

Replication methods for financial indexes

B. Rémillard, B. Nasri,
M. Ben-Abdellatif

G-2017-58

July 2017

Cette version est mise à votre disposition conformément à la politique de libre accès aux publications des organismes subventionnaires canadiens et québécois.

Avant de citer ce rapport, veuillez visiter notre site Web (<https://www.gerad.ca/fr/papers/G-2017-58>) afin de mettre à jour vos données de référence, s'il a été publié dans une revue scientifique.

This version is available to you under the open access policy of Canadian and Quebec funding agencies.

Before citing this report, please visit our website (<https://www.gerad.ca/en/papers/G-2017-58>) to update your reference data, if it has been published in a scientific journal.

Les textes publiés dans la série des rapports de recherche *Les Cahiers du GERAD* n'engagent que la responsabilité de leurs auteurs.

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2017
– Bibliothèque et Archives Canada, 2017

The authors are exclusively responsible for the content of their research papers published in the series *Les Cahiers du GERAD*.

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2017
– Library and Archives Canada, 2017

Replication Methods for Financial Indexes

Bruno Rémillard^a

Bouchra Nasri^b

Malek Ben-Abdellatif^c

^a CRM, GERAD & HEC Montréal, Montréal (Québec),
Canada H3T 2A7

^b GERAD & HEC Montréal, Montréal (Québec),
Canada H3T 2A7

^c HEC Montréal, Montréal (Québec), Canada,
H3T 2A7

bruno.remillard@hec.ca

bouchra.nasri@hec.ca

malek.ben-abdellatif@hec.ca

July 2017

Les Cahiers du GERAD

G–2017–58

Copyright © 2017 GERAD

Abstract: In this paper, we first present a review of statistical tools that can be used in asset management either to track financial indexes or to create synthetic ones. More precisely, we look at two important replication methods: the strong replication, where a portfolio of very liquid assets is created and the goal is to track an actual index with the portfolio, and weak replication, where a portfolio of very liquid assets is created and used to either replicate the statistical properties of an existing index, or to replicate the statistical properties of a custom asset. In addition, for weak replication, the target is not an index but a payoff, and the replication amounts to hedge the portfolio so it is as close as possible to the payoff at the end of each month. For strong replication, the main tools are predictive tools, so filtering techniques and regression play an important role. For weak replication, which is the main topic of this paper, in order to determine the target payoff, the investor has to find or choose the distribution function of the target index or custom index, as well as its dependence with other assets, and use a hedging technique. Therefore, the main tools for weak replication are modeling (estimation and goodness-of-fit) and optimal hedging. For example, an investor could wish to obtain Gaussian returns that are independent of some ETFs replicating the Nasdaq and S&P 500 indexes. In order to determine the dependence of the target and a given number of indexes, we introduce a new class of easily constructed models of conditional distributions called B-vines. We also propose to use a flexible model to fit the distribution of the assets composing the portfolio and then hedge the portfolio in an optimal way. Examples are given to illustrate all the important steps required for the implementation of this new asset management methodology.

Keywords: ETF, hedge funds, replication, smart beta, copulas, B-vines, HMM, hedging

Résumé: Dans cet article, nous présentons d'abord une analyse des outils statistiques qui peuvent être utilisés dans la gestion des actifs soit pour reproduire des indices financiers, soit pour créer des actifs synthétiques. Plus précisément, nous examinons deux méthodes de répliation importantes: la répliation forte, où un portefeuille d'actifs très liquides est créé et l'objectif est de reproduire un index réel avec le portefeuille, et la répliation faible, où un portefeuille d'actifs très liquides est créé, et utilisé pour reproduire les propriétés statistiques d'un index existant ou indice personnalisé. En outre, pour la répliation faible, la cible n'est pas un indice, mais une valeur intrinsèque comme dans une option, et la répliation équivaut à se couvrir contre cette option à chaque échéance mensuelle. Pour la répliation forte, les outils principaux sont des outils prédictifs, comme les techniques de filtrage et la régression. Pour la répliation faible, qui est le sujet principal de cet article, afin de déterminer le rendement cible, l'investisseur doit trouver ou choisir la fonction de distribution de l'indice cible ou de l'indice personnalisé, ainsi que sa dépendance avec d'autres actifs et utiliser une stratégie de couverture. Par conséquent, les principaux outils pour l'impantation de la répliation faible sont la modélisation (estimation et tests d'ajustement) et des techniques de couverture optimale. Par exemple, un investisseur pourrait souhaiter obtenir des rendements gaussiens indépendants de certains fonds indiciels reproduisant les indices Nasdaq et S&P 500. Afin de déterminer la dépendance de la cible et un nombre donné d'indices, nous introduisons une nouvelle classe de modèles de distributions conditionnelles faciles à construire appelés B-vignes. Nous proposons également d'utiliser un modèle flexible qui s'ajuste à la distribution des actifs composant le portefeuille et ensuite couvrir le portefeuille d'une manière optimale. Des exemples sont donnés pour illustrer toutes les étapes requises à la mise en oeuvre de cette nouvelle méthodologie de gestion d'actifs.

Mots clés: FNB, fonds de couverture, répliation, copules, B-vignes, HMM, couverture

Acknowledgments: Partial funding in support of this work was provided by the Natural Sciences and Engineering Research Council of Canada, by the Fonds Québécois de Recherche sur la Nature et les Technologies, and by the Groupe d'études et de recherche en analyse des décisions (GERAD).

1 Introduction

Historically, hedge funds have been an important class of alternative investment assets for diversifying portfolios.

Mainly based on the work on Fung and Hsieh (2001, 2004) and Hasanhodzic and Lo (2007), major investors like financial institutions looked for more efficient and affordable methods to generate the same kind of returns. This was mainly done by strong replication, i.e., by constructing portfolios of very liquid assets tracking a hedge fund index. Nowadays, smart beta methods, a new brand name for replication techniques, offer even more flexibility to small investors as well, through ETFs. For example, Horizons HFF (hhf.to) is an ETF targeting the Morningstar Broad Hedge Fund Index SM, while State Street SPDR ETF (spy) tracks S&P 500 index.

In addition to strong replication, weak replication, based of the payoff distribution model of Dybvig (1988), was proposed by Amin and Kat (2003) and extended by Kat and Palaro (2005). This innovative approach consists in constructing a dynamic strategy to track a payoff, in order to reproduce the statistical properties of hedge fund returns together with their dependence with a selected investor portfolio. It can also be used to construct synthetic indexes with tailor-made properties, which is an advantage over strong replication since the latter can only replicate an existing index.

In Section 2, we review the main statistical techniques that can be use to replicate indexes, including a new “Smart Beta” approach that can be used to diversify investors portfolios. In order to implement the proposed methodology, a new family of conditional distribution called B-vines are introduced in Section 3. The essential steps of modeling and hedging are discussed in Section 4. Examples of applications are then given in Section 5.

2 Replication methods

There are basically two replication approaches: strong replication, where the target is the index (naive or imitative method, and factor-based method), and weak replication, where the target is a payoff determined by the distribution of an existing index or a custom index, also called synthetic index. In both cases, the idea is to construct a portfolio of liquid assets with end of the month values as close as possible to the target.

Strong replication is divided in two sub-groups. On one hand, there is the “naive replication”, where the investor try to imitate the hedge fund manager investment strategy or the index composition. This is kind of easy for indexes when their composition is known, but it is far from obvious when the strategy or composition is unknown. For example, for a Merger Arbitrage Fund index, the idea is to long (potential) sellers and short (potential) buyers.

On the other hand, the factorial approach attempts to reproduce hedge fund returns or indexes by investing in a portfolio of assets that provide similar end of month returns.

Alternative beta funds based on the factorial approach have been launched by several institutions including Goldman Sachs, JP Morgan, Deutsche Bank, and Innocap, to name a few. According to Wallerstein et al. (2010), the short version Verso of Innocap, based on filtering methods, performed best in the turbulent period 2008–2009. Note also that Laroche and Rémillard (2008) showed that factor-based replicators produce independent returns over time, which might be interesting from an investor’s perspective. Furthermore, an investor can easily track the performance of a given replicator. However, in a recent study, Towsey (2013) found very high correlations between factor-based replicators and indexes like S&P 500. This undesirable dependence show that these replicators cannot really be used for diversification purposes, contrary to synthetic indexes that can be built with weak replication techniques. An illustration of this powerful technique is given in Section 5.4.

Before presenting the mathematical framework defining strong and weak replication, we summarize in Table 1 the main differences between the two approaches. Tracking is possible for weak replication if the value of the payoff is posted at the end of the month. In this case, the analog of the tracking error is the

RMSE (root mean square error) of the hedging error. This important value appears in our examples of implementation in Section 5. For synthetic indexes, it is possible to control the dependence.

Table 1: Main differences between strong replication and weak replication.

