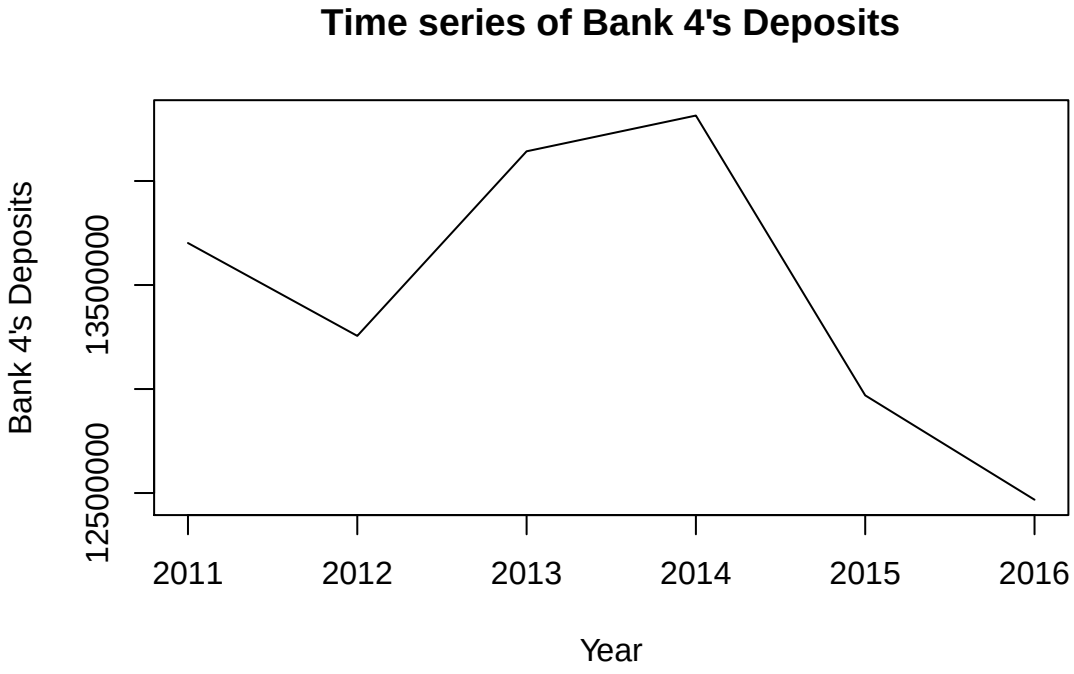


Figure 3.4: Historical time series of Bank 4's deposits

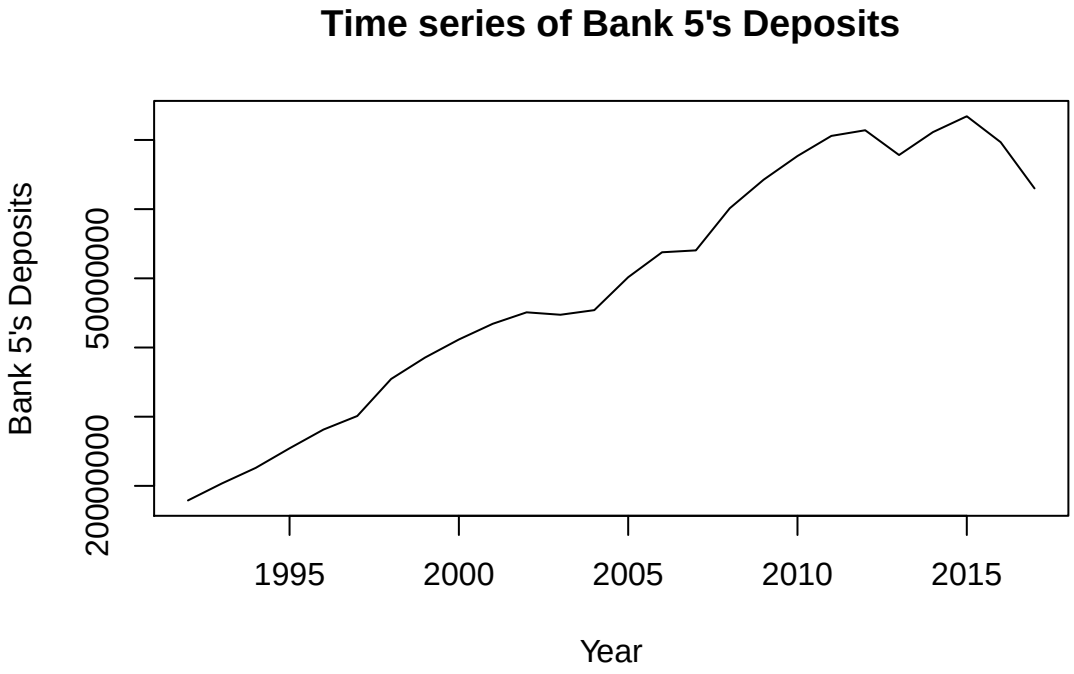


Figure 3.5: Historical time series of Bank 5's deposits

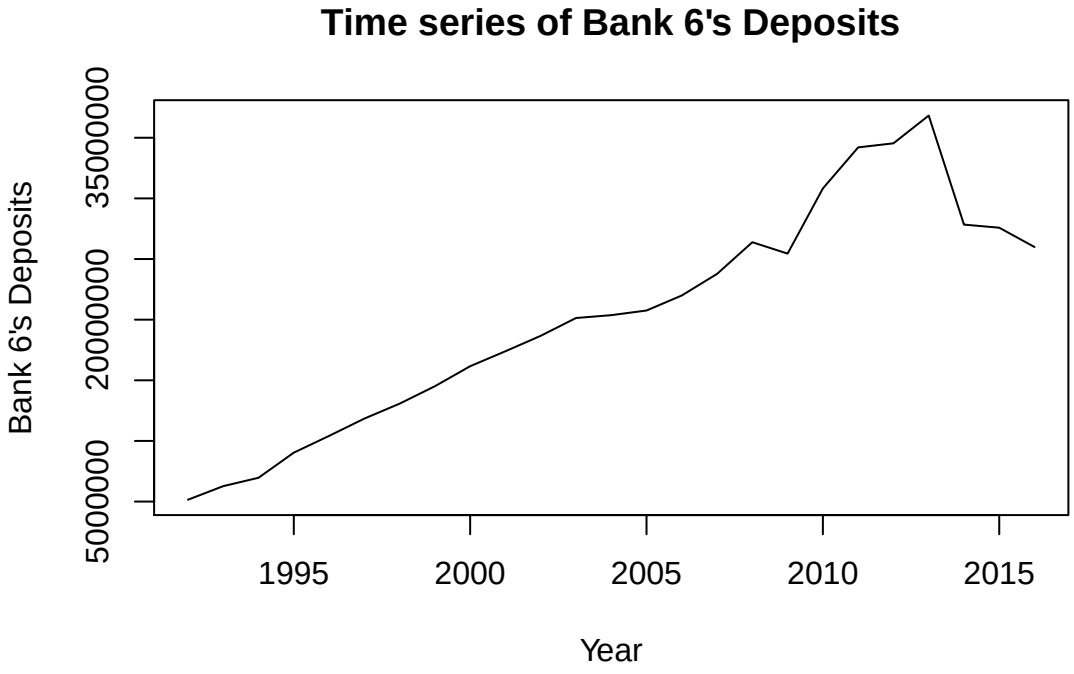


Figure 3.6: Historical time series of Bank 6's deposits

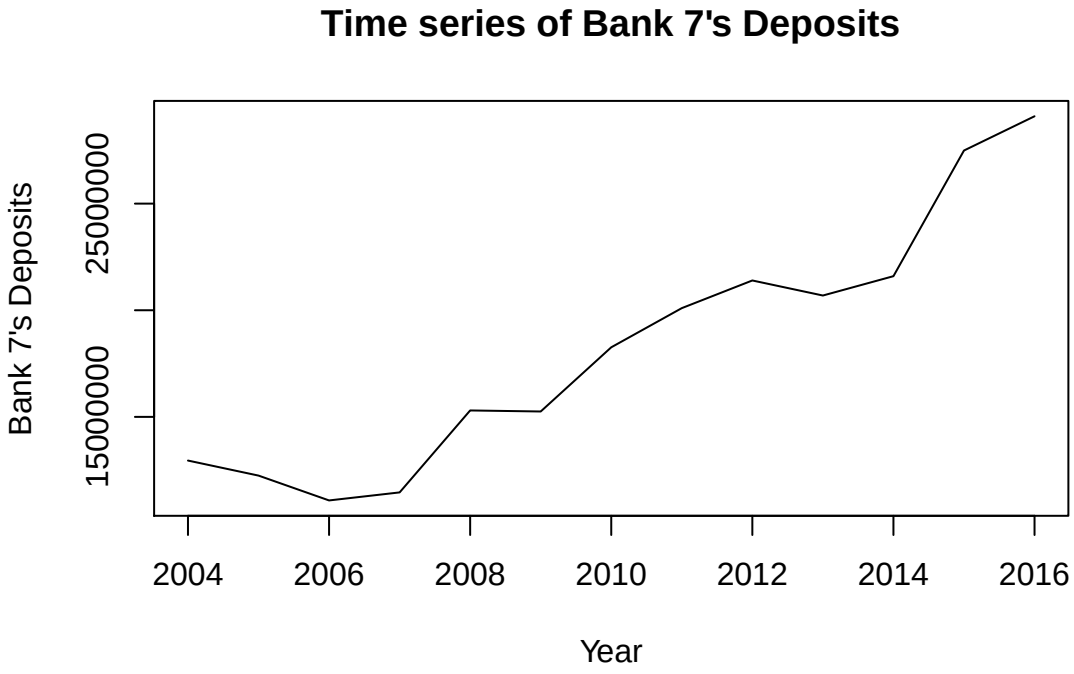


Figure 3.7: Historical time series of Bank 7's deposits

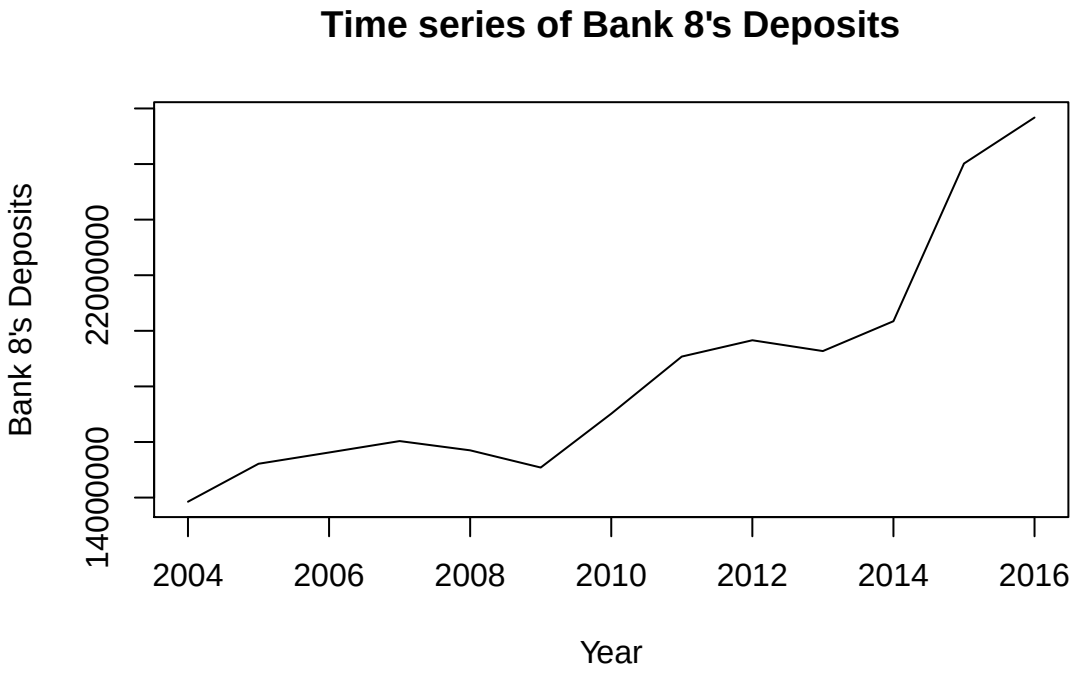


Figure 3.8: Historical time series of Bank 8's deposits

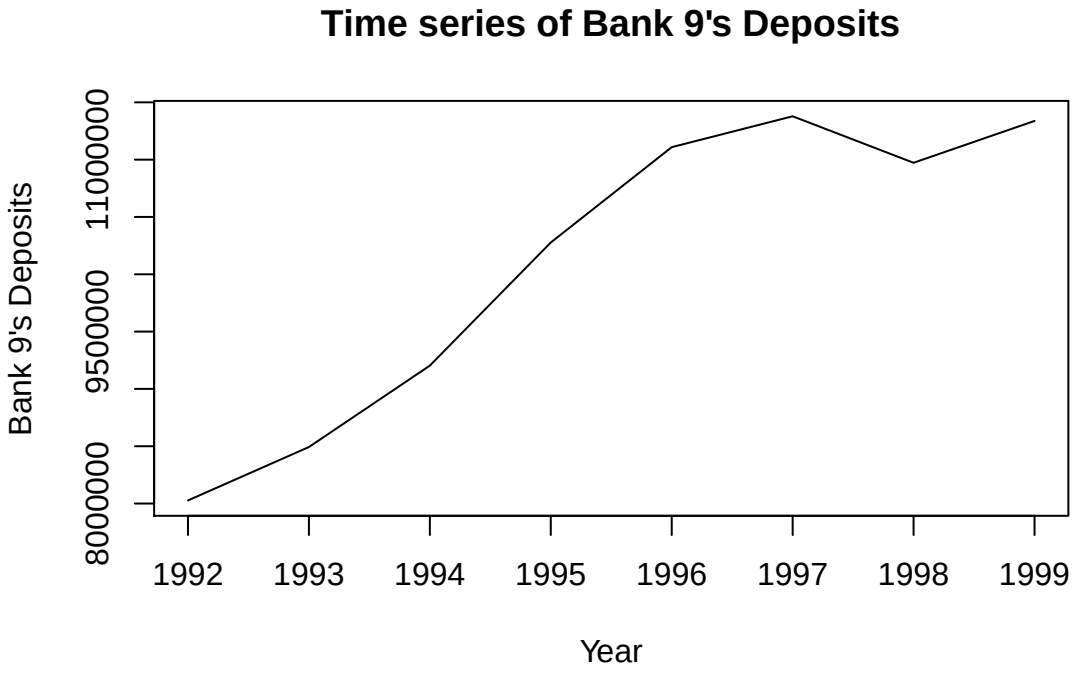


Figure 3.9: Historical time series of Bank 9's deposits

In these graphics, it is possible to observe that the time series of the clients deposits are mainly growing. The decaying periods presented in the time series are usually associated to crisis episodes that happened in Portugal, such as in the years of 2011 and 2014.

In general, these banks have more than 10 observations, which is important since we are going to use an AR(2) model and it will cut on two estimated residuals for each bank. Furthermore, it is also relevant to notice that many of these banks show downfalls at some point of its time series, which will help to validate our simulations. That is, by considering banks that have falls in their clients deposits somewhere in their time series, it will contribute for the simulations to exhibit falls as well. If we would consider only one time series, the simulations would result in characteristics similar to those, whether it included falls or not. This is why panel data is relevant in this matter.

