FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

LUISA PASSEBON MENDONÇA

ALTERNATIVE BETA MODEL IN PRACTICE: THE BRAZILIAN FUND INDUSTRY CASE

SÃO PAULO

2021

LUISA PASSEBON MENDONÇA

ALTERNATIVE BETA MODEL IN PRACTICE: THE BRAZILIAN FUND INDUSTRY CASE

Dissertação apresentada ao Programa de Mestrado Profissional da Escola de Economia de São Paulo da Fundação Getulio Vargas, como requisito para a obtenção do título de Mestre em Economia.

Área de concentração:
Engenharia Financeira.

Orientador:
Prof. Dr. Afonso de Campos Pinto

SÃO PAULO

2021

Mendonça, Luisa Passebon.
Alternative Beta Model in Practice: the Brazilian Fund Industry Case / Luisa Passebon Mendonça. – 2021.
74 f.

Orientador: Afonso de Campos Pinto.
Co-orientador: Alessandro Martim Marques.
Dissertação (mestrado profissional MPFE) – Fundação Getulio Vargas, Escola de Economia de São Paulo.

1. Fundos de investimento - Brasil. 2. Investimentos - Análise. 3. Investimentos - Administração. 4. Fundos hedge. 5. Kalman, Filtragem de. I. Pinto, Afonso de Campos. II. Marques, Alessandro Martim. III. Dissertação (mestrado profissional MPFE) – Escola de Economia de São Paulo. IV. Fundação Getulio Vargas V. Título.

CDU 336.767(81)

Ficha Catalográfica elaborada por: Isabele Oliveira dos Santos Garcia CRB SP-010191/O

Biblioteca Karl A. Boedecker da Fundação Getulio Vargas - SP

LUISA PASSEBON MENDONÇA

ALTERNATIVE BETA MODEL IN PRACTICE: THE BRAZILIAN FUND INDUSTRY CASE

Dissertação apresentada ao Programa de Mestrado Profissional da Escola de Economia de São Paulo da Fundação Getulio Vargas, como requisito para a obtenção do título de Mestre em Economia.

Área de concentração:
Engenharia Financeira.

Data da Aprovação: 21 / 09 / 2021

Banca Examinadora:

Prof. Dr. Afonso de Campos Pinto
(Orientador)
FGV - EESP

Prof. Dr. Alessandro Martim Marques
(Co-Orientador)
FGV - EESP

Prof. Dr. Alan de Genaro Dario
FGV - EAESP

ACKNOWLEDGEMENTS

Agradeço aos meus colegas de turma e aos meus professores do curso de mestrado pela troca de conhecimentos e experiências que tornaram este curso ainda mais interessante e enriquecedor, ao professor Alessandro Marques, pela orientação deste trabalho, ao meu irmão, Gustavo Mendonça, pelo apoio e motivação ao longo desta jornada, e aos meus pais por todo apoio.

ABSTRACT

The present work introduces practical aspects into the *alternative beta* methodology in order to bring its implementation closer to a real life experience. The literature focus mostly on theoretical aspects with little attention to practical usage. The methodology developed in this work touches several practical aspects of the investment process in order to be closer to real life replication experience. We implemented the algorithm using the Kalman filter in order to replicate eight Brazilian Multimarket funds between January 2015 and December 2020. We were able to closely track and, in half of the period, even outperform the original funds while observing the impact of the real life details on the model performance.

Keywords: Alternative Beta. Return Replication. Hedge Fund Tracking. Kalman Filter in Finance.

RESUMO

O presente trabalho introduz aspectos práticos na metodologia do *alternative beta* de modo a aproximar a sua implementação de uma experiência real. A literatura se concentra principalmente em aspectos teóricos, com pouca ênfase ao uso prático. A metodologia desenvolvida neste trabalho aborda diversos aspectos práticos do processo de investimento de forma a torná-la mais próxima de uma experiência de replicação real. Implementamos o algoritmo utilizando o filtro de Kalman para replicar oito fundos Multimercado brasileiros entre janeiro de 2015 e dezembro de 2020. Conseguimos acompanhar de perto e, na metade do período, superar os fundos originais, enquanto observamos o impacto dos elementos práticos na performance do modelo.

Palavras-chave: Alternative Beta. Replicação de Retornos. Tracking de Hedge Fund. Filtro de Kalman em Finanças.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

Over the past decade, the fund industry has developed with the introduction of new products and strategies, providing a wider range of investment alternatives in the market. In this context, the hedge funds were particularly prominent due to their ability to implement financial technologies and unconventional and more complex strategies.

The fund's return behavior depends on the asset allocation in the portfolio. The portfolio, however, works as a black box with unknown elements, we cannot observe in which assets the fund is investing. By studying the return's behavior, the Capital Asset Pricing Model (CAPM) determined that it can be explained by mainly two sources. The main source would be the exposures to systematic risk, the market influence on the performance, known as the beta component. The remaining part is the non-systematic risk, which is attributed to the managers' skills when determining the investment strategies and referred to as alpha.

Based on the CAPM conclusion that the main source for the returns variation would come from beta, Sharpe (1992) proposed the style analysis methodology, a way of making it possible to unveil the fund's portfolio composition and interpret the fund's trading strategy. By applying a style regression analysis, Sharpe demonstrated how to decompose a mutual fund's returns in order to identify which risk exposures are present in its portfolio.

The literature later evolved to the *alternative beta* concept, the idea that, given that we can identify the risk factors that generates beta, by investing on these factors, respecting their proportion in the portfolio, we are able to reproduce the fund's return. It is important to note that this methodology does not seek to replicate the fund's investment strategy nor invest on the same assets, but to replicate the fund's risk-return profile.

The *alternative beta model* becomes particularly interesting once we consider how it would be able to address important issues identified in the fund industry. Firstly we can highlight the lower costs. Funds must charge high performance fees in order to cover for their managing costs. The *alternative beta*, on the other hand, does not have managing costs, since there is no investment decision outside the designed algorithm.

Secondly, by implementing the model, the risk-return profile delivered would be similar, however the investor would not have to deal with high lock-up periods derived out of the investment in illiquid assets. The *alternative beta*'s fundamental hypothesis in to invest only on exchange-traded and highly liquid assets, leading to an easier investment and disinvestment process for the client.

The last advantage is the higher transparency. This issue is specially important for pension and sovereign funds, given the regulatory investment restrictions that prevents them from concentrating on certain market risks, making specific strategies, like day trades, and investing on certain jurisdictions. Not knowing which exposures are present on the fund's portfolio may inhibit them from making this investment. The *alternative beta* addresses this matter once it identifies the risk factors from the fund's portfolio in order to make the replication.

Since the first works on the *alternative beta*, the literature on the theme has developed mostly on the theoretical framework, specially concerning the models used for the implementation. Going from linear regression to linear regression with rolling window, and later to the Kalman filter approach, the literature aimed to study how the model's choice could affect the results. Still little has been explored on the issues of an actual implementation and how practical aspects could impact the replication performance.

The objective of this work is to fill this gap in the literature by designing a replication algorithm that will create similar conditions to a practical experience and use it to simulate an investment for the Brazilian fund industry by replicating Brazilian Multimarket funds. We aim to understand how much the introduction of practical aspects can impact the replication performance and how well the *alternative beta* track the chosen funds.

This work is divided in the following way: in Chapter 2, we review the development of the academic literature up to the creation of the *alternative beta model* and the works about the concept. Chapter 3 introduces the theoretical framework behind the *alternative beta model* and the Kalman filter model. Chapter 4 discusses the methodology used to implement the model for the funds replication. Chapter 5 presents the estimation results and the discoveries made in the replication process. Chapter 6 presents our conclusion and final remarks.

2 LITERATURE OVERVIEW

In this chapter we discuss the evolution of the literature from the first works on style analysis and mutual funds returns to the concept of the *alternative beta model* as we will approach in the present work.

2.1 Style Analysis

The main concept that gives rise to the *alternative beta model* was first presented by Sharpe (1992). The idea behind this paper is that a hedge fund's portfolio works as a black box, we don't truly know which assets are part of its portfolio, but having this knowledge means understanding how investment decisions are made. Sharpe intended to demonstrate it is possible to apply a style regression analysis to decompose a mutual fund's returns in order measure how much of the portfolio's returns each asset class factor can explain.

Sharpe then introduced what he called the Factor Model, a linear regression that models the relationship between asset classes factors and the observed hedge fund's returns. He tested the model using twelve asset classes representing a market capitalization-weighted returns' index built with a large number of securities. The out-of-sample performance concluded the style analysis can be successful in capturing a substantial part of the portfolio's returns.

Although Sharpe's model is able to explain the risk exposures inside a hedge fund's portfolio, it isn't possible to directly implement it on an investment. The academic works that followed the factor model can be divided into two different groups: (i) The style analysis literature, which aimed to develop Sharpe's model, and (ii) The *alternative beta* methodology, which attempted to evolve another direction by creating the necessary links between the style analysis and the investment in practice, making it possible for the replication process.

Still in the style analysis literature, Schutt and Caldeira (2016) implemented the idea with a Kalman Filter model introducing time-varying exposures in Brazilian hedge funds for the 2006 to 2011 period. The selected factors were based on onshore assets and the results showed the fund's strategies were centered in stock market but there was a rising exposure to interest rates. The model was able to explain 50% of the fund's returns and the authors discussed that the remaining part that could not be explained was attributed to the manager's abilities.

2.2 Alternative Beta Methodology's Development

Amenc et al. (2003) attempt to prove that, by applying a factor model, it is possible to predict the hedge fund's returns. Differently from Sharpe (1992), the authors seek to extend the Factor Model by introducing market variables that represents the financial risks in the model and allowing for a real investment strategy. This was an important step towards the concept of *alternative beta*, since the proposed model understands the fund's exposures and replicates these risk factors with a group of selected assets.

The authors go further to analyze if tactical style allocation and transaction costs could affect the results, both critical matters for the model's employment. They concluded that rebalancing the portfolio asset weights can be beneficial for the replication results and that transaction costs are not a major concern and can be overruled.

Kat (2007) claims that the Factor Model isn't the only approach on hedge fund replication, since the *alternative beta* could be implemented with other two methodologies, the mechanical trading rule and the FundCreator. Both alternatives are described as ways to address the issues identified in the Factor Model, such as not accessing the liquidity factor given the model's hypothesis of only investing in liquid assets, the lack of dynamic trading derived from the dependency on historical data and the assumption that the relationship between the risk factors and fund returns is not always linear and normal.

The author explains that the mechanical trading rule is an approach based on the fact that it is possible to reproduce some of the most basic hedge fund strategies, yet they are widely used and subject to overcrowding issues. The FundCreator follows the idea that the replication should focus on the risk factor exposure in order to reproduce the same statistical characteristics of the return series instead of the same returns. The methodology allows the investor to generate an original portfolio similar to the hedge fund's, which could be attractive as an investment. However, the replication would not be exact in the return point of view considering their sequence of arrival could differ from the observed ones in the hedge fund.

Tancar and Viebig (2008) summarizes the development of the *alternative beta* framework with an overview of the methodologies and discusses which are the advantages and disadvantages of it's implementation. The authors deepen the discussion on how the replication can help solving problems present on hedge fund industry. Important aspects such as high lock-up periods derived out of the investment in illiquid assets traded in OTC market or exotic derivatives, high performance fees to cover management costs along with the lack of investment and risk transparency, which could be a great issue to fiduciary agents and pension funds. As explained in the paper, the model's fundamental hypothesis guarantee investments solely in liquid assets, lowers fees since there are no remunerations nor due diligence processes and finally the replication methodology is focused on finding the

underlying risk factors, providing important information about the investment's exposures.

2.3 The Linear Approach

The empirical branch of academic works about *alternative beta* has expanded with works intending to take a step forward from the model introduced by Sharpe (1992). Jaeger and Wagner (2005) adopted a similar model as the one described by Amenc et al. (2003), a linear model with rolling window. This approach introduces time variation in the parameters discovery process, that related the hedge fund returns to directly observable market price. Apart from explaining the portfolio's returns with risk exposures, there was an attempt to divide the returns into its two components following the Capital Asset Pricing Model (CAPM) approach. Thus, the authors identified which part could be attributed to the *alternative beta*, the exposure on risk factors that includes non linear strategies such as short selling, leverage and the usage of derivatives, and which came from alpha, a part that derives from the manager's abilities.

The authors tested the model on hedge fund indexes following different investment strategies in order to understand if the methodology could be implemented independently of how the investment decision was taken. They concluded that the model does not work with the same efficiency depending on the strategy. This result was most evident when the investment was not based solely in risk factors but on the performance of an specific assets.

Hasanhodzic and Lo (2007) studied the improvement on the replication performance solely by applying a rolling window regression instead of a fixed weights one. By implementing both models, the authors show evidence of how allowing for time variation can enhance the out-of-sample performance and better tracking to the returns. However, they discuss that, even though they were able to achieve better results, this linear model is still unable to capture non-linearities on the portfolio's strategy, an important aspect of an hedge fund's investment.

The only *alternative beta* adaptation for the Brazilian market so far was Lima (2007), who implemented both linear models proposed by Hasanhodzic and Lo (2007) on 10 hedge funds and 2 hedge fund indexes. The author finds that the fixed weight regression leads to poor results with a low explanation capacity and few statistically significant coefficients. On the other hand, the results on the rolling window model with daily rebalancing showed similar statistical characteristics to the replicated funds.

Lima (2007) came to the same conclusion as Hasanhodzic and Lo (2007), some strategies cannot be replicated because of the usage of arbitrage and diversified strategies, and that non-linearity and illiquidity premia are aspects that cannot be replicated. Moreover, replicating methodology led to greater cumulative losses and bigger kurtosis on the

returns' distribution than the respective fund.

2.4 The Kalman Filter Usage

Roncalli and Teiletche (2008) take a step forward to demonstrate how the substitution of linear approaches on the hedge fund replication with a state space model as the Kalman filter can result in much more efficient tracking. As the authors explained, when the weights oscillation frequency is high, the rolling window regression can reproduce the dynamic allocation with a delay. Increasing the frequency does not solve the issue, since the model becomes unable to estimate the weights properly. The approach proposed also allows for the time variation on the parameters, but the dynamic allocation performs much closer to the hedge fund's.

Apart from that, the authors test other benefits deriving from this new model. First, the state space representation can lead to a higher proportion of positive returns and lower drawdowns. Likewise, they were able to prove the obtained results produced Sharpe ratios above buy-and-hold strategies on standard assets.

Roncalli and Teiletche (2008) also introduce an important conclusion to the *alternative beta* framework, the intangibility of alpha reproduction. As explained in the paper, alpha is not only composed by the managers performance fee, but also by the factors the replication is unable to capture. The illiquidity premia and the high frequency trading are essential components on the hedge fund's investment strategy that cannot be included in the model given the hypothesis on both solely investing in liquid assets and the usage of historical data to calibrate the weights. Therefore, the alpha still carries part of the *alternative beta* that could not be reproduced.

Amenc et al. (2010) seek to understand if the introduction of non-linear factors could enhance the replication's performance. Firstly, the authors extend both linear models with fixed weights and rolling window developed by Hasanhodzic and Lo (2007) and include among the risk factors a piecewise linear approximation of options payoff structure. Furthermore, they test the Markov regime switching model and the Kalman filter in order to obtain a comparison between their results.

They find that going beyond the linear case does not necessarily enhance the out-of-sample replication power, but improves the in-sample fitting. A procedure capable of improving the out-of-sample replication grounds the factors' selection on economic analysis. Apart from that, the authors find that Kalman filter can deliver better results than both linear models and the Markov regime switching model, even though the out-of-sample performance shows there is still much to work to do to achieve good results.

Roncalli and Weisang (2011) deepen the discussion proposed by Amenc et al. (2010)

in which there are still many issues on the hedge fund replication that must be addressed in order to get better results. The authors focus on the main problems in the replication process, the lack of reactivity, a disability to capture tactical allocation, and lack of access to the hedge fund alpha. Even when facing these obstacles, the Kalman filter outperforms other models when we attempt to capture the dynamic nature of the hedge fund and hence try to improve the explanation power of the in-sample. However, as exposed by Roncalli and Teiletche (2008), the model shows difficulties in accessing some aspects of the non-linearity in returns, illiquidity premia and reproducing a dynamic strategy.

Throughout the *alternative beta*'s evolution, literature has focused mostly on the model used for the implementation, going from linear regression to linear regression with rolling window, and later to the Kalman filter approach, in order to study its effect on the results. Still little has been explored on the issues of an actual implementation and how the introduction of practical aspects on the methodology could impact the replication's performance. This work seeks to fill this gap by designing an algorithm that aims to create similar conditions to a real investment strategy.

3 THEORETICAL FRAMEWORK

The goal of the present work is to describe in detail the practical implementation of the *alternative beta model* for the Brazilian market case by using the Kalman filter model. This chapter will discuss in depth the theoretical framework behind both models.

3.1 State Space Model

The state space model is a unified methodology applied to time series analysis to understand the relation between a set of observed variables and an unobserved parameter, which describes the evolution in the state of the underlying system. The representation was first built to be implemented in control engineering to describe a dynamical system with multiple inputs and outputs, however, its usage was later extended to other applications and brought by Harvey (1989) to the financial context.

The state space representation is very general and can be applied to a variety of different models. The formulation can be adapted in order to fit both univariate and multivariate problems, gaussian and non-gaussian approaches, time variant and time invariant models, and linear and non-linear systems. The Kalman filter model is a special case in the state space models, since it is a linear and gaussian approach with time variant parameters, that adds the dynamics in the estimation process.

As Durbin and Koopman (2012) have explained, the state space analysis aims to infer the properties from the unobserved variable of interest by using the information available up to the moment in the observed ones. The representation is usually applied to problems involving smoothing of a time series, forecasting, filtering the noise in the series and extracting signals.

Harvey (2006) detailed that the general linear state space formulation consists in two equations: (i) the *transition equation*, which models the dynamics of the random unobserved variable over time; and (ii) the *measurement equation*, responsible for relating the observed and unobserved variables.

The relation between the vector of observable variables $\mathbf{y}_t$ of dimension $N \times 1$ and the $m \times 1$ vector of unobservable variables, known as the state vector α_t, is given by the *measurement equation*

$$\mathbf{y}_t = \mathbf{Z}_t \, \alpha_t + \mathbf{d}_t + \varepsilon_t, \qquad t = 1, \ldots, T \;, \tag{3.1}$$

where $\mathbf{Z}_t$ is an $N \times m$ matrix, $\mathbf{d}_t$ is an $N \times 1$ vector and ε_t is an $N \times 1$ vector of serially uncorrelated disturbances with mean zero and covariance matrix $\mathbf{H}_t$.

Since the elements from the state vector α_t are not observable, we must model their random behavior through a first-order Markov process, in which the probability of a future state depends only on the current state and not on the path followed up to that time. This dynamics is given by the *transition equation*

$$\alpha_t = \mathbf{T}_t \, \alpha_{t-1} + \mathbf{c}_t + \mathbf{R}_t \, \eta_t, \qquad t = 1, \ldots, T \ , \tag{3.2}$$

where $\mathbf{T}_t$ is an $m \times m$ matrix, $\mathbf{c}_t$ is an $m \times 1$ vector, $\mathbf{R}_t$ is an $m \times g$ matrix and η_t is an $g \times 1$ vector of serially uncorrelated disturbances with mean zero and covariance matrix $\mathbf{Q}_t$.

Harvey (2006) explains that, to obtain a complete formulation, we must assume that the initial vector α_0 has mean $\mathbf{a}_0$ and covariance given by a positive semi-definite matrix $\mathbf{P}_0$. Apart from that, we assume that both equation's terms of error, ε_t and η_t are uncorrelated with the initial state α_0 and are uncorrelated with each other in all the periods, although this last premise may be relaxed.

The matrices and vectors $\mathbf{Z}_t$, $\mathbf{d}_t$, $\mathbf{H}_t$, $\mathbf{T}_t$, $\mathbf{c}_t$, $\mathbf{R}_t$ and $\mathbf{Q}_t$ are known as the system matrices and, together with the mean and covariance matrix of the initial vector, $\mathbf{a}_0$ and $\mathbf{P}_0$, are assumed to be known throughout all the periods in the estimation. Therefore, there is no need for them to be included in the information set.

3.2 Kalman Filter Model

The Kalman filter is a linear and gaussian model that computes recursively the optimal estimator of the state vector at time t using the information available in the set of observable variables up to, and including, time t.