Method	Target	Tracking	Synthetic index	Controlled dependence
Strong	Index	Yes	No	No
Weak	Payoff	Possible	Possible	Possible

2.1 Factorial approach for strong replication

To implement the factorial approach, one needs the returns¹ R_t^* of the target fund S^* and one needs to select appropriate liquid assets (factors) $\mathbf{S} = (S^{(1)}, \dots, S^{(p)})$ composing the replication portfolio. The returns of \mathbf{S} are denoted by $\mathbf{R}_t = (R_t^{(1)}, \dots, R_t^{(p)})$, and the associated weights are denoted by $\beta_t = (\beta_{t,1}, \dots, \beta_{t,p})$. The model is written in the linear form

$$R_t^* = \beta_t^\top \mathbf{R}_t + \varepsilon_t, \quad (1)$$

where the ε_t 's are non-observable tracking error terms.

The unknown weights β_t are then evaluated from a predictive method using relation (1), e.g., by using a rolling-window regression over the last 24 months, or by using filtering methods. Note that for filtering, one must also define the (Markovian) dynamics of the weights β_t ; see, e.g., Roncalli and Teïletche (2007).

To measure the performance of a replicating method, one use the tracking error (TE), defined in the in-sample case by

$$\text{TE}_{in} = \left\{ \frac{1}{n} \sum_{t=1}^n \left(R_t^* - \hat{\beta}_t^\top \mathbf{R}_t \right)^2 \right\}^{1/2},$$

while for the out-of-sample, it is defined by

$$\text{TE}_{out} = \left\{ \frac{1}{n} \sum_{t=1}^n \left(R_t^* - \hat{\beta}_{t-1}^\top \mathbf{R}_t \right)^2 \right\}^{1/2},$$

where $\hat{\beta}_t$ is the vector of predicted weights using returns $(R_t^*, \mathbf{R}_t), (R_{t-1}^*, \mathbf{R}_{t-1}), \dots$. The out-of-sample tracking error is a more realistic measure of performance, since the error $R_t^* - \hat{\beta}_{t-1}^\top \mathbf{R}_t$ is the one monitored by investors. As seen in the example below, filtering usually yields better results than regression in terms of tracking error.

Example 1 *Rémillard (2013, Chapter 10). The target is HFRI Fund Weighted Composite Index, and the factors are S&P500 Index TR, Russel 2000 Index TR, Russell 1000 Index TR, Eurostoxx Index, Topix, US 10-year Index, 1-month LIBOR.*² Here, two methods were used to compute the dynamic weights β : a regression with a 24-month window, and a Kalman filter, where the dynamics of the β 's is a random walk, meaning that $\beta_t = \beta_{t-1} + \eta_t$, where the innovations η_t are assumed to be independent and identically distributed.

This is a very basic and unrealistic model, but the model can be improved, e.g., by adding dependence in the increments or adding constraints of the portfolio compositions. In this case, the Kalman filter assumptions are no longer met, and one should use for example a particles filter (Rémillard, 2013, Chapter 9). However, even with a simple model and the Kalman filter, the results are surprisingly good, better than the rolling-window regression. In-sample and out-of-sample statistics for our example are displayed in Table 2.

In general, the β_t are much less variable in the Kalman filter case, leading to less expensive transactions, in addition to being a better tracking method. See, e.g., Rémillard (2013, Chapter 10).

¹Typically monthly returns, especially in the case of hedge fund indexes.

²Data, from April 1997 to October 2008, were provided by Innocap.

Table 2: In-sample and out-of-sample statistics.

	TE	Corr	Mean	Std	Skew	Excess kurt
Portfolio	In-sample statistics					
Target	-	1.00	8.12	7.72	-0.59	2.45
Regression	10.58	0.93	8.79	8.32	-0.69	2.22
Kalman	8.54	0.95	9.68	7.75	-0.59	2.53
Portfolio	Out-of-sample statistics					
Target	-	1.00	8.12	7.72	-0.59	2.45
Regression	19.27	0.83	9.30	9.86	-0.11	3.34
Kalman	14.71	0.86	9.97	8.20	-0.40	2.63

NOTE: Tracking error, mean and volatility are expressed in annual percentage. Recall that the excess kurtosis of the Gaussian distribution is 0.

Before ending this section, it is worth noting that one could also use machine learning methods for tracking purposes. It would be interesting to compare the performance of machine learning vs filtering. This will be done in a forthcoming work.

2.2 Weak replication

Weak replication is an alternative replication method proposed by Amin and Kat (2003) and later extended by Kat and Palaro (2005) based on the payoff distribution model of Dybvig (1988). The aim was to replicate hedge fund returns or hedge fund indexes not by identifying the return generating betas as in the factor-based approach, but by building a trading strategy that can be used to generate the (statistical) distribution of the hedge fund returns or indexes. The implementation proposed in Kat and Palaro (2005) is subject to several shortcomings and inconsistencies. Improvements of the Kat-Paloro method were proposed in Papageorgiou et al. (2008) for a start.

In view of applications to asset management, and mainly for diversification purposes, it is desirable to generalize the Kat-Paloro approach (limited to only one reference asset). To this end, it was suggested in Ben-Abdellatif (2010) to consider a multivariate asset \mathbf{S} of $p = d + 1$ components, where $S^{(1)}, \dots, S^{(d)}$ represent the value of reference portfolios of the investor, and the so-called reserve asset $S^{(d+1)}$ ³. As before, S^* is the index one seeks to replicate. The aim is not to reproduce the monthly values of S^* , which might not even exist, but rather reproduce its statistical properties.

The steps required to implement the proposed weak replication method are given next.

2.2.1 Implementation steps

1. Determine the joint distribution of the (daily) returns \mathbf{R}_k of \mathbf{S}_k .
2. Find a compatible distribution for the monthly returns $\mathbf{R}_{0,T}$. In particular, find the marginal distributions F_1, \dots, F_d of $\mathbf{R}_{0,T}^{(ref)} = (R_{0,T}^{(1)}, \dots, R_{0,T}^{(d)})$, find the copula of $\mathbf{R}_{0,T}^{(ref)}$, and find the conditional distribution $\mathcal{F}(\cdot, \mathbf{x})$ of $R_{0,T}^{(res)}$ given $\mathbf{R}_{0,T}^{(ref)} = (R_{0,T}^{(1)}, \dots, R_{0,T}^{(d)}) = \mathbf{x}$.

This can be done by simulation from daily returns, as suggested in Section 4.2. Again, we suggest to use a Gaussian HMM. We strongly advise against using real monthly returns to complete this step since in general the sample size for estimation purposes is not long enough, and in addition, there is a lack of compatibility between the distribution of the daily and monthly returns, thus creating a bias.

3. Find or choose the distribution function F_* of the return $R_{0,T}^*$ of the target index S^* .
4. Find or choose the conditional distribution function $\mathcal{H}(\cdot, \mathbf{x})$ of $R_{0,T}^*$ given $\mathbf{R}_{0,T}^{(ref)} = \mathbf{x}$, which can be expressed as

$$\mathcal{H}(y, \mathbf{x}) = \mathcal{C}\{G(y), \mathbf{F}(\mathbf{x})\},$$

³E.g., equal weighted portfolio of highly liquid futures contracts

where $\mathbf{F}(\mathbf{x}) = (F_1(x_1), \dots, F_d(x_d))$, and $\mathcal{C}(\cdot, \mathbf{v})$ is the conditional distribution of $U = F_\star(R_{0,T}^\star)$ given $\mathbf{V} = \mathbf{F}(\mathbf{R}_{0,T}^{(ref)}) = \mathbf{v}$.

5. Compute the return function g given by

$$g(\mathbf{x}, y) = \mathcal{Q}\{\mathcal{F}(y, \mathbf{x}), \mathbf{x}\}, \quad (2)$$

where $\mathcal{Q}(\cdot, \mathbf{x})$ is the conditional quantile function (Rémillard et al., 2017), defined as the inverse of $\mathcal{H}(\cdot, \mathbf{x})$. The function g can also be expressed as

$$g(\mathbf{x}, y) = F_\star^{-1}[\mathcal{C}^{-1}\{\mathcal{F}(y, \mathbf{x}), \mathbf{F}(\mathbf{x})\}]. \quad (3)$$

6. Compute the payoff function G defined by

$$G(\mathbf{S}_T) = 100 \exp\{g(\mathbf{R}_{0,T})\}.$$

7. Construct a dynamic portfolio $\{\mathcal{V}_k(\mathcal{V}_0, \boldsymbol{\varphi})\}_{k=0}^n$ of the assets \mathbf{S} , traded daily, in order to generate the payoff $G(\mathbf{S}_T)$ at the end of the month. More precisely, letting $\beta_k = e^{-rkT/n}$ be the discounting factors, the discounted value of the portfolio at the end of the month is

$$\beta_n \mathcal{V}_n = \mathcal{V}_0 + \sum_{k=1}^n \boldsymbol{\varphi}_k^\top (\beta_k \mathbf{S}_k - \beta_{k-1} \mathbf{S}_{k-1}),$$

where $\boldsymbol{\varphi}_k^{(j)}$ is number of shares of asset $S^{(j)}$ invested during $((k-1)T/n, kT/n]$, and $\boldsymbol{\varphi}_k$ may depend only on $\mathbf{S}_0, \dots, \mathbf{S}_{k-1}$. Initially, the portfolio initial value is \mathcal{V}_0 .