Concerning the sample itself, considering all the observations from the 9 banks regardless of the dates, we were able to obtain the data description presented in Table 3.2, shown below. Here, it is important to observe that even though the minimum value of the deposits is under 10 billion euros, the average of the respective bank's deposits is not. Furthermore, the mean value of all banks is the one taken into consideration in the model estimation as $\bar{D}$, as follows.

Table 3.2: Large Bank's summary statistics

Minimum	Median	Maximum	Mean ($\bar{D}$)	Standard Deviation
5 152 522	21 597 821	73 426 264	27 251 747	16 779 694

In this table, we have that the clients deposits vary from 5 billion euros to 73 billion euros. Thus, the chosen banks have a wide range of values to be studied, which makes it possible to make diverse simulations with the obtained model. Furthemore, since the median of the sample is 20 billion euros, meaning that half of the observations of the sample are below that value, and its mean is 30 billion euros, we have that most of the observations are below the latter value. This is a similar case to the whole sample, as there is a compensation of the deposits for some banks that have the highest records. Finally, the standard deviation indicates that the overall deposits fluctuate in approximately 17 billion euros.

3.2 Model Estimation

A usually considered approach in finance includes the use of the momentum (see, for example, Crombez (2001) and Jegadeesh and Titman (2001)). The momentum is, essentially, the empirically observed trend that the data evidence. That is, the momentum shows that if, in our case, the deposits tend to grow, they will continue to grow, and if they tend to decay, they will keep decaying. This feature can be introduced by means of an AR(2) model, thus that's our proposal for the deposits evolution.

3.2.1 The Model

Considering a linear AR(2) model and representing by $D_{i,t}$ the clients deposits for bank i at time t, with $i = 1,...,9$ and $t = 1,...,T$, equations (2.14) and (2.21) can be rewritten as:

$$D_{i,t} = \alpha + \beta_1 D_{i,t-1} + \beta_2 D_{i,t-2} + \varepsilon_{it}. \tag{3.1}$$

For the data treatment we consider the *plm* package (Croissant and Millo, 2008) that is available for the *R* software (R Core Team, 2014). This package adapts the estimation methods to unbalanced panels, which is a feature that wasn't provided by other econometrics packages. The respective code is presented in Figure A.1.

Remark 3.2.1. *Notice that the function plm with the option "pooling", centers the data using the overall mean of the deposits. That is, the model is adjusted to the data* $D^*_{i,t} = D_{i,t} - \bar{D}, \quad i = 1,...n, \quad j = 1,...,T$, *with* $\bar{D} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} D_{i,t}$ *representing the overall mean of the deposits.*

By using the *p*lm function with the referred option, we get the adjusted AR(2) pooled model, estimating the model parameters through the OLS method without considering

unobserved effects. This means that, not only there is not a specific intercept for each bank, as also the estimates of the coefficients apply to all banks. This is important for us, because our final goal is to obtain a plausible model (calibrated with data from different banks, incorporating distinct scenarios), that we can use to simulate the clients deposits evolution for a general bank. The estimation results are presented in Table 3.3, below.

Table 3.3: Estimated AR(2) coefficients

	Estimate	**Standard Error**	**P-value**
α	1 229 300	427 830	0.00489
β_1	1.2082	0.10166	2×10^{-16}
β_2	−0.22016	0.10328	0.03529

Since the p-values of all variables are smaller than 5%, there is statistical evidence that all the considered variables are significant to this research. Furthermore, it is also possible to verify that the parameters satisfy the stationarity conditions for an AR(2) process:

* $1.2082 + (-0.22016) < 1$
* $(-0.22016) - 1.2082 < 1$
* $-1 < -0.22016 < 1$.

The results of the estimation also evidence an R^2 of 0.98217, with the respective adjusted value of 0.981984, which means that the chosen independent variables can almost completely explain the considered dependent variable. This allows the results to have an economic interpretation. Therefore, we were able to work with the following model equation:

$$D^*_{i,t} = 1\,229\,300 + 1.2082 D^*_{i,t-1} - 0.22016 D^*_{i,t-2} + \varepsilon_{it}. \tag{3.2}$$

Thus, we obtained an intercept of more than 1 billion euros, which is the independent term of the model. It has a standard error of approximately 428 million euros, which influences the variation of the demeaned deposits in that amount, in a given year, together with the standard errors of the independent variables' coefficients. Since β_1 and β_2 are the parameters that measure the effect of the value of the demeaned deposits from the previous time periods in the present, we have that: as β_1 is positive and greater than one, a value greater than the demeaned deposits in the previous time period will be added to the intercept; as β_2 is negative and approximately 20%, around one fifth of the value of the demeaned deposits two time periods before will be subtracted to the previously described sum. Finally, an error term with unknown distribution is added. However, the interpretation of a model with demeaned deposits sounds very unnatural and it is hard to account.

Thus, we will now introduce the momentum representation, which, removing the effect of the overall mean, is the following:

$$D_{i,t} = \bar{D}(1 - 0.98804) + 1\,229\,300 + 0.98804D_{i,t-1} + 0.22016(D_{i,t-1} - D_{i,t-2}) + \varepsilon_{it} \tag{3.3}$$

and, considering that the overall mean is $\bar{D} = 27\,251\,747$, it can also be represented by:

$$D_{i,t} = 1\,555\,231 + 0.98804D_{i,t-1} + 0.22016(D_{i,t-1} - D_{i,t-2}) + \varepsilon_{it}. \tag{3.4}$$

This equation might be easier for the reader to understand than the previous ones since the momentum term already represents the trend that the time series has been following. That is, if the deposits time series has been growing, this will be a positive term, otherwise it will most certainly be negative. Therefore, since the first coefficient of the independent variables is almost equal to one, we can conclude that the present value of the clients deposits will be the value of the intercept, plus nearly the value assumed by the clients deposits in the previous time period, plus approximately 20% of the trend of the series in the two previous time periods, plus an error term.

With these estimates, we needed to test whether this model was accurate or not. In order to get to valid conclusions, the residuals of the model had to be checked first.

3.2.2 Residuals

For the adjusted model we have the residuals histogram in Figure 3.10 and qq-plot in Figure 3.11.

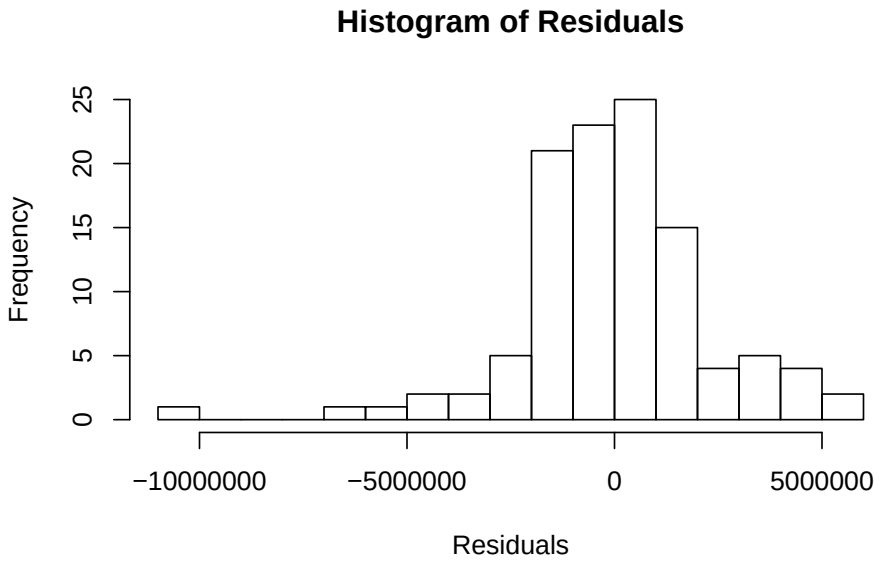


Figure 3.10: AR(2) residuals for the bank deposits

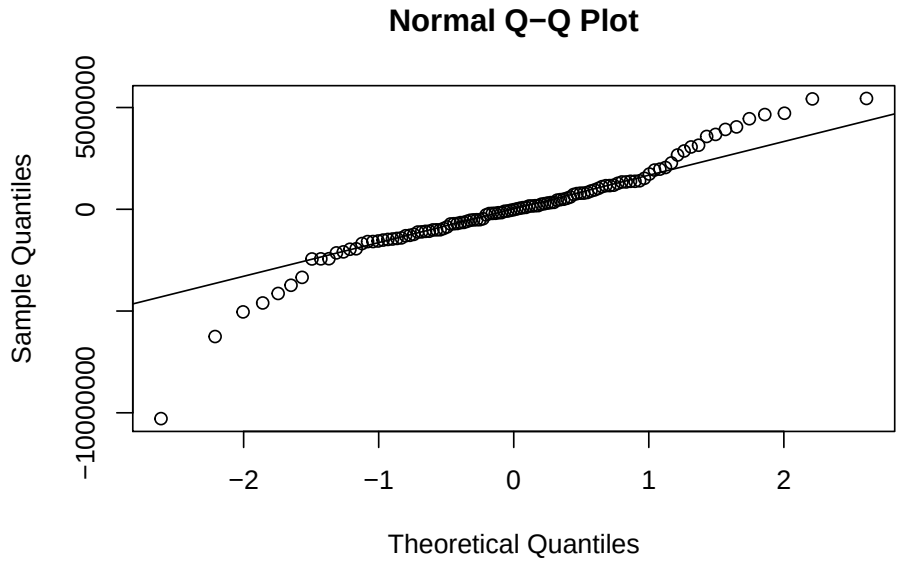


Figure 3.11: QQ-plot for the AR(2) residuals for the bank deposits

In these graphics, the residuals didn't seem to follow a Normal distribution, as also suggested by the residuals skewness and kurtosis that are presented in Table 3.4.

Table 3.4: Residuals skewness and kurtosis

Skewness	**Kurtosis**
-0.6479653	6.517575

In order to validate our conclusions, we used both the Jarque-Bera and the Shapiro-Wilk tests for normality. Furthermore, since the autocorrelation of the residuals also needed to be tested, we present all the results in Table 3.5.

Table 3.5: P-values of the tests applied to the AR(2) model's residuals

Test	**AR(2)**
Jarque-Bera	2.2×10^{-16}
Shapiro-Wilk	4.946×10^{-5}
Ljung-Box	0.7315

According to the p-values of the normality tests, the null hypothesis of normality is rejected in both cases. This confirms that the residuals did not follow a Normal distribution, both by the ordered statistics and by the skewness and kurtosis results.

Finally, we preferred the Ljung-Box test for autocorrelation because it can be used in the presence of non-normal disturbances, as an opposition to the Box-Pierce's test-statistic. This test returned a p-value of 0.7315, not rejecting the null hypothesis, which confirms an absence of autocorrelation among the residuals of the AR(2) model.

3.2.3 Pooled Model

With non-normal residuals, some "pooling" tests, that is, procedures to test if individual banks or time effects should be introduced in the model, couldn't be applied to our model. Consequently, we used a Lagrange Multipliers test by Breusch and Pagan (Breusch and Pagan, 1980), which, as Honda, 1985 shows, is robust to non-normal disturbances while testing both effects simultaneously. The result was a p-value of 0.1896, which means that the null hypothesis of non-significant effects is not rejected. This represents statistical evidence that there were neither individual nor time effects, which validates our model.

By using a pooled model, we are neither in the presence of random effects nor of fixed effects. This implies that the estimation method does not have to consider unobserved effects, which makes this model a generic model for banks of large dimension. It considers both successful and failed banks, some with a few and others with lot of observations, as well as with downfalls in their time series or constantly growing ones.

Chapter 4

Computational simulation and results

In this Chapter, we aim to discuss the results of the simulations that we were able to generate with the estimated model from the previous chapter. Therefore, we present their trajectories and compare them to simulations originated by the most commonly used models at explaining clients deposits' evolution. These are based in one single time series, that is, these models consider individual banks.

4.1 Panel data results

With the model presented in equation (3.4), we simulated 9 trajectories for clients deposits for the following 30 years from the last recorded observation (the first 2 years in the trajectories, correspond to initial given values for starting the AR(2) process simulation). The considered mean value for all cases (including the individual banks that will be studied in the following sections) was $\bar{D} = 10\,000\,000$ because it was the criterion for choosing our sample, even though it doesn't influence the simulated trajectories.

The fact that we didn't choose the actual mean of the considered banks to perform the simulations doesn't change the results in meaning. That is, the graphics would look the same except for a vertical translation of the graphics for the difference between the actual mean and the value we chose for each case (see equation (3.3)). Therefore, we chose our reference to select our sample as the initial position.

According to the final results in Chapter 3, regarding the residuals of the adjusted model, its residuals did not follow a Normal distribution. Thus, we used a resampling technique with replacement, which is known as the Bootstrap method, to simulate them. That is, we used the residuals from the original time series in the simulations, randomly distributed among the observations. The original residuals could be used more than once.

Therefore, our simulations are presented from Figure 4.1 to Figure 4.3.

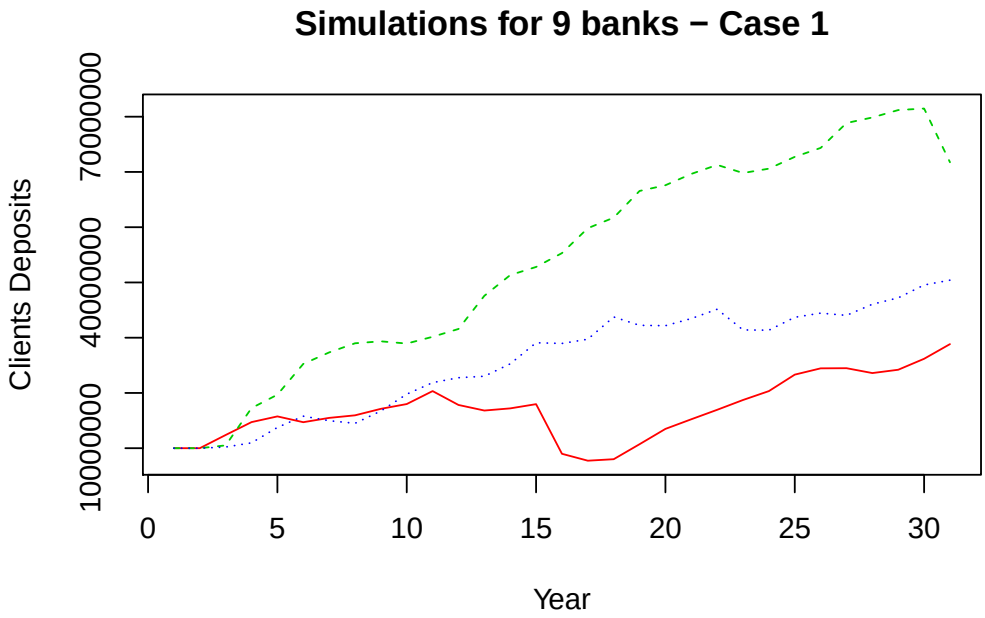


Figure 4.1: Simulated trajectories with the model in equation (3.3) for 9 Banks - Case 1