Since the model is gaussian, the error components present in both the *measurement* and the *transition equation*, ε_t and η_t, as well as the initial state, α_0, are normally distributed. Bearing in mind that the normal distribution is characterized by its first two moments, the Kalman filter algorithm works by estimating at each step the optimum values for the mean and covariance matrix of the conditional distribution of the state vector as new information are incorporated into the system.

The concept behind the filter's process is that it uses the information available up to the moment in the observable variables and projects forward the value of the state vector in order to predict the next state. Since we are estimating an unobservable variable, we will have an estimation error that can only be corrected once we have evolved in time and obtained more information about the state vector. After introducing the new information, the algorithm will then compare the measurement with the prediction and update the value assigned for the state vector. Thus, we can divide the implementation into these two

steps, which we will refer to as *prediction* and *update*. For the explanation of the model, we will be adopting the specification used in Harvey (2006).

3.2.1 Prediction Step

For the time $t = 1$ we are assuming the parameters for the model's initialization are known, we have both the initial state α_0 and its distribution moments, $\mathbf{a}_0$ and $\mathbf{P}_0$. For $t = 2, \ldots, T$, we will be basing our estimation on the information available up to $t - 1$.

We are considering that $\mathbf{a}_t$ is the best predictor to α_t. Therefore, the best *prior* estimator of α_t given all the information available up to $t-1$, $\mathbf{Y}_{t-1} = \{\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}\}$, will be given by $\mathbf{a}_{t\,|\,t-1} = E[\alpha_t \mid \mathbf{Y}_{t-1}]$.

We then replace α_t by its expression in the conditional expected value to get

$$\begin{aligned} \mathbf{a}_{t\,|\,t-1} &= E_{t-1}\left[\mathbf{T}_t\,\mathbf{a}_{t-1} + \mathbf{c}_t + \mathbf{R}_t\,\eta_t\right] \\ &= \mathbf{T}_t\,E_{t-1}\left[\mathbf{a}_{t-1}\right] + \mathbf{c}_t + \mathbf{R}_t\,E_{t-1}\left[\eta_t\right]\,. \end{aligned} \tag{3.3}$$

Since $E_{t-1}[\eta_t] = 0$, that reduces to

$$\mathbf{a}_{t\,|\,t-1} = \mathbf{T}_t\,\mathbf{a}_{t-1} + \mathbf{c}_t\,. \tag{3.4}$$

The prediction error of the state vector will be then given by:

$$\begin{aligned} \mathbf{a}_t - \mathbf{a}_{t\,|\,t-1} &= (\mathbf{T}_t\,\alpha_{t-1} + \mathbf{c}_t + \mathbf{R}_t\,\eta_t) - (\mathbf{T}_t\,\mathbf{a}_{t-1} + \mathbf{c}_t) \\ &= \mathbf{T}_t\,(\alpha_{t-1} - \mathbf{a}_{t-1}) + \mathbf{R}_t\,\eta_t\,. \end{aligned} \tag{3.5}$$

Our error could have been caused by an error in the prediction process of the state vector or an idiosyncratic shock, external to the model's variables.

The covariance matrix is given by

$$\begin{aligned} \mathbf{P}_{t\,|\,t-1} &= E_{t-1}\left[(\alpha_t - \mathbf{a}_{t\,|\,t-1})(\alpha_t - \mathbf{a}_{t\,|\,t-1})'\right] \\ &= \mathbf{T}_t\,E_{t-1}\left[(\alpha_{t-1} - \mathbf{a}_{t-1})(\alpha_{t-1} - \mathbf{a}_{t-1})'\right]\mathbf{T}_t' \\ &\quad + \mathbf{R}_t\,E_{t-1}\left[\eta_t\,\eta_t'\right]\mathbf{R}_t' - 2\,\mathbf{T}_t\,E_{t-1}\left[(\alpha_{t-1} - \mathbf{a}_{t-1})\,\eta_t'\right]\mathbf{R}_t'\,. \end{aligned} \tag{3.6}$$

That, since $E_{t-1}\left[(\alpha_{t-1} - \mathbf{a}_{t-1})\,\eta_t'\right] = 0$, reduces to

$$\mathbf{P}_{t\,|\,t-1} = \mathbf{T}_t\,\mathbf{P}_{t-1}\,\mathbf{T}_t' + \mathbf{R}_t\,\mathbf{Q}_t\,\mathbf{R}_t'\,. \tag{3.7}$$

We can obtain the predictor of the vector of observations, $\mathbf{y}_t$, using the information available up to time $t-1$

$$\begin{aligned}\tilde{\mathbf{y}}_{t\,|\,t-1} &= E_{t-1}\left[\mathbf{y}_t\right] \\ &= E_{t-1}\left[\mathbf{Z}_t\,\mathbf{a}_t + \mathbf{d}_t + \varepsilon_t\right] \\ &= \mathbf{Z}_t\,E_{t-1}\left[\mathbf{a}_t\right] + \mathbf{d}_t + E_{t-1}\left[\varepsilon_t\right]\,.\end{aligned} \tag{3.8}$$

Since $E_{t-1}[\varepsilon_t] = 0$,

$$\tilde{\mathbf{y}}_{t\,|\,t-1} = \mathbf{Z}_t\,\mathbf{a}_{t\,|\,t-1} + \mathbf{d}_t\,. \tag{3.9}$$

Therefore, the prediction error will be

$$\begin{aligned}\mathbf{y}_t - \tilde{\mathbf{y}}_{t\,|\,t-1} &= (\mathbf{Z}_t\,\alpha_t + \mathbf{d}_t + \varepsilon_t) - (\mathbf{Z}_t\,\mathbf{a}_{t\,|\,t-1} + \mathbf{d}_t) \\ &= \mathbf{Z}_t\,(\alpha_t - \mathbf{a}_{t\,|\,t-1}) + \varepsilon_t\,,\end{aligned} \tag{3.10}$$

and the covariance matrix given by

$$\begin{aligned}\mathbf{F}_t &= E_{t-1}\left[(\mathbf{y}_t - \tilde{\mathbf{y}}_{t\,|\,t-1})(\mathbf{y}_t - \tilde{\mathbf{y}}_{t\,|\,t-1})'\right] \\ &= E_{t-1}\left[\mathbf{Z}_t\,(\alpha_t - \mathbf{a}_{t\,|\,t-1})(\alpha_t - \mathbf{a}_{t\,|\,t-1})'\,\mathbf{Z}_t' + \varepsilon_t\,\varepsilon_t' + 2\,\mathbf{Z}_t\,(\alpha_t - \mathbf{a}_{t\,|\,t-1})\,\varepsilon_t'\right]\,.\end{aligned} \tag{3.11}$$

Since $E_{t-1}\left[(\alpha_t - \mathbf{a}_{t\,|\,t-1})\,\varepsilon_t'\right] = 0$ and $\varepsilon_t\,\varepsilon_t' = \mathbf{H}_t$,

$$\mathbf{F}_t = \mathbf{Z}_t\,\mathbf{P}_{t\,|\,t-1}\,\mathbf{Z}_t' + \mathbf{H}_t\,. \tag{3.12}$$

3.2.2 Update Step

Now we have information about time t, meaning the value for $\mathbf{y}_t$ observed. We wish to update the state vector using the new information. The update of the *posterior* estimation $\mathbf{a}_{t\,|\,t} = E\left[\alpha_t \mid \mathbf{Y}_{t-1},\, \mathbf{y}_t\right]$ is characterized by a bayesian model of conditional mean, meaning a gaussian joint distribution of α_t and $\mathbf{y}_t$

$$\mathbf{a}_{t\,|\,t} = \mathbf{T}_t\,\mathbf{a}_{t\,|\,t-1} + \mathbf{c}_t + \mathbf{K}_t\,(\mathbf{y}_t - \tilde{\mathbf{y}}_{t\,|\,t-1})\,. \tag{3.13}$$

The algorithm will be basing the correction procedure on a linear combination between the *prior* estimation of the state vector and the error between the observed value of $\mathbf{y}_t$ and the expected value of the prediction, $\tilde{\mathbf{y}}_{t\,|\,t-1}$. We add a component in the equation called the Kalman gain, $\mathbf{K}_t$, which is defined as the relative weight given to the measure

and current state estimate in order to calibrate our model. We wish to minimize the difference between the observed value and the predicted *posterior,* $\mathbf{a}_{t\,|\,t}$.

Using equation (3.10), we can expand equation (3.13) as

$$\mathbf{a}_{t\,|\,t} = \mathbf{T}_t\,\mathbf{a}_{t\,|\,t-1} + \mathbf{c}_t + \mathbf{K}_t\,\mathbf{Z}_t\,\alpha_t - \mathbf{K}_t\,\mathbf{Z}_t\,\mathbf{a}_{t\,|\,t-1} + \mathbf{K}_t\,\varepsilon_t\;. \tag{3.14}$$

Now using the *measurement equation* (3.1), we finally obtain

$$\mathbf{a}_{t\,|\,t} = \left(\mathbf{T}_t - \mathbf{K}_t\,\mathbf{Z}_t\right)\mathbf{a}_{t\,|\,t-1} + \mathbf{K}_t\,\mathbf{y}_t + \left(\mathbf{c}_t - \mathbf{K}_t\,\mathbf{d}_t\right)\;. \tag{3.15}$$

The expression for $\mathbf{K}_t$ will be given by

$$\mathbf{K}_t = \mathbf{T}_t\,\mathbf{P}_{t\,|\,t-1}\,\mathbf{Z}_t'\,\mathbf{F}_t^{-1}\;, \tag{3.16}$$

and the covariance matrix will be

$$\mathbf{P}_{t\,|\,t} = \mathbf{T}_t\left(\mathbf{P}_{t\,|\,t-1} - \mathbf{P}_{t\,|\,t-1}\,\mathbf{Z}_t'\,\mathbf{F}_t^{-1}\,\mathbf{Z}_t\,\mathbf{P}_{t\,|\,t-1}\right)\mathbf{T}_t' + \mathbf{R}_t\,\mathbf{Q}_t\,\mathbf{R}_t'\;. \tag{3.17}$$

Up to this point, we have assumed that the system matrices were known, however, they may partially depend on the vector of unobserved parameters. Therefore, we will need to estimate the unknown parameters using the maximum likelihood estimation (MLE).

As Mergner (2009) explained, in order to estimate the model using MLE, we must guarantee first that it is fully specified through the joint probability density function, and also that the parameters contain all the needed information for the estimation, which can be obtained from the dependent variables.

We are assuming that our set of observable variables, $\mathbf{y}_t$, is i.i.d. and that the joint density, $L(\mathbf{y}, \alpha)$, is given by the product of individual densities

$$L\left(\mathbf{y}, \alpha\right) = p\left(\mathbf{y}_1, \ldots, \mathbf{y}_t\right) = \prod_{t=1}^{T} p\left(\mathbf{y}_t\right)\;. \tag{3.18}$$

Also, we will be working with the natural logarithm of the likelihood function,

$$\ln L\left(\mathbf{y}, \alpha\right) = \sum_{t=1}^{T} \ln p\left(\mathbf{y}_t\right)\;. \tag{3.19}$$

Since the observations for the time series models are usually not independent, we must replace the density functions with the probability density functions in the likelihood function. Therefore, we will be conditioning the distribution of $\mathbf{y}_t$ to the set of information available up to time $t-1$, which we refer to as $\mathbf{Y}_{t-1}$. Hence

$$\ln L\left(\mathbf{y}, \alpha\right) = \sum_{t=1}^{T} \ln p\left(\mathbf{y}_t \mid \mathbf{Y}_{t-1}\right)\;. \tag{3.20}$$

If the disturbances and the initial vector in the general state space formulation have a multivariate normal distribution, we can show that the conditional distribution of $\mathbf{y}_t$ is normal with conditional mean

$$E\left(\mathbf{y}_t \mid \mathbf{Y}_{t-1}\right) = \mathbf{Z}_t\, \alpha_t \ , \tag{3.21}$$

and conditional covariance

$$Var\left(\mathbf{y}_t \mid \mathbf{Y}_{t-1}\right) = \mathbf{F}_t \ . \tag{3.22}$$

For gaussian models, $\mathbf{y}_t$ is has a conditional probability density given by

$$p\left(\mathbf{y}_t \mid \mathbf{Y}_{t-1}\right) = \frac{1}{2\pi} \mid \mathbf{F}_t \mid^{\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\mathbf{y}_t - \tilde{\mathbf{y}}_{t\,|\,t-1}\right)' \mathbf{F}_t^{-1} \left(\mathbf{y}_t - \tilde{\mathbf{y}}_{t\,|\,t-1}\right)\right) \ . \tag{3.23}$$

Thus, we will have a log likelihood function given by

$$\ln L\left(\mathbf{y}, \alpha\right) = -\frac{TN}{2} \ln\left(2\pi\right) - \frac{1}{2}\sum_{t=1}^{T} \ln \mid \mathbf{F}_t \mid - \frac{1}{2}\sum_{t=1}^{T} \left(\mathbf{y}_t - \tilde{\mathbf{y}}_{t\,|\,t-1}\right)' \mathbf{F}_t^{-1} \left(\mathbf{y}_t - \tilde{\mathbf{y}}_{t\,|\,t-1}\right) \ . \tag{3.24}$$

The log likelihood is referred to as the *prediction error decomposition*.

If the vector of parameters is identifiable, the maximum likelihood estimate of the parameters, $\hat{\alpha}$, is found by maximizing the likelihood with respect to α.

3.3 Alternative Beta Model

3.3.1 Context and Concept

The Style Analysis literature evolved around the concept that there is an evidence of predictability in the returns involving traditional allocation strategies. It is argued that, with the usage of econometric models, it is possible to understand the sources for the fund's returns and how the investment decisions are made.

Sharpe (1992) started the studies on these sources and came to the conclusion that stands until today that the returns can be explained mainly by two different sources. The first would be the exposure to systematic risk, which is the component we commonly refer to the market influence, and known as beta. The remaining part is the non-systematic risk, which would be explained by an specific behavior, the hedge fund manager's specific skills. This is the alpha component.

It is a consensus that identifying how much of the return comes from the alpha is not an easy task. Usually, it is defined as the part that could not be explained by the exposures to systematic risks, and, thus, should be considered a consequence of the manager's skills. This would then be a component that cannot be directly modeled nor estimated.

However, progress in academic research identified that distinguishing alpha from beta in the hedge fund's returns wasn't the only challenge, since they would also have different systematic risks exposures. Their strategies involves unconventional techniques such as short selling, leverage and the use of derivatives, creating non linear payoffs. Therefore, the hedge fund's beta would be considered to be different from the traditional one, more complex. This is why the academic literature chose to refer to the hedge fund's beta as the *alternative beta*.

This idea also extends to the alpha component. As stated, alpha was always considered to be what beta couldn't explain from the return, being attributed to the manager's skills. However, Roncalli and Teiletche (2008) diverged from this opinion by introducing the idea that there is more on the alpha than just the manager's abilities. Every aspect of the hedge fund's trading dynamics that could not be captured by the model would be present in that component. This would include the illiquidity premia and the ultra-high frequency trading, which are not reproducible through low frequency trading and liquid instruments.

The idea behind the *alternative beta model* is to benefit from the fact that we can model the beta component in order to discover the fund's exposures to systematic risk factors and use this information to replicate its returns. Once we know how much each factor represents from the portfolio composition, we can use this as the proportion of the wealth we must invest in order to obtain the same return generated by the fund's strategies and exposures.

As explained by Tancar and Viebig (2008), the concept behind the *alternative beta* isn't to replicate the hedge fund's strategies in a one to one reproduction, specially because its payoffs are complex and non linear, and our model wouldn't be able to make such approximation using instruments with linear payoffs. The methodology aims to emulate the average returns generated by these strategies by reproducing the same risk exposures identified in order to resemble the risk-return profile.

Hence, the replication isn't as complex as to attempt to reproduce the same strategies structured by the fund's manager, nor as simple as to deliver the same number obtained on the return series. The answer to the real *alternative beta model*'s objective lies in between, it is a replication of the returns distribution's parameters so that we can have the same risk-return profile identified in the fund.

Once said that, it is important to note that, even though the estimation will get as close as possible to the fund's real risk exposures, we run our model without truly knowing what is inside its portfolio and by using general risk factors commonly used in investments, the model's output portfolio won't be the same and not even close to the fund's real one. Apart from that, there are parts of a hedge fund's trading dynamics that cannot be captured and modeled, such as the high frequency trading, illiquidity premia

and non linear payoffs. We don't expect to explain all the return by using the *alternative beta model*, but most of it.

The model's first assumption is that most of the returns can be explained by the exposures to the systematic risk factors. Since the alpha component cannot be directly modeled and estimated, we are relying on the fact that, by choosing the right risk factors, we can understand how the strategies were structured and, thus, replicate the fund's risk-return profile.

This second assumption is important to the methodology viability. Since it is not possible to invest in the same assets as the fund we are attempting to replicate, but on risk factors we estimate the fund is investing in, the daily rebalancing to adjust possible mistakes in our estimations and keep a good tracking is essential. We must only invest in instruments which allows us to easily buy and sell our position on a daily basis. No lock-up periods can be accepted in this model, we must guarantee liquidity and transparency in our dynamics. Therefore, differently from the hedge funds, we can only consider exchange-traded liquid assets as inputs for the *alternative beta model*.

The replication mechanics begins with the selection of the risk factors that should be representative of the investment strategies followed by the fund. We wish to capture most of the possible exposures in the investment universe, therefore, the factor selection is a crucial step in the implementation process. The more general the factor, the more types of strategies it can captures in the estimation. We call these general risk factor the *primary factors*.

After choosing the factors which we wish to include in our model, to help explain the fund's returns, we must find exchange-traded liquid instruments that offers exposure to these risk factors. These assets will introduce in our system all the information needed for the estimation process and will later make it possible for the replication investment to happen.

We run our model with all the information required and have as output the weights for each risk factor representing the fund's portfolio exposure to it. The last step is the investment in those assets using the same weight generated by the model for each period of time as the proportion needed to replicate the hedge fund's risk-return profile.

3.3.2 Alternative Beta using Kalman Filter

We have based the model's implementation in the state space representation described by Roncalli and Teiletche (2008).

The fund's return $\mathbf{R}$ can be explained by the exposures to the m systematic risk factors' returns $\mathbf{F}$, our model's observable variables. We wish to find out the weights for each risk factor in order to understand the fund's portfolio exposure and be able to

replicate its risk-return profile. However, the vector of weights that will dictate how much we must invest in each factor on the replication process, given by the state vector β, is unobserved and its dynamics through time has to be modeled by our model.

Given this context, the *measurement equation* in the *alternative beta model* will relate the return with the risk factors, finding the values for the weights:

$$\mathbf{R}_t = \sum_{i=1}^{m} \beta_t^{(i)} \, \mathbf{F}_t^{(i)} + \varepsilon_t \; . \tag{3.25}$$

The *transition equation* will model the behavior of the vector of weights over time through a first-order Markov process. Since we have m risk factors and, therefore, m weight parameters to model, the we must have an equation for each factor's parameter, leaving us with m *transition equations*,

$$\begin{cases} \beta_t^{(1)} & = \beta_{t-1}^{(1)} + \eta_t^{(1)} \; , \\ & \vdots \\ \beta_t^{(m)} & = \beta_{t-1}^{(m)} + \eta_t^{(m)} \; , \end{cases} \tag{3.26}$$

where $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and $\eta \sim \mathcal{N}(0, \sigma_i^2)$ are uncorrelated processes.

If we consider the state space representation described in Sections 3.1 and 3.2, the matrices $\mathbf{T}$ and $\mathbf{R}$, that multiply respectively the state vector and the disturbance in the *transition equation*, will be equal to the identity matrix. The vectors $\mathbf{c}$ and $\mathbf{d}$ are null in the model, meaning we are not considering any exogenous variables in our estimation.

The covariance matrices from both disturbances will be $\mathbf{H}_t = \sigma_\varepsilon^2$ and $\mathbf{Q}_t$ the diagonal matrix

$$\mathbf{Q}_t = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_m^2 \end{pmatrix} \; .$$

Roncalli and Teiletche (2008) assumed that the initial vector's covariance matrix would be $\mathbf{P}_0 = 0_m$, meaning that they know the initial allocation, and that the estimation of the best estimate for β, given by $\mathbf{b}_t$ and $\mathbf{b}_{t|t-1}$, is homogeneous with respect to the vector of parameters.