This hedging problem is typical in financial engineering, where \mathcal{V}_0 can be interpreted as the value of an option on \mathbf{S} having payoff G at maturity T , and one wants to replicate the payoff. Usually, we are more interested in the price of the option, while here the emphasis is on the hedging portfolio, which is the object of the investment.

For hedging, we suggest to use the discrete time hedging method defined in Section 4.3. This strategy, adapted for a continuous time model, is optimal with respect to minimizing the square hedging error.

2.2.2 K-P measure

If the goal is attained, i.e., $\mathcal{V}_n = G(\mathbf{S}_T)$, then the return of the portfolio is

$$\log(\mathcal{V}_n/\mathcal{V}_0) = \log(100/\mathcal{V}_0) + g(\mathbf{R}_{0,T}),$$

which has the same distribution as $\alpha + S^\star$, where $\alpha = \log(\mathcal{V}_0/100)$ can be used to estimate manager's alpha or the feasibility of the replication. Kat and Palaro (2005), in the context of replicating hedge funds, suggested that the initial amount \mathcal{V}_0 to be invested in the portfolio be viewed as a measure of performance of the hedge fund manager. Here we prefer to use α which we call the K-P measure. It can be interpreted as follows:

- If $\alpha = 0$, i.e., $\mathcal{V}_0 = 100$, the strategy generates the same returns as S^\star (in distribution);
- If $\alpha < 0$, i.e., $\mathcal{V}_0 < 100$, it is worth replicating, generating superior returns (in distribution), while if $\alpha > 0$, i.e., $\mathcal{V}_0 > 100$, it may be not worth replicating.

Note that centered moments like standard deviation, skewness, kurtosis, are not affected by the value of the K-P measure α . However, the expected value of the portfolio is $\alpha + E(S^\star)$.

Example 2 *A simple example in risk management is an investor interested in creating a portfolio S^\star with a specific distribution function F_\star , which would be independent of several reference indexes, so that the return of the hedging portfolio will not be affected by extreme behavior of the reference indexes. In this case, g is given by*

$$g(\mathbf{x}, y) = F_\star^{-1}\{\mathcal{F}(y, \mathbf{x})\}. \quad (4)$$

An example of implementation of this model is given in Section 5.4.

Remark 1 It makes sense that $\alpha > 0$, especially if the target distribution of S^* is not realistic. For example, one could wish to generate Gaussian returns with annual mean of 30% and a volatility of 1% that is independent of $S^{(1)}$, but the real distribution would be Gaussian with mean $.3 - 12\alpha$. In fact, if the joint distribution of the monthly returns is Gaussian, with annual means μ_1, μ_2, μ_3 , annual volatilities $\sigma_1, \sigma_2, \sigma_3$ and correlations ρ_{12}, ρ_{13} , then, according to Equation (2),

$$g(x, y) = \frac{1}{12} \left\{ \mu_3 - r + \sigma_3 \left(\frac{x - \mu_1}{\sigma_1} \right) \left(\rho_{13} - \rho_{12} \sqrt{\frac{1 - \rho_{13}^2}{1 - \rho_{12}^2}} \right) + \sigma_3 \left(\frac{y - \mu_2}{\sigma_2} \right) \sqrt{\frac{1 - \rho_{13}^2}{1 - \rho_{12}^2}} \right\},$$

so using the Black-Scholes setting with associated risk neutral measure \mathbb{Q} ,

$$\mathcal{V}_0 = 100e^{-r/12} E^{\mathbb{Q}} \left\{ e^{g(R_{0,1/12}^{(1)}, R_{0,1/12}^{(2)})} \right\} = 100e^\alpha,$$

with

$$\begin{aligned} \alpha = & \frac{\mu_3}{12} - \frac{r}{12} - \frac{1}{12} \left\{ \mu_1 \frac{\sigma_3}{\sigma_1} + \mu_2 \frac{\sigma_3}{\sigma_2} \sqrt{\frac{1 - \rho_{13}^2}{1 - \rho_{12}^2}} - \frac{\sigma_3^2}{2} \right\} \\ & + \frac{1}{12} \left\{ \frac{\sigma_3}{\sigma_1} \left(r - \frac{\sigma_1^2}{2} \right) \left(\rho_{13} - \rho_{12} \sqrt{\frac{1 - \rho_{13}^2}{1 - \rho_{12}^2}} \right) \right. \\ & \left. + \frac{\sigma_3}{\sigma_2} \left(r - \frac{\sigma_2^2}{2} \right) \sqrt{\frac{1 - \rho_{13}^2}{1 - \rho_{12}^2}} \right\}. \end{aligned}$$

As a result, the genuine mean of the target is independent of μ_3 ! For example, if $r = 1\%$, $\mu_1 = 8\%$, $\mu_2 = 6\%$, $\sigma_1 = 10\%$, $\sigma_2 = 8\%$, $\sigma_3 = 1\%$, $\rho_{12} = 0.25$ and $\rho_{13} = 0$, then $\alpha = \frac{\mu_3}{12} - \frac{.02499}{12}$, and we would get a Gaussian distribution with an annual mean of 2.499% and an annual volatility of 1% that is independent of $S^{(1)}$. It is interesting to look at the real annual mean of the portfolio (assuming perfect hedging) as a function of ρ_{13} . This is illustrated in Figure 1. Note that the maximum value 2.501% is attained for $\rho_{13} = -0.072$.

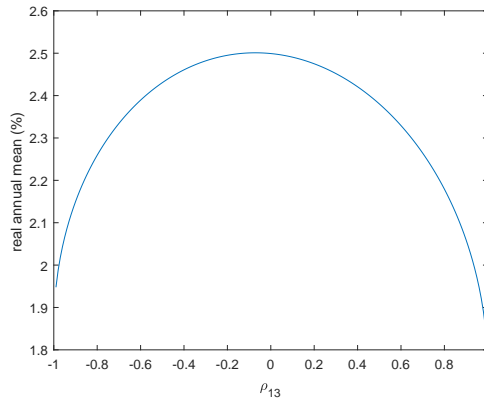


Figure 1: Real annual mean in percent of the Gaussian distribution of the monthly return R^* as a function of the correlation ρ_{13} with monthly return $R^{(1)}$.

2.2.3 Choice of \mathcal{C}

First, note that \mathcal{C} is a function of the copula C of (U, \mathbf{V}) viz.

$$\mathcal{C}(u, \mathbf{v}) = \frac{\partial_{v_1} \cdots \partial_{v_d} C(u, v_1, \dots, v_d)}{c_{\mathbf{V}}(1, v_1, \dots, v_d)}, \quad (u, \mathbf{v}) \in (0, 1)^{1+d}, \quad (5)$$

where $c_{\mathbf{V}}$ is the density of the copula $C_{\mathbf{V}}(\cdot) = C(1, \cdot)$. When $d = 1$, we can take $\mathcal{C}(u, v) = \partial_v C(u, v)$ for any copula C . However, if $d \geq 2$, then the copula of \mathbf{V} matters. One cannot just take any $d + 1$ -dimensional copula C . To solve this intricate problem, we propose to use a construction similar to the one used for vine copulas. This new construction is described in Section 3, after we discuss why the choice of \mathcal{C} matters.

To this end, let $\tilde{\mathcal{C}}$ be an arbitrary conditional distribution function of U given a d -dimensional random vector $\tilde{\mathbf{V}}$ associated with the copula \tilde{C} of $(U, \tilde{\mathbf{V}})$, and define

$$\tilde{g}(\mathbf{x}, y) = F_{\star}^{-1} \left[\tilde{\mathcal{C}}^{-1} \{ \mathcal{F}(y, \mathbf{x}), \mathbf{F}(\mathbf{x}) \} \right].$$

Setting $\mathcal{Z} = \mathcal{F} \left(R_{0,T}^{(res)}, \mathbf{R}_{0,T}^{(ref)} \right)$, one gets

$$\begin{aligned} P \left[\tilde{g}(\mathbf{R}_{0,T}) \leq y, \mathbf{R}_{0,T}^{(ref)} \leq \mathbf{x} \right] &= P \left[\mathcal{Z} \leq \tilde{\mathcal{C}} \left\{ F_{\star}(y), \mathbf{F} \left(\mathbf{R}_{0,T}^{(ref)} \right) \right\}, \mathbf{R}_{0,T}^{(ref)} \leq \mathbf{x} \right] \\ &= E \left[\tilde{\mathcal{C}} \{ F_{\star}(y), \mathbf{V} \} \mathbb{I}(\mathbf{V} \leq \mathbf{F}(\mathbf{x})) \right] \\ &= \int_{(0, \mathbf{F}(\mathbf{x}))} \tilde{\mathcal{C}} \{ F_{\star}(y), \mathbf{v} \} c_{\mathbf{V}}(\mathbf{v}) d\mathbf{v}, \end{aligned}$$

since \mathcal{Z} is uniformly distributed and is independent of $\mathbf{R}_{0,T}^{(ref)}$, according to Rosenblatt (1952). So, in general, $\tilde{F}_{\star}(y) = E \left[\tilde{\mathcal{C}} \{ F_{\star}(y), \mathbf{V} \} \right]$ is not the target distribution function F_{\star} . However, $\tilde{F}_{\star} = F_{\star}$ if $\tilde{\mathcal{C}}(1, \mathbf{v}) = C_{\mathbf{V}}(\mathbf{v})$. One then must be careful with the choice of \mathcal{C} in order to have compatibility.