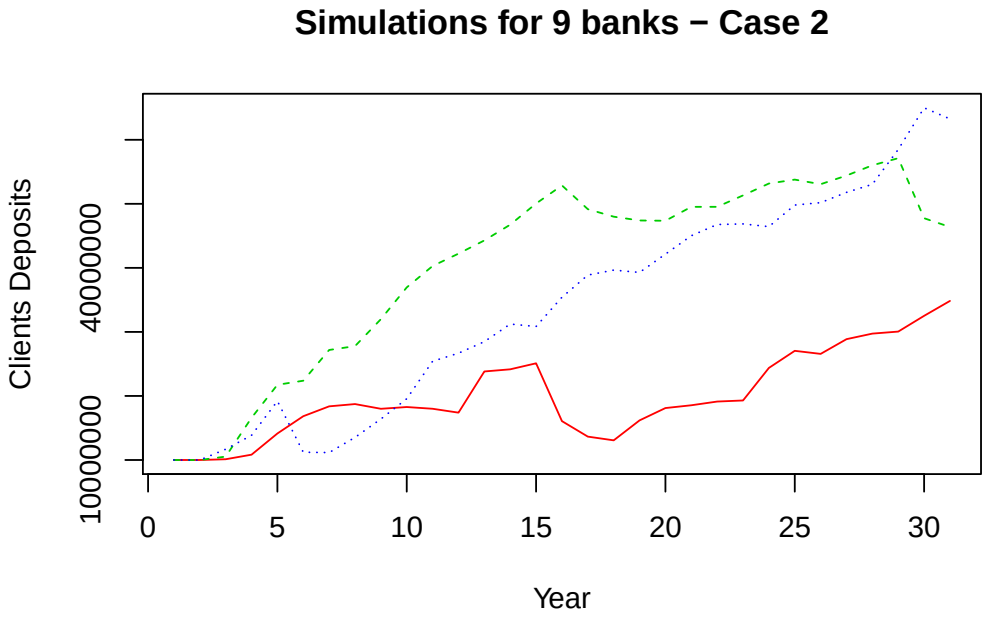


Figure 4.2: Simulated trajectories with the model in equation (3.3) for 9 Banks - Case 2

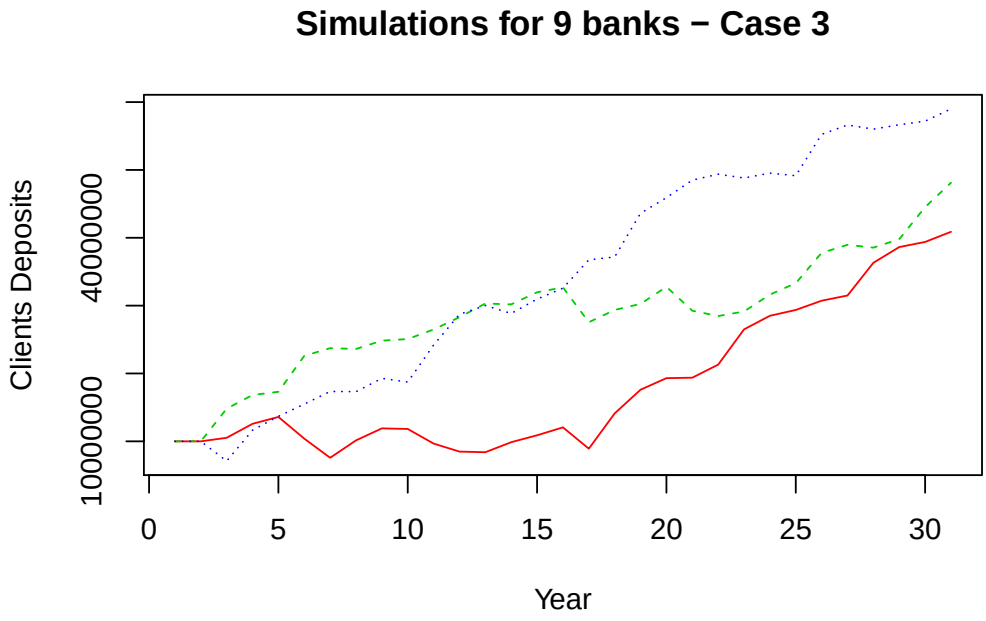


Figure 4.3: Simulated trajectories with the model in equation (3.3) for 9 Banks - Case 3

In these simulations we are able to observe that there are always at least two distinct scenarios: one in which the time series has almost a constant value until it reaches approximately 30 billion euros in the end (corresponding to the red line in the three graphics); another in which the time series is mainly growing, while either having a few accidents or a slower growth. Here, it is possible to notice that there are never extremely fast growths, as in the first 15 years of the simulations, approximately, the trajectories of the deposits don't evolve quickly to greater values, except for two paths in Case 2. We get a wide range of possibilities that derives from the use of panel data, obtaining both good and bad scenarios.

Considering the values that these time series can reach, we should recall that our data include deposits from 5 billion euros to 73 billion euros, approximately. Therefore, it is ordinary that the simulated time series can register both low and high values, such as reaching 30 billion euros or 70 billion euros, respectively.

Thus, these simulations represent the various possible cases that a bank can face: prosperity, cautiousness or the risk of bankruptcy. This means that the model presented in equation (3.3) can help banks to manage their activity, taking into account the many risks that they may need to face.

Furthermore, we were able to obtain more indicators of the advantages of panel data modelling, such as the Maximum Drawdown, which will be discussed in Section 4.3.

4.2 Analysis of Individual Banks

In order to compare our model with the most commonly used by the authors referred in Chapter 1, the AR(1) and the AR(2) models applied to individual banks, we chose 2 different banks from our sample: Bank 1 and Bank 5. We selected these banks mainly for two factors: the number of observations of their time series was reasonable; their time series had not big sudden falls.

Since we are considering autoregressive models it was not recommended to use a bank with only a few observations (such as Bank 3 or Bank 4) because the estimated residuals would have even less values and, as tested, the estimated coefficients of the respective models would not be significant. Moreover, a consistently growing time series would allow us to test the AR(1) simulations. That is, if this model would result in both good and bad scenarios with a successful bank, then the use of panel data would be pointless. However, as discussed before, with only one time series, the autoregressive models would not be expected to simulate significantly different cases from the ones observed in the respective time series.

4.2.1 Bank 1

To start with, we had to estimate the AR(1) and AR(2) models with Bank 1's data in order to check whether these models could be compared to ours. Using the *R* software, as shown in Figure A.2 and in Figure A.3, we obtained the results of the coefficients' estimations presented in Table 4.1 for the AR(2) case and in Table 4.2 for the AR(1) case. Let us focus on the AR(2) model first.

Table 4.1: Estimated AR(2) coefficients for Bank 1

	Estimate	**Standard Error**	**P-value**
α	2306783	2015825	0.2768
β_1	1.02367	0.30087	0.0059
β_2	−0.08804	0.30788	0.7802

Since β_2 is not significant in the AR(2) model estimation, we checked both the ACF and PACF of this model to check whether the AR(2) model was actually not appropriate for this data. The results are presented in Figure 4.4 and in Figure 4.5.

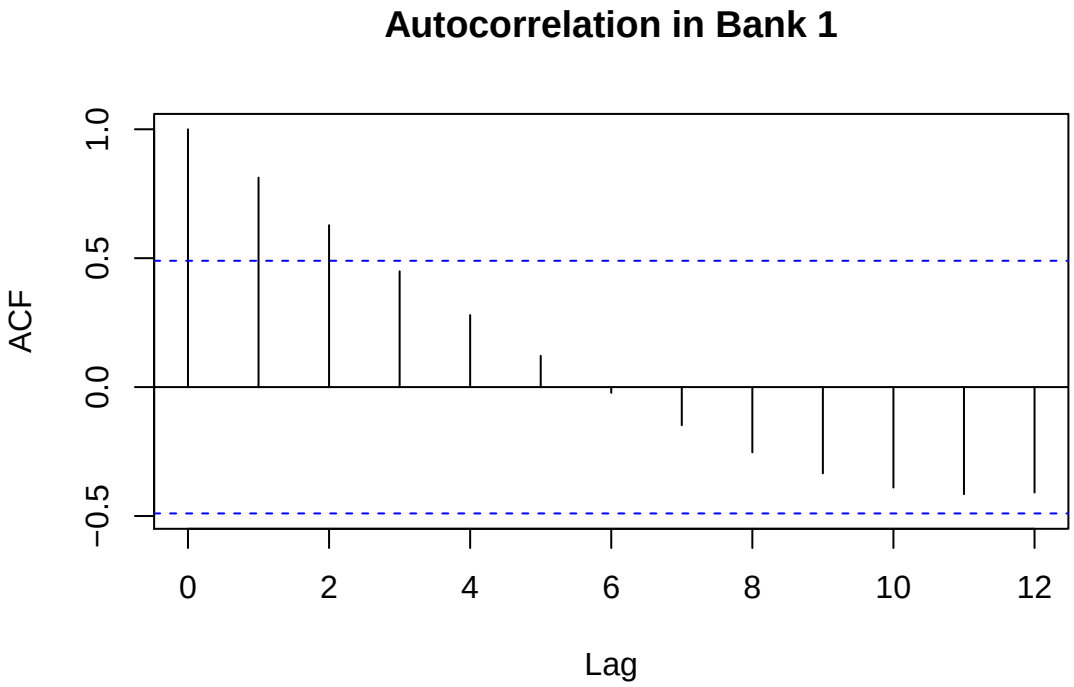


Figure 4.4: Autocorrelation Function in Bank 1's time series

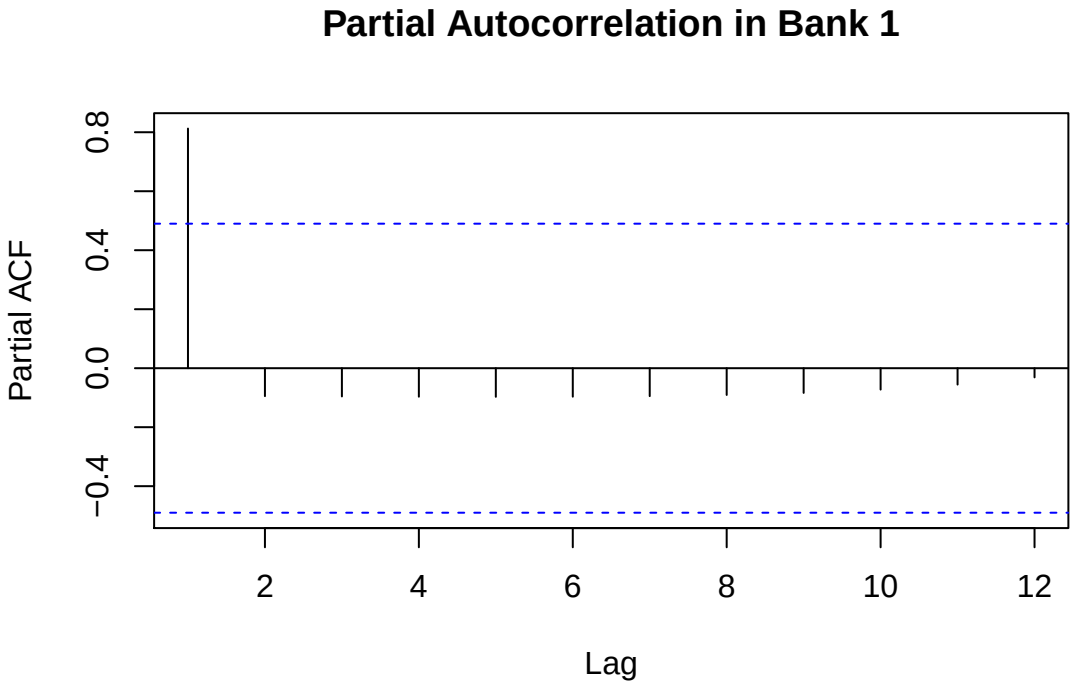


Figure 4.5: Partial Autocorrelation Function in Bank 1's time series

As shown in the graphics above, the time series of Bank 1's deposits is stationary and the most appropriate autoregressive model to study them is the AR(1) model. Therefore, there would be no point in proceeding with this model's analysis. It is thus excluded from our research. Let us now focus on the AR(1) case.

Table 4.2: Estimated AR(1) coefficients for Bank 1

	Estimate	**Standard Error**	**P-value**
α	2 038 634	1 705 459	0.253
β_1	0.95267	0.08542	5×10^{-8}

According to the PACF graphic, the AR(1) model seemed to be relevant, which was proven by the significance of β_1 in its estimation, as presented in Table 4.2. Furthermore, the estimate of the respective coefficient is smaller than 1, which verifies that the data are stationary. Finally, even though the intercept of this model is not significant, it won't alter the nature of the model because it is a constant and it is not associated to any variable, thus it is not necessary to exclude it. Furthermore, this estimation resulted in an R^2 of 0.9054, and in the respective adjusted value of 0.8981, which shows a good adjustment.

Before proceeding to the simulations, we only had to check the residuals of the estimated model for normality and autocorrelation. For normality, we had some preliminary results:

Table 4.3: Skewness and Kurtosis of the residuals of Bank 1 model

Moments	**AR(1)**
Skewness	0.3499915
Kurtosis	3.280727

In Table 4.3, the values for the skewness and kurtosis of the residuals indicate normality, since these are close to the reference values of the Normal distribution. However, as we did in Chapter 3, we performed both Jarque-Bera and Shapiro-Wilk's tests for normality and the Ljung-Box test for autocorrelation. The results are the following:

Table 4.4: P-values of the tests applied to the AR(1) model's residuals

Test	**AR(1)**
Jarque-Bera	0.81
Shapiro-Wilk	0.4825
Ljung-Box	0.804

Since all p-values are greater than 0.05, none of the null hypothesis is rejected. That means that the residuals of the estimated AR(1) model are considered to follow a Normal distribution and not to be autocorrelated. Thus, considering an AR(1) model with its coefficient estimated at 0.95267 for its first lag and normal residuals, we obtained the simulations presented from Figure 4.6 to Figure 4.8 for Bank 1's clients deposits. The respective graphics are shown below.