4 METHODOLOGY

In this chapter we will detail the steps followed in the implementation of the model and which indicators were selected for the results analysis.

4.1 Fund Selection

The Brazilian Securities and Exchange Commission (CVM) has divided the fund industry into four classes to help regulate this market and to guide the investment decision. As stated in the regulation, a fund is firstly classified according to its investment asset class, being designated as Fixed Income, Stocks, Multimarket, or Foreign Exchange Rate, and later subcategorized in line with its strategies. For each of the four primary classifications, there is a minimum percentage allocation rule. All the information on the fund industry classification and the specific allocation rules for each one are described in Appendix A.

With the exception of the Multimarket category, all the other classifications have a minimum allocation focused mainly on one risk factor. Therefore, all the funds classified as Fixed Income, Stocks or Foreign Exchange Rate will not have an investment diversification among primary risk factors, but will base their investment decision on subfactors that derives from one risk exposure. Fixed Income funds will have strategies based on the interest rate market; Stock funds, on the other hand, will be concentrated on companies characteristics; as for the Foreign Exchange Rate funds, they will base their investment decision specially on exchange rate derivatives.

As approached in Section 3.3.1, we intend on working with primary risk factors observed in the Brazilian market in order to make our model's input more general and able to capture a greater variety of strategies. Selecting a fund we know that has a concentrated exposure in one factor and not exploring the possible subfactors means we will not be explaining most of its returns with the model's implementation. Thus, we have decided to only work with the Multimarket category.

In order to eliminate arbitrariness and add transparency and independence to the fund selection process, we have decided to choose a ranking that elected the best Brazilian Multimarket funds. Most of the published rankings have two important hurdles we tried to avoid in our selection, they either focused their analysis on a short period of time, which could add a bias in the comparison, or they only evaluated the profitability of the fund.

When the methodology considers a short period of time, we cannot guarantee we are not comparing new funds with those who have been in the market in a longer period of time. Recent funds are trying to establish their brand in the industry, therefore, they

tend to be more aggressive in their strategies and take more risk aiming for a higher profitability. Apart from that, when analyzing a longer period we can contemplate a greater number of scenarios in which the manager had to rethink his strategies, enriching the comparison. Another relevant point would be for the methodology to evaluate the funds not by their profitability alone, but also by their risk adjusted return, which would consider the consistency in delivering positive outcomes without compromising their risk indicators.

Therefore, our objective was to find a ranking that contemplated both factors. We based our selection on the "Top 10 Brazilian Multimarket Funds of the Decade" ranking developed by Morningstar by request from Easynvest, a Brazilian broker dealer. In order to build the ranking, Morningstar firstly built a dataset containing all the Brazilian Multimarket funds with inception prior to January 2010 and that existed over the whole decade, avoiding the survivorship bias. Latter, they applied four filters to this dataset, being them:

i Fund must have a net worth of at least BRL 25 million;

ii Fund must have more than 50 investors;

iii Funds' shareholders must be retail or qualified investors;

iv Only the funds from independent manager would be accepted. All the funds from large banks were withdrawn from the dataset.

After filtering the set of Multimarket funds, they applied the Morningstar Risk-Adjusted Return (MRAR) methodology, defined as follows:

$$MRAR\,(\gamma) = \left[\frac{1}{T}\sum_{t=1}^{T}\left(1 + r_{Gt}\right)^{-\gamma}\right]^{-\frac{12}{\gamma}} - 1\;, \tag{4.1}$$

where,

- γ represents the investor's level of risk aversion. Morningstar uses $\gamma = 2$,
- T is the number of months in a period,
- r_{Gt} is the geometric excess return in month t and defined as $\frac{1+TR_t}{1+R_{bt}} - 1$,
- TR_t is the return of the fund in month t,
- R_{bt} is the return on the risk free asset in month t.

The returns were then adjusted by the risk taken by each fund. In the comparison between the set of Multimarket funds, Morningstar selected the ones that were able to deliver positive returns without having had major negative oscillations over the analyzed period.

The Top 10 Brazilian Multimarket Funds in the Decade selected funds is shown on Table 1.

Table 1 – Top 10 Brazilian Multimarket funds in the decade

Ranking Position	**Fund**	**Fund Manager**
1	ARX Extra FIC FIM	ARX Investimentos
2	Verde CSHG FIC FIM	Verde Asset
3	Bresser Hedge Plus FIM	Bresser Administração de Recursos
4	Sharp Long Short FIM	Sharp Capital Gestora Recursos
5	JGP Equity FIC FIM	JGP Gestão de Recursos
6	Mapfre Inversion FIM	Mapfre Investimentos
7	Safra Galileo AB FIC FIM	Safra Asset Management
8	Bahia AM FIC FIM	Bahia Asset Management
9	BTG Discovery FIM	BTG Asset Management
10	Seival FGS Agressivo FIC FIM	Seival Investimentos

After analyzing the ranking, we decided to withdraw two funds from the selection, namely Sharp Long Short FIM and Bahia AM FIC FIM.

In the case of the Sharp Long Short FIM, as the name points out, the fund follows a Long/Short strategy, in which the investor take a long position on an underpriced stock and a short position on an overpriced one, both on the same market sector. This strategy depends on the relationship between two assets and a careful selection of each pair. Knowing that one of the premises of the *alternative beta model* is replicating the funds' returns based on primary risk factors, the algorithm wouldn't be able to capture this dynamics involving an intra-factor strategy. Therefore, we decided not to include the fund in the implementation.

For the Bahia AM FIC FIM, the withdrawn from the selection was based on the principle of independence. Even though no information has ever been released about the investment strategy or the portfolio's composition, the fact of having family relation with a professional from this asset management could produce the idea of access to the portfolio or investment strategy. Hence, to avoid questioning about the results and the quality of the implemented model, we have decided not to consider the Bahia AM FIC FIM.

The final eight funds selected for the replication process are the following:

1. ARX Extra FIC FIM (ARX)[1]
2. Verde CSHG FIC FIM (VERDE)
3. Bresser Hedge Plus FIM (BRESSER)
4. JGP Equity FIC FIM (JGP)
5. Mapfre Inversion FIM (MAPFRE)
6. Safra Galileo AB FIC FIM (SAFRA)
7. BTG Discovery FIM (BTG)
8. Seival FGS Agressivo FIC FIM (SEIVAL)

It is important to note that all the selected funds have a large number of investors, are open to investments, and have a large net worth, considering the Brazilian mutual funds industry. Having a great amount of wealth to invest represents a challenge when we consider that the managers must decide the best strategy and the most efficient investment diversification among the assets in order to avoid the excess of concentration in a market or factor, as well as to avoid the price impact caused by large orders size, which would happen each time the strategy changed.

This means these funds don't often alter their portfolio composition and, when they do it, all changes are made slowly to avoid the price impact. This represents an important point for the replication, if the fund's trading strategy is very dynamic and its positions and exposures change almost daily, the replication may become less efficient and much harder. However, when the fund keeps its positions by implementing long-term strategies, the *alternative beta model* can capture a better picture of the risk factor allocation throughout the time.

We also highlight that, when a fund has the initials "FIC" in its name, it means that it is a Fund of Funds (FoF) and it invests its capital in other funds. When applying the *alternative beta model* to a FoF, we aim to capture the risk factor exposures from the master fund by observing the return series this investment has produced. Even though the fund doesn't have the assets in its portfolio, we must keep in mind that it is benefiting from the other fund's strategies and profitability. Therefore, the model's implementation should work the same way.

[1] The names placed between the parenthesis will be used to refer to each fund from this point forward in the study

4.2 Factor Selection

4.2.1 Challenges in Factor Selection

According to Weisang (2014), the factor selection is an important step in the Hedge Fund replication methodology. Both the minimization of the tracking error criterion and the explainability of the results depends on the quality of the chosen factors. If we choose a set of risk factors that diverges from the fund's portfolio exposures, we would be providing the wrong input for our model's optimization and, thus, forcing the algorithm to find an spurious relationship between the returns and the factors where there was none. This would lead to results detached from the reality and with no investment sense. In conclusion, we wouldn't be identifying the fund's exposures to the risk factors and, thus, wouldn't be able to explain the source of return that comes from that.

In the search for the best set of risk factors, we also face an important tradeoff: we believe that adding as many factors as possible would help improve the model's adherence, since, by doing that, we would be including more investment options for the algorithm to consider and, therefore, the estimation would result in a more detailed output about the fund's exposures. However, adding more factors in the model would not only introduce a difficulty in the Kalman filter model's optimization, but also would enlarge the rebalancing costs of the replication. Daily changes in the position of a small number of assets can be considered a small cost to the fund replication process, however, rebalancing a great amount of assets daily would significantly raise the costs, which could even make the replication infeasible.

Weisang (2014) proposed a methodology to help selecting an appropriate set of factors for the dynamic replication model. The author remarks that the usual selection factor models, such as Principal Component Analysis (PCA), wouldn't be effective in this case. When applying the PCA methodology to equally important latent factors, the estimation would lead to the false inference that the principal component would be an equally weighted vector of all the variables. The estimation would only run a correct identification with a minimum amount of 40,000 observations.

To overcome this difficulty, Weisang (2014) suggested adopting a bayesian approach and exploring the posterior distribution of the possible models using observable variables. The proposed algorithm combines an adaptive sampling with Reversible Jump Markov Chain Monte Carlo (RJ - MCMC) technique.

Even though there is a solution in the literature for the factor selection challenge, the methodology described is complex and of difficult implementation. Therefore, we decided to select the risk factors for the model based on the investment possibilities available to the funds and which are observable in the Brazilian fund industry, as usually chosen by fund managers. We will have to limit the selection of our options to the main

risk exposures in order to avoid the tradeoff we have mentioned above.

4.2.2 Selecting the Factors

Given that the selected funds are large enough to trade on the offshore market, we must consider both onshore and offshore exposures on our risk factor selection. Keeping in mind the tradeoff explained in the previous section, we had to limit our choices to the exposures that better represent the investments from a fund in the Brazilian market. The primary factors we chose to include in our model were:

1. Brazilian stock market,
2. Brazilian nominal compounding interest rate,
3. Brazilian inflation protected interest rate,
4. The exchange rate of Brazilian Real (BRL) per U.S. Dollar (USD),
5. US stock market,
6. US interest rate,
7. Emerging market foreign exchange rate.

Even though we acknowledge that there are other factors that could have been included, such as exchange rates from other countries, stock and interest rate exposures from Europe and Asia and even the corporate bonds in the Brazilian market, we focused on choosing the most relevant factors that could be represented in the model by an investable and exchange-traded liquid instrument.

The choice for the offshore exposures to be focused on the US financial market was linked to a recent trend in the Brazilian institutional investment, which search a diversification from the regional and political risks, mainly using the financial instruments traded in US, as well as emerging market and G10 foreign exchange rates.

Differently from the other countries, the Brazilian fund regulation[2] requires to all investment funds the full disclosure of the portfolio composition within a maximum period of 180 days. The asset managers must unveil the portfolio composition with the percentage representation of each asset. We considered this as an opportunity to evaluate if the chosen factors were representative of the real exposures in the portfolio. Apart from that, we would be able to make our assumptions whether the estimation with those factors

[2] Brazilian Securities and Exchange Commission Instruction number 522 from May 8th, 2012, items IV and V.

would lead to a good yield performance replication given how much of the risk exposure we were able to capture prior to the estimation.

As noted in section 4.1, some of the selected funds are classified as Fund of Funds (FoF), meaning they invest most of their wealth in another fund, known as the master one. In order to analyze the portfolio composition of these funds and understand if the chosen factors are representative of their exposures, we must consider the master fund, which would bring more information about the real allocation strategy the manager is applying.

The Table 2 bellow describes the funds' portfolio composition in November 2020. The information is available to the public in the Brazilian Securities and Exchange Commission's (CVM) website.[3]

Analyzing the information in Table 2, we could make parallels between the categories in the Table and the chosen risk factors. The Brazilian stock market factor would correspond to the exposures in stocks and its securities lending market; both Brazilian interest rate factors would represent the investment on government bonds and its securities lending market; both foreign exchange rate factors could partially correspond to future contract exposure, which is usually how this investment is made; and the offshore stock market and interest rate factors would represent the offshore investment.

Option and forward contracts will not be captured in our model, firstly because they depend on expiration date and strike price selection, which would introduce a new selection challenge, and secondly because they are used in more complex strategies. The replication model wouldn't be able to capture and structure the same payoffs generated by an investment in these derivatives only by running an optimization. Apart from that, we cannot replicate the investment on corporate bonds, since their market shows low liquidity and the trades usually are in the over-the-counter market, with lower price transparency, as well as there is no representative index that would allow for asset diversification in our investment.

As we can see in the Table 2, some funds are diversified among the chosen factors, however, others show a great concentration on some exposures. This is the case of JGP, which is concentrated in the stock markets, and SEIVAL, which strategy is focused on the interest rate exposure.

As we have discussed before, the concentration in a small number of factors is not an ideal condition for the replication model, unless we leave the primary factor level and deepen the selection on the sub factors. In this case, the funds strategy would not only consider sector in stock market and the tenors in the interest rate curve, but also the company's situation and the rate level of a bond. Our hypothesis about that will be tested in the estimation output, where the tracking error criterion would worsen the more

[3] http://conteudo.cvm.gov.br/menu/regulados/fundos/consultas/fundos.html

Table 2 – Fund's portfolio composition

When a fund is classified as FoF, the initials "FIC" are present in its name; in these cases, the master fund's name was related and the information presented refers to the master fund's portfolio. The Stocks classification includes stocks, receipts and ETFs; Government Bonds includes all bonds issued by the National Treasury; Future Contracts includes both Interest Rate and Stocks Future Contracts traded on the Brazilian Exchange; Options refer to all option contracts traded both in the Exchange and OTC markets; Forward Contracts refers to all forward agreements traded both in the Exchange and OTC markets; Government Bonds Lending and Stocks Lending refers to those securities lending market; Offshore Investments includes all the investments made on other countries; Corporate Bonds refers to bonds issued by Brazilian companies and traded on the OTC market; Fund Shares refers to investments made in other funds; and Others includes obligations, receivables, payables, among other accounting factors.

Fund	**Master Fund**	**Portfolio Composition of the Fund or Master Fund**										
		Stocks	**Government Bonds**	**Future Contracts**	**Options**	**Forward Contracts**	**Government Bonds Lending**	**Stocks Lending**	**Offshore investment**	**Corporate Bonds**	**Fund Shares**	**Others**
ARX Extra FIC FIM	ARX Extra Master FIM	19.38%	37.69%	0.2%	11.70%	-	-	-	5.70%	-	3.89%	21.44%
Verde CSHG FIC FIM	Verde Master FIM	15.31%	30.80%	-0.3%	0.12%	0.37%	23.47%	0.10%	16.03%	0.12%	2.43%	11.28%
Bresser Hedge Plus FIM	-	41.59%	36.70%	0.02%	0.42%	-	-	-	-	-	4.12%	17.15%
JGP Equity FIC FIM	JGP Equity Master FIM	68.25%	-	0.02%	-	-	9.60%	-	18.15%	0.96%	0.87%	2.15%
Mapfre Inversion FIM	-	8.88%	32.45%	0.13%	-	-	40.53%	-	6.06%	-	9.72%	2.23
Safra Galileo AB FIC FIM	Safra Galileo Master FIM	9.70%	47.24%	0.16%	-	-	27.93%	-	14.28%	0.04%	-	0.65%
BTG Discovery FIM	-	19.31%	37.46%	0.01%	1.77%	-	12.53%	0.59%	17.29%	-	6.81%	4.23%
Seival FGS Agressivo FIC FIM	Seival FGS Master Agressivo FIM	-	68.53%	-	-	-	-	-	-	-	19%	12.47%

concentrated in a smaller number of factors the fund is.

Since our objective was to make the replication process possible to be implemented in a real market condition, we selected for each factor a representative instrument and an investable asset, as follows:

- Brazilian stock market - It will be represented by the Bovespa Index (Ibovespa), the main performance indicator of the stocks traded in the Brazilian exchange. The instrument chosen to represent this factor was the BOVA11, an ETF traded in B3 that passively follows the Ibovespa.

- Brazilian nominal compounding interest rate - It will be represented by the Brazilian Financial and Capital Markets Association's (ANBIMA) Index IRF-M P2, which is composed by fixed rate government bonds. The instrument chosen to represent this factor was the IRFM11, an Exchange Traded Fund (ETF) traded in B3 that passively follows the IRF-M P2 Index.

- Brazilian inflation protected interest rate - It will be represented by the Brazilian Financial and Capital Markets Association's (ANBIMA) Index IMA-B, which is composed by inflation-indexed government bonds linked to the consumer inflation rate. The instrument chosen to represent this factor was the IMAB11, an ETF traded in B3 that passively follows the IMA-B Index.

- The exchange rate of Brazilian Reals (BRL) per U.S. Dollar (USD) - It will be represented by the spot exchange rate of BRL per USD. The instrument chosen to represent this factor was the first contract month of the Futures contracts traded in the Brazilian exchange, B3, (DOL) with financial settlement using the Brazilian Central Bank foreign exchange rate PTAX.

- US stock market - It will be represented by the S&P 500 Index, the main performance indicator of the 500 largest U.S. publicly traded companies. The instrument chosen to represent this factor was the first contract month of the mini S&P 500 futures contract, the contract traded in the Chicago Mercantile Exchange (CME), in the USA, based on the S&P 500 Index.

- US interest rate - It will be represented by the U.S. Treasury Bonds exposure. Since the U.S. interest rate curve is divided by its tenors, we chose the most frequently traded one, the 10 year bond. The instrument chosen to represent this factor was the iShares 7-10 Year Treasury Bond ETF, an ETF traded the in NASDAQ, in the USA, based on ICE's US Treasury 7-10 Year Index.

- Emerging market foreign exchange rate - It will be represented by a basket of emerging market foreign exchange rates. The exchange-traded instrument chosen to

represent this factor was the WisdomTree Emerging Currency Strategy Fund, an actively managed ETF traded in the NYSE, in the USA, which portfolio is composed by future contracts on the emerging market foreign exchange rates.

Even though we are considering the ETFs IRFM11 and IMAB11 as the investable assets for the Brazilian interest rate factors, these ETFs have only been created in 2019 and their price history wouldn't cover for most of the period we are analyzing in this work. Knowing that those are passive ETFs, with a tracking error in comparison with their reference index smaller than 1% by regulatory obligation, for the whole analysis period, we will use, as a proxy for these assets, the reference index they are representing.

4.3 Data Treatment

4.3.1 Selected Dataset

The dataset used in this work was obtained from Blomberg between January 2015 and December 2020.

When extracting the data, we have treated these series in order to deal with the holiday gaps. We consider relevant to observe this aspect which is frequently faced by fund managers and which will be significant to bringing the *alternative beta model*'s implementation from theory to the reality. When there is a holiday in a location, the fund will continue to trade in the other countries which are negotiating. We reproduced this behavior in the replication process.

Hence, when there was a holiday in Brazil, but the American market was open, the Brazilian funds would be trading only offshore assets. When the opposite happened, it was a holiday in the American market, but the Brazilian market was open, the funds would keep trading only Brazilian assets. Therefore, we decided to fill the holiday gaps by replicating the same price as the previous day only when one of the markets was still open and trading. When both markets were closed there would be no trading and those days were withdrawn from the dataset.

These holiday gap filling procedure didn't cause any impact on the replication nor the investment process since we worked with the return series instead of the price series for both the funds and the factors. When we gap filled a holiday, the price in that day is the same from the previous one and the return would become zero. On that day, those factors would not be considered in the calculation of the replication return and no investment could have been made in that market, since their weights would be multiplied by zero.

As all funds are Brazilian, the price series extracted were in BRL. For the factors, it depended on where the asset was traded, if the asset was traded in Brazil the price was

in BRL, if it was in the USA, it was in USD. This would make the investment easier later in the analysis.

We consider that the six years we have selected for the analysis will represent an interesting period to understand how well the *alternative beta model* would work in reality, since there were great changes in the investment scenarios over this period. All fund managers were challenged by those events and had to rethink their funds' strategies in order to adapt to all the political and economic adjustments. Thus, we will be able to test the models adherence in a variety of situations and analyze how well our model can capture those changes.