3 B-vines models

The aim of this section is to find a flexible way to construct a conditional distribution of a random variable Y given a d -dimensional random vector \mathbf{X} . Using the representation of conditional distributions in terms of copulas, this problem amounts to constructing the conditional distribution \mathcal{C} of a uniform random variable U given a random vector \mathbf{V} (with uniform margins) that is coherent with the distribution function $C_{\mathbf{V}}$ of \mathbf{V} .

As noted before, when $d = 1$, the compatibility condition is not a constraint at all since $C_{\mathbf{V}}(v) = v$, $v \in [0, 1]$, and the solution is simply to take $\mathcal{C}(u, v) = \partial_v C(u, v)$, for a copula C that is smooth enough.

Next, in the case $d = 2$, if D_1 and D_2 are bivariate copulas, with conditional distributions $\mathcal{D}_j(u, t) = \partial_t D_j(u, t)$, $j \in \{1, 2\}$, and $C_{\mathbf{V}}$ is the copula of $\mathbf{V} = (V_1, V_2)$, then

$$\mathcal{C}(u, \mathbf{v}) = \mathcal{D}_2 \{ \mathcal{D}_1(u, v_1), \partial_{v_1} C_{\mathbf{V}}(v_1, v_2) \}, \quad \mathbf{v} = (v_1, v_2) \in (0, 1)^2, \quad (6)$$

defines a conditional distribution for U given $\mathbf{V} = v$, compatible with the law of \mathbf{V} . This construction is a particular case of a D -vine copula (Joe, 1996; Aas et al., 2009).

Guided by formula (6), let D_j , $j \in \{1, \dots, d\}$ be bivariate copulas and let $\mathcal{D}_j(u, t) = \partial_t D_j(u, t)$ be the associated conditional distributions. For $j \in \{1, \dots, d\}$, further let $\mathcal{R}_{j-1}(v_1, \dots, v_j)$ be the conditional distribution of V_j given $V_1 = v_1, \dots, V_{j-1} = v_{j-1}$, with $\mathcal{R}_0(v_1) = v_1$, and for $(u, \mathbf{v}) \in (0, 1)^{d+1}$, set $\mathcal{C}_0(u) = u$, and

$$\mathcal{C}_j(u, v_1, \dots, v_j) = \mathcal{D}_j \{ \mathcal{C}_{j-1}(u, v_1, \dots, v_{j-1}), \mathcal{R}_{j-1}(v_1, \dots, v_j) \}. \quad (7)$$

Note that $E \{ \mathcal{C}_j(u, v_1, \dots, v_{j-1}, V_j) | V_1 = v_1, \dots, V_{j-1} = v_{j-1} \}$ is given by

$$\begin{aligned}
\int_0^1 \mathcal{C}_j\{u, v_1, \dots, \mathcal{R}_{j-1}(v_1, \dots, v_j)\} d\mathcal{R}_{j-1}(v_1, \dots, v_j) &= \int_0^1 \mathcal{D}_j\{\mathcal{C}_{j-1}(u, v_1, \dots, v_{j-1}), t\} dt \\
&= D_j\{\mathcal{C}_{j-1}(u, v_1, \dots, v_{j-1}), 1\} \\
&= \mathcal{C}_{j-1}(u, v_1, \dots, v_{j-1}).
\end{aligned}$$

It follows that \mathcal{C}_j is the conditional distribution of U given V_1, \dots, V_j . The conditional quantile of U given V_1, \dots, V_j is also easy to compute, satisfying a recurrence relation similar to (7). In fact, if the conditional quantile of \mathcal{C}_j is denoted by Γ_j , then for any $j \in \{1, \dots, d\}$, and for any $u, v_1, \dots, v_d \in (0, 1)$,

$$\Gamma_j(u, v_1, \dots, v_j) = \Gamma_{j-1} \left[\mathcal{D}_j^{-1} \{u, \mathcal{R}_{j-1}(v_1, \dots, v_j)\}, v_1, \dots, v_{j-1} \right]. \quad (8)$$

In general, this construction does not lead to a proper vine copula since all copulas involved are not bivariate copulas, the copula of \mathbf{V} being given. In fact, it is more general than the pair-copula construction method used in vines models. Nevertheless, this type of model will be called *B-vines* and its construction is illustrated below, where the underlined variables (in red) mean that their distributions $\mathcal{R}_0, \dots, \mathcal{R}_{d-1}$ are known,⁴ and the conditional copulas $\mathcal{D}_1, \dots, \mathcal{D}_d$ have to be chosen, in order to determine $\mathcal{C}_1, \dots, \mathcal{C}_d$.

$$\begin{array}{rcl}
\text{Level 1:} & \overset{\mathcal{D}_1}{U} \mid \underline{V_1}^{\mathcal{R}_0} & \implies \mathcal{C}_1 \\
\text{Level 2:} & \overset{\mathcal{D}_2}{U|V_1} \mid \underline{V_2|V_1}^{\mathcal{R}_1} & \implies \mathcal{C}_2 \\
& \vdots & \ddots \\
\text{Level } j: & \overset{\mathcal{D}_j}{U|V_1, \dots, V_{j-1}} \mid \underline{V_j|V_1, \dots, V_{j-1}}^{\mathcal{R}_{j-1}} & \implies \mathcal{C}_j \\
& \vdots & \ddots \\
\text{Level } d: & \overset{\mathcal{D}_d}{U|V_1, \dots, V_{d-1}} \mid \underline{V_d|V_1, \dots, V_{d-1}}^{\mathcal{R}_{d-1}} & \implies \mathcal{C}_d
\end{array}$$

Note that *B-vines* can be particularly useful in conditional mean regression (OLS, GAM, GLM, etc.) and conditional quantile settings, where the distribution of the covariates is often given; see, e.g., Rémillard et al. (2017). It can also be used in our replication context when the target S^* exists; in this case, we could look at *B-vines* constructed from popular bivariate families like Clayton, Gumbel, Frank, Gaussian and Student, and find the ones that fit best the data, in the same spirit as the choice of vines for copula models in the R packages *CDVine* or *VineCopula*. In a future work we will propose goodness-of-fit tests for these models.

4 Modeling and hedging

In what follows, building on Papageorgiou et al. (2008), we propose a model to fit the data and deal with numerical problems arising from using a larger number of assets for hedging.

To implement successfully the proposed replication approach, one needs to model the distribution of the returns \mathbf{R}_t and $\mathbf{R}_{0,T}$. Once this is done, we will have as a by-product the conditional distribution \mathcal{F} and the Rosenblatt's transforms $\mathcal{R}_0, \dots, \mathcal{R}_{d-1}$ used for computing the conditional distribution \mathcal{C} , as in Section 3. For replicating an existing asset S^* , one further needs the joint distribution of $(R_{0,T}^*, \mathbf{R}_{0,T}^{(ref)})$. To do this, we propose to use Gaussian Hidden Markov Models (HMM) as defined in Rémillard et al. (2017). This model is described next in Section 4.1. Next, one needs to find a distribution of the month returns compatible with the distribution of the daily returns. A solution to this problem is proposed in Section 4.2. Finally, a replication method is proposed in Section 4.3.

⁴ $\mathcal{R}_0, \dots, \mathcal{R}_{d-1}$ are called the Rosenblatt's transforms and are particularly important in simulating copulas or for testing goodness-of-fit. See, e.g., Rémillard (2013).

4.1 Gaussian HMM

Regime-switching models are quite intuitive. First, the regimes $\{1, \dots, l\}$ are not observable and are modeled by a finite Markov chain with transition matrix \mathbf{Q} . At period t , given that the previous regime τ_{t-1} has value i , the regime $\tau_t = j$ is chosen with probability Q_{ij} , and given $\tau_t = j$, the log-returns \mathbf{R}_t has a Gaussian distribution with mean $\boldsymbol{\mu}_j$ and covariance matrix \mathbf{B}_j .

The law of most financial time series can be modeled adequately by a Gaussian HMM, provided the number of regimes is large enough. Indeed, the serial dependence in regimes propagates to returns and captures the observed autocorrelation in financial time series. Also, the conditional distribution is time-varying, leading to conditional volatility, as well as conditional asymmetry and kurtosis. Finally, the Black-Scholes framework is a particular case of this model when the number of regimes is 1. Parameters are quite easy to estimate and there is also an easy way to choose the number of regimes, depending on the results of goodness-of-fit tests; see, e.g., Rémillard et al. (2017) for more details.

In the next section, we introduce the continuous time limit of a Gaussian HMM, the main reason being that for this limiting process, one can show that there exists an equivalent martingale measure that is optimal in the sense of Schweizer (1992) and can be used for pricing and hedging; see, e.g., Rémillard and Rubenthaler (2016).