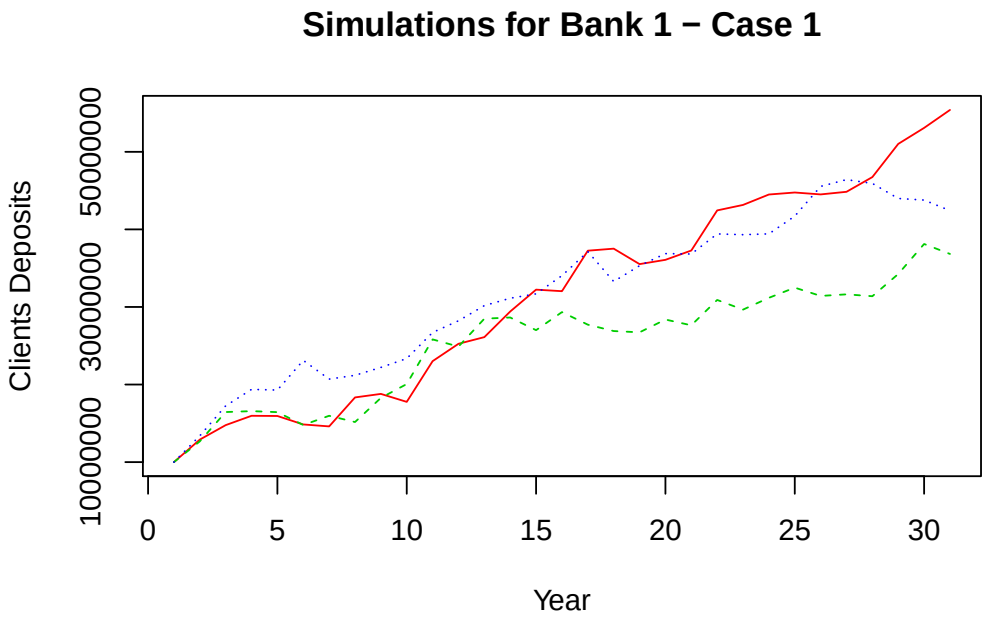


Figure 4.6: Simulated trajectories with an AR(1) for Bank 1 - Case 1

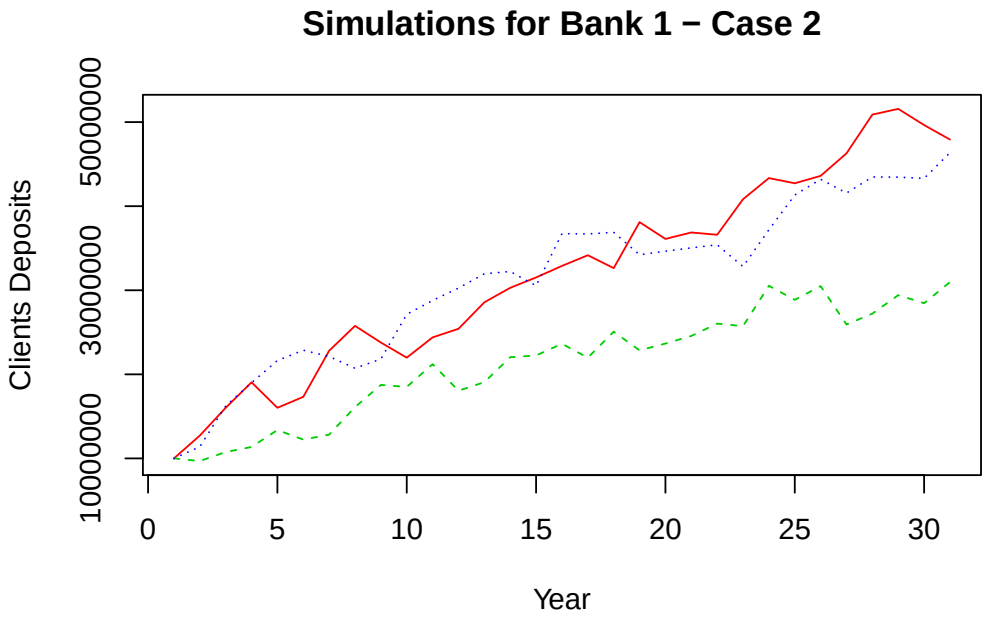


Figure 4.7: Simulated trajectories with an AR(1) for Bank 1 - Case 2

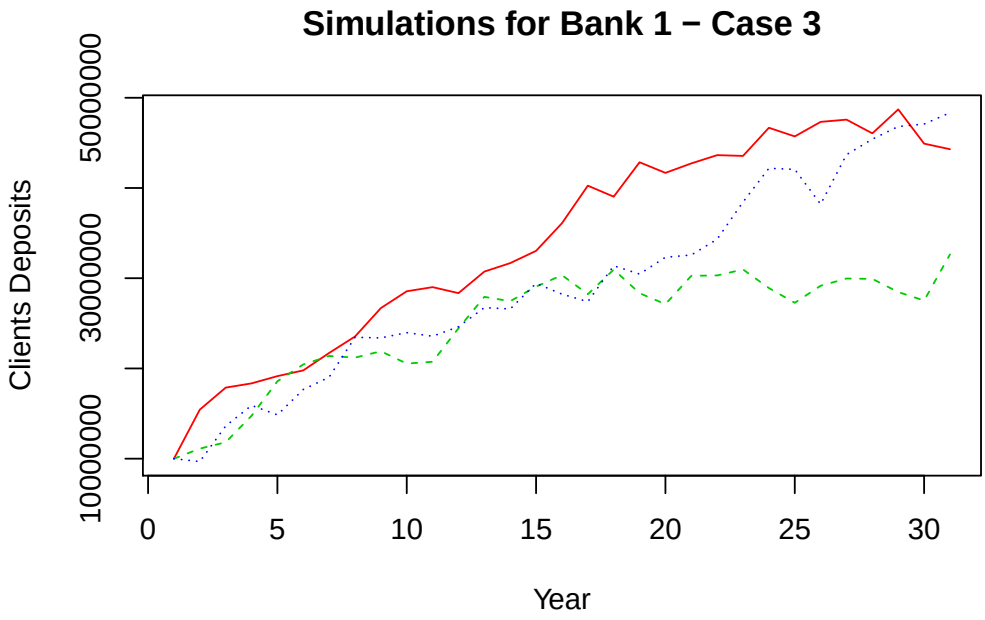


Figure 4.8: Simulated trajectories with an AR(1) for Bank 1 - Case 3

These simulations were obtained for the bank whose time series is represented in Figure 3.1 (Bank 1): a growing time series that starts at approximately 10 billion euros and reaches a value close to 30 billion euros, except for an interval of two periods of time, in which it decays.

In the Figures above, it is possible to notice the constantly growing trend, except for the green dashed line in the third case. Therefore, all the simulations are optimistic because none of them represents big losses and neither downward trajectories. Furthermore, the worst case scenario is maintaining the deposits at a roughly plain level of 30 billion euros after a successful period.

The described graphics don't evolve too quickly, which is reasonable, but since those don't show scenarios that are different from one another, those are not the most viable source. That is, if a bank only takes these simulations into account when managing resources, it will only consider good scenarios, which may lead to desastrous consequences since there may be sudden "bank runs" that weren't planned. This is one visible example of the effects of using only one time series in simulations with an AR(1) model.

4.2.2 Bank 5

We'll repeat the procedure applied to Bank 1. Thus, considering those obtained results, we'll start by estimating the AR(2) model with Bank 5's deposits. The code used in the *R* software is shown in Figure A.5 and the estimation results are presented below, in Table 4.5.

Table 4.5: Estimated AR(2) coefficients for Bank 5

	Estimate	**Standard Error**	**P-value**
α	3 888 176	1 887 173	0.052
β_1	1.333	0.2505	2.82×10^{-5}
β_2	−0.3923	0.2417	0.12

As we verified in Bank 1, the second coefficient of the AR(2) model is not significant, thus, once again, we will confirm this result with both the Autocorrelation Function and the Partial Autocorrelation Function:

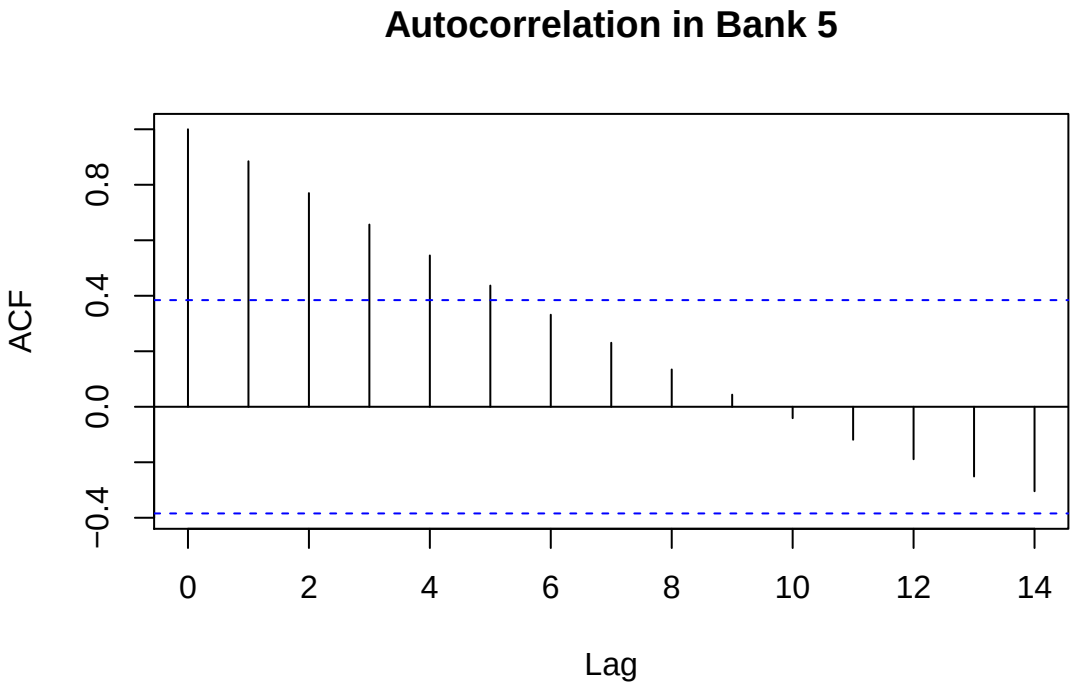


Figure 4.9: Autocorrelation Function in Bank 5's time series

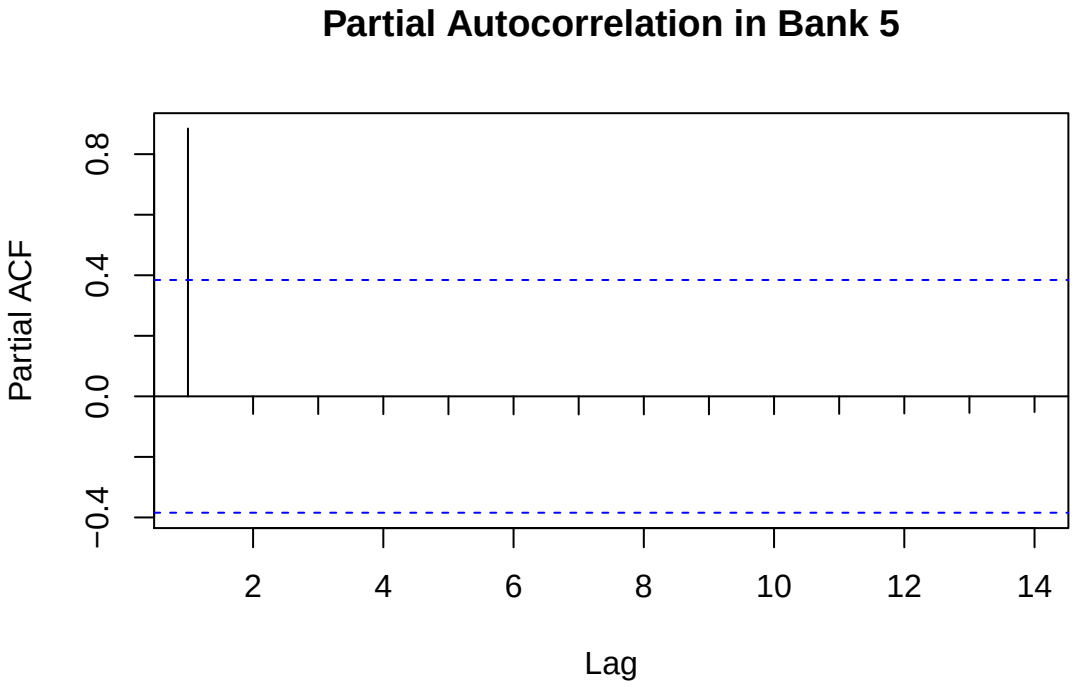


Figure 4.10: Partial Autocorrelation Function in Bank 5's time series

The results are similar to the ones found in Bank 1: the data appear to be stationary but only the AR(1) model seems to be relevant for our research. Therefore, we will now analyze the AR(1) model estimates:

Table 4.6: Estimated AR(1) coefficients for Bank 5

	Estimate	**Standard Error**	**P-value**
α	4 918 310	1 576 701	0.00482
β_1	0.9358	0.03054	2×10^{-16}

In Table 4.6, it is shown that both the intercept and the coefficient of the first lag of the model are significant at a significance level of 5%. Furthermore, looking at the estimates, we can confirm that the data are stationary, as initially revealed by the ACF. Finally, this estimation resulted in an R^2 of 0.9761, and in the respective adjusted value of 0.9751. These values also indicate a good adjustment of the model to the sample.

We also checked the normality of the residuals of Bank 5 in order to decide the best method in the simulations. Hence, we had to verify the respective skewness and kurtosis for the considered model, with the respective results presented in Table 4.7.

Table 4.7: Skewness and Kurtosis of the residuals of Bank 5 models

Moments	**AR(1)**
Skewness	−0.6371799
Kurtosis	3.441407

The skewness and kurtosis of the residuals in this model are close to the values of the Normal distribution. Thus, in order to validate these results, we performed the same tests that we referred in the Bank 1's case, including the verification of autocorrelation. The results for all tests are presented in Table 4.8.

Table 4.8: P-values of the tests applied to the AR(1) model's residuals

Test	**AR(1)**
Jarque-Bera	0.175
Shapiro-Wilk	0.264
Ljung-Box	0.3265

All of the obtained p-values are greater or equal than 0.05, which means that the null hypothesis is never rejected. Therefore, we can assume that the residuals of the model AR(1) for Bank 5 also follow a Normal distribution and do not verify autocorrelation. Hence, we are now in the conditions to simulate this model.