Among many other scenarios, we can highlight that, between January 2015 and December 2020, Brazil has faced:

1. Central Bank reference interest rates going from 12.25% in January 2015 to 2.00% in December 2020;
2. Foreign exchange rate going from a mean of BRL 2.63/USD in January 2015 to a mean of BRL 5.15/USD in December 2020, with a peak of BRL 5.94 in May 15th, 2020;
3. Changes in the profile of the Brazilian public debt caused by debt rolling problems;
4. Three different governments: Dilma Rousseff (January 2011 - August 2016), Dilma Rousseff's impeachment process, followed by the vice-president Michel Temer's government (August 2016 - December 2018), and the election of Jair Bolsonaro (January 2019 - current);
5. Biggest corruption investigation in Brazil, also known as "Operation Car Wash", which led to a great volatility in the financial market; and
6. Social security and spending ceiling reforms, which led to a positive market perception and forced changes in the impact in the investment strategies.

We can also highlight a few events and scenario changes in the international markets that had a major impact on the fund's investment decisions and strategies, such as:

1. The G10 decrease in the sovereign interest rates, reaching the scenarios of low or negative levels;
2. The Brexit, the withdrawal of the United Kingdom from the European Union; and
3. Global pandemic of SARS-COV-19 in 2020, which was considered to be a rare Black Swan that impacted all markets in the world, causing prices to fall and even become negative, in the oil futures, and leading to a great volatility in all markets.

4.3.2 Calculating the Fund's Gross NAV

The Net Asset Value (NAV) is the minimum fraction of the fund's net worth that can be purchased by an investor. The NAV value obtained from Bloomberg is considered to be net since it has been treated in order to remove the administration and performance fees charged by the funds. Keeping in mind that the *alternative beta* methodology aims to capture all results provided by the investment in the fund's strategy, in order to find the part of it that comes from the exposures to systematic risk factors and use this information to replicate the risk-return profile by assuming similar risk exposure, the model would only be effective if the return series provided as input contains all the information on the fund's results.

Once we remove the fees from the returns value, we are not fully representing the fund's performance. If we applied the model using the NAV, we would be capturing the exposures of only part of the performance, and the replication wouldn't reach its objective. Thus, we must reproduce the reverse process in order to obtain the gross NAV (G-NAV) and this new series must be used as the input of the *alternative beta model*.

For this process, we start with the NAV and the series for the Brazilian interbank deposit interest rate, CDI. We must also obtain the administration and performance fees charged by each selected fund.

We start the G-NAV estimation by adding back the administration fee to the fund's return. In order to do so, we must first calculate the following inputs: the fund's returns, the one-day Brazilian interbank deposit interest rate, and the one-day administration fee. Once we have all these values, we remove the one-day Brazilian interbank deposit interest rate from the fund's return and, then, add the one-day administration fee, obtaining the series of the returns with the administration fee.

Adding back the performance fee is a more complex process. According to the Brazilian fund regulation[4], the performance fee is charged semiannually and depends on scenarios involving two important concepts:

- *The fund's high-water mark performance criteria* (F-HWM): This concept is based on the idea that the manager can only charge for the performance fee if he was able to add value to the fund's NAV. Therefore, for every new period, we establish a reference value for comparison, the last NAV from the previous period. This is called the *fund's high-water mark*. The fund is said to be above the high-water mark if its NAV on the new period is higher than the reference NAV.

- *The benchmark's high-water mark performance criteria* (B-HWM): This concept is based on the idea that the manager can only charge for the performance fee if he

[4] Brazilian Securities and Exchange Commission Instruction number 555 from December 17th, 2014, Article 86.

was able to deliver a better performance than the benchmark, otherwise its strategy didn't add any value to the investor. Every day both performances are compared and the fund is said to be above the high-water mark if it outperformed the benchmark.

The performance comparison in the B-HWM is made by putting both the fund's returns and the benchmark series on the same basis. In the beginning of the period, we start with the value 100 and incorporate for each day the return, making a daily accrual and producing two comparable series.

In order not to cause an impact on the NAV in the end of the period when the performance fee is charged, the fund relies on a mechanism of daily provision. When the fund meet the requirements, the provision can be taken from that day's returns and will become the performance fee in the end of the period.

The requirements for the charge of the fee are divided into the provision made over the period and the final charge in the last day of the semester. They are described bellow.

Provisioning the performance fee over the period: every day we must check the fund's and the benchmark's high-water marks to understand which is the corresponding scenario

- If the fund meet the requirement of being above the F-HWM and the B-HWM, we should provision the performance fee from that day's return.
- If the fund is bellow the B-HWM and there has been a prior provision for performance fee, the difference between the performances must be removed from what has been provisioned and added to the return.
- If the fund is bellow the B-HWM, but there is no prior provision for performance fee, nothing can be added to the return.

At the end of the period: in the last day of the semester, we must check the fund's and the benchmark's high-water marks to understand which is the corresponding scenario

- If the fund meet the requirement of being above the F-HWM and the B-HWM, the performance fee is charged. In that case, we must reset the accrual in order to begin a new period.
- If the fund is bellow one of the high-water marks, the performance fee cannot be charged and the accrual cannot be reset. We must continue both the accrual and the fee's provision until the fund reaches the end of a new period and finds itself above both high-water marks.

To include the performance fee in the return, we had to follow the reverse process than we have described. When the requirements were attended and the fund's performance was above the water marks, we added the fee to the return, meaning that on that day it would have been removed and provisioned by the fund. When the fund was bellow the benchmark's high-water mark, we removed the performance difference that would have been added on that day's return. Following this logic, we were able to recompose the fund's G-NAV that will be used as the input for our model's estimation.

4.4 Model's Implementation

The main objective of this work is to bring the *alternative beta model*'s implementation to practice and deal with the hurdles this experience would introduce. In this section, we describe how the model was implemented and which adaptations had to be done in order to incorporate the practical experience into a theoretical model.

4.4.1 Implementing the Kalman Filter Model

When bringing the *alternative beta model* into practice, the academic literature usually runs the designed algorithm on the whole selected period at once and then models how the investments would have occurred each day. By running the estimation at once, we face two issues, one practical and one theoretical, that makes our model lose touch with reality.

The practical aspect is connected to the fact that, since we aim to reproduce the real implementation for the model, we would have to run our estimations and make our investments daily in order to attempt to replicate the fund's returns. Running the estimation for the whole period and later stating that we would have invested a determined proportions on each risk factor is unreal and only works in theory. The practice requires the model to be estimated every day only adding the latest information available in the market, in order to revise our replication strategy, and real daily investments to be made.

The theoretical aspect is that, when we estimate the whole period at once, the Kalman filter provides better outputs for our risk factor weights. The model's estimation is made in two steps, the prediction and the update, as explained in Section 3.2. Each time we introduce a new information, our system will automatically update the estimation for the previous periods, but now using future knowledge of the fund's and factors returns. Running a single estimation of all the periods at once will always provide us with outputs corrected to adjust to the new scenarios given by the future states.

However, we observe that the last observation won't have the same dynamics. Considering we haven't provided to our system any information about its future, the weights generated by this last estimation won't be corrected and could be worse compared

to the updated ones, specially for the replication purpose. This theoretical aspect is directly linked to the practical one, once we need to run our estimations every day and make our investments, we will always be using the output related to the last observation. If it is worse for not having been updated using future information, our weights will not be the best we could get. Even if we add new information to the model and obtain better estimation to the risk factors weights in the previous periods, they have stayed in the past and our investments have already been made. This cannot be changed, we can only work with the best output we have up to that moment.

In Section 4.3.1 we have mentioned that the selected period for our analysis was between January 2015 and December 2020. We have selected a 252 days window as the base period to run our estimations, which will work as a historical base to start the model. Therefore, we won't have any replication results for those first 252 days in our dataset.

For each day, we will be considering as our model's input all the data from January 2015 up to that day and we will run the *alternative beta model*. The estimation will provide us with the output for all the days present in the input except for the first one, however, the only information that is important to obtain from the results will be the last observation, which will guide new investment decisions. We must only know what to do now that we have those new information about the fund's and the risk factors returns and how to adjust our strategies.

By following this procedure, we must have, in the selected period and dataset, 1,296 estimations, each one using the information from a new day.

The *alternative beta model* was implemented using the R software. In order to run the Kalman filter, we used the package *dlm* (*Dynamic Linear Model*), which provides routines for maximum likelihood estimation, Kalman filtering and smoothing, and Bayesian analysis of normal linear state space models.

4.4.2 Impact of Using Lagged NAV

In order to achieve the ideal results in the replication process, the NAV used in the estimation would have to be released on the same day and in time for the investments to still be done before the market's closing. However, we know this dynamics is unreasonable since the manager must calculate daily the fund's net worth using the closing prices for each asset in the portfolio. The NAV could not be disclosed before the markets closing and the replication investment paired with the fund's investment would be impossible.

Since the Multimarket funds often invest in many asset classes, the NAV computation takes more than one day. The disclosure on Bloomberg, where we obtained the data used in the work, has a lag of two days. This means that on day t the only NAV available for our estimation will be the one from $t-2$. This is often disregarded in the literature,

but it is an important aspect in order to understand if the *alternative beta model* is feasible to be implemented.

When attempting to bring the real life experience to the work, we must consider that, when making our daily investments, we would have to use the weight given by the estimation from two days prior. The prices used for the investment would be from day t, however the weights for the investment in each factor will be given by the values from $t-2$.

This will have a direct impact on the replication results, specially because we will be losing the trading dynamics the funds have. For every change in the allocation strategy we capture in our estimation, we will be two days behind when actually investing our wealth. This will have a certain effect on our replication's tracking error, making it worse.

In this case, we will be counting with the fact explained in Section 4.1 that all the selected fund are very big and take more time than smaller funds to change their strategies on risk exposures. Since they have considerable positions on each asset and each factor, selling the position and buying another one could have a significant price impact on the trading and on the market. Therefore, all these movements are slow and diluted over a few days. Even if we have a disadvantage given by the lagged NAV, the difficulty in major strategy changes will counterbalance the impact.

4.4.3 Introducing Transaction Cost

If implemented in practice, the replication process would involve transaction costs when investing and daily rebalancing the portfolio. Since we want to compare the performance of the replication we have developed with the fund we are tracking, we must include these costs, otherwise we would be biasing our results. The fund has expenses when making its investments and all the costs are deduced from its returns. When we ignore these spendings, not only we move away from reality, but also our results show an unfair advantage with respect to the fund.

Amenc et al. (2003) introduced a transaction cost component in order to test if the optimal tactical allocation strategy generated by the model would turn into a very costly sub-optimal dynamic trading strategy. The costs were then modeled as an affine function of the volume of transactions

$$TC_{t+1} = a + b\sum_{i=1}^{m} W_{i,t} \mid w_{i,t+1} - w_{i,t} \mid , \tag{4.2}$$

where a is the fixed cost component, b the proportional cost component, $W_{i,t}$ is the wealth invested in the asset i at the time t, and $w_{i,t}$ is the proportion of wealth invested in the asset i at the time t. The authors chose the values of the cost components as $a = 0$ and

$b = 5\%$.

We opted to adapt this methodology in order to remove the dependency on the wealth invested initially on each asset. The adjustment would be done in the proportional cost so that variations in the weight of each factor, meaning a change in the investment of that asset, would represent a cost and the final value for our function would be directly subtracted from the replication return. This new approach results in

$$TC_{t+1} = a + b \sum_{i=1}^{m} \mid w_{i,t+1} - w_{i,t} \mid . \tag{4.3}$$

When calculating which would be the real values for our cost function components, we had to take into account that the trading activity involves two types of costs, the direct and the indirect. The direct costs are the exchange, settlement, custody and brokerage fees. All of them depending on the chosen exchange and broker.

The indirect costs are not given by fees, but are the changes in the price produced by market imperfections when trading. This is referred to as slippage - the difficulty in trading big orders or the need to execute them fast and, as a consequence, penetrating in the order book and causing a price impact on the trade - or low liquidity issues, forcing the investor to take the orders available in the book in other price levels.

In our case, the direct costs cannot be avoided, since all the investors must pay the required fees. However, as for the indirect costs, as we have explained in section 4.2.2, all the instruments selected to represent the risk factors in our model have a market maker. This represents an advantage in our trading possibilities since orders of great size or trading in moments in lower liquidity would be absorbed by the market maker, reducing or even removing this imperfections. Apart from that, big funds adopt order fragmentation rules along the day in order to optimize their execution in the book. Thus, we will be considering that we won't need to model the indirect costs, they can be withdrawn from our model.

The selected factors are negotiated in four different exchanges and in two different countries, and analyzing each one of those costs in order to apply the perfect function wouldn't be feasible. Therefore, we will make an assumption that, since the exchanges are competitors in the segment they act, their costs would be standardized. This is not unreasonable to consider since the methodology followed to calculate the fees for each product is very similar.

We will be basing our proportional cost component on the fee charged by the Brazilian Exchange, B3, for the ETFs specifically for the fund segment. It is a composition of the negotiation fee, of 0.005%, and the settlement fee, of 0.018%, resulting in a fee of 0.023%[5]. The broker fee must also be included in this proportional cost, however it varies

[5] The reference for the cost was taken from the B3's website http://www.b3.com.br/pt_

according to the institutions. As a proxy for this cost, we will consider the Broker will charge a mean fee of half of what the Exchange is charging, 0.0115%. The sum of this two costs would give us $b = 0.0345\%$.

The custody fee is charged as a fixed cost and depends on the value under custody. We will be considering that our replication fund would be included in the range of custody above BRL 17 million, a value reached by even small funds, and would be charged a annual fee of 0.0005%[6]. Since our cost function will be applied daily, the fixed cost will be removed from the return each day. When calculating the daily custody fee, we reach the value of 0.000002%. Therefore, by simplicity, we will approximate the fixed cost to zero, making $a = 0$.

The final cost function with the value of the components is given by

$$TC_{t+1} = 0.0345\% \sum_{i=1}^{m} \mid w_{i,t+1} - w_{i,t} \mid . \tag{4.4}$$

4.4.4 Result Indicators

To evaluate the replication results in comparison to the fund, we will calculate four indicators.

The first indicator will be the *tracking error* (TE). It is an adherence measure and determines if the replication's return behave similarly to the fund's returns by measuring the difference between both values at all times. It can be calculated by the following function

$$TE = \sqrt{\frac{\sum_{t=1}^{T}(R_{replication} - R_{fund})^2}{T-1}} . \tag{4.5}$$

The second indicator is called *hit ratio*, (HR) and estimates the percentage of the time in which the replication's return was greater then the fund's returns. This will provide a complement to the *tracking error* analysis since, if the tracking wasn't good and our returns didn't behave as the fund's, we can evaluate if our investments yield were inferior to the fund's or if we have consistently managed to overcome the fund's profitability.

Keeping in mind that our objective was to incorporate the practical experience into the theoretical model, we introduced the lagged NAV and the transaction costs into our methodology. As we have explained, our assumption was that both procedures would have an important impact on the replication results, therefore, these first two indicators

br/produtos-e-servicos/tarifas/listados-a-vista-e-derivativos/renda-variavel/tarifas-de-acoes-e-fundos-de-investimento/a-vista/

[6] The reference for the cost was taken from the B3's website http://www.b3.com.br/pt_br/produtos-e-servicos/tarifas/servicos-da-central-depositaria/tarifas-de-servicos-de-custodia/

were calculated in three different scenarios to measure how great the effect of the lagged NAV and the transaction cost were.

In the first scenario, the *tracking error* and the *hit ratio* will be calculated as if the NAV disclosure and the investment had been made in the same day, t, in order to demonstrate if we were able to capture with the chosen factors the systematic risk exposures that explains the source of the fund's returns. On the second scenario we introduce the lag in the NAV. This forces our replication process to reproduce the exposure from two days prior, losing the dynamic in trading. The last one is the introduction of both the lag in the NAV and the transaction costs. We wish to understand whether the costs of a daily portfolio rebalancing could have such impact in the results that the model would become unfeasible.

The *tracking error* and the *hit ratio* will also be calculated for two different periods. The first will be referred to as the *entire replication period* and consists of all the dates we have estimated an allocation and may produce a replication. Since we have used a window containing the initial 252 days as a base period to run our estimations, we won't have generated the weights for those first 252 days, therefore, the *entire replication period* comprehends the period between December 24th 2015 to December 31st 2020.

For the second period, we kept in mind that, since there were many scenario changes on our selected period, we expect both adherence indicators to vary across the years. Therefore, we will calculate them also for each year in order to better understand if periods of lower and higher volatility can impact our ability to track the fund's returns and exposures.

The third indicator will be an accrual from both returns. We want to compare them on the same basis, therefore they must begin at the same time with value 100 and accrual for each day the obtained return. With this new series, we will be able to plot both performances together and compare how well the replication process went.

The last indicator is the *maximum drawdown* (MDD), a risk measure frequently used to minimize significant negative variations. The *drawdown* concept is usually used to refer to a significant loss in the value of an asset or a portfolio. It is measured as the negative return calculated between a local maximum and a local minimum during a specific period. The *maximum drawdown* is defined for a period of time as the greatest observed *drawdown*.

In order to calculate the *maximum drawdown*, we will consider a rolling window of 252 days. For every new observation, we make an analysis in the selected window and identify the greatest *drawdown*. The objective here is to test an *alternative beta* drawback highlighted by Tancar and Viebig (2008), in which the replication does not structures strategies based on risk exposures as the fund manager does. The methodology is simply

running an optimization in order to obtain a similar risk-return profile observed in the fund by investing on instruments that offers exposure to the primary risk factors. The *alternative beta model* is not able to capture hedge positions nor complex strategies and, therefore, control its risk.

5 RESULTS

The goal of this chapter is to discuss the *alternative beta model*'s results and the performance indicators obtained with the implementation of the methodology described in Chapter 4 for each of the eight selected funds.

The procedures described in Section 4.4 were applied to the dataset of daily returns from the chosen funds and risk factors between January 2015 and December 2020. The Kalman filter estimation was implemented following the procedure described in Section 4.4.1, in which we begin with a window of 252 days and run the estimation multiple times, each time adding the information from a new day.[1]

5.1 Analyzing the Replication's Tracking Performance

The analysis of the return accrual graph allows us to compare the performances between the fund and the *alternative beta model* results obtained for the replication period. An example is shown by Figure 1, which demonstrates the comparative performance between VERDE and its replication. Both series begins at 100 and accrual the fund's daily return. The graphs for all funds are displayed in Appendix B.

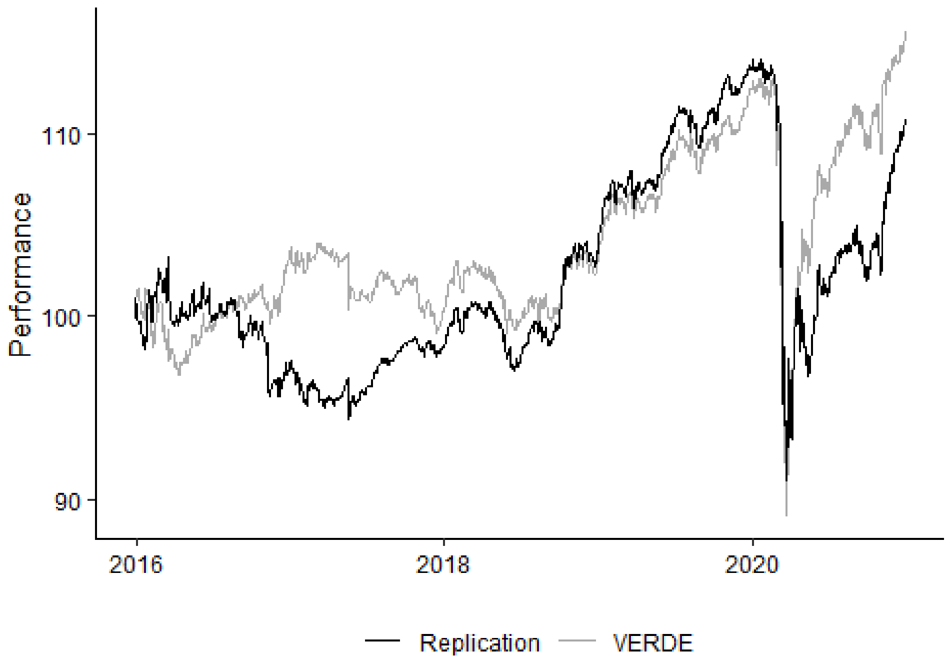

Figure 1 – Performance comparison for VERDE.

[1] For each fund there was a total of 1,296 estimations, leading to an average processing time of 16 hours with the computer's eight processors running in parallel.