4.1.1 Continuous time limiting process

Under weak conditions, the continuous time limit of a Gaussian HMM is a regime-switching geometric Brownian motion (RSGBM). Using the same notations as in Rémillard and Rubenthaler (2016), let \mathcal{T} be a continuous time Markov chain on $\{1, \dots, l\}$, with infinitesimal generator $\boldsymbol{\Lambda}$. In particular, $P(\mathcal{T}_t = j | \mathcal{T}_0 = i) = P_{ij}(t)$, where the transition matrix \mathbf{P} can be written as $\mathbf{P}(t) = e^{t\boldsymbol{\Lambda}}$, $t \geq 0$. Then, the (continuous) price process \mathbf{X} modeled as a RSGBM satisfies the stochastic differential equation

$$d\mathbf{X}_t = D(\mathbf{X}_t)\boldsymbol{v}(\mathcal{T}_t)dt + D(\mathbf{X}_t)\boldsymbol{\sigma}(\mathcal{T}_t)d\mathbf{W}_t, \quad (9)$$

where $D(\mathbf{s})$ is the diagonal matrix with diagonal elements $(s_j)_{j=1}^d$ and \mathbf{W} is a d -dimensional Brownian motion, independent of \mathcal{T} . Note that the time scale is in years, and we assume that $a(j) = \boldsymbol{\sigma}(j)\boldsymbol{\sigma}(j)^\top$ is invertible for any $j \in \{1, \dots, l\}$.

4.1.2 Relationship between discrete time and continuous time parameters

The relationship between the continuous-time parameters $(\boldsymbol{v}, \mathbf{a}, \boldsymbol{\Lambda})$ of the limiting RSGBM and the parameters of the Gaussian HMM is the following: if the parameters $\boldsymbol{\mu}_h, \mathbf{B}_h, \mathbf{Q}_h$ of the discrete time model are obtained from data sampled $1/h$ times a year, then $\boldsymbol{v}(j) \approx [\boldsymbol{\mu}_h(j) + \frac{1}{2}\text{diag}\{\mathbf{B}_h(j)\}]/h$, where $\text{diag}(\mathbf{B})$ is the vector of the diagonal elements of a matrix \mathbf{B} , $\mathbf{a}(j) \approx \mathbf{B}_h(j)/h$, and $\boldsymbol{\Lambda} \approx (\mathbf{Q}_h - I)/h$. For example, for daily data, one usually takes $h = 1/252$.

Note that if we define $\mathbf{X}_{h,t} = \mathbf{S}_{\lfloor t/h \rfloor}$ and $\mathcal{T}_{h,t} = \tau_{\lfloor t/h \rfloor}$, where $\lfloor a \rfloor$ stands for the integer part of $a \in \mathbb{R}$, then the processes $(\mathbf{X}_{h,t}, \mathcal{T}_{h,t})$ converge in law to $(\mathbf{X}, \mathcal{T})$. Note also that the optimal hedging strategy converges as well (Rémillard and Rubenthaler, 2009).

4.2 Monthly returns compatibility

Compatibility means that the distribution of the monthly returns $\mathbf{R}_{0,T}$ is the same as the distribution of the sum of typically $n = 21$ consecutive daily returns. Since the hedging will be done under a continuous time RSGBM, there is no compatibility problem. However, since we need the distribution of $\log(\mathbf{X}_T)$ to construct the payoff, and the latter is not known explicitly, we propose to simulate a large number of monthly returns $\log(\mathbf{X}_T)$, say 10000, which is impossible to get in practice, and then fit a Gaussian HMM to these simulated data.

The joint distribution of the monthly return is then approximated by a mixture of (multivariate) Gaussian distributions, and the conditional distribution function \mathcal{F} is also a (univariate) Gaussian mixture. See, e.g., Rémillard et al. (2017) for more details.

4.3 Discrete time hedging

Since we fitted a Gaussian HMM to the daily returns, an obvious solution of the hedging problem would be to use the results of Rémillard and Rubenthaler (2013) for optimal hedging in discrete time; see also Rémillard (2013). However, implementing this methodology requires interpolating functions on a $(d + 1)$ -dimensional grid. Since we are aiming for applications with $d \geq 2$, this approach leads to too much imprecision. For example, a (too) small grid of 100 points for each asset would require computing and storing $10^{2(d+1)}$ points, while a relatively precise grid of 1000 points for each asset requires $10^{3(d+1)}$ points. Even with $d = 2$, this means storing 10^9 points, which is too much.

This is why we consider a continuous-time approximation, which does not require any interpolation or grid construction and works in any dimension. It is easy to show, see, e.g., Rémillard and Rubenthaler (2009) that many interesting discrete time models can be approximated by continuous time models. In particular, this is true for the Gaussian HMM whose continuous time limit is the RSGBM. Option pricing and optimal quadratic hedging have been studied recently for this process (Rémillard and Rubenthaler, 2016), and it turns out that the optimal hedging strategy and option price can be deduced from an equivalent martingale measure. Under this equivalent martingale measure, assets still follow a RSGBM, with the additional feature that the distribution of the regimes is now an inhomogeneous continuous time Markov time. Nevertheless, this model is quite easy to simulate and does not require any calibration to option prices.

4.3.1 Continuous time approximation

Because we have possibly more than 2 risky assets, and based on the results in Rémillard and Rubenthaler (2009, 2016), we approximate φ_k by $\phi_{\frac{k-1}{n}T}$, where ϕ is the optimal hedging strategy of the RSGBM obtained from Rémillard and Rubenthaler (2016, Lemma 4.1).

To get nearly optimal hedging strategies in discrete time, we first use Monte Carlo methods by simulating the process \mathbf{X} under the optimal martingale measure, as given by Equation (17), to obtain the values $C_{kT/n}(\mathbf{s}, i)$ and $\nabla_{\mathbf{s}} C_{kT/n}(\mathbf{s}, i)$ given by formulas (18) and (19). Then we simply discretize the continuous time optimal hedging values (20)–(21) to get, for $k \in \{1, \dots, n\}$,

$$\varphi_k = \nabla_{\mathbf{s}} C_{(k-1)T/n}(\mathbf{S}_{k-1}, \hat{\tau}_{k-1}) + G_{k-1} D^{-1}(\mathbf{S}_{k-1}) \rho(\hat{\tau}_{k-1}) / \beta_{k-1}, \quad (10)$$

$$\tilde{\mathcal{V}}_k = \tilde{\mathcal{V}}_{k-1} + \varphi_k^\top (\beta_k \mathbf{S}_k - \beta_{k-1} \mathbf{S}_{k-1}), \quad (11)$$

$$G_k = \beta_k C_{kT/n}(\mathbf{S}_k, \hat{\tau}_k) - \tilde{\mathcal{V}}_k, \quad (12)$$

where $\tilde{\mathcal{V}}_0 = \mathcal{V}_0 = C_0(\mathbf{S}_0, \hat{\tau}_0)$, $G_0 = 0$, and $\tilde{\mathcal{V}}_k = \beta_k \mathcal{V}_k$ are the discounted portfolio values. In particular, $\varphi_1 = \nabla_{\mathbf{s}} C_0(\mathbf{S}_0, \hat{\tau}_0)$.

Remark 2 One could replace $C_{kT/n}(\mathbf{S}_k, \hat{\tau}_k)$ by the weighted average $\sum_{j=1}^l C_{kT/n}(\mathbf{S}_k, j) \eta_k(j)$, where $\eta_k(j)$ is the predicted probability of $\tau_k = j$, given the past observations.

We now have the necessary tools to tackle the implementation problem. Two examples of application are presented next.

5 Examples of application

In this section, we provide some empirical evidence regarding the ability of the model to replicate a synthetic index. In the implementation of the replication model, we consider a 3-dimensional problem.

5.1 Assets

The first step is to select the two reference portfolios $P^{(1)}$ and $P^{(2)}$ and the reserve asset $P^{(3)}$. These 3 portfolios are dynamically traded on a daily basis, so we choose very liquid instruments with low transaction

costs. We therefore restrict the components of the portfolios to be Futures contracts. The cash rate is the BBA Libor 1-month rate. Log-returns on futures are calculated from the reinvestment of a rolling strategy in the front contract. the front contract is the nearest to maturity, on the March/June/September/December schedule and is rolled on the first business day of the maturity month at previous close prices. Each future contract is fully collateralized, so that, the total return is the sum of the rolling strategy returns and the cash rate.

The first investor portfolio is related to equities while the second is related to bonds. The reserve asset is a diversified portfolio. The composition of these portfolios is detailed in Table 3. As in Ben-Abdellatif (2010), we use daily returns from 01/10/1999 to 30/04/2009 (115 months). Table 4 presents some descriptive statistics of the daily returns $R^{(1)}$, $R^{(2)}$, $R^{(3)}$.

Table 3: Portfolios' composition.

$P^{(1)}$	60%	S&P/TSE 60 IX future
	40%	S&P500 EMINI future
$P^{(2)}$	100%	CAN 10YR BOND future
$P^{(3)}$	10%	E-mini NASDAQ-100 futures
	20%	Russell 2000 TR
	20%	MSCI Emerging Markets TR
	10%	GOLD 100 OZ future
	10%	WTI CRUDE future
	30%	US 2YR NOTE (CBT)

Table 4: Summary statistics for the three portfolios on an annual basis.