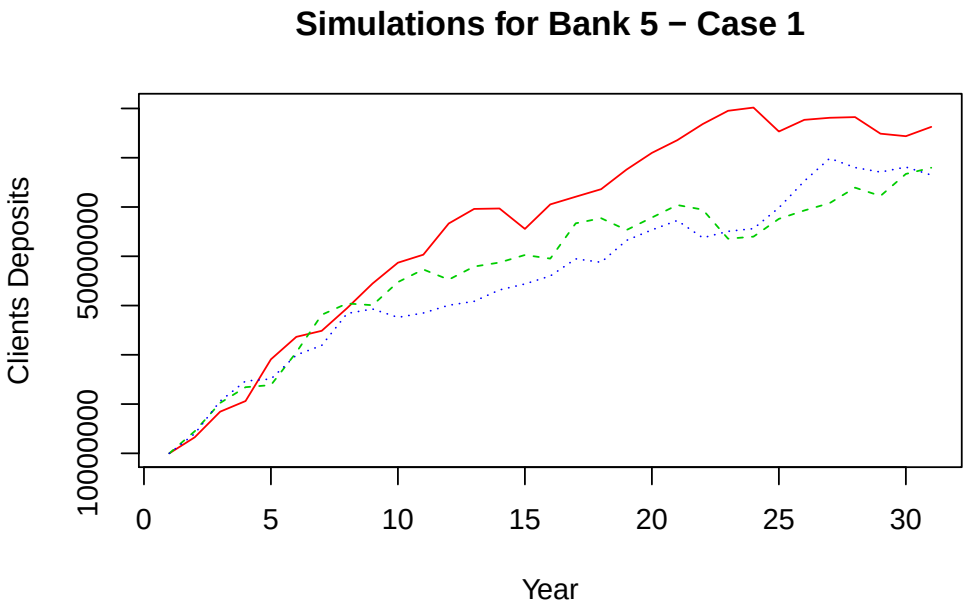


Figure 4.11: Simulated trajectories with an AR(1) for Bank 5 - Case 1

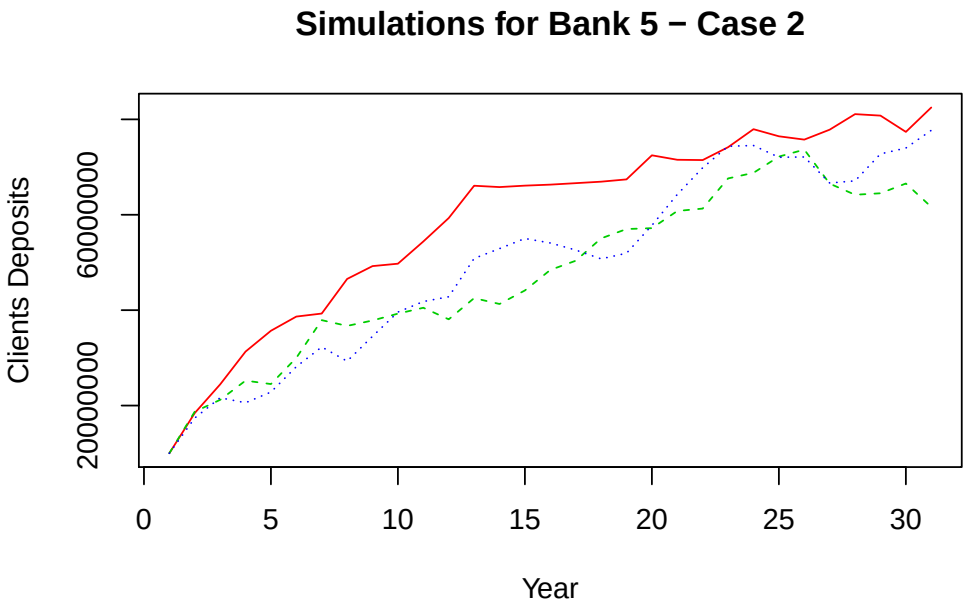


Figure 4.12: Simulated trajectories with an AR(1) for Bank 5 - Case 2

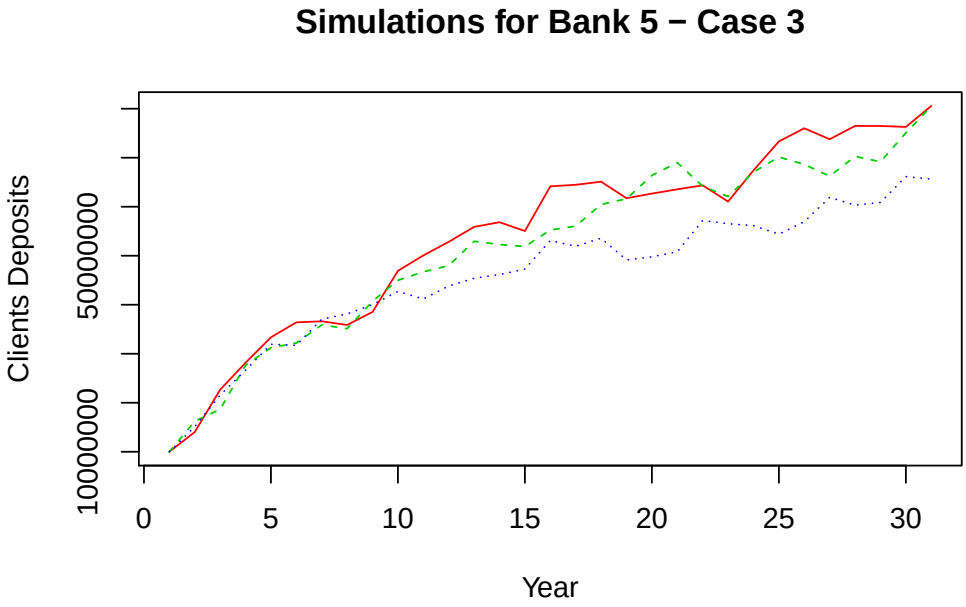


Figure 4.13: Simulated trajectories with an AR(1) for Bank 5 - Case 3

The simulations were now obtained for the bank whose deposits are represented in Figure 3.5 (Bank 5): a series that starts near the 18 billion euros to grow until it reaches approximately 70 billion euros, when it falls back to the 60 billion euros.

In the trajectories found from Figure 4.11 to Figure 4.13, we can observe a very significant growth in the first 10 years. Furthermore, even more visibly than in Bank 1, we can notice that every simulated trajectory is extraordinarily optimistic; not only the values grow to extremely large values (as the series from which these derive) but also there are never remarkable falls. These facts emphasize our previous points: with panel data there are more realistic scenarios and the simulations with only one time series reflect the sample itself.

4.3 Comparisons

As shown in the previous section, the model in equation (3.4) shows more diversified case scenarios, including worse ones, than the AR(1) model applied to two different banks (one at a time). That is, while the simulations with just one bank seem to be constantly growing, and faster, according to the reached values, the results with panel data show the possibilities of lacking resources and of sudden losses of money.

4.3.1 Maximum Drawdown

While discussing the simulations of the 9 Banks, we referred an indicator called Maximum Drawdown (MDD). This is an indicator of downside risk, which we used to check whether the obtained simulations were actually simulating bad scenarios. A Maximum Drawdown is, in an observed evolution of a time series, the maximum fall of the values, starting from a local maximum of the time series, before it is achieved a new local maximum. That is, when the time series reaches a maximum value and starts decaying, that maximum is recorded until the series starts growing again and a new maximum is attained. In the period of time contained between the two maximums, the lowest achieved value is also recorded. The first maximum is subtracted to this minimum and the result is divided by the first maximum as well and multiplied by 100 in order to register the fall in percentage. Mathematically:

$$MDD = \frac{Min - Max}{Max} \times 100 \tag{4.1}$$

where Min is the minimum value attained between two maximums and Max is the first local maximum of every two consecutive considered.

Through the analysis of the simulations, we noticed that the first 10 years were decisive in the trend that the trajectories would take. Hence, we thought that it would be reasonable to separate the analysis of the 9 trajectories of each case in two different situations: considering the 30 years; only assessing the first 10 years. Thus, we would be able to highlight the main problem of the AR(1) model with just one time series: the beginning of the simulations never show episodes of sudden losses of money. Additionally, since our goal is to have realistic simulations, we also chose to compare the referred MDD's with the ones from the historical data.

Since there are many simulations (or banks, in the case of historical data), we wouldn't use more than one MDD for each of them. Thus, we chose the maximum MDD of all simulations of each category: the 9 Banks of our sample, Bank 1 and Bank 5. Also, for comparing matters, we also calculated the average MDD in each case.

We'll start by the historical data, since it's our source of comparison. There's a summary of the several results in Table 4.9.

Table 4.9: Maximum Drawdowns of the Historical Data

Sample	**Maximum**	**Average**
9 Banks	29.435 %	11.34914 %
Bank 1	11.76489 %	-
Bank 5	14.199638 %	-

These results will be our references. The 9 Banks have a maximum MDD of approximately 29% of the maximum considered in the calculus, while Bank 1 registered 11% and Bank 5 recorded 14%. The latter two are represented in the corresponding time series, exhibited in Figure 4.14, in Bank 1's case, and in Figure 4.15, in Bank 5's case.

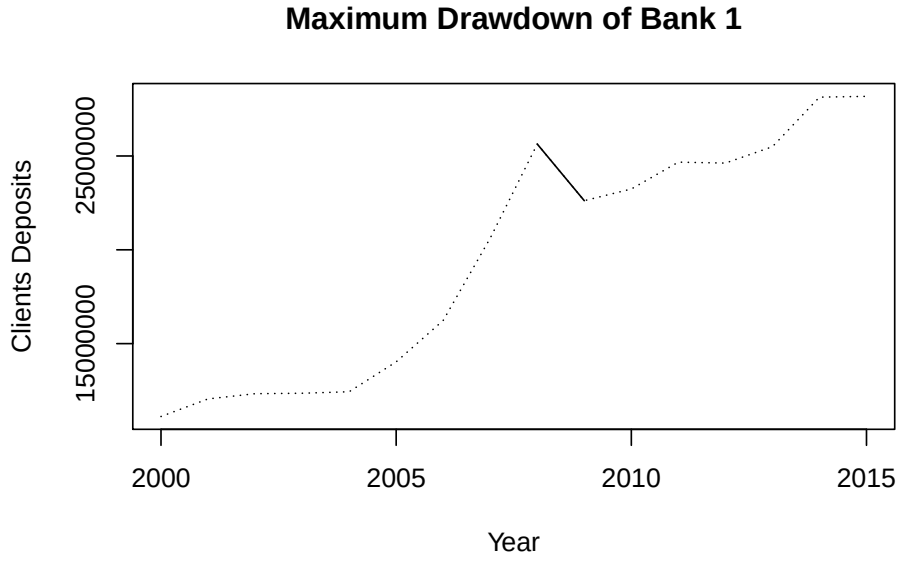


Figure 4.14: Bank 1's deposits time series (dotted line) and its values considered in the Maximum Drawdown calculus (solid line)

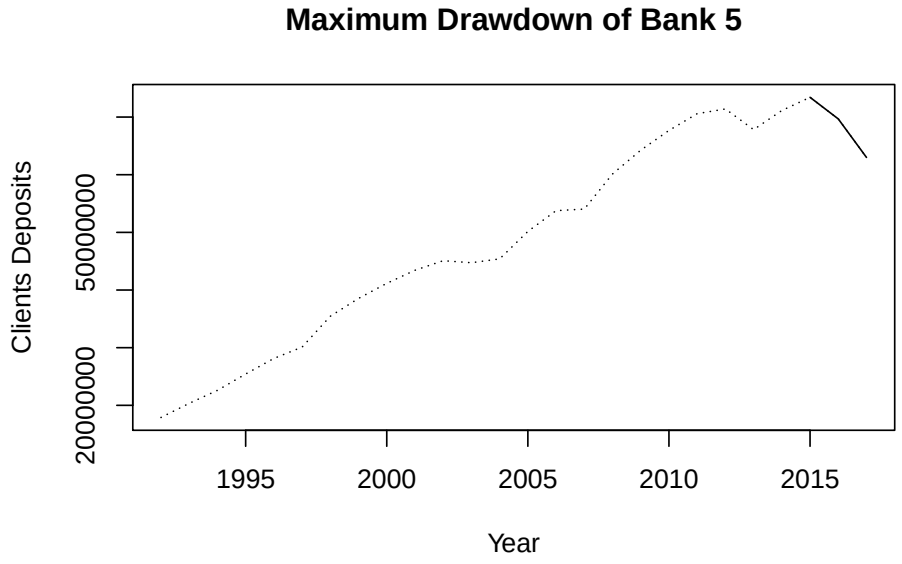


Figure 4.15: Bank 5's deposits time series (dotted line) and its values considered in the Maximum Drawdown calculus (solid line)

It is not intuitive to compare both MDD's: while the one from Bank 5 is in the end of the time series, thus representing a higher percentagem of a greater value (around 70 billion euros), the one from Bank 1 is approximately in the middle of the time series, representing a smaller percentage of a smaller value. However, we can notice that, since deposits time series tend to be growing, the further they are from the origin of the graphic, the smaller tend to be the percentage of the fall, because the same absolute value in comparison to two different maximums, will have a bigger impact in the smallest value. Thus, 14% of 70 billion euros is a much bigger fall than 11% of 25 billion euros.

Regarding the average MDD's, in the historical case, the individual banks only have one time series, hence, it is not reasonable to calculate its average. However, the set of the 9 Banks allows us to obtain that quantity, which is approximately 11%. This value is very close to the maximum MDD of Bank 1, which makes it an even more interesting bank to consider in the comparisons.

Regarding the simulated values, we obtained the MDD's for the referred distinct cases presented in Table 4.10 and in Table 4.11, respectively considering 30, and specifically the first 10, years from the simulations.

Table 4.10: Maximum Drawdowns of the Simulated Trajectories for 30 years

Sample	**Maximum**	**Average**
9 Banks	61.64756 %	31.42079 %
Bank 1 (AR(1))	15.92397 %	11.43346 %
Bank 5 (AR(1))	16.32471 %	8.964539 %

In the simulated trajectories for the 9 Banks, the worst case scenario had an MDD of approximately 61.65%, which is clearly greater than the remaining registered values. This happens because our model makes it possible to simulate critical events sooner. Since the trajectories don't reach high values quickly, the biggest drawdowns happen earlier, which causes the percentage of the fall to be bigger. However, this doesn't mean that there won't be bad events afterwards; they just won't have an effect as noticeable as in the first periods of the simulations. Furthermore, since our trajectories for the 9 Banks register more bad events throughout the whole series, its average is also greater in this case, with a result of approximately 31%.

Regarding the first 10 years of the simulations, we noticed that Bank 1 registered greater values than Bank 5, even though it had been the opposite case in the 30-year simulations. This happens because the previously recorded MDD had been registered in the first 10 years of the simulated trajectories.

Table 4.11: Maximum Drawdown of the first 10 years of the Simulations

Sample	Maximum	Average
9 Banks	44.36276 %	15.00251 %
Bank 1 (AR(1))	15.92397 %	7.294522 %
Bank 5 (AR(1))	9.236793 %	2.49694 %

It is also possible to notice that the tipically used model, the AR(1), has the smallest values both in average and in maximum. Its results don't seem to be realistic, since it can't simulate any crisis or sudden massive losses of money. Thus, it doesn't seem to be as reliable as the panel data AR(2) model.

4.3.2 Densities of the Simulated Deposits

Another form of comparison of the results of the simulations is by presenting them in density plots. This could either be by the graphical representation of the probabilities of the obtained values (hence density plots) or of the values themselves in histograms.

4.3.2.1 Histograms

We focused on simulations for 5 and 10 years, performing a total of 10 000 trajectories for panel data and for each of the individual banks and recorded the last observation from each trajectory for each case. The histograms are presented from Figure 4.16 to Figure 4.18.

Remark 4.3.1. *Before proceeding to the analysis, we should notice that the negative values of the deposits could be interpreted as bankruptcy or debts (for example, in case of external financing), as well as simply the loss of that amount in clients deposits (because, as it was already referred, the initial value of the trajectories is mainly illustrative).*

The simulations results show that the tails of the density plots are more pronounced with panel data. This represents a better adjustment of the models to the data. For instance, Bank 1 and Bank 5's simulations never consider events of bankruptcy, since the simulations never reach the zero level. This reinforces the idea that the simulations of the AR(1) model with only one bank are always too optimistic in the first time periods of the simulations. Once again, this is not a reliable assumption because, otherwise, every time the simulations would be created and there was a successful bank in study, there would never be a bad case in the simulations. There should always be pessimistic simulations as well when banks are stress testing so that they can better manage their activity.