It is possible to see that there are moments in which the replication was able to closely track and even outperform the fund, specifically in the beginning of 2016 and during the year of 2019. We highlight the impact of the SARS-COV-19 crisis, in March 2020, in both funds investments, in which the greater market volatility and falling asset prices lead to a significant loss. Even though they were able to recover from this negative result in the months that followed, VERDE fund showed a better performance during this period. We can attribute this result to the effect of the lagged NAV on the replication methodology. Considering this scenario of greater volatility and rising prices, the delay of two days between the fund's and the replication's investment could implicate in the loss of the window of opportunity.

The impact of the SARS-COV-19 crisis can also be observed in the performance replication of other funds. Figure 2 shows the comparative performance between ARX and its replication. Even though in most of the analyzed period the algorithm was able to closely track the fund, when we reached the crisis event in 2020, the replication's behavior became very different from the fund's and their performances decoupled. The original fund was able to continuously improve its performance; the replication, on the other hand, not only had a significant loss, but also showed difficulty in recovering period.

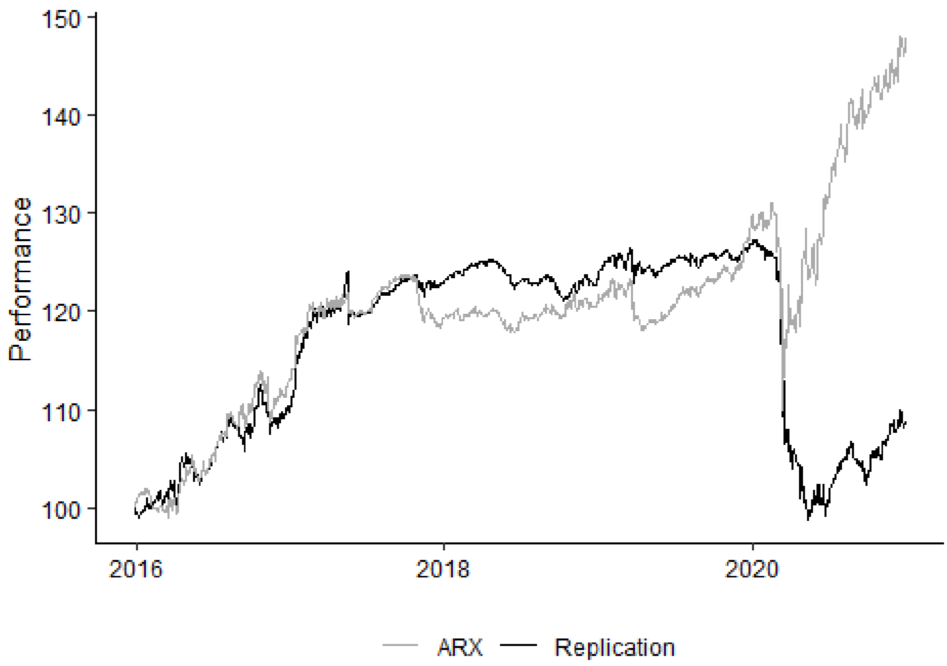

Figure 2 – Performance comparison for ARX.

In order to measure the *alternative beta model*'s adherence when replicating the fund's risk-return profile, we use the *tracking error* and *hit ratio* indicators to analyze the obtained results. Table 3 summarizes these performance indicators calculated for the entire replication period for each selected fund considering the three scenarios (i) paired NAV, (ii) lagged NAV, and (iii) lagged NAV and the transaction cost.

Table 3 – *Tracking error* and *hit ratio* indicators obtained for the entire replication period

Fund	Paired NAV		Lagged NAV		Lagged NAV and TC	
	TE	***HR***	***TE***	***HR***	***TE***	***HR***
ARX	0.163%	49.5%	0.478%	52.7%	0.479%	52.7%
VERDE	0.086%	51.8%	0.267%	52.5%	0.268%	52.5%
BRESSER	0.232%	52.1%	0.477%	51.4%	0.475%	51.4%
JGP	0.679%	50.0%	1.206%	49.5%	1.210%	49.5%
MAPFRE	0.114%	53.3%	0.207%	53.0%	0.207%	53.0%
SAFRA	0.066%	52.4%	0.166%	56.0%	0.165%	56.0%
BTG	0.107%	49.2%	0.251%	49.4%	0.251%	49.4%
SEIVAL	0.697%	50.2%	1.287%	50.1%	1.287%	50.1%

In order to understand whether the *alternative beta model* we have designed was able to explain most of the funds return variation with the chosen systematic risk factors, we must analyze the *tracking error* calculated without the lagged NAV. Without any friction, we wish to know how well did our model capture the fund's risk exposures and track its return. As we can observe, the values obtained for this indicator were less than 1% for all the funds, which is considered to be low. This result confirms that the chosen risk factors were appropriate for the replication process.

Even though the *tracking error* was low, we can highlight that JGP and SEIVAL had the highest values for this indicator, which reveals that the model was efficient, but did not have a great adherence for them. As explained in Section 4.2.2, these two funds had its portfolio exposures concentrated mainly in one risk factor, JGP invested 68% of its net worth in stock and SEIVAL 68% in government bonds. Obtaining a worse result for these funds in comparison to the other ones corroborates the assumption that, when using primary risk factors for the model's implementation, a diversified allocation will lead to a better tracking. Once we wish to replicate a fund whose exposure is concentrated in one risk factor, we should consider deepening the selection to sub factors in order to explore other investment aspects and improve our results.

Regarding the effect of the lagged NAV, we can observe there was an impact on the adherence of the replication process. The *tracking errors* became higher once the replication process had to be done using the weights estimated for two day prior. The transaction costs, however, did not have the same effect, since we can observe the *tracking error* almost did not change once the cost was introduced. Our conclusion would be similar from the one described by Amenc et al. (2003), the transaction costs are not a major concern in terms of replication adherence.

The *hit ratio* indicator oscillated around 50% for all the funds independently of the addition of the lagged NAV and the transaction costs. This means that only on half of the analyzed period our replication was able to provide a better return than the fund.

We have also calculated these indicators for each year in order to understand if periods of lower and higher price volatility could impact our model's ability to track the fund's returns and risk exposures. The results by year for each fund are summarized in Table 4.

The *tracking error* variation over the years leads to the conclusion that the macroeconomic scenarios can impact the *alternative beta*'s performance. We highlight how, comparatively, the values for this indicator became higher during the SARS-COV-19 crisis, in 2020 - when there was a greater price volatility in the market - specially after we introduced the lagged NAV in the methodology. In a moment of greater changes in the asset prices, quick and strategic responses are decisive. Knowing that our replication would have faced a delay in the investment process in comparison with the fund, changes in the allocation would only be made after the disclosure of the NAV. This could mean that, in a scenario of greater volatility, the prices would have already moved and the window of opportunity would be lost.

Even though we were able to notice an impact on the model's adherence caused by the changes in the macroeconomic scenario by observing the changes on the *tracking error* over the years, the same cannot be applied to the *hit ratio*. As in the results obtained for the entire replication period, the ratio oscillated around 50% for all the funds. We can observe that for each fund some years showed a higher ratio, in which the replication outperformed the fund almost 60% of the time. However, we were not able to find a consistency among the funds for an specific year in which this happened, nor a correlation between *tracking error* value and *hit ratio*.

Table 4 – *Tracking error* and *hit ratio* indicators obtained for each year.

Fund	**Year**	**Paired NAV**		**Lagged NAV**		**Lagged NAV and TC**	
		TE	***HR***	***TE***	***HR***	***TE***	***HR***
ARX	2016	0.016%	41.9%	0.358%	52.7%	0.359%	52.7%
	2017	0.111%	57.1%	0.317%	60.6%	0.316%	60.6%
	2018	0.103%	53.7%	0.143%	53.7%	0.143%	53.7%
	2019	0.150%	46.1%	0.192%	46.9%	0.192%	46.9%
	2020	0.297%	49.6%	0.927%	50.4%	0.932%	50.4%
VERDE	2016	0.004%	45.3%	0.388%	51.9%	0.388%	51.9%
	2017	0.087%	59.5%	0.193%	59.1%	0.195%	59.1%
	2018	0.098%	52.9%	0.139%	51.8%	0.139%	51.8%
	2019	0.070%	49.6%	0.093%	50.4%	0.093%	50.4%
	2020	0.125%	51.6%	0.375%	49.6%	0.376%	49.6%
BRESSER	2016	0.006%	52.3%	0.544%	53.1%	0.542%	53.1%
	2017	0.232%	55.6%	0.458%	53.7%	0.454%	53.7%
	2018	0.214%	52.5%	0.309%	52.1%	0.308%	52.1%
	2019	0.316%	50.8%	0.432%	50.4%	0.431%	50.4%
	2020	0.267%	48.8%	0.595%	48.4%	0.594%	48.4%
JGP	2016	0.633%	47.7%	1.145%	47.7%	1.162%	47.7%
	2017	0.569%	55.2%	1.092%	53.7%	1.093%	53.7%
	2018	0.815%	49.4%	1.241%	48.6%	1.244%	48.6%
	2019	0.678%	52.3%	1.004%	52.3%	1.004%	52.3%
	2020	0.691%	45.7%	1.506%	45.7%	1.505%	45.7%
MAPFRE	2016	0.003%	54.3%	0.116%	50.0%	0.116%	50.0%
	2017	0.066%	51.4%	0.129%	52.1%	0.129%	52.1%
	2018	0.143%	55.6%	0.266%	57.6%	0.265%	57.6%
	2019	0.131%	56.6%	0.183%	57.0%	0.182%	57.0%
	2020	0.153%	48.8%	0.284%	48.8%	0.286%	48.8%
SAFRA	2016	0.004%	46.9%	0.204%	56.2%	0.203%	56.2%
	2017	0.080%	52.5%	0.164%	58.3%	0.163%	58.3%
	2018	0.107%	57.6%	0.198%	58.4%	0.198%	58.4%
	2019	0.041%	58.9%	0.080%	58.5%	0.079%	58.5%
	2020	0.047%	47.3%	0.154%	48.8%	0.154%	48.8%
BTG	2016	0.004%	47.7%	0.363%	48.4%	0.362%	48.4%
	2017	0.110%	50.2%	0.193%	51.0%	0.194%	51.0%
	2018	0.136%	48.6%	0.197%	49.8%	0.197%	49.8%
	2019	0.121%	50.4%	0.160%	51.2%	0.161%	51.2%
	2020	0.111%	48.4%	0.286%	46.9%	0.287%	46.9%
SEIVAL	2016	0.616%	50.4%	1.217%	50.0%	1.215%	50.0%
	2017	0.574%	53.3%	1.119%	54.1%	1.121%	54.1%
	2018	0.819%	49.4%	1.240%	49.4%	1.241%	49.4%
	2019	0.682%	52.3%	1.009%	51.9%	1.008%	51.9%
	2020	0.778%	46.1%	1.746%	45.3%	1.746%	45.3%

5.2 Practical Aspects of the Implementation

When bringing the implementation into practice, the *alternative beta model* must guide the daily investment decision in order to allow for the fund's risk-return profile replication. Thus, by running the designed algorithm, the Kalman filter model will produce as an output the daily optimal exposure for each risk factor, meaning how much of the replication fund's net worth has to be invested on each factor to achieve good tracking results.

The literature regarding the *alternative beta model* usually includes a restriction on the model's optimization so that the sum of the obtained weights is equal to one. This restrains the replication fund from investing more than its net worth. Given our intention of bringing the model's implementation closer to a real life situation, we have not forced the weights to add to one, since the funds we aim to replicate are commonly leveraged.

Figure 3 shows an example of the optimal exposures for each risk factor over the entire replication period for VERDE, the graphs for all funds are displayed in Appendix C. We can observe that up to April 2017 there was a significant volatility in the exposure, reaching a leverage level of 2.5 times the replication fund's net worth on nominal compounding interest rate. After this period, the weight curves became smoother and the investment process would have become easier, given there would not be considerable changes in the fund's positions.

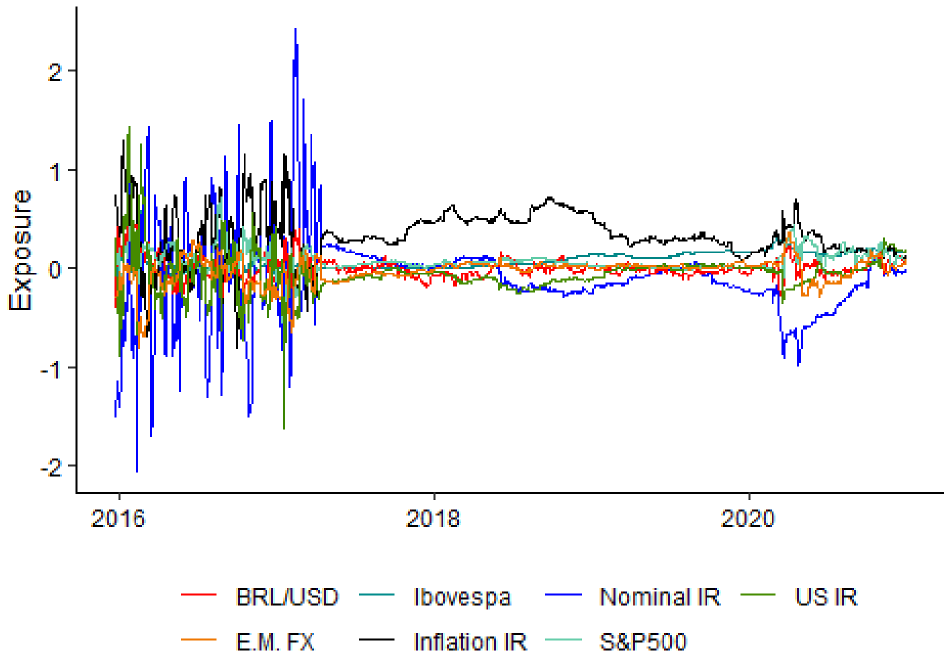

Figure 3 – Risk factor exposures for VERDE.

This greater volatility and change of behavior in the fund allocation can also be identified in other funds. Figure 4 shows the optimal exposures obtained for ARX. As

observed in VERDE, in the beginning of the analyzed period there is a greater allocation's volatility specially for the nominal interest rate factor. The exposures evolve for a period of stabilization and later, during the SARS-COV-19 crisis, there is a new and more significant period of volatility.

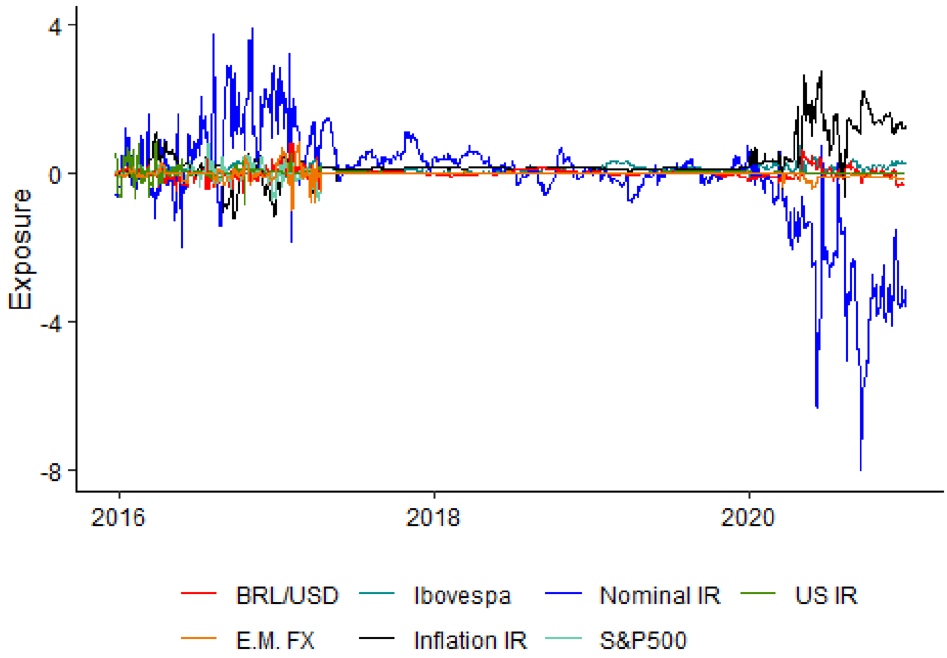

Figure 4 – Risk factor exposures for ARX.

In order to explain this volatility on the beginning of the period, we have run a series of analysis testing if there was a particularity in the fund's return series near the date or if it was a period of time necessary for the model to reach stabilization during the Kalman filter implementation. However, none of the tests came to a conclusion and this turning date seemed to be inherent to the dataset and model's optimization.

The observed volatility and leverage in the exposures could represent a challenge in practical terms. The instruments chosen to represent the risk factors and make the implementation possible in a real market condition are mainly Exchange Traded Funds (ETFs) NAV. An important characteristic of this asset is that it is cash, its settlement involves the exchange of a financial amount between buyer and seller. Considering that our *alternative beta model* allows for leverage and that we cannot guarantee that the replication will compensate the financial value invested in a long position with the amount on a short position, the investment scenario using cash assets would involve borrowing multiples of the fund's net worth. This would probably be unfeasible in reality.

The replication becomes even more difficult when the optimal exposure involves a short position and we face the impossibility of a naked short selling in the Brazilian market. After taking a short position on the ETF, our fund would be obligated to borrow

the asset on the market and pay a fee during the period. Not only this market is not liquid enough to allow for considerable short position, as the leveraged ones, but also this would mean an additional cost to the process.

The solution proposed that would make the implementation possible for moments in which the optimal allocation requires the fund to take a leveraged position is to use a future contract with the same exposure as the ETF. Future contracts do not involve a financial settlement in the trading, but only on the expiration date or on the position closing, therefore the replication fund would have the possibility of not leaving the money standing and investing it on the one-day interbank deposit overnight, improving its profitability during those periods.

In this scenario, we would be considering the usage of the following future contracts:

- Brazilian stock market: The first contract month of the Ibovespa future contract.
- Brazilian nominal compounding interest rate: The one-day interbank deposit (DI) future contract with the closest duration to the IRFM11.
- Brazilian inflation protected interest rate: DI x IPCA future contract with the closest duration to the IMAB11.
- US interest rate: The first contract month of the 10-year Treasury future contract.

The ETF representing the Emerging Market Foreign Exchange Rate does not have a future contract with the same exposure that would make the replacement possible. Thus, even on moments of greater exposure volatility, we would still be representing this risk factor with the WisdomTree's ETF.

Another challenge in the replication process is whether the leverage positions will be accepted on both ETFs and future contracts considering the exchange and clearing house risk controls and if the margin call for the derivatives would not be as great as to make the investment unfeasible. The optimal exposure generated by the model could lead to a great leverage and the need of multiples of the fund's net worth to be invested on a single asset. This may turn out to be impractical because the exchange and clearing house risk control would not allow such an investment to be done.

In a scenario in which there is a great leverage and the replication fund is not allowed to make all necessary investment because of risk precautions, we would have to attempt to fulfill most of the designated investment and deal with not achieving a full replication on that day.

5.3 Comparing the Risk Measure

Concerning the risk measure analysis, the comparison between the obtained *maximum drawdown* indicators for the fund and the *alternative beta model* allows us to understand whether the replication process lack of risk control could be a significant drawback in real life implementation. In Figure 5 we show an example of the *maximum drawdown* comparison graph between VERDE and its replication. In order to make the calculations, we have used a rolling window of 252 days, therefore the comparison period begins in 2017. The graphs for all funds are displayed in Appendix D.