Statistics	$R^{(1)}$	$R^{(2)}$	$R^{(3)}$
	Daily returns		
Mean	0.0198	0.0209	0.0363
Volatility	0.1327	0.0592	0.1238
Skewness	-0.6447	-0.3261	-0.4418
Excess kurtosis	8.5478	2.0583	5.1415

5.2 Modeling

As discussed in Section 4.1, we use a Gaussian HMM to model the joint distribution of the returns of the 3 portfolios. The choice of the number of regimes is done as suggested in Rémillard et al. (2017): we choose the lowest number of regimes m so that the goodness-of-fit test for m regimes has a P -value larger than 5%. This leads to a selection of 6 regimes for the daily returns. The large number of regimes for the daily returns is due to the fact that the sample period contains the last financial crisis. Usually, for non-turbulent periods, 4 regimes are sufficient for fitting daily returns. The estimated parameters are given in Table 5. The associated transition matrix for daily returns of the Gaussian HMM is

$$\mathbf{Q}_{daily} = \begin{pmatrix} 0.9608 & 0.0000 & 0.0181 & 0.0000 & 0.0000 & 0.0211 \\ 0.0160 & 0.1494 & 0.3384 & 0.0000 & 0.4962 & 0.0000 \\ 0.0000 & 0.0579 & 0.6746 & 0.0108 & 0.2567 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.9823 & 0.0177 & 0.0000 \\ 0.0176 & 0.0993 & 0.2753 & 0.0175 & 0.5882 & 0.0021 \\ 0.0599 & 0.0000 & 0.0000 & 0.0000 & 0.0071 & 0.9330 \end{pmatrix},$$

and the infinitesimal generator associated with the limiting RSGBM is

$$\mathbf{\Lambda}_{daily} = \begin{pmatrix} -9.8765 & 0.0000 & 4.5658 & 0.0000 & 0.0000 & 5.3107 \\ 4.0402 & -214.3435 & 85.2680 & 0.0000 & 125.0353 & 0.0000 \\ 0.0000 & 14.5863 & -81.9990 & 2.7205 & 64.6922 & 0.0000 \\ 0.0000 & 0.0052 & 0.0004 & -4.4624 & 4.4569 & 0.0000 \\ 4.4414 & 25.0201 & 69.3636 & 4.4157 & -103.7633 & 0.5226 \\ 15.0866 & 0.0000 & 0.0000 & 0.0000 & 1.7858 & -16.8724 \end{pmatrix}.$$

Table 5: Estimated parameters for the Gaussian HMM fitted on daily returns.

Regime	Daily returns			
	$\boldsymbol{\mu}_j$		\mathbf{B}_j	
1	-0.0182	0.0250	-0.0026	0.0157
	0.0409	-0.0026	0.0028	-0.0021
	-0.1706	0.0157	-0.0021	0.0200
2	0.1709	0.0131	-0.0000	0.0114
	-1.6439	-0.0000	0.0040	0.0009
	0.1790	0.0114	0.0009	0.0170
3	0.6694	0.0050	0.0002	0.0018
	0.0619	0.0002	0.0036	-0.0006
	0.9667	0.0018	-0.0006	0.0040
4	0.1486	0.0042	-0.0002	0.0028
	0.0286	-0.0002	0.0018	0.0000
	0.2178	0.0028	0.0000	0.0047
5	-0.6934	0.0084	-0.0013	0.0049
	0.2548	-0.0013	0.0023	-0.0009
	-0.9222	0.0049	-0.0009	0.0067
6	-0.4565	0.1169	-0.0115	0.0788
	0.0749	-0.0115	0.0099	-0.0110
	-0.4082	0.0788	-0.0110	0.0889

NOTE: The values are expressed on an annual basis.

Finally, for the last observation, corresponding to the beginning of the hedging, the estimated probability of occurrence of each regime is

$$\boldsymbol{\eta}_{daily} = (0.9433, 0.0003, 0.0246, 0.0000, 0.0006, 0.0312).$$

Therefore, we will take for granted that at time $t = 0$, we are in regime 1.

5.2.1 Monthly returns

As suggested in Section 4.2, we simulated 10 000 values of monthly returns under the estimated RSGBM. We fitted a Gaussian HMM and found that 3 regimes were necessary, which is larger than usual, but we have to remember that we are fitting 10 000 values.

The estimated parameters are given in Table 6, and the associated transition matrix is

$$\mathbf{Q}_{monthly} = \begin{pmatrix} 0.1209 & 0.6788 & 0.2003 \\ 0.1719 & 0.6184 & 0.2097 \\ 0.1926 & 0.5846 & 0.2229 \end{pmatrix}.$$

Finally, for the last observation, corresponding to the beginning of the hedging, the estimated probability of occurrence of each regime is $\boldsymbol{\eta}_{monthly} = (0.1796, 0.7635, 0.0569)$. In particular, it means that the probability $\pi_{monthly}$ of being in each regime next month is

$$\boldsymbol{\pi}_{next} = \boldsymbol{\eta}_{monthly} \mathbf{Q}_{monthly} = (0.1639, 0.6273, 0.2088). \quad (13)$$

It then follows that the conditional distribution $\mathcal{F}(\cdot, \mathbf{x})$ is mixture of 3 Gaussian distributions, with mean $\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x}$ and standard deviation σ_j , $j \in \{1, 2, 3\}$, and weights given by (13), where the values of the parameters are given in Table 7. More precisely,

$$\mathcal{F}(y, \mathbf{x}) = \sum_{j=1}^3 \pi_{next}(k) \Phi \left(\frac{y - \alpha_j - \boldsymbol{\beta}_j^\top \mathbf{x}}{\sigma_j} \right), \quad (y, \mathbf{x}) \in \mathbb{R}^3, \quad (14)$$

where Φ is the distribution function of the standard Gaussian.

Table 6: Estimated parameters for the Gaussian HMM fitted on 10 000 simulated monthly returns under RSGBM.

Regime	μ_j		B_j	
1	0.0728	0.0085	-0.0006	0.0067
	0.0320	-0.0006	0.0027	-0.0004
	0.1081	0.0067	-0.0004	0.0096
2	-0.4201	0.0726	-0.0117	0.0396
	0.0050	-0.0117	0.0067	-0.0067
	-0.2813	0.0396	-0.0067	0.0421
3	-0.4201	0.0726	-0.0117	0.0396
	0.0050	-0.0117	0.0067	-0.0067
	-0.2813	0.0396	-0.0067	0.0421

NOTE: The values are expressed on an annual basis.

Table 7: Parameters of the conditional distribution of $R_{0,T}^{(3)}$ given $(R_{0,T}^{(1)}, R_{0,T}^{(2)})$.

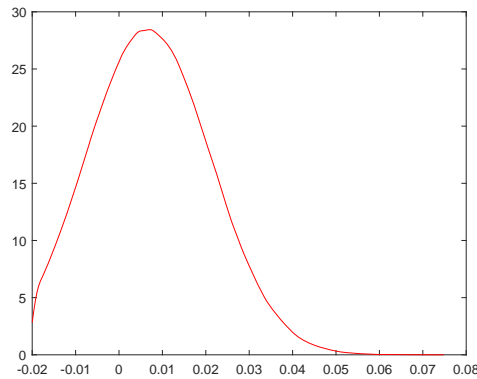
Regime	α_j	β_j	σ_j	π_j
1	0.0037	(0.6343 , -0.3353)	0.0463	0.1639
2	-0.0014	(0.6090 , -0.1828)	0.0296	0.6273
3	-0.0063	(0.6876 , 0.1661)	0.0231	0.2088

5.3 Target distribution function

For this example, the target distribution F_\star is a truncated Gaussian distribution at $-a$, with (annual) parameters μ_\star and σ_\star , meaning that

$$F_\star(y) = \begin{cases} 0, & y \leq -a; \\ \frac{\Phi\left(\frac{y-\mu_\star/12}{\sigma_\star/\sqrt{12}}\right) - \Phi\left(\frac{-a-\mu_\star/12}{\sigma_\star/\sqrt{12}}\right)}{\Phi\left(\frac{a+\mu_\star/12}{\sigma_\star/\sqrt{12}}\right)}, & y \geq -a. \end{cases} \quad (15)$$

Setting $z = \frac{a+\mu_\star/12}{\sigma_\star/\sqrt{12}}$ and $\kappa = \Phi'(z) / \Phi(z)$, the mean of this distribution is $\frac{\mu}{12} + \frac{\sigma}{\sqrt{12}}h$, while the standard deviation is $\frac{\sigma_\star}{\sqrt{12}}\sqrt{1-h^2-2hz}$. With $a = 0.02$, $\mu_\star = 0.08$ and $\sigma_\star = 0.05$, one gets an annual mean of 0.0842, and an annual volatility of 0.0477. Note that $F_\star(0) = 1 - \Phi\left(\frac{\mu_\star}{\sqrt{12}\sigma_\star}\right) / \Phi(z) = 0.3$. The density is displayed in Figure 2.