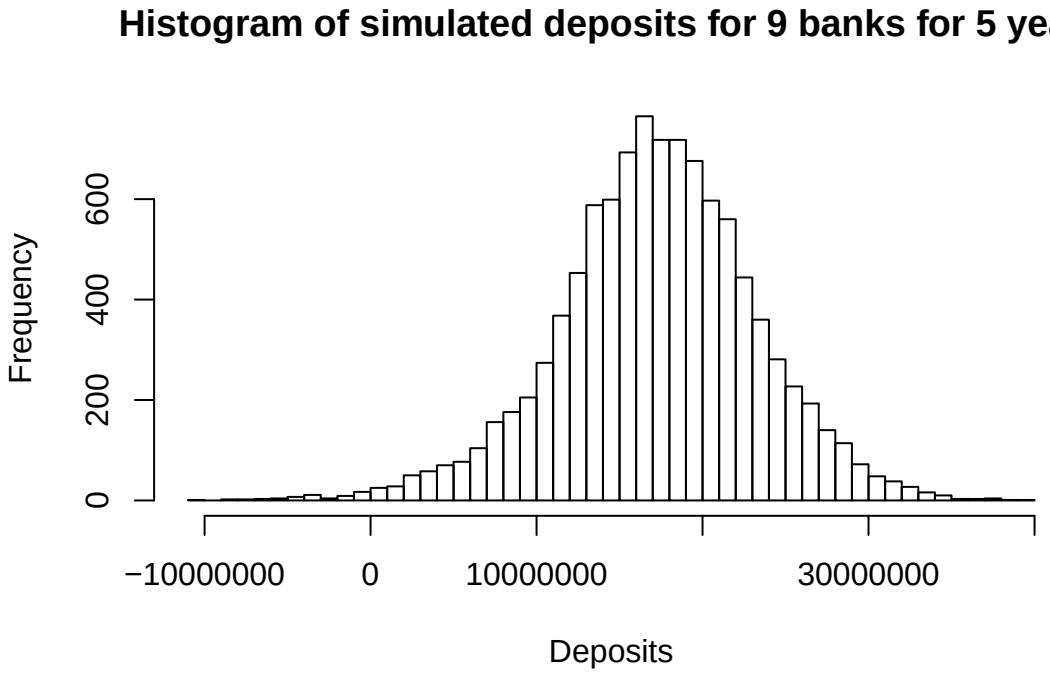


Figure 4.16: 9 Banks' histogram for 5 years

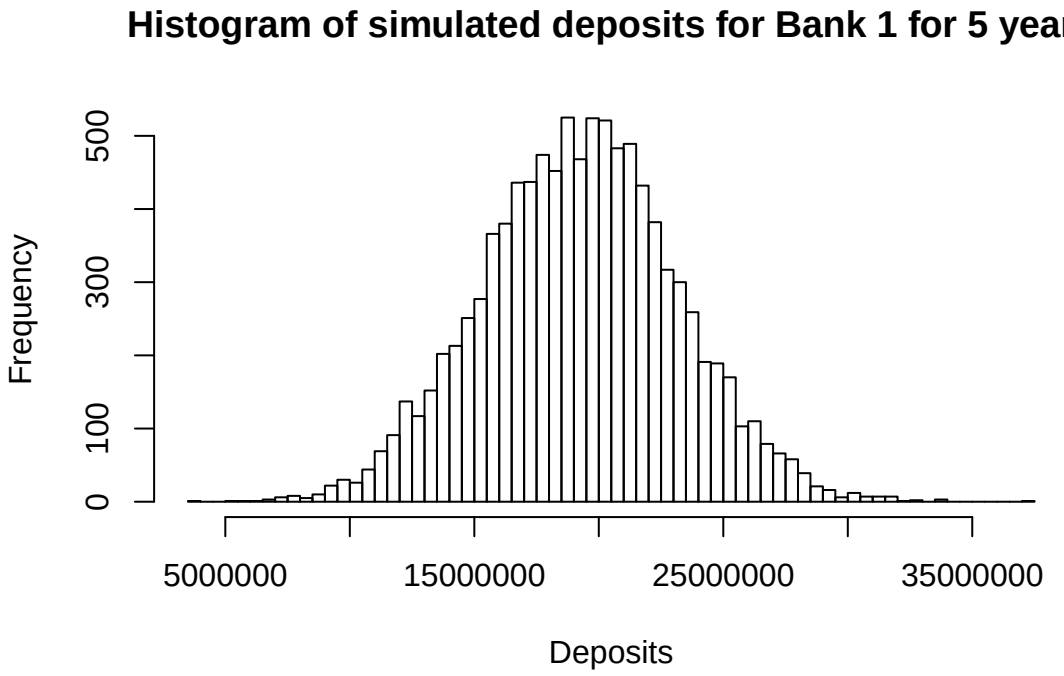


Figure 4.17: Bank 1's histogram for 5 years

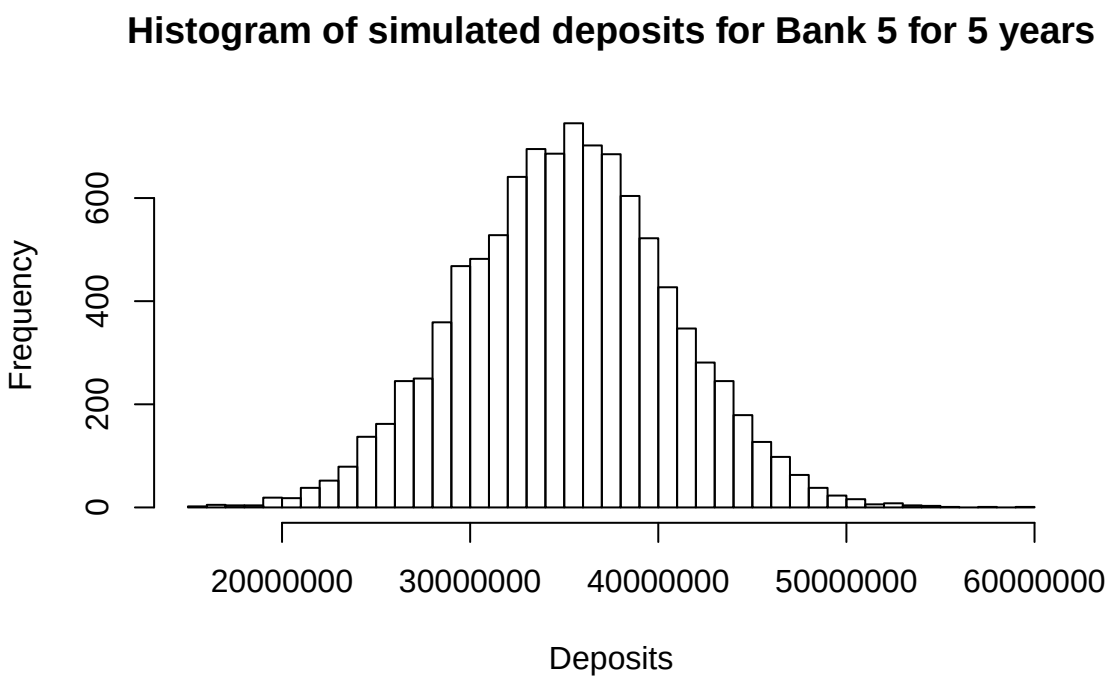


Figure 4.18: Bank 5's histogram for 5 years

In order to have a more detailed view of the results, we recorded the quantiles regarding the tails of the densities of the simulated deposits. These are presented in Table 4.12, below.

Table 4.12: Quantiles for the 5-year simulations of clients deposits

Sample	**1%**	**5%**	**95%**	**99%**
9 Banks	1 588 987	7 131 494	26 900 452	31 056 477
Bank 1	10 259 667	12 689 774	25 672 700	28 259 616
Bank 5	22 255 994	25 914 712	44 278 112	48 026 598

This table, together with the histograms, shows that the deposits in the first quantile have significant differences between each other. The 9 Banks register the minimum value with approximately 1.6 billion euros, followed by Bank 1 with 10 billion euros. This shows not only that the 9 Banks have the lowest values, reaching negative ones, but also the previously referred optimism of the AR(1) model. Regarding the medium quantiles, we can notice that even though the 9 Banks have the wider range, along with Bank 5, the first also has the longest tails. The centered mass has higher frequencies for each value while the tails evidence very small frequencies for a lot of other values. This evidences the different scenarios presented in the simulations. Finally, the highest quantiles also

highlight the spread of the tails, since even though the 9 Banks register the smallest 99% quantile, they reach values that are over 10 billion euros from that record.

By analysing these results, we calculated the probability of the banks reaching zero euros in clients deposits. With the 10 000 simulations for 5 years, this probability is approximately 1.23% in the 9 banks case, and 0% in the remaining cases. This shows that there is significant risk involved in the management of liquid assets. A 10 billion euros fall in clients deposits will certainly influence the ability of a bank to respond to such a financial crisis. Thus, the panel data model exhibits financial distress situations even in the beginning of the simulations.

We will now present the histograms for the 10-year simulations below, from Figure 4.19 to Figure 4.21.

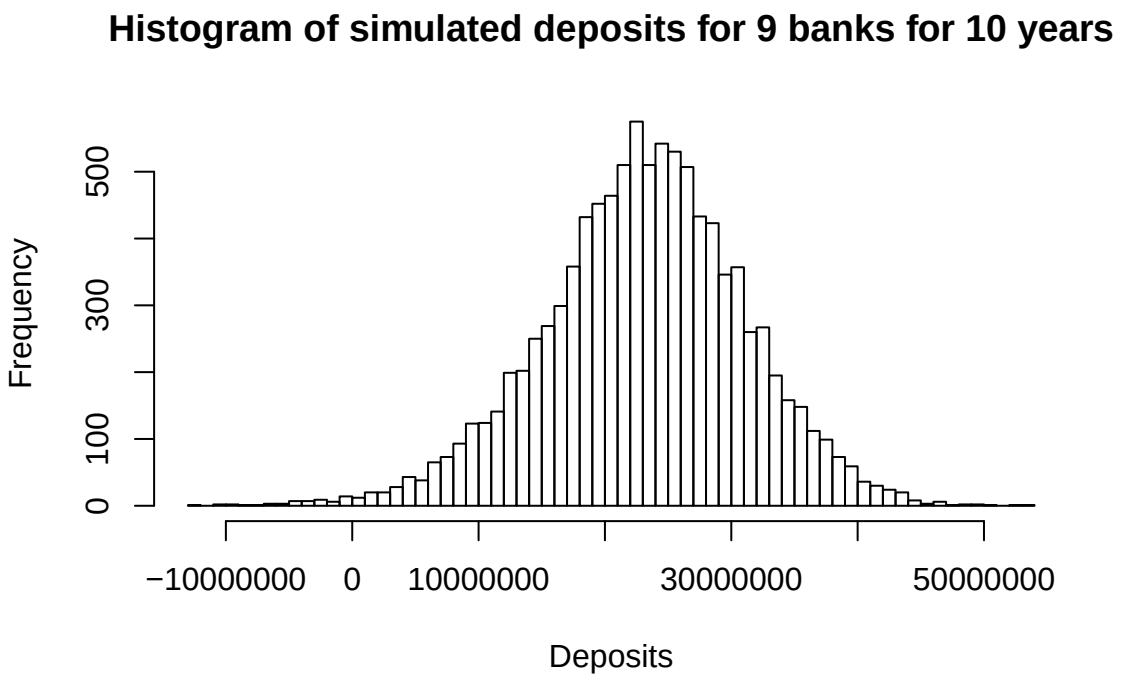


Figure 4.19: 9 Banks' histogram for 10 years

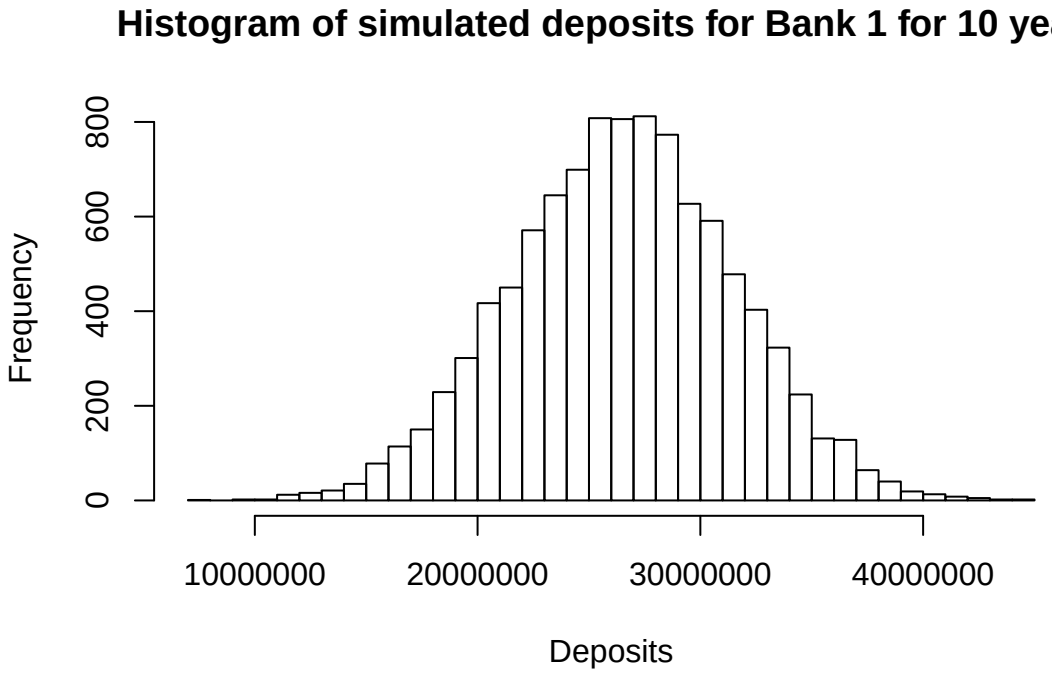


Figure 4.20: Bank 1's histogram for 10 years

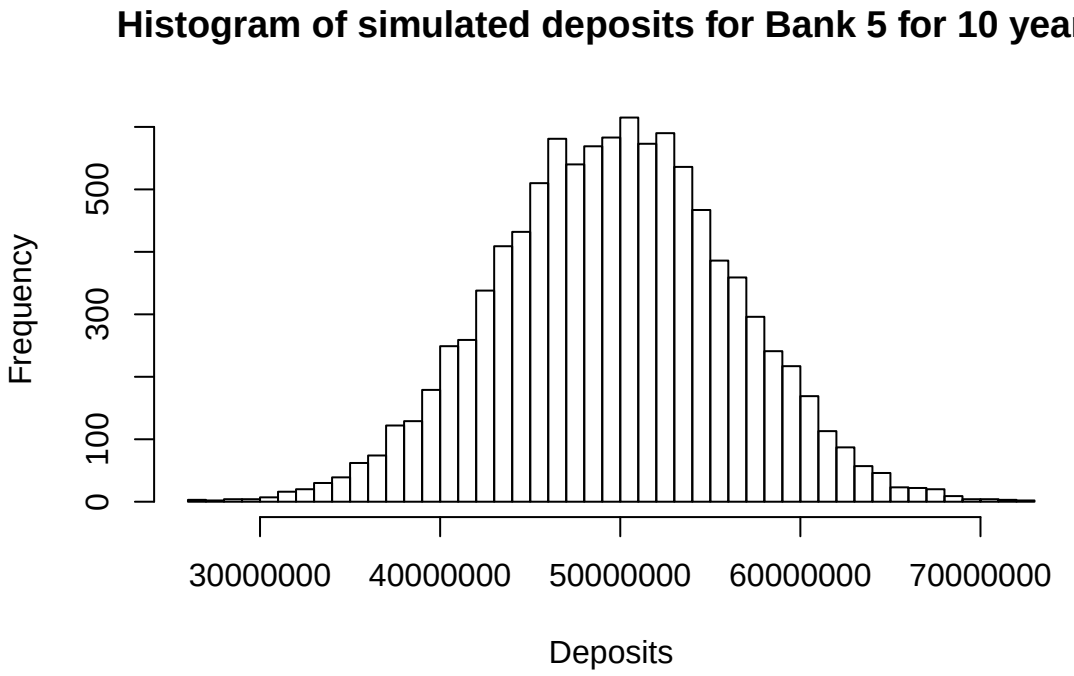


Figure 4.21: Bank 5's histogram for 10 years

In the graphics above, there is a reinforcement of the discussed points from the corresponding records for the 5-year horizon. That is, the densities of the values along the graphics are more concentrated in the panel data's case. There appears to be a centered mass of the most observed values, both in the panel data and in the individual banks cases. Moreover, the individual banks, whose residuals were considered to follow a Normal distribution, show a greater evidence of their distribution in the 10-year simulations as well.

These results also evidence what we have been stating so far: the AR(1) model with individual banks reflects the original time series. That is, the range of simulated values is similar to the original sample. This justifies the large shape of Bank 1's and Bank 5's histograms. Thus, panel data returns a more realistic result and we present the quantiles that evidence the tails of the densities shown in the histograms above. These details are exhibited in Table 4.13, below.