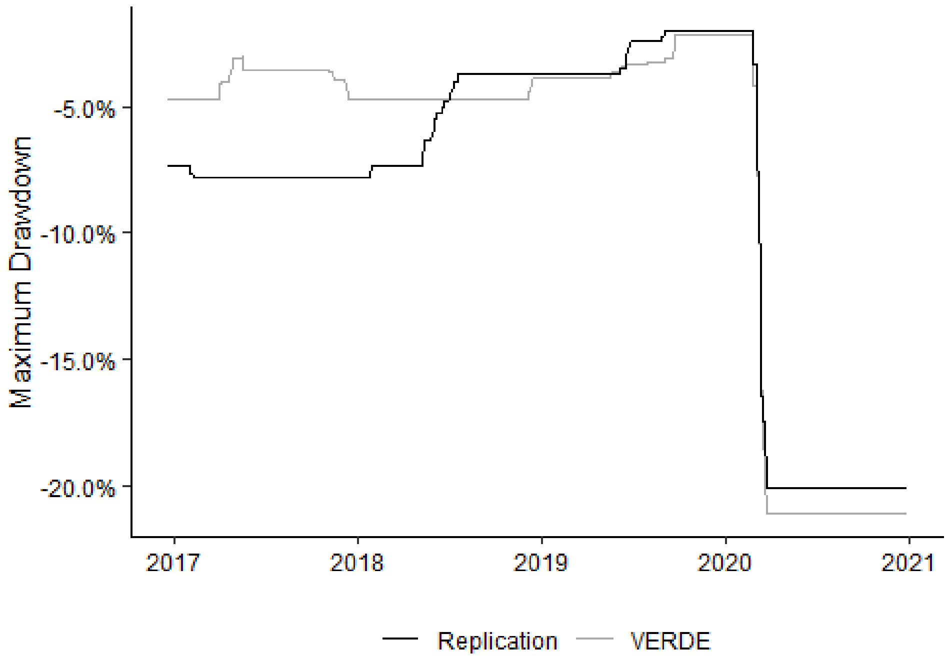

Figure 5 – Maximum Drawdown comparison for VERDE.

We can observe that VERDE had a better risk indicator in the beginning of the comparison period, when its *maximum drawdown* was smaller than the replication's one. We highlight how both funds risk controls were affected by the SARS-COV-19 crisis, when their maximum loss in the portfolio value reached the level of 20%.

When analyzing the *maximum drawdown* comparison between ARX and its replication, shown in Figure 6, we can observe that the allocation's volatility in the beginning of the analyzed period did not impact the risk measure as happened with VERDE. During this period both funds' risk indicator showed similar values and behavior. After the 2020 crisis, however, when there was a greater volatility in the replication's exposures, the replication's *maximum drawdown* became worse than the fund's. Even though we cannot prove by using the *alternative beta model*, this behavior may be inferred as an efficient hedge strategy performed by the original fund.

The greater volatility in the replication's exposures seems to have a correlation with

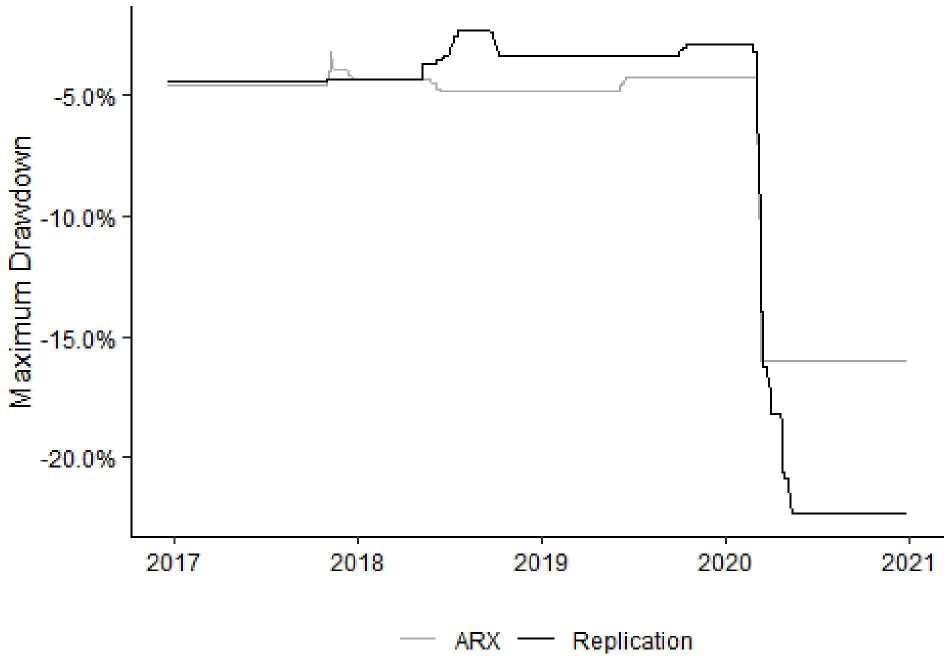

Figure 6 – Maximum Drawdown comparison for ARX.

worse risk indicators. Moments of considerable changes in the positions were those when the replication process showed higher values for the *maximum drawdown* in comparison to the fund. This could be indicative of the model's lack of risk control once it allows for considerable changes in the allocations.

6 CONCLUSIONS

The purpose of this work was to introduce practical aspects into the *alternative beta* methodology in order to bring its implementation closer to what would happen in a real life experience. The concept of risk-return profile replication is still recent in the academic literature and the developments we have observed over the years are mostly theoretical and surrounds the model used, little is explored on the issues of an actual implementation and how they impact the replication performance.

Seeking to address this constraint, we have introduced three aspects in the designed algorithm that would create a similar condition to a real implementation. Firstly, we considered the fact that, in order to obtain a good adherence on the fund's returns replication, we would have to make daily investments and portfolio rebalancing. As a consequence, we would need to run the optimization every day, only adding the latest information available in the market, which could lead to a worse estimation of the risk factors weights.

We have also have considered the fact that, even though not being the ideal situation for the replication process, the NAV available for the estimation would always be lagged, since its pricing is not immediate. This aspect introduces a challenge to the implementation, the fund would have already established its strategy and positions and the replication would have a delay of two days to make its investments, facing the risk of losing opportunity windows or dealing with market adversities.

The last practical aspect was the introduction of transaction costs. This element is often left out of academic works for being considered to have a marginal effect on the model's results. However, not considering these costs could be biasing our results, specially when thinking the fund we aim to replicate discounts its investment expenses from its returns.

We have chosen to include in our model seven primary risk factors that would better represent the main exposures observed for a Brazilian fund with access to the offshore market in order to guarantee we would be able to explain most of the funds return variation. For each selected factor, we have designated an exchange-traded asset that would have enabled a real investment situation and that was used as the data input for the estimation. We implemented the designed replication algorithm to eight Brazilian Multimarket funds between January 2015 to December 2020.

We were able to obtain good adherence on all eight funds replication with low values for the *tracking error* indicator. The algorithm was able to closely replicate and, in half of period, even outperform the original funds results. By analyzing the results

considering the paired NAV scenario, we were able to conclude that the selected risk factors were appropriate for the replication process and did capture most of the funds risk exposures. Although also showing a good tracking performance, the model was not as efficient for JGP and SEIVAL, funds whose portfolio exposure was concentrated mainly in one risk factor. This corroborates for the assumption that, when using primary risk factors for the model's implementation, a diversified allocation will lead to a better tracking.

We observed that practical aspects of the *alternative beta model*'s implementation did impact its replication performance. Both the introduction of the lagged NAV and the transaction costs had an effect on the tracking adherence indicators, however the costs impact was almost imperceptible compared to the impact on the *tracking error* caused by the lagged NAV. Even though the replication became worse after the inclusion of these factors, the impact was not as great as to make a real life implementation unfeasible. Another feature we wish to highlight is the influence macroeconomic scenarios and price volatility can have on the model's replication performance.

For the purpose of trading, the greater exposure volatility combined with leverage could indicate a practical challenge for the *alternative beta model*'s implementation, specially when using ETFs as the chosen investment instrument. We proposed as a solution the usage, in those moments, of future contracts with the same risk exposure as the ETF, allowing for the necessary leverage and short selling strategies.

The greater volatility in the replication's exposures did also impact the risk indicator, since events of considerable changes in the allocation were those when the replication process showed higher values for the *maximum drawdown* in comparison to the fund. On other periods, both funds risk measures were very close, even showing similar behaviour during the SARS-COV-19 crisis.

Extensions of this work could include: improvement on the investment dynamics with the usage of the future contracts when necessary and with the inclusion of a stop loss mechanism that would prevent major drawdowns for the replication and lead to better results; the development of an approach using using subfactors, such as sector indexes or different bond durations, instead of the primary risk factors in order to analyze whether they could provide better results for the replication of funds whose portfolio is concentrated in few exposures; and the implementation of the methodology using linear regression with rolling window in order to understand if the Kalman filter introduces effiency gains in the replication process even with the practical aspects.

REFERENCES

Amenc, N., S. El Bied, and L. Martellini (2003). Predictability in hedge fund returns (corrected). Financial Analysts Journal 59(2), 32–46.

Amenc, N., L. Martellini, J. Meyfredi, and V. Ziemann (2010). Passive hedge fund replication: beyond the linear case. European Financial Management 16(2), 191–210.

Durbin, J. and S. Koopman (2012). Time series analysis by state space methods. Oxford University Press.

Harvey, A. (1989). Forecasting, structural time series models and the Kalman Filter. Cambridge University.

Harvey, A. (2006). Forecasting with unobserved components time series models. North-Holland.

Hasanhodzic, J. and A. Lo (2007). Can hedge-fund returns be replicated? the linear case. Journal of Investment Management 5(2), 5–45.

Jaeger, L. and C. Wagner (2005). Factor modeling and benchmarking of hedge funds: can passive investments in hedge fund strategies deliver? Journal of Alternative Investments 8(3), 9–36.

Kat, H. (2007). Alternative routes to hedge fund return replication. Journal of Wealth Management 10(3), 25–39.

Lima, R. (2007). Replicando Retornos de Hedge Funds - Uma Aplicação ao Mercado Brasileiro. Master's thesis, IBMEC - RJ.

Mergner, S. (2009). Application of state space models in finance. Universitatsverlag Gottingen.

Roncalli, T. and J. Teiletche (2008). An alternative approach to alternative beta. Journal of Financial Transformation 24, 43–52.

Roncalli, T. and G. Weisang (2011). Tracking problems, hedge fund replication and alternative beta. Journal of Financial Transformation 31, 19–29.

Schutt, I. and J. Caldeira (2016). Análise de estilo dinâmica de fundos multimercados: Aplicação para o mercado brasileiro. Análise Econômica 34.

Sharpe, W. (1992). Asset allocation: Management style and performance measurement. Journal of Portfolio Management 18, 7–19.

Tancar, R. and J. Viebig (2008). Alternative beta applied: an introduction to hedge fund replication. Financial Markets and Portfolio Management 22(3), 259–279.

Weisang, G. (2014). Factor selection in dynamic hedge fund replication models: A bayesian approach. Bayesian Model Comparison 34, 181–222.

APPENDICES

APPENDIX A – BRAZILIAN FUND INDUSTRY CLASSIFICATION

The Brazilian Securities and Exchange Commission (CVM) has divided the fund industry into four classes according its investment policy. The segmentation is based on the funds portfolio's primary risk factor, being the four categories Fixed Income, Stocks, Multimarket and Foreign Exchange Rate.

Funds classified as Fixed Income have as the main risk factor the interest rate or price index movements. The regulation specifies to them a minimum allocation rule of 80% in assets related to these two risk factors. Fixed Income funds can also receive four different subclassification:

i Short Term: The fund's portfolio assets have an weighted average term of less than 60 days;

ii Referenced: The fund invests at least 95% of its equity in assets that reference its benchmark;

iii Foreign Debt: The fund invests at least 80% of its equity in Fixed Income assets issued abroad;

iv Simple: The fund invests at least 95% of its equity in Government Bonds.

In Foreign Exchange Rate funds, the portfolio's main risk factor is the fluctuation in the price of the foreign currency or the variation of an interest rate called the foreign exchange coupon. They must keep at least 80% of their net worth invested in assets that are directly or indirectly related to these risk factors.

Stock funds are funds created with the objective of investing in the stock market, therefore, their main risk factor is the variation in the prices of shares. They must invest at least 67% of their equity in shares admitted to trading on the organized market or in related assets, such as bonuses or subscription receipts, share deposit certificates, shares of stock funds, shares of stock index funds, or Brazilian Depositary Receipts (BDR) classified as levels II or III.

Multimarket funds have greater management freedom, since they have an investment policy that involves several risk factors, without a minimum allocation rule on any particular factor. They can invest in assets from different markets — fixed income, foreign exchange rate, equities — and may use derivatives for both leverage and portfolio protection.

APPENDIX B – REPLICATION PERFORMANCE COMPARISON

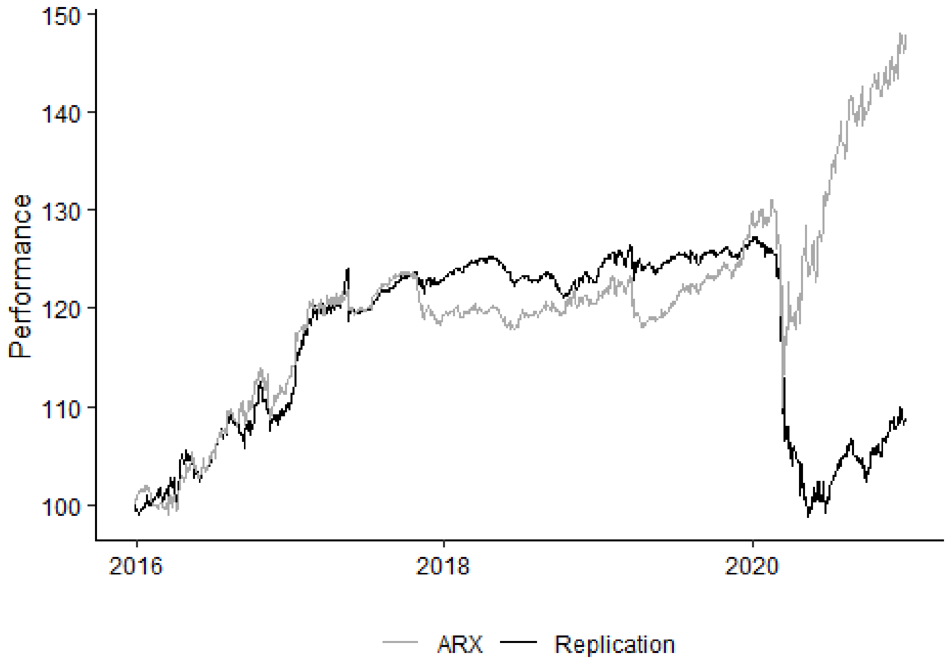

Figure 7 – Performance comparison for ARX.

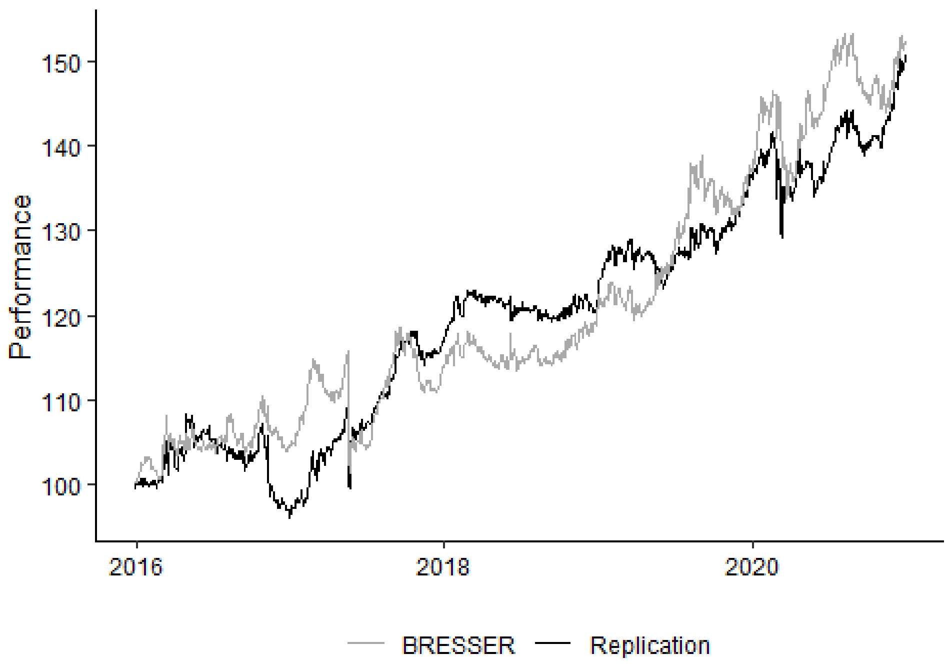

Figure 8 – Performance comparison for BRESSER.

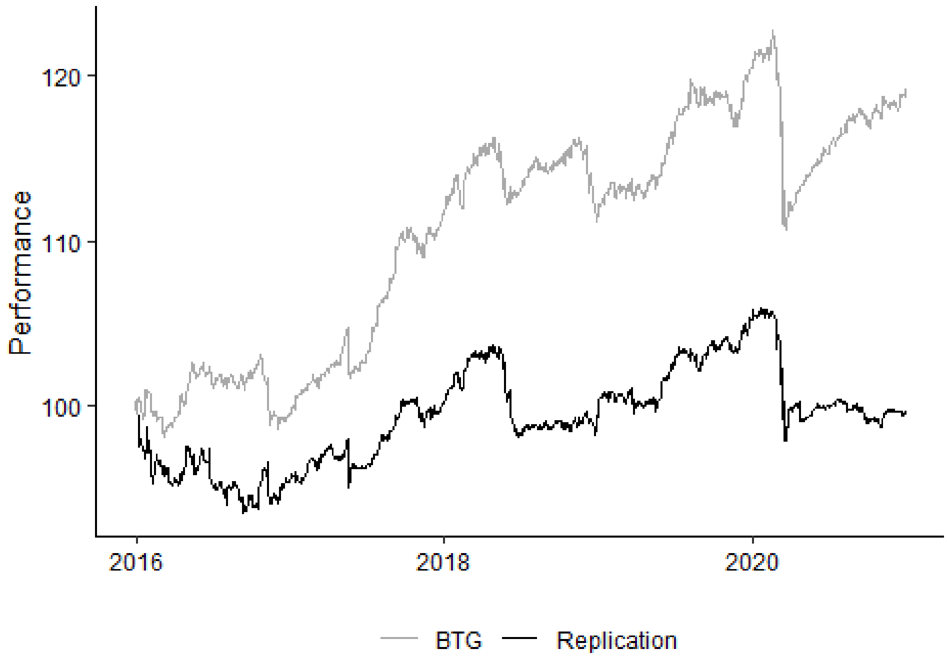

Figure 9 – Performance comparison for BTG.

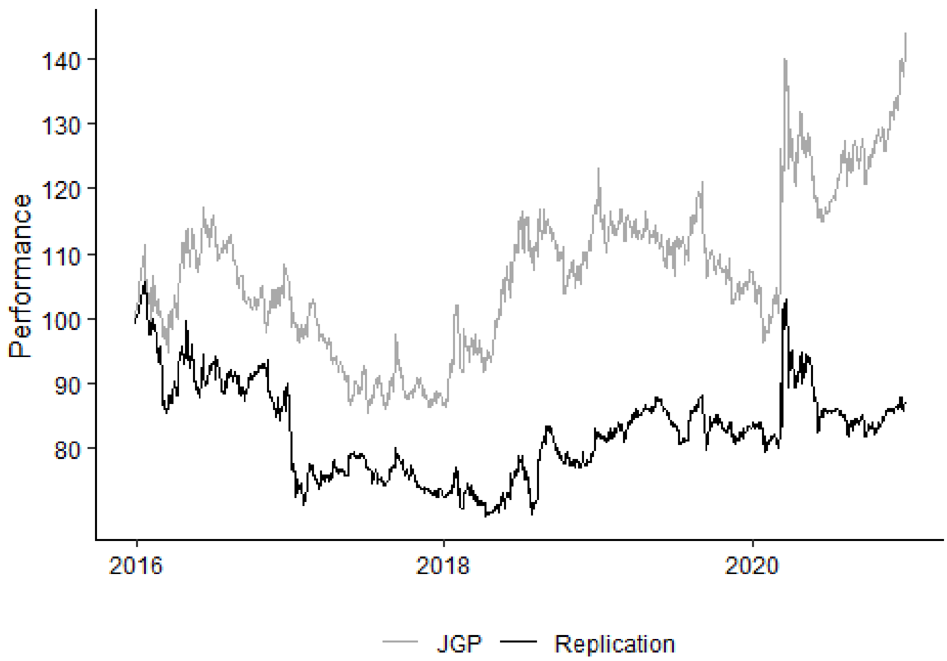

Figure 10 – Performance comparison for JGP.