**Figure 2: Target density for the monthly returns.**

In the remaining of the section, we try to replicate the monthly returns of a synthetic hedge fund having distribution F_\star given by (15). We will rebalance the portfolio once a day, so $n = 21$. For simplicity, we

take $\mathbf{S}_0 = (1, 1, 1)$ and $r = 0.01$. We will consider two models: the first one is the independence model, meaning that $\mathcal{C}(u, v_1, v_2) = u$, so that the return function g is given by (4). This model is studied in Section 5.4. We consider another model, called the Clayton model, define using the B-vines representation by $D_1(u, t) = [\max\{0, u^{-\theta} + t^{-\theta} - 1\}]^{-1/\theta}$, which is the so-called Clayton copula of parameter $\theta \in (-1, 1)$, with Kendall's $\tau = \frac{\theta}{\theta+2}$, and $D_2(u, t) = ut$, the independence copula. For this case, we take $\theta = -2/3$, leading to a Kendall's tau of -0.5 . This means that we require a negative dependence with asset $P^{(1)}$.

Finally, for each model, we simulated 1 000 replication portfolios.

5.4 Synthetic index independent of the reference portfolios

The results of this first experiment are quite interesting, as can be seen from the results displayed in Table 8, especially the tracking error given by the RMSE. Note also that the mean of the hedging error is significantly smaller than 0, meaning that the portfolio is doing better on average than the target payoff, even if the K-P measure $\alpha = 0.0078$ is positive. The target distribution is also quite well replicated. The distribution of the hedging errors is also quite good, as can be seen from the estimated density displayed in Figure 3. Finally, letting $\tau^{(1)}$ and $\tau^{(2)}$ represent the estimated Kendall's tau between the variable and the returns of portfolio $P^{(1)}$ and $P^{(2)}$ respectively, one can see from that the returns of the hedged portfolio are independent of the returns of the reference portfolios, as measured by Kendall's tau, meaning that the synthetic asset has the desired properties.

Table 8: Descriptive statistics for the hedging error $HE = G(\mathbf{S}_{21}) - \mathcal{V}_{21}$, target payoff $G(\mathbf{S}_{21})$, portfolio \mathcal{V}_{21} , target return $g(\mathbf{R}_{21})$ and portfolio return $\log(\mathcal{V}_{21}/\mathcal{V}_0)$ in the independence model, based on 1000 replications.

Statistics	HE	$G(\mathbf{S}_{21})$	\mathcal{V}_{21}	$g(\mathbf{R}_{21})$	$\log(\mathcal{V}_{21}/\mathcal{V}_0)$	Target
Average	-0.012	100.770	100.782	0.0076	-0.0001	0.0078
Median	-0.012	100.741	100.760	0.0074	-0.0003	0.0073
Volatility	0.035	1.299	1.290	0.013	0.013	0.013
Skewness	0.431	0.201	0.192	0.172	0.162	0.267
Kurtosis	7.939	2.581	2.614	2.559	2.593	2.760
Minimum	-0.145	98.083	98.013	-0.019	-0.028	-0.02
Maximum	0.241	104.987	104.926	0.049	0.040	
RMSE	0.037					
$\tau^{(1)}$				0.023	0.024	0
$\tau^{(2)}$				-0.061	-0.060	0

NOTE: Here $\mathcal{V}_0 = 100.786$ and $\alpha = \log \mathcal{V}_0/100 = 0.0078$. The statistics for the target distribution are also displayed for sake of comparison. Note also that $\varphi_1 = (-26.464, 5.630, 42.050)$, showing that we are short of the first asset at the beginning.

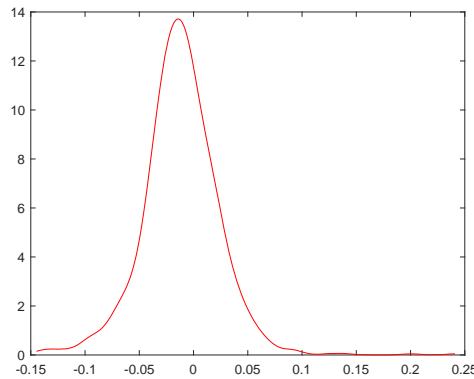


Figure 3: Estimated density of the hedging error $G(\mathbf{S}_{21}) - \mathcal{V}_{21}$ for the independence model based on 1000 replications. Here $\mathcal{V}_0 = 100.7864$ and $\alpha = \log \mathcal{V}_0/100 = 0.007833$.

5.5 Synthetic index with Clayton dependence level-1 dependence

The results of this second experiment are also quite interesting, but for different reasons. As can be seen from the results displayed in Table 9, our goal of replicating the distribution is not achieved. The tracking error given by the RMSE is too large, the average gain of the portfolio is negative and its volatility is too large to be interesting for an investor, even if the K-P measure $\alpha = 0.0064$ is smaller than in the independence model. This might be due to the fact that initially, the weight of the assets in the portfolio are quite large, since $\varphi_1 = (-724.845, 84.394, 648.811)$. Furthermore, The distribution of the hedging errors is not good at all, as can be seen from the estimated density displayed in Figure 4. The conclusion is that the target distribution is not quite well replicated, and one should not invest in this strategy. The only positive point is that the dependence between the returns of the payoff and portfolio seems to match the theoretical one, as measured by Kendall's tau.

Table 9: Descriptive statistics for the hedging error $HE = G(S_{21}) - V_{21}$, target payoff $G(S_{21})$, portfolio V_{21} , target return $g(\mathbf{R}_{21})$ and portfolio return $\log(V_{21}/V_0)$ in the Clayton model with $\tau = -0.5$ based on 1000 replications.

Statistics	HE	$G(S_{21})$	V_{21}	$g(\mathbf{R}_{21})$	$\log(V_{21}/V_0)$	Target
Average	-1.152	100.689	101.842	0.00680	-0.0271	0.0078
Median	3.399	100.608	97.071	0.0061	-0.0362	0.0073
Volatility	27.739	1.235	28.917	0.0122	0.2784	0.0133
Skewness	-0.772	0.268	0.753	0.240	0.077	0.267
Kurtosis	3.697	2.610	3.616	2.586	2.525	2.760
Minimum	-140.008	98.126	48.336	-0.019	-0.733	-0.02
Maximum	50.675	104.836	244.844	0.047	0.889	
RMSE	27.763					
$\tau^{(1)}$				-0.443	-0.461	-0.5
$\tau^{(2)}$				0.093	0.111	

NOTE: Here $V_0 = 100.645$ and $\alpha = \log V_0/100 = 0.0064$. The statistics for the target distribution are also displayed for sake of comparison.

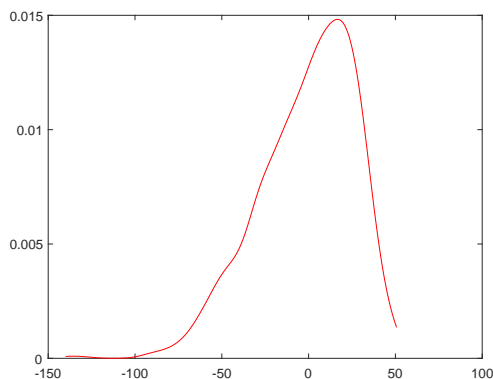


Figure 4: Estimated density of the hedging error $G(S_{21}) - V_{21}$ for the Clayton model with $\tau = -0.5$ based on 1000 replications.

To conclude this section, we computed the K-P measure for Clayton models as a function of Kendall's τ . This is illustrated in Figure 5 and it is coherent with the fact that the conditional distribution $\mathcal{D}_{1,\tau}$, with $\tau = \frac{\theta}{\theta+2}$, are ordered according to Lehmann's order. It then follows from (3) that the payoff are ordered as well, so the value of the option increases with τ .

5.6 Discussion

Before deciding to replicate an asset S^* , we should always perform a Monte Carlo experiment as we did in Sections 5.4–5.5. Using simulations, we can decide in advance if an asset S^* is worth replicating. For example, for our data, it is worth using the independence model, but it is not worth using the Clayton model. Simulation scan also be useful in tracking a more realistic P&L since transactions costs can be included in the simulations.

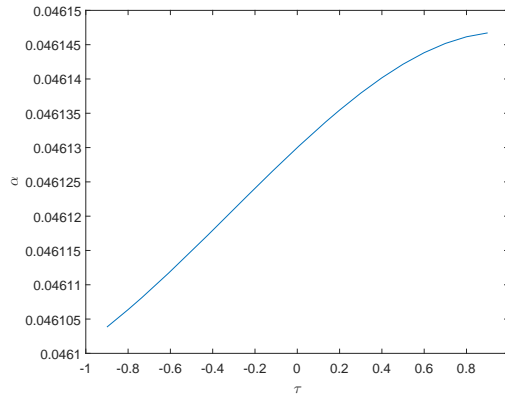


Figure 5: Graph of the K-P measure $\alpha = \log(\lambda_0/100)$ as a function of Kendall's τ for the Clayton model.