Table 4.13: Quantiles for the 10-year simulations of clients deposits

Sample	**1%**	**5%**	**95%**	**99%**
9 Banks	2 489 109	9 383 741	35 822 954	40 897 027
Bank 1	15 237 647	18 343 916	34 526 930	37 886 178
Bank 5	34 421 072	38 910 852	60 348 692	64 671 942

The results for the 10-year simulations, also considering the histograms, indicate the aspects already referred. That is, the 9 Banks register the minimum value in the lowest quantile again, revealing that these represent the only case in which the simulations attain negative values, emphasizing the worst case scenarios. Furthermore, in the medium quantiles, the results exhibit similar ranges in the panel data and Bank 5 cases again and evidence the largest spread of the tails in the 9 Banks case. That is, the medium quantiles are expected to show a certain variety of levels, while the tails of the densities should present a greater spread of values in the panel data case. This is evidenced when reaching the last considered quantile, where we can observe that the spread of the tails is bigger in the 9 Banks case, expanding in more than 10 billion euros. This might be justified by the fact that the simulations don't reach the highest values in all the trajectories, which is reasonable, since the survivorship bias is noticeable in the individual banks cases.

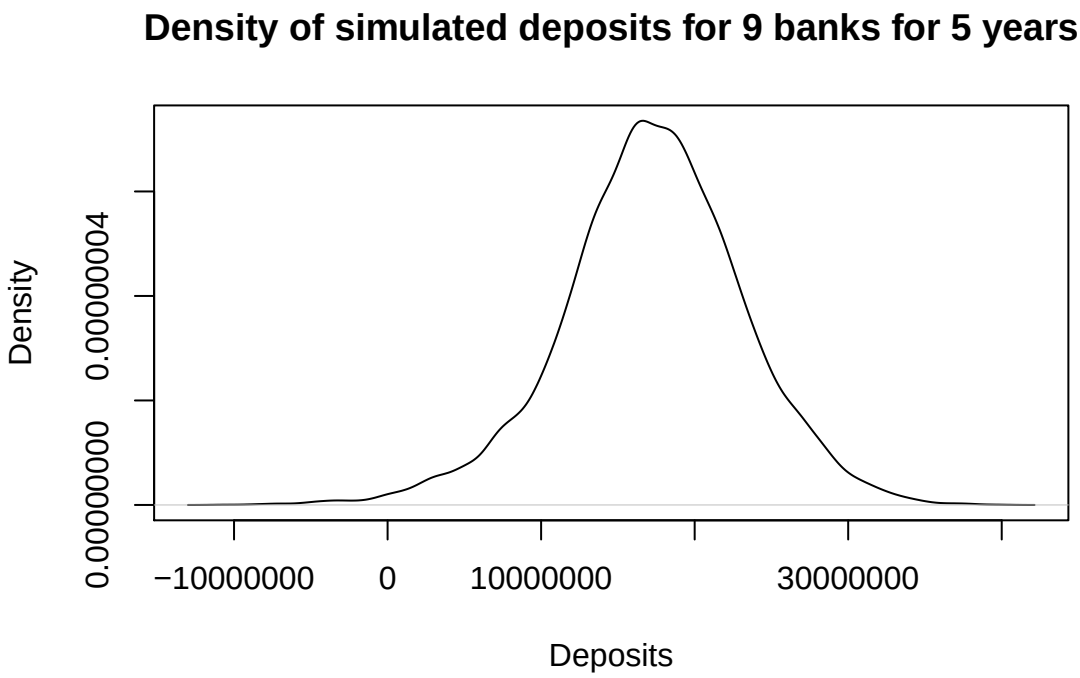
Therefore, the probability of the deposits reaching zero euros was calculated again for the three cases. In the 9 Banks, comparing to the 5-year simulations, this value rose to 1.77%, and maintained its level as 0% in the individual banks cases. Thus, the 10-year simulations also reinforced that the AR(2) model, considering the momentum, shows not only that bad events can happen, as well as that those are more likely to happen as time evolves.

4.3.2.2 Density Plots

The density plots help to see the shape of the histograms. Thus, we have grouped them in pairs, by category, considering the 5 and 10-year simulations, in order to compare their evolution in time. We present them from Figure 4.22 to Figure 4.24, where it is easier to notice the shapes of the densities of the simulated deposits. Consequently, the simulations of the 9 Banks show a greater concentration of the deposits in a smaller range, having a bigger spread of the tails, especially the left one, while in Bank 1 and in Bank 5 there is a greater expansion of the values in the center of the respective densities.

Regarding the evolution of each case from the 5-year to the 10-year simulations, it is possible to verify that the 9 Banks reach a higher mean of their clients deposits and not a higher minimum value in the left tail. This evidences the ability of the model to simulate both good and bad scenarios. Furthermore, the assumed values of the simulated deposits are more spread even in the center of mass of the distribution. That is, the values around its mean have smaller densities but include more values with higher densities.

The description associated to the 9 Banks is not completely relatable to the individual banks' cases, as in both Bank 1 and Bank 5 the graphics seem to translate to the right. Noticing by the tails and the means of the respective densities, the simulations only reach higher values, both in minimum as in maximum. Thus, their means are also greater, but their left tails don't reach similar values as in the 5-year simulations. Moreover, the densities are even similar from the 5-year to the 10-year simulations. This reflects the consequences of using an AR(1) model with a single time series, specifically a constantly growing one.



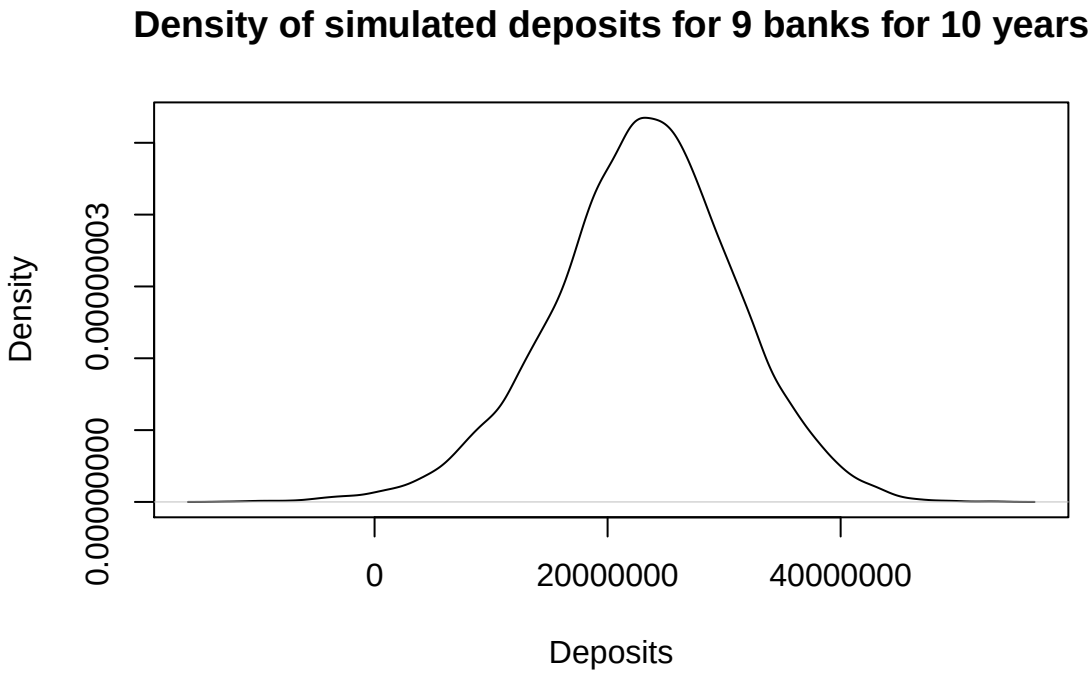
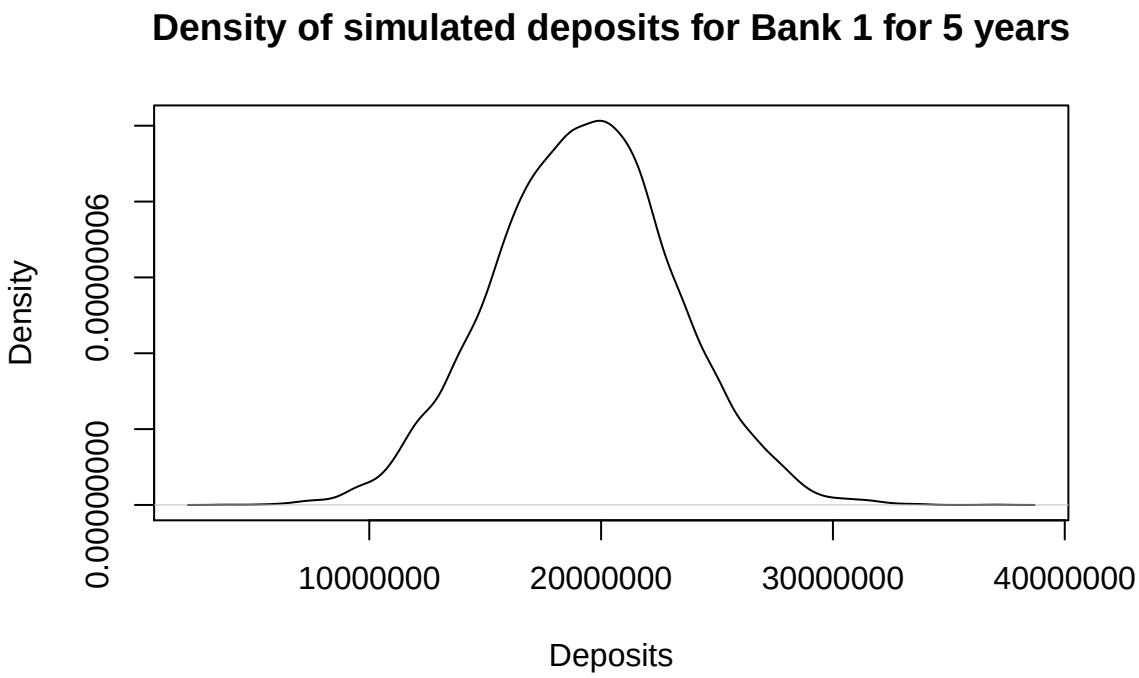


Figure 4.22: 9 Banks' density plots for 5 and 10 years



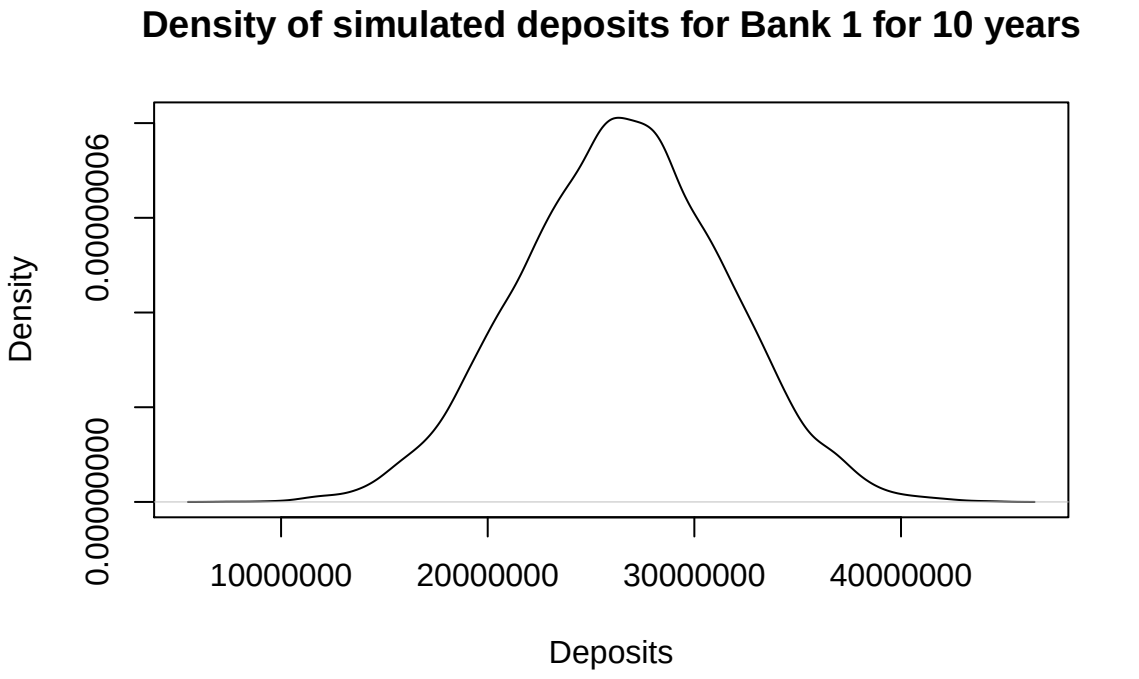
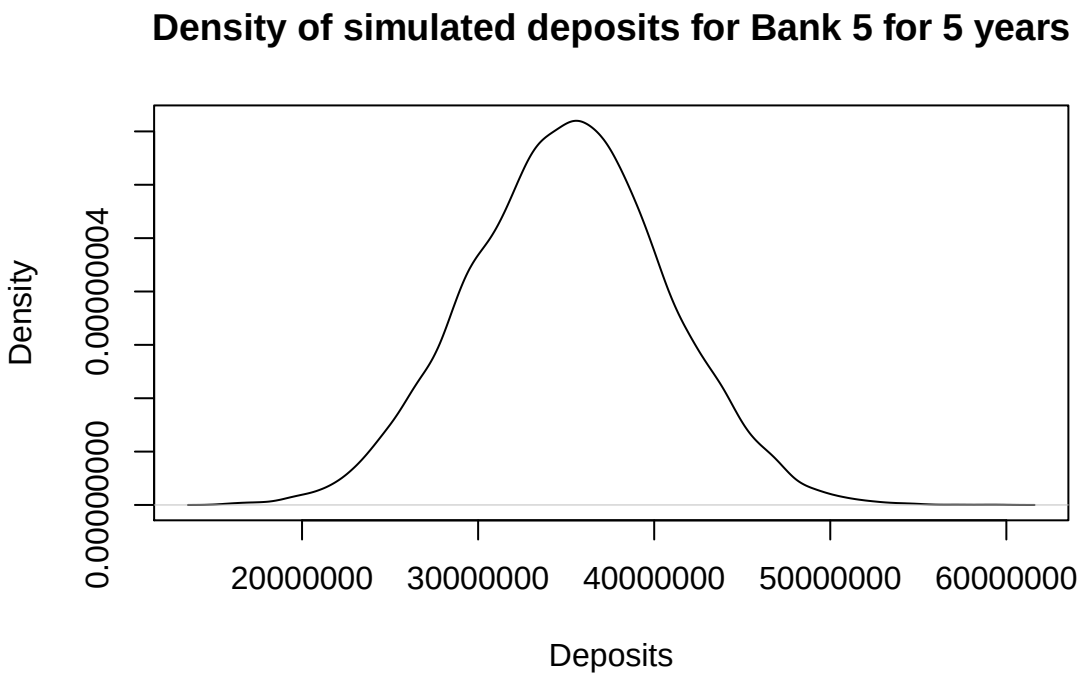


Figure 4.23: Bank 1's density plot for 5 and 10 years



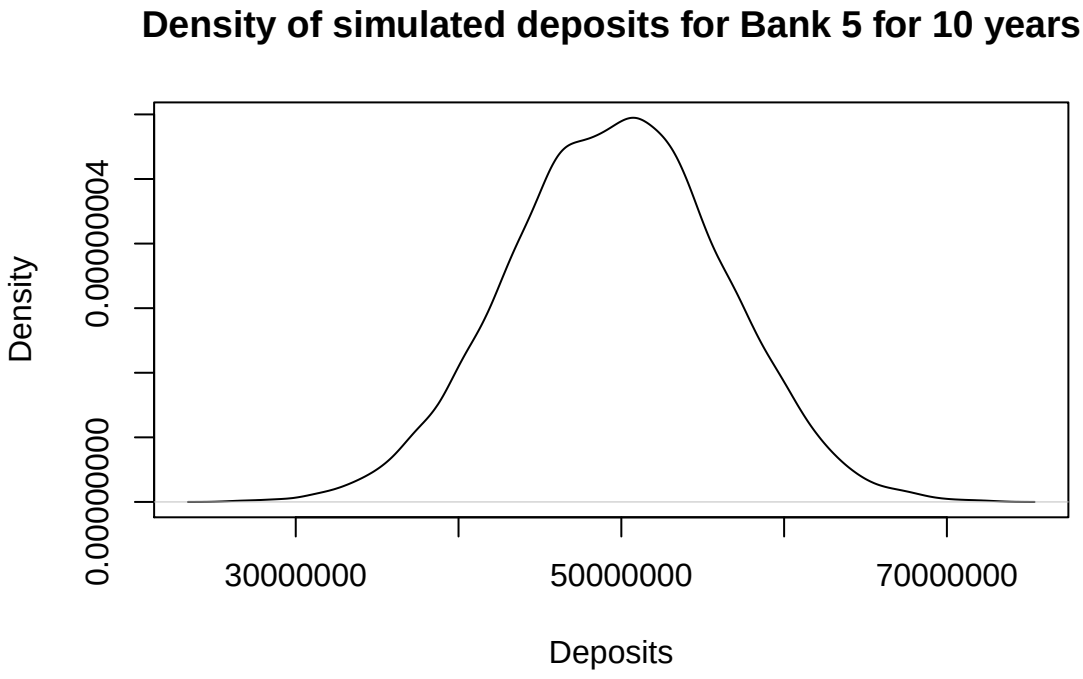


Figure 4.24: Bank 5's density plot for 5 and 10 years

Chapter 5

Conclusions

The goal of this research was to develop a panel data model that could explain the evolution of banks' clients deposits using the momentum. This would be an innovation in opposition to the usually referred AR(1) model with only one time series and it would be important because it is an arising issue to study liquidity risk. In this subject, it is relevant to try to understand how non-maturity deposits evolve throughout time, so that banks can optimize their activity management.