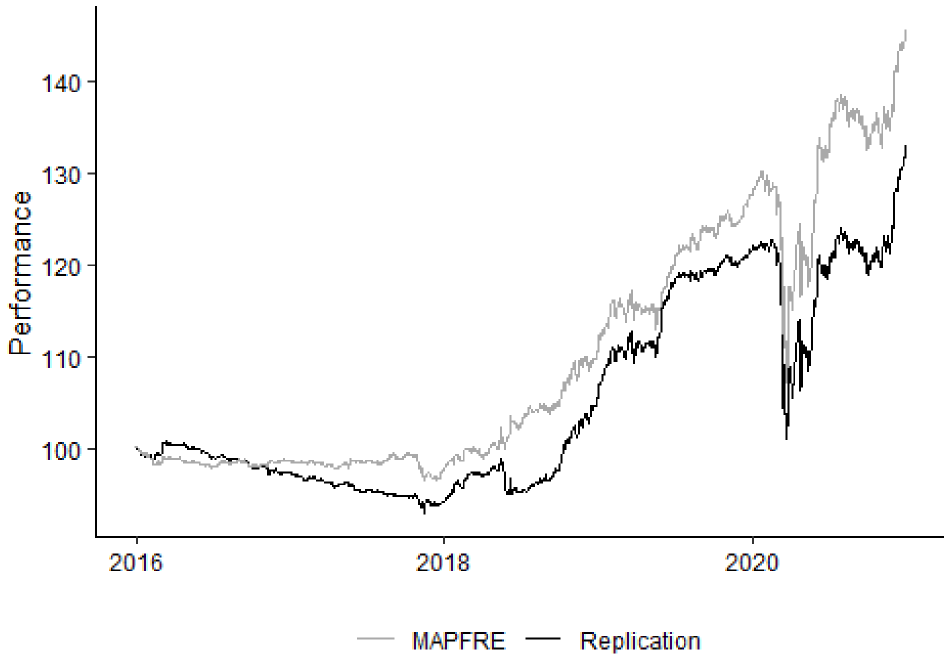

Figure 11 – Performance comparison for MAPFRE.

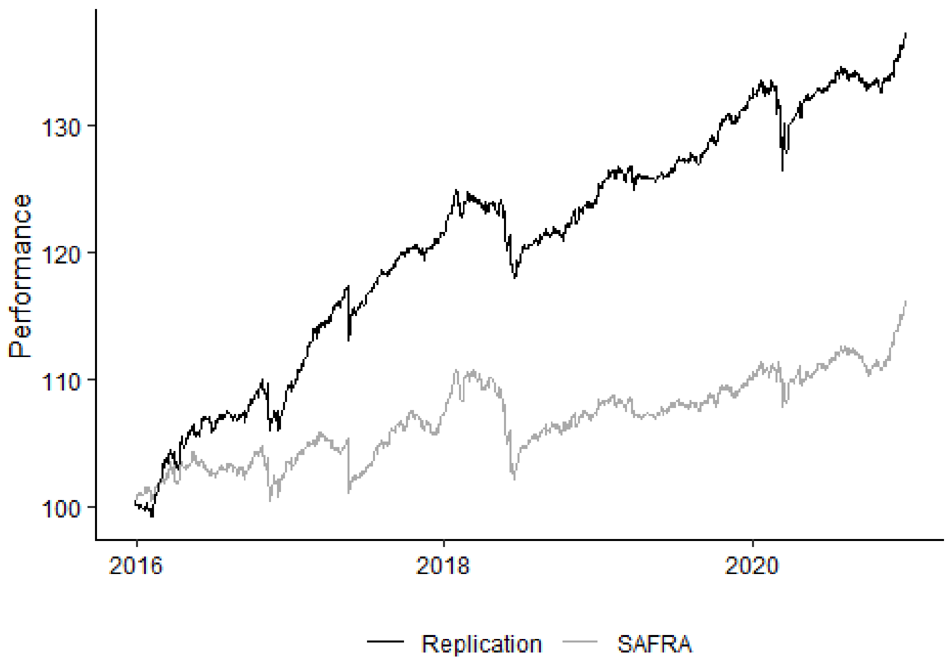

Figure 12 – Performance comparison for SAFRA.

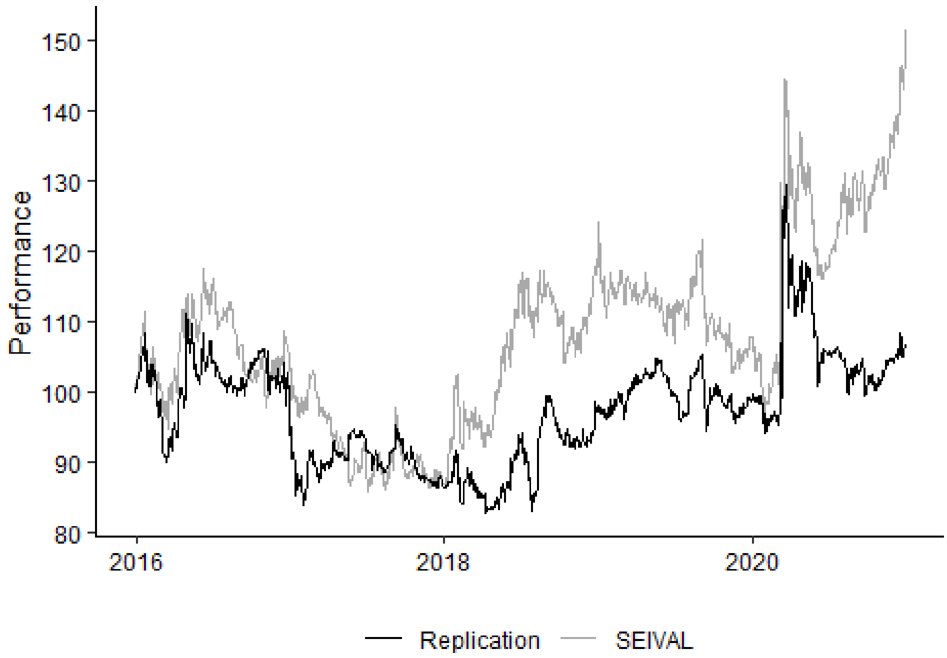

Figure 13 – Performance comparison for SEIVAL.

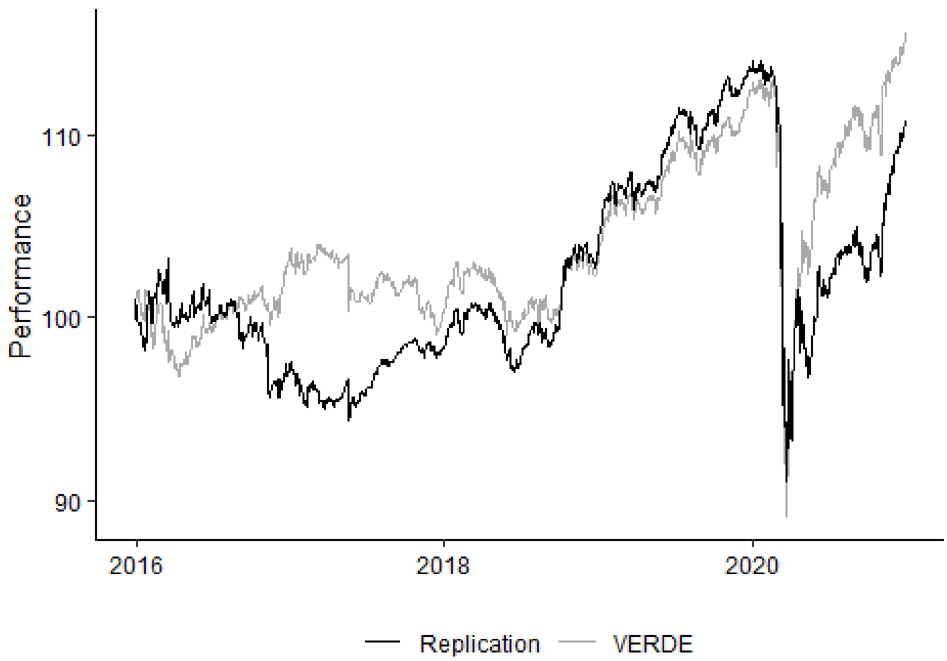

Figure 14 – Performance comparison for VERDE.

APPENDIX C – RISK FACTOR EXPOSURES

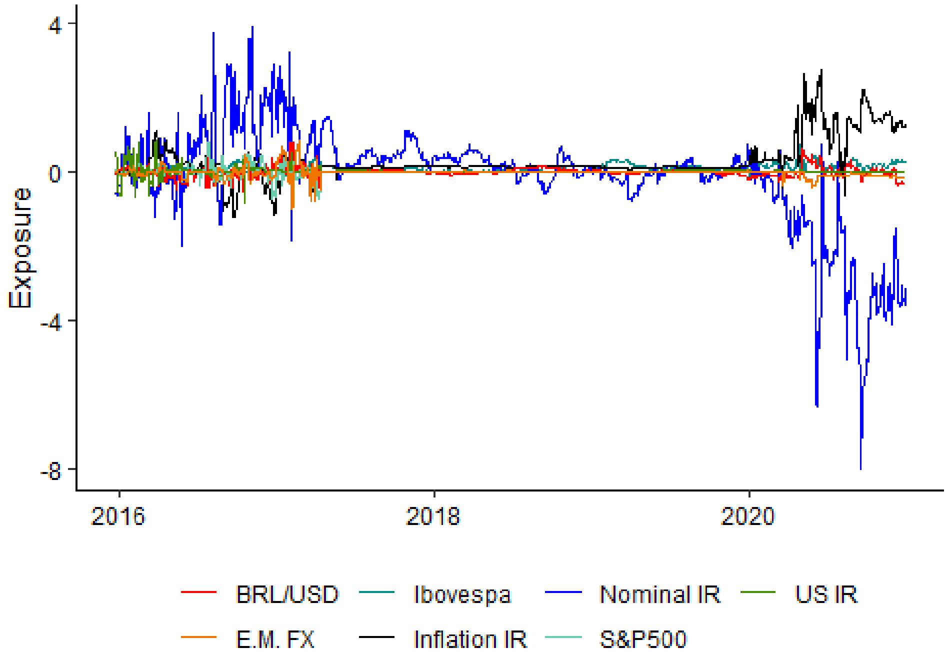

Figure 15 – Risk factor exposures for ARX.

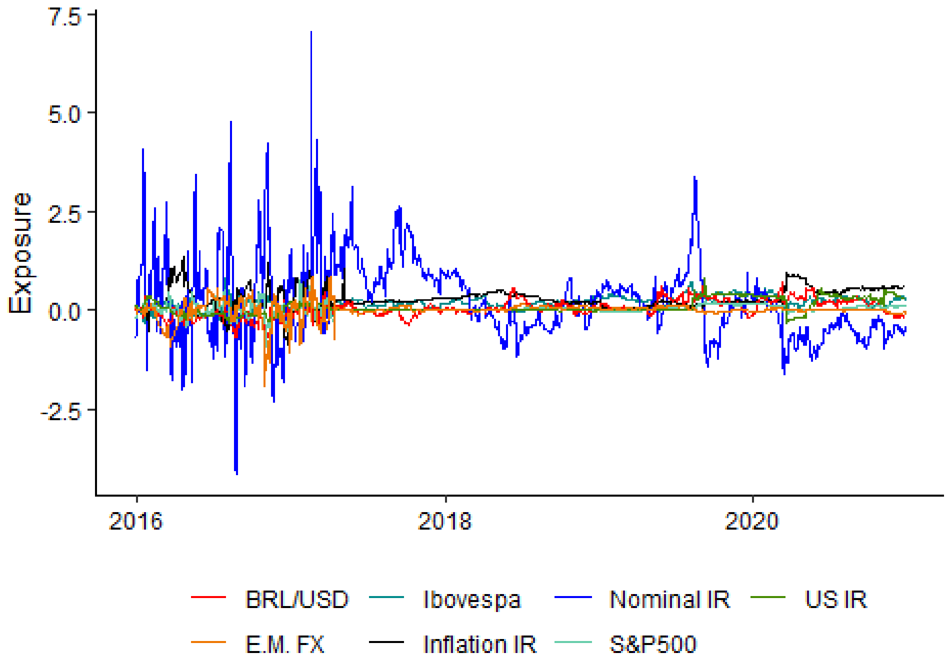

Figure 16 – Risk factor exposures for BRESSER.

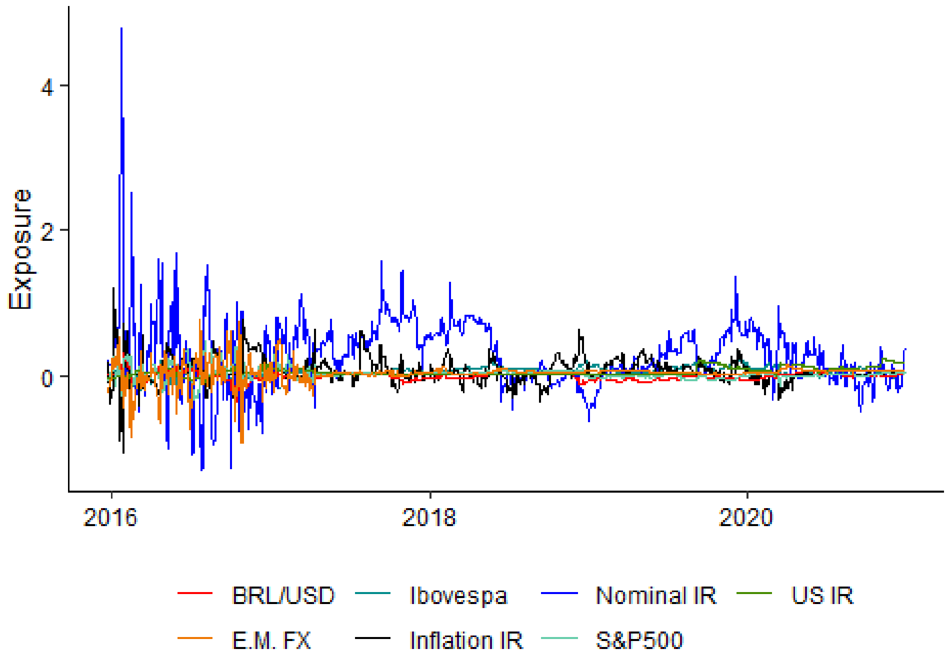

Figure 17 – Risk factor exposures for BTG.

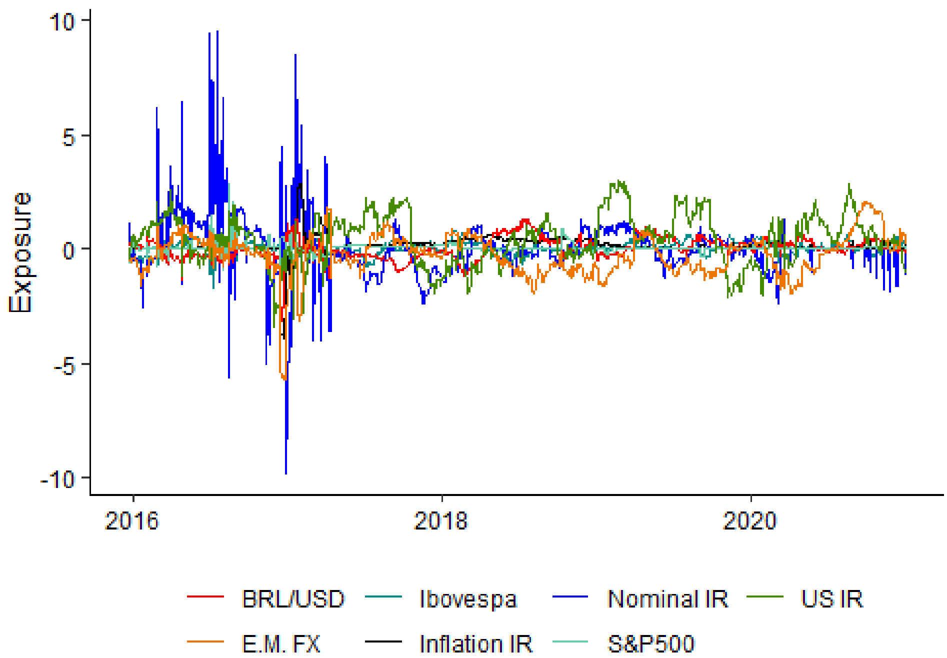

Figure 18 – Risk factor exposures for JGP.

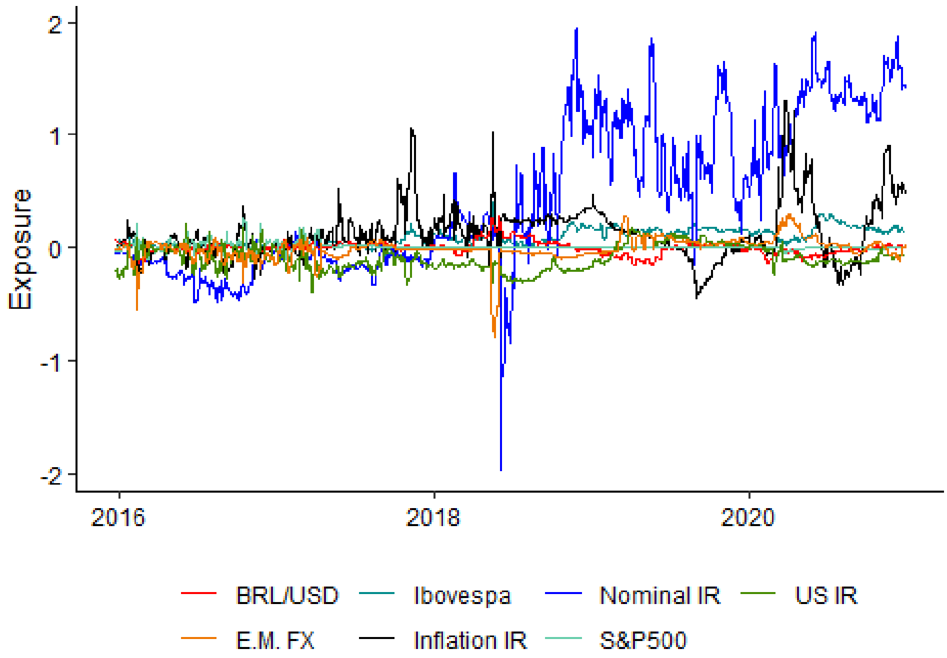

Figure 19 – Risk factor exposures for MAPFRE.

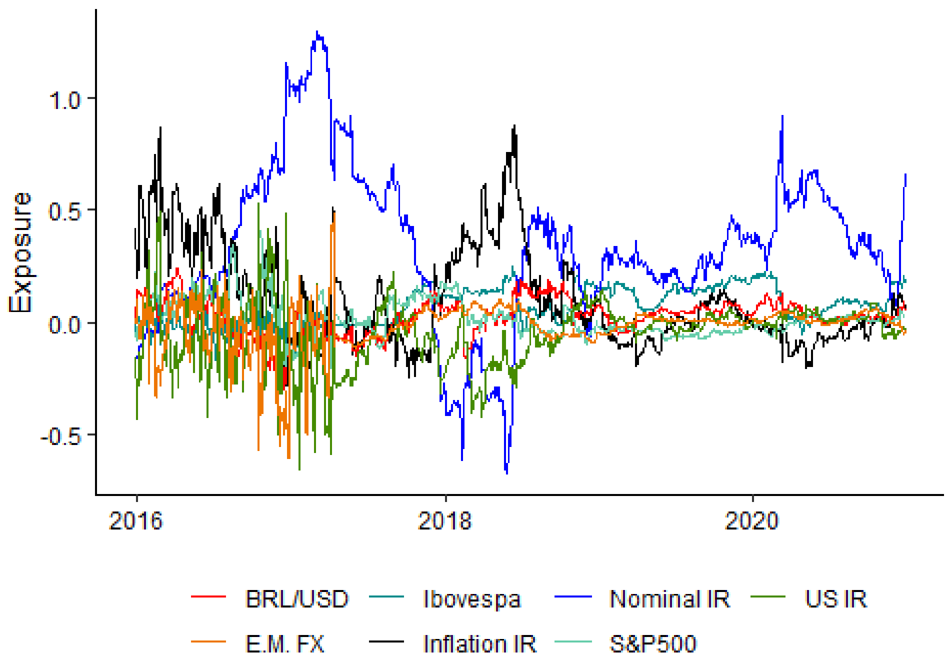

Figure 20 – Risk factor exposures for SAFRA.