We notice that in all cases, the initial investment is more than 100, meaning that the K-P measure is positive. This can be attributed to the choice of the reserve asset. Indeed, Papageorgiou et al. (2008) showed that the choice of the reserve asset can affect the replication results especially the mean return, which depends linearly on the K-P measure. Nevertheless, at least in the case of the independence, we were able to achieve our goal.

It is also worth mentioning that due to (3), if two dependence models \mathcal{C}_1 and \mathcal{C}_2 are ordered according to Lehmann's order, i.e., for any $\mathbf{v} \in (0, 1)^d$, $\mathcal{C}_1(u, \mathbf{v}) \leq \mathcal{C}_2(u, \mathbf{v})$, for all $u \in [0, 1]$, then the K-P measures are also ordered.

6 Conclusion

We looked at two important methods of replication of indexes: strong and weak replication. For strong replication, the aim is to construct a portfolio of liquid assets that is as close as possible to an existing index, so statistical methods related to prediction like regression and filtering play an important role. For weak replication, the aim is to construct a portfolio of liquid assets that is as close as possible to a payoff constructed in such a way that the portfolio returns have predetermined distributional properties, such as the marginal distribution and the conditional distribution relative to some reference assets entering in the construction of the portfolio.

We also introduced a new family of conditional distribution models called B-vines that can be useful in many fields, not just weak replication of indexes.

We showed how to implement weak replication in general framework, and we showed that it is possible to construct efficiently a synthetic asset that is independent of prescribed asset classes, with a predetermined distribution. Using simulations, we can decide in advance if an asset S^* is worth replicating. For example, for our data, it is worth using the independence model, but it is not worth using the Clayton model.

For future work, we plan to investigate the performance of machine learning methods compared to filtering methods for strong replication purposes. We will also propose goodness-of-fit tests for the B-vines models introduced in Section 3.

Appendix

A Optimal hedging in continuous time

For $j \in \{1, \dots, l\}$, let $\mathbf{m}(j) = (\mathbf{v}(j) - r\mathbf{1})$, where $\mathbf{1}$ is the vector with all components equalled to 1, $\boldsymbol{\rho}(j) = \mathbf{a}(j)^{-1}\mathbf{m}(j)$, and set $\ell_j = \boldsymbol{\rho}(j)^\top \mathbf{m}(j) = \boldsymbol{\rho}(j)^\top \mathbf{a}(j)\boldsymbol{\rho}(j) \geq 0$. Further set $\boldsymbol{\gamma}(t) = e^{(T-t)\{\boldsymbol{\Lambda} - D(\ell)\}}\mathbf{1}$. Next, define

$$(\tilde{\boldsymbol{\Lambda}}_t)_{ij} = \boldsymbol{\Lambda}_{ij}\boldsymbol{\gamma}_j(t)/\boldsymbol{\gamma}_i(t), \quad i \neq j, \quad (\tilde{\boldsymbol{\Lambda}}_t)_{ii} = -\sum_{j \neq i} (\tilde{\boldsymbol{\Lambda}}_t)_{ij}. \quad (16)$$

Then $\tilde{\boldsymbol{\Lambda}}_t$, $t \in [0, T]$, is the infinitesimal generator of a time inhomogeneous Markov chain.

In Rémillard and Rubenthaler (2016), it is shown that the optimal hedging problem is related to an equivalent martingale measure \mathbb{Q} , in the sense that under the risk neutral measure \mathbb{Q} , if the price process \mathbf{X} satisfies

$$d\mathbf{X}_t = rD(\mathbf{X}_t)dt + D(\mathbf{X}_t)\boldsymbol{\sigma}(\mathcal{T}_{t-})d\mathbf{W}_t, \quad (17)$$

and \mathcal{T} is a time inhomogeneous Markov chain with generator $\tilde{\boldsymbol{\Lambda}}_t$, then the value of an option with payoff Φ at maturity T is given by

$$C_t(\mathbf{s}, i) = e^{-r(T-t)}E^{\mathbb{Q}}[\Phi(\mathbf{X}_T)|\mathbf{X}_t = \mathbf{s}, \mathcal{T}_t = i]. \quad (18)$$

If the payoff is smooth enough so that it is differentiable almost everywhere, then

$$\nabla_{\mathbf{s}}C_t(\mathbf{s}, i) = e^{-r(T-t)}D^{-1}(\mathbf{s})E^{\mathbb{Q}}[\Phi'(\mathbf{X}_T)\mathbf{X}_T|\mathbf{X}_t = \mathbf{s}, \mathcal{T}_t = i], \quad i \in \{1, \dots, l\}. \quad (19)$$

Since C_t and $\nabla_{\mathbf{s}}C_t$ are related to expectations, one can use Monte Carlo methods to obtain unbiased estimates of these values.

Next, setting $\boldsymbol{\alpha}_t(s, i) = \nabla_{\mathbf{s}}C_t(s, i) + C_t(s, i)D^{-1}(\mathbf{s})\boldsymbol{\rho}(i)$, and $\mathcal{G}_t = e^{-rt}C_t(\mathbf{X}_t, \mathcal{T}_t) - \mathcal{V}_t$, with $\mathcal{G}_0 = 0$, where \mathcal{V}_t is the discounted value of the (continuous time) hedging portfolio at time t , then the optimal hedging strategy is

$$\boldsymbol{\phi}_t = \boldsymbol{\alpha}_t(X_t, \mathcal{T}_{t-}) - e^{rt}\mathcal{V}_{t-}D^{-1}(\mathbf{X}_t)\boldsymbol{\rho}(\mathcal{T}_{t-}) \quad (20)$$

$$= \nabla_{\mathbf{s}}C_t(\mathbf{X}_t, \mathcal{T}_{t-}) + e^{rt}\mathcal{G}_{t-}D^{-1}(\mathbf{X}_t)\boldsymbol{\rho}(\mathcal{T}_{t-}). \quad (21)$$

References

- Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance Math. Econom.*, 44(2):182–198.
- Amin, G. and Kat, H. M. (2003). Hedge fund performance 1990-2000: Do the “money machines” really add value. *Journal of Financial and Quantitative Analysis*, 38(2):251–275.
- Ben-Abdellatif, M. (2010). Réplication des fonds de couverture. Master’s thesis, HEC Montréal.
- Dybvig, P. H. (1988). Distributional analysis of portfolio choice. *The Journal of Business*, 61(3):369–393.
- Fung, W. and Hsieh, D. (2001). The risk in hedge fund strategies: Theory and evidence from trend followers. *Review of Financial Studies*, 14:313–341.
- Fung, W. and Hsieh, D. (2004). Hedge fund benchmarks: A risk-based approach. *Financial Analysts Journal*, 60(5):65–80.
- Hasanhodzic, J. and Lo, A. W. (2007). Can hedge-fund returns be replicated?: The linear case. *Journal of Investment Management*, 5(2):5–45.
- Joe, H. (1996). Families of m -variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In Rüschendorf, L., Schweizer, B., and Taylor, M. D., editors, *Distributions with fixed marginals and related topics*, volume 28 of *Lecture Notes–Monograph Series*, pages 120–141. Institute of Mathematical Statistics, Hayward, CA.

- Kat, H. M. and Palaro, H. P. (2005). Who needs hedge funds? A copula-based approach to hedge fund return replication. Technical report, Cass Business School, City University.
- Laroche, P. and Rémillard, B. (2008). On the serial dependence of hedge fund indices returns. Technical report, Innocap.
- Papageorgiou, N., Rémillard, B., and Hocquard, A. (2008). Replicating the properties of hedge fund returns. *Journal of Alternative Investments*, 11:8–38.
- Rémillard, B. (2013). *Statistical Methods for Financial Engineering*. Chapman and Hall/CRC Financial Mathematics Series. Taylor & Francis.
- Rémillard, B., Hocquard, A., Lamarre, H., and Papageorgiou, N. A. (2017). Option pricing and hedging for discrete time regime-switching model. *Modern Economy*.
- Rémillard, B., Nasri, B., and Bouezmarni, T. (2017). On copula-based conditional quantile estimators. *Statistics & Probability Letters*, 128:14–20. September 2017.
- Rémillard, B. and Rubenthaler, S. (2009). Optimal hedging in discrete and continuous time. Technical Report G-2009-77, Gerad.
- Rémillard, B. and Rubenthaler, S. (2013). Optimal hedging in discrete time. *Quantitative Finance*, 13(6):819–825.
- Rémillard, B. and Rubenthaler, S. (2016). Option pricing and hedging for regime-switching geometric Brownian motion models. Working paper series, SSRN Working Paper Series No. 2599064.
- Roncalli, T. and Teïletche, J. (2007). An alternative approach to alternative beta. Technical report, Société Générale Asset Management.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Stat.*, 23:470–472.
- Schweizer, M. (1992). Mean-variance hedging for general claims. *Ann. Appl. Probab.*, 2(1):171–179.
- Towsey, M. (2013). Hedge fund replication. Technical report, Aon Hewitt.
- Wallerstein, E., Tuchschnid, N. S., and Zaker, S. (2010). How do hedge fund clones manage the real world? *The Journal of Alternative Investments*, 12(3):37–50.