Benbachir and Hamzi (2016) had already showed that the Autoregressive model of order 2 could be relevant when explaining the non-maturity deposits. However, this idea never got to be developed. Thus, we decided to include a momentum term in the model, which is commonly used in the financial area, in order to generate a better interpretation and comprehension of the deposits models.

In order to develop our model, we collected data on 51 Portuguese banks, restricting it to the largest ones. That is, we only used a sample comprised of banks that had an average of clients deposits of at least 10 billion euros, which were only 9. From this sample we were able to use 129 observations, although only 111 residuals resulted from the estimation. However, having a small sample wasn't an issue because, since we were using panel data, the diversified information from several banks would compensate its quantity.

While developing our model, after checking its significance, we had to verify that certain requirements were being met. For instance, the data had to be stationary and the residuals of our model couldn't suffer from autocorrelation. In order to validate our model, we implemented several tests, that included the Shapiro-Wilk test for normality and the Ljung-Box test for autocorrelation. Furthermore, we used the *R* software to

perform inference, validate our model and analyze our sample. This procedure was also applied to the models used in the comparisons, both an AR(1) and an AR(2) models with a single time series for each.

Regarding the computational simulations, we generated 9 trajectories for our model and 9 trajectories for each AR(1) model concerning the individual banks. These were selected from our sample in order to make comparisons. The AR(2) models with just one bank were excluded from the research because their second coefficient weren't significant in the estimations.

The results showed that the panel data model, in addition to showing a growing evolution of the deposits, would present more diversified case scenarios, including worse ones, in the simulations, which goes against the frequently used models and samples. That is, the problem with using AR(1) models with only one time series, is that the simulations won't be significantly different from the observed data. Therefore, if that model only uses a constantly growing series in the sample, for example, then it won't be able to simulate falls for the future.

When simulating models with stress testing purposes or when trying to manage banks' activity efficiently, a wide range of possibilities is pursued. That is, the dynamics of liquidity risk must be taken into account and that can't be achieved when considering the actions of only one institution. A variety of cases must be considered when trying to understand what could go wrong in financial operations.

When comparing the different models, we considered two indicators: the Maximum Drawdown and the probability density of the simulated deposits. The Maximum Drawdown is a commonly used indicator in finance, as it highlights the downfalls of the time series. It considers the falls from each local maximum to the following local minimum that arises before a new maximum is attained. This indicator is useful because it helps to compare bad scenarios and to identify whether these happened in the beginning, in the middle or in the end of the trajectories, facilitating comparisons in this matter. The probability density of the simulated deposits helps to understand whether the simulations are realistic or not. By comparing the results, it is possible to infer which models present the best results, both in the range of simulated possibilities and in the concentration of the deposits.

The panel data model showed greater drawdowns, and also simulated episodes of significant financial distress, as reinforced in the histograms and density plots. Therefore, it helps to simulate more diversified scenarios and it is a better tool to manage banks' activity, since it accounts the various possibilities associated with liquidity risk.

For future work, it might also be interesting to analyze the results of a momentum model with small and medium banks, also using a panel data sample in order to create diversified scenarios.

Bibliography

Benbachir, S., & Hamzi, M. M. E. (2016). Non-maturity deposit modeling in the framework of asset liability management. *International Journal of Economics and Financial Research*, *2*(5), 79–98. Retrieved from https://EconPapers.repec.org/RePEc:arp:ijefrr:2016:p:79-98

Birge, J. R., & Júdice, P. (2013). Long-term bank balance sheet management: Estimation and simulation of risk-factors. *Journal of Banking & Finance*, *37*(12), 4711–4720. doi:10.1016/j.jbankfin.2013.0

Box, G., & Ljung, G. (1978). On a measure of lack of fit in time series models. *Oxford University Press*, *65*(2), 297–303. Retrieved from https://www.jstor.org/stable/2335207

Box, G., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, *65*(332), 1509–1526. Retrieved from https://www.jstor.org/stable/2284333

Box, J. E. P., Jenkins, G. M., & Reinsel, G. C. (1970). *Time Series Analysis - Forecasting and Control* (Third). Prentice-Hall International, Inc.

Breusch, T. S., & Pagan, A. (1980). The lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies*, *47*(1), 239–253. Retrieved from https://EconPapers.repec.org/RePEc:oup:restud:v:47:y:1980:i:1:p:239-253.

Croissant, Y., & Millo, G. (2008). Panel data econometrics in r: The plm package. *Journal of Statistical Software, Articles*, *27*(2), 1–43. doi:10.18637/jss.v027.i02

Crombez, J. (2001). Momentum, rational agents and efficient markets. *Journal of Behavioral Finance*, *2*, 190–200. doi:10.1207/S15327760JPFM0204_3

Fu, P. H., & Feng, A. K. (1985). Prediction and analysis of the balance of savings deposits of residents based on arima. *International Journal of Research and Reviews in Applied Sciences*, *19*(1), 114–118.

Hałaj, G. (2016). *Dynamic balance sheet model with liquidity risk* (Working Paper Series No. 1896). European Central Bank. Retrieved from https://ideas.repec.org/p/ecb/ecbwps/20161896.html

Hansen, B. E. (1994). *Time Series Analysis*. Princeton University Press.

Hansen, B. E. (2000). *Econometrics* (July 10, 2019). University of Wisconsin.

Honda, Y. (1985). Testing the error components model with non-normal disturbances. *The Review of Economic Studies*, *52*(4), 681–690. Retrieved from http://www.jstor.org/stable/2297739

Janosi, T., Jarrow, R. A., & Zullo, F. (1999). An empirical analysis of the jarrow-van deventer model for valuing non-maturity demand deposits. *The Journal of Derivatives*, *7*(1), 8–31. doi:10.3905/jod.1999.319107. eprint: https://jod.iijournals.com/content/7/1/8.full.pdf

Jarque, C., & Bera, A. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, *55*(2), 163–172.

Jarrow, R., & van Deventer, D. (1998). The arbitrage-free valuation and hedging of demand deposits and credit card loans. *Journal of Banking & Finance*, *22*(3), 249–272. Retrieved from https://EconPapers.repec.org/RePEc:eee:jbfina:v:22:y:1998:i:3:p:249-272

Jegadeesh, N., & Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance*, *56*(2), 699–720. Retrieved from http://www.jstor.org/stable/222579

Lipton, A. (2015). Modern monetary circuit theory, stability of interconnected banking network, and balance sheet optimization for individual banks. *International Journal of Theoretical and Applied Finance*. doi:10.1142/S0219024916500345

O'Brien, J. M. (2000). *Estimating the value and interest rate risk of interest-bearing transactions deposits* (Finance and Economics Discussion Series No. 2000-53). Board of Governors of the Federal Reserve System (US). Retrieved from https://EconPapers.repec.org/RePEc:fip:fedgfe:2000-53

R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.R-project.org/

Ribeiro, C. S. (2014). *Econometria*. Escolar Editora.

Royston, J. (1982). An extension of shapiro and wilk's w test for normality to large samples. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *31*(2), 115–124.

Shapiro, S., & Wilk, M. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591–611.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). MIT Press Books. The MIT Press.

Appendix 1

Table A.1: Data - Part 1

Year	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5
1992					17893172.45
1993					20344784.07
1994					22597769.38
1995			16655819.48		25427858.86
1996			17845118.26		28136545.92
1997			18848445.25		30099764.57
1998			18489051.39		35450000.5
1999			19669012.68		38545789.65
2000	11115848	28920221			41160378
2001	12053116	29441050			43425050
2002	12330930	27088044			45083857
2003	12355632	30623978			44733023
2004	12435609	33608210			45403221
2005	14028451	34395431			50161963
2006	16235505	33244197			53767835
2007	20621866	39246611			54038767
2008	25633620	44907168			60127756
2009	22617852	46307233			64255685
2010	23240863	45609115			67680045
2011	24671328	47516110		13701919	70587491
2012	24621139	49389866		13255447	71404154
2013	25494961	48959752		14142828	67824469
2014	28134617	49816736		14314659	71134176
2015	28177814	51538583		12969431	73426264
2016		48797647		12467819	69680130
2017					63000000

Table A.2: Data - Part 2

Year	Bank 6	Bank 7	Bank 8	Bank 9
1992	5152522.421			8026965.014
1993	6272662.882			8493011.841
1994	6964061.612			9203145.42
1995	9033978.113			10277147.08
1996	10414964.93			11108872.62
1997	11841078			11378418.01
1998	13073507.85			10972645.92
1999	14511936.23			11337721.09
2000	16159751			
2001	17394740			
2002	18667656			
2003	20136614			
2004	20371090	12953161	13851659	
2005	20753083	12247389	15217252	
2006	21993671	11082844	15622396	
2007	23775030	11459761	16033144	
2008	26386754	15301954	15700248	
2009	25446450	15253588	15081297	
2010	30819220	18262476	17018297	
2011	34206162	20098566	19073613	
2012	34540323	21395469	19659923	
2013	36830893	20690967	19271178	
2014	27838824	21597821	20345997	
2015	27582142	27488734	26017806	
2016	25989719	29094675	27672590	
2017				

```
> Depositos.AR2pooled <- plm(Depositos ~ lag(Depositos,1) + lag(Depositos,2),
data=Depositos, model = "pooling")
> summary(Depositos.AR2pooled)
Pooling Model

Call:
plm(formula = Depositos ~ lag(Depositos, 1) + lag(Depositos,
    2), data = Depositos, model = "pooling")

Unbalanced Panel: n = 9, T = 3-24, N = 111

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-10286046  -1097245    -20666   1135762   5445063

Coefficients:
                      Estimate  Std. Error t-value Pr(>|t|)
(Intercept)        1.2293e+06  4.2783e+05  2.8734  0.00489 **
lag(Depositos, 1)  1.2082e+00  1.0166e-01 11.8851  < 2e-16 ***
lag(Depositos, 2) -2.2016e-01  1.0328e-01 -2.1318  0.03529 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    3.1957e+16
Residual Sum of Squares: 5.6964e+14
R-Squared:      0.98217
Adj. R-Squared: 0.98184
F-statistic: 2975.38 on 2 and 108 DF, p-value: < 2.22e-16
```

Figure A.1: Results of the estimation of an AR(2) model for the 9 banks

```
> modelounico = lm(valores ~ lag(valores, 1), data= valores)
> summary(modelounico)

Call:
lm(formula = valores ~ lag(valores, 1), data = valores)

Residuals:
     Min       1Q   Median       3Q      Max
-3841117 -1055705  -345080   662137  3949189

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     2.039e+06  1.705e+06   1.195    0.253
lag(valores, 1) 9.527e-01  8.542e-02  11.152 5.01e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1988000 on 13 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.9054,    Adjusted R-squared:  0.8981
F-statistic: 124.4 on 1 and 13 DF,  p-value: 5.005e-08
```

Figure A.2: Results of the estimation of an AR(1) model for Bank 1

```
> ar2model <- lm(valores ~ lag(valores, 1) + lag(valores,2), data= valores)
> summary(ar2model)

Call:
lm(formula = valores ~ lag(valores, 1) + lag(valores, 2), data = valores)

Residuals:
     Min       1Q   Median       3Q      Max
-4113688 -1225391    58608   638535  3646275

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.307e+06  2.016e+06   1.144   0.2768
lag(valores, 1)  1.024e+00  3.009e-01   3.402   0.0059 **
lag(valores, 2) -8.804e-02  3.079e-01  -0.286   0.7802
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2145000 on 11 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.8929,    Adjusted R-squared:  0.8734
F-statistic: 45.83 on 2 and 11 DF,  p-value: 4.623e-06
```

Figure A.3: Results of the estimation of an AR(2) model for Bank 1

```
> modelounico51 = lm(valores51 ~ lag(valores51, 1), data= valores51)
> summary(modelounico51)

Call:
lm(formula = valores51 ~ lag(valores51, 1), data = valores51)

Residuals:
     Min       1Q   Median       3Q      Max
-7124085 -1317728   -10616  2335060  4640658

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       4.918e+06  1.577e+06   3.119  0.00482 **
lag(valores51, 1) 9.358e-01  3.054e-02  30.644  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2706000 on 23 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.9761,    Adjusted R-squared:  0.9751
F-statistic: 939.1 on 1 and 23 DF,  p-value: < 2.2e-16
```

Figure A.4: Results of the estimation of an AR(1) model for Bank 5

```
> modelounico52 <- lm(valores52 ~ lag(valores52, 1) + lag(valores52,2),
data= valores52)
> summary(modelounico52)

Call:
lm(formula = valores52 ~ lag(valores52, 1) + lag(valores52, 2),
    data = valores52)

Residuals:
     Min       1Q   Median       3Q      Max
-4958622 -1333722  -312917  1492367  5305413

Coefficients:
                  Estimate Std. Error t value  Pr(>|t|)
(Intercept)       3.888e+06  1.887e+06   2.060     0.052 .
lag(valores52, 1) 1.333e+00  2.505e-01   5.321 0.0000282 ***
lag(valores52, 2) -3.923e-01  2.417e-01  -1.623     0.120
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2653000 on 21 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.9758,    Adjusted R-squared:  0.9735
F-statistic: 423.5 on 2 and 21 DF,  p-value: < 2.2e-16
```

Figure A.5: Results of the estimation of an AR(2) model for Bank 5