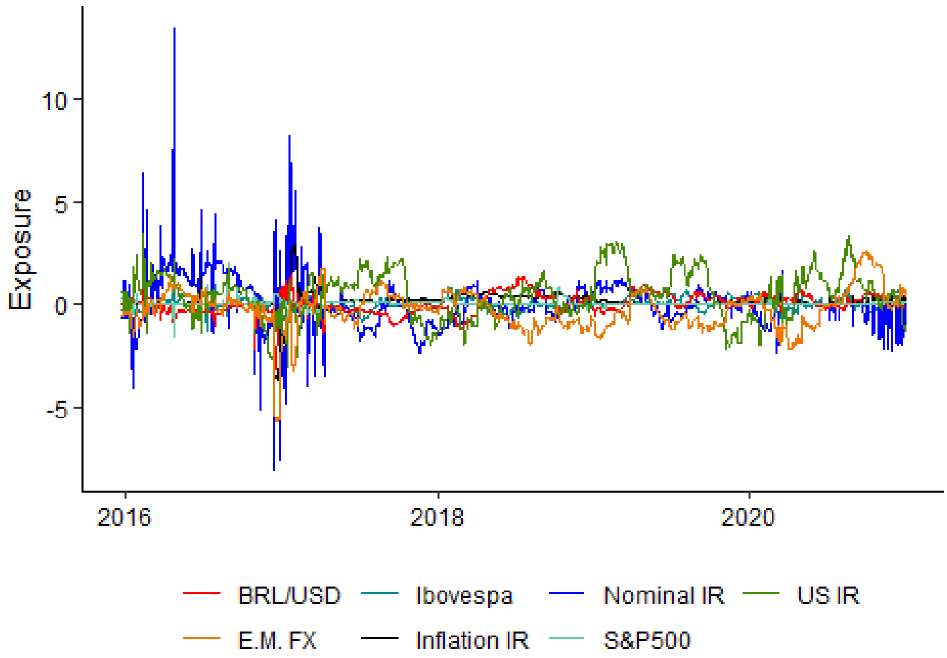

Figure 21 – Risk factor exposures for SEIVAL.

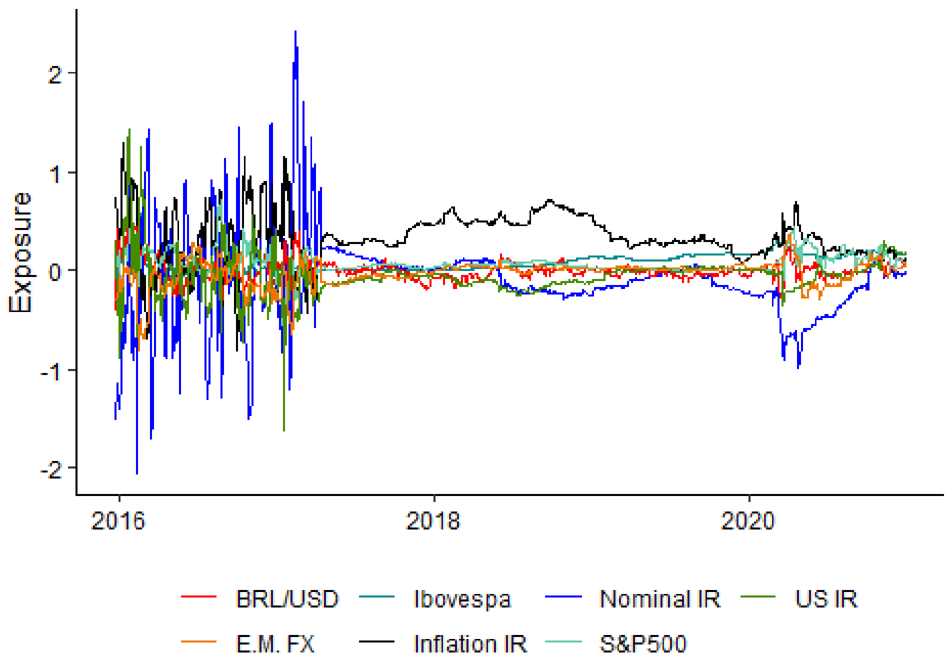

Figure 22 – Risk factor exposures for VERDE.

APPENDIX D – MAXIMUM DRAWDOWN COMPARISON

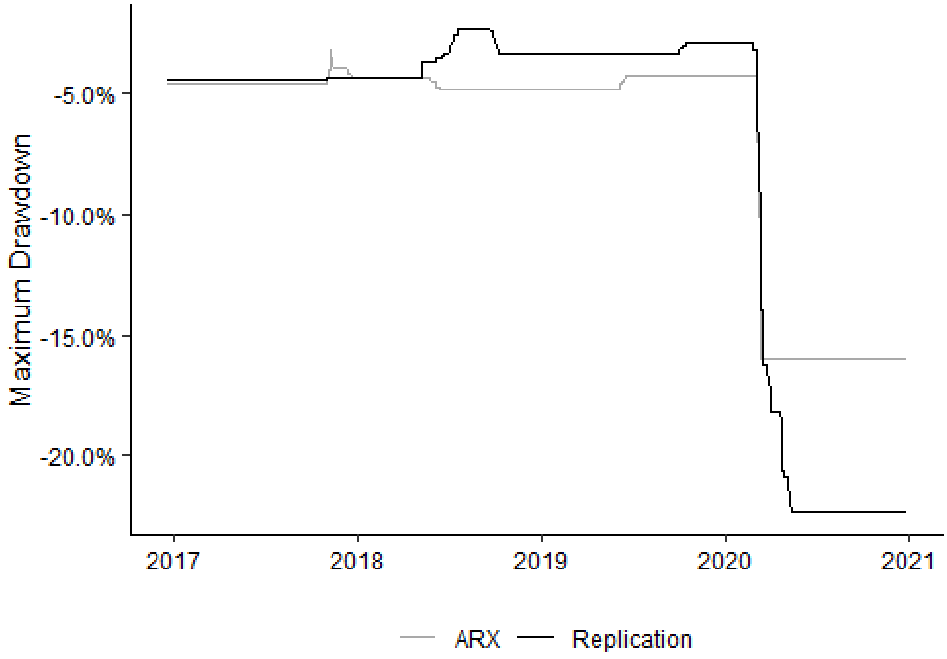

Figure 23 – Maximum Drawdown comparison for ARX.

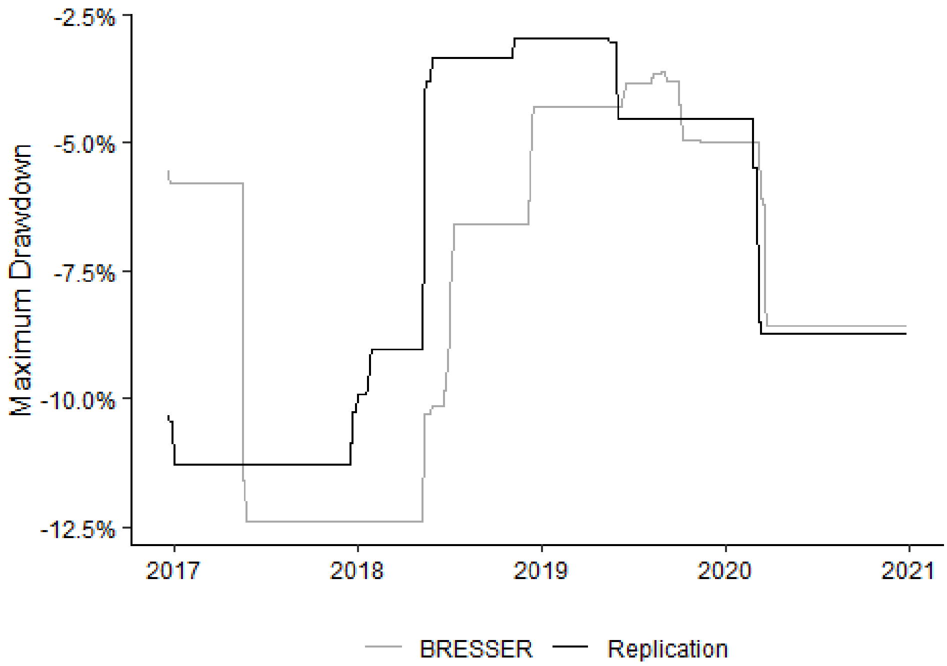

Figure 24 – Maximum Drawdown comparison for BRESSER.

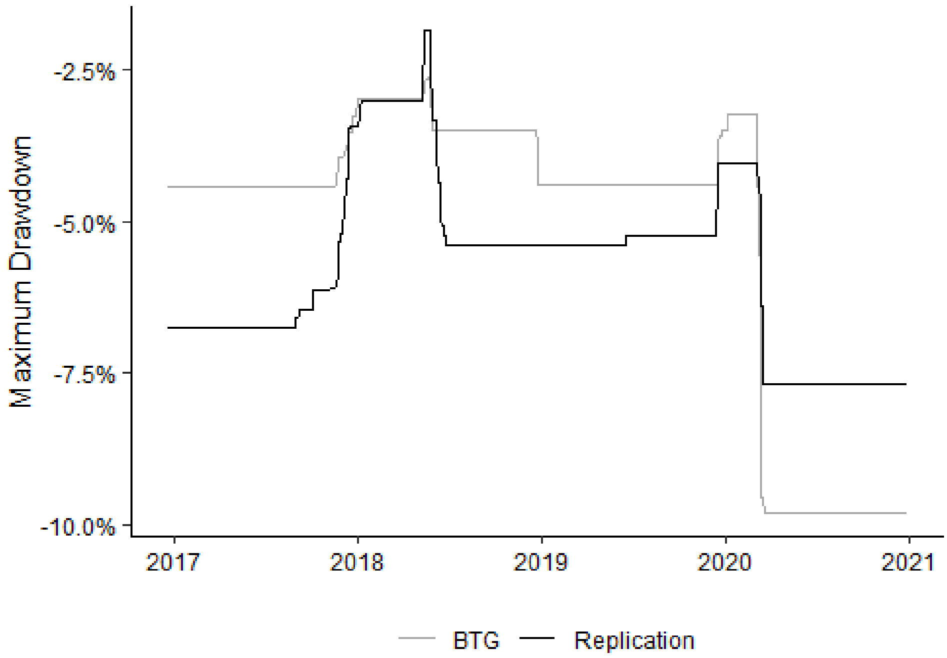

Figure 25 – Maximum Drawdown comparison for BTG.

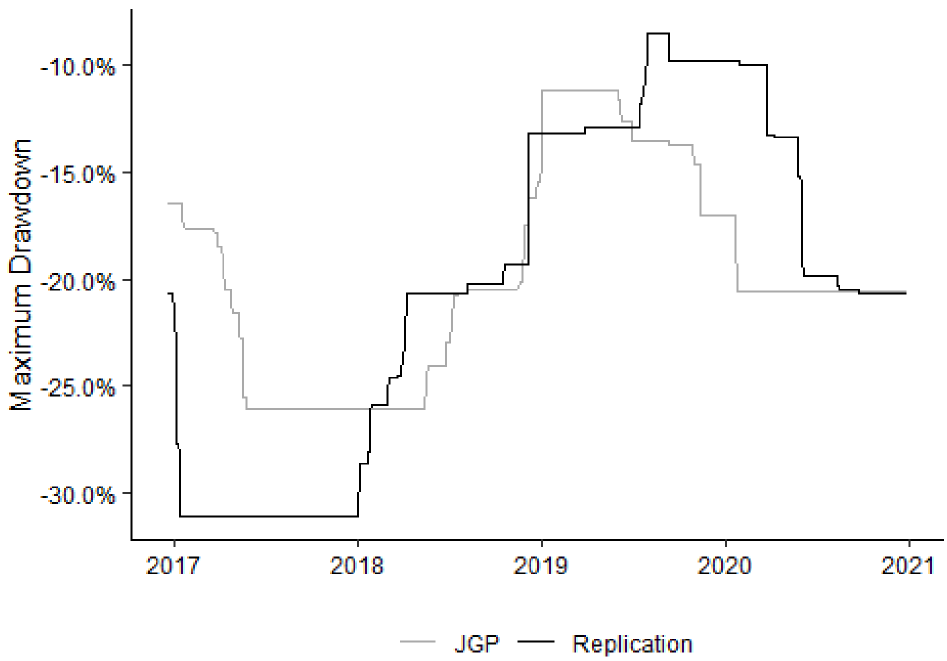

Figure 26 – Maximum Drawdown comparison for JGP.

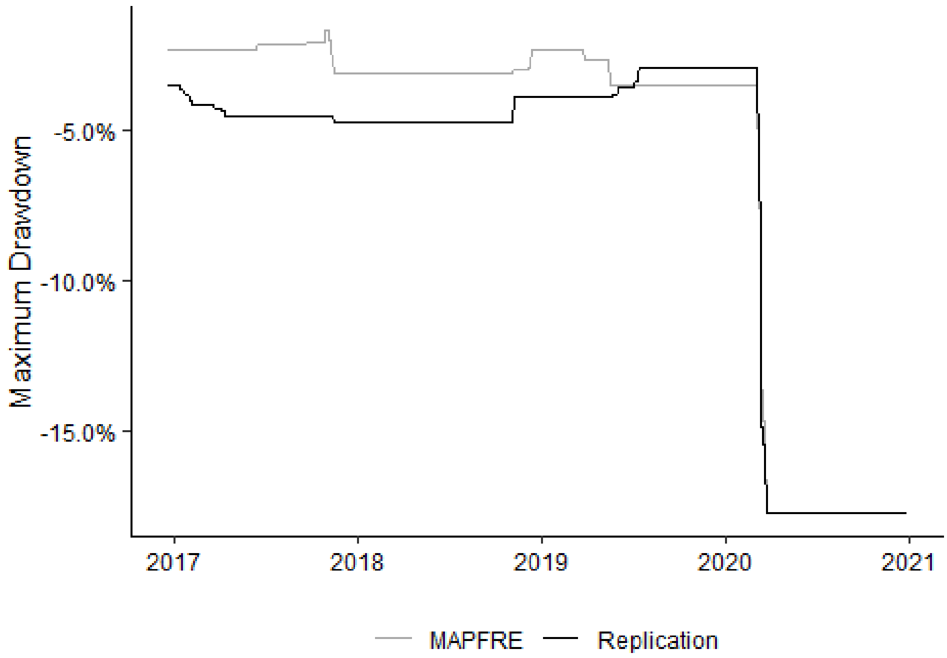

Figure 27 – Maximum Drawdown comparison for MAPFRE.

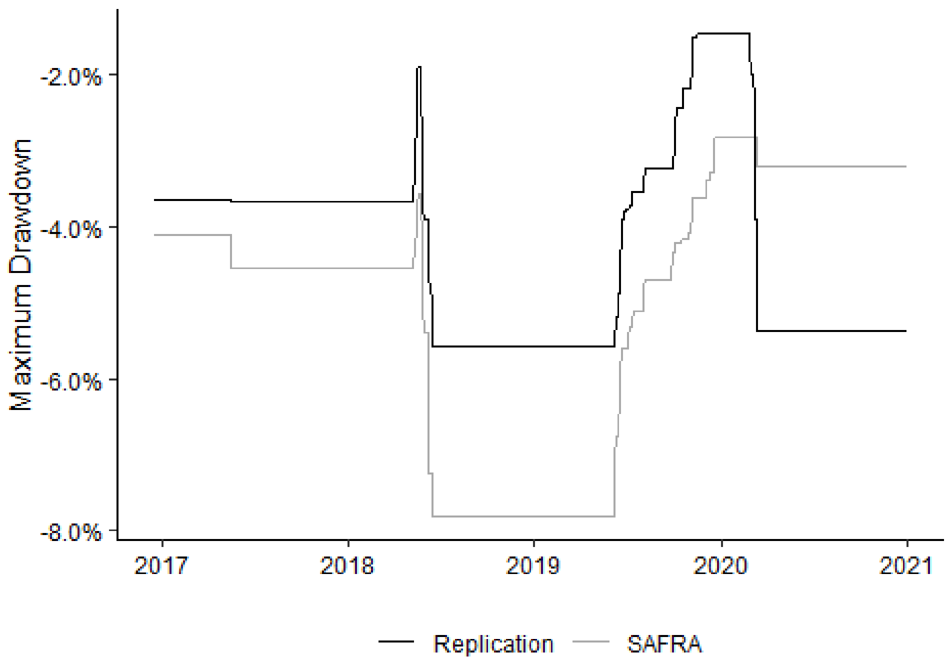

Figure 28 – Maximum Drawdown comparison for SAFRA.

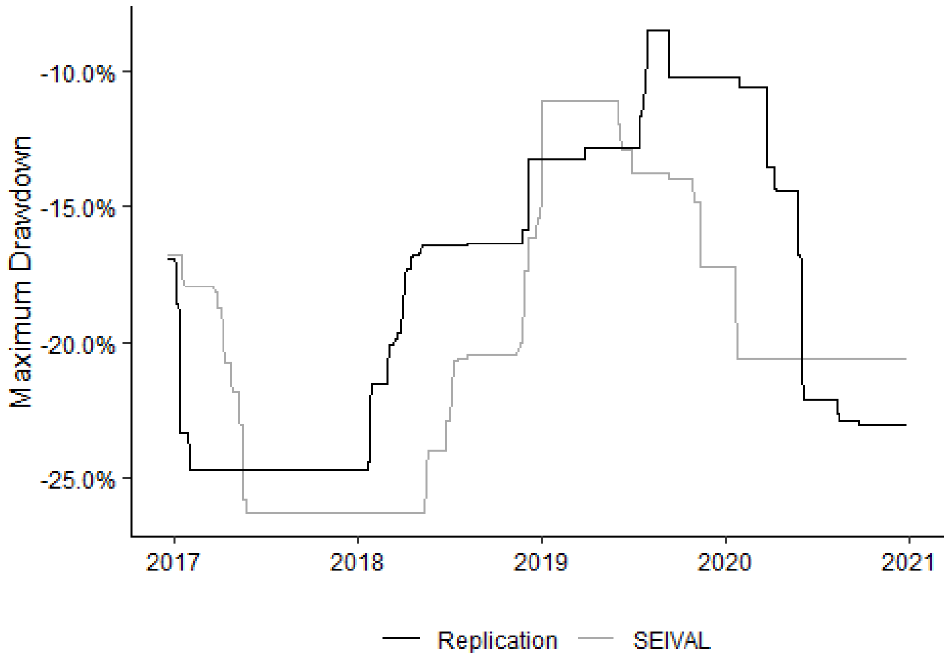

Figure 29 – Maximum Drawdown comparison for SEIVAL.

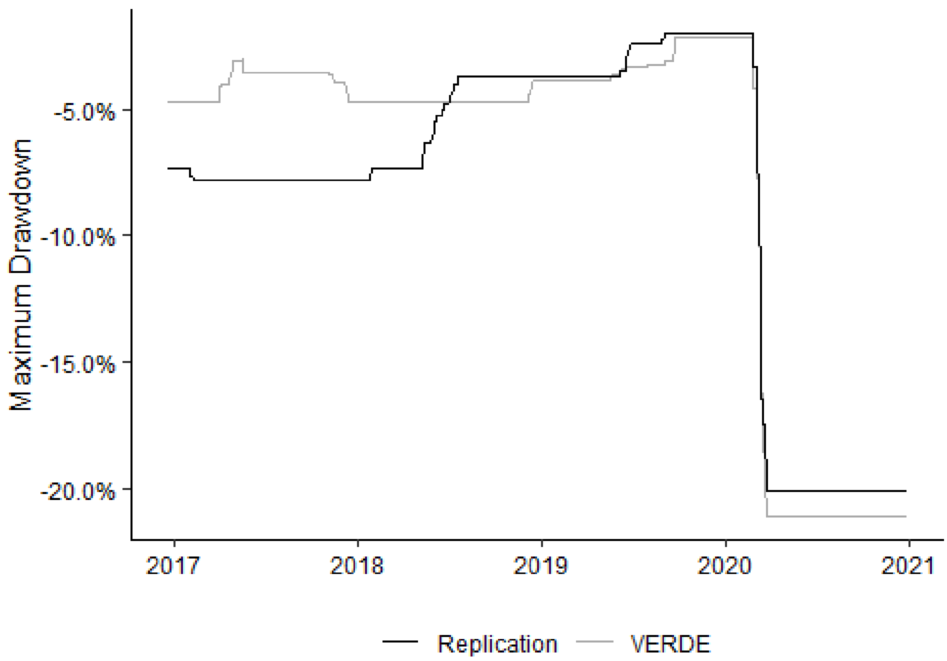

Figure 30 – Maximum Drawdown comparison for VERDE.