Applications of Statistical Learning in Quantitative Finance

Dissertation

submitted to the

Faculty of Business, Economics and Informatics

of the University of Zurich

to obtain the degree of

Doktor der Wirtschaftswissenschaften, Dr. oec.

(corresponds to Doctor of Philosophy, PhD)

presented by

Urban Ulrych

from Slovenia

approved in April 2022 at the request of

Prof. Dr. Erich Walter Farkas

Prof. Dr. Markus Leippold

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, 06.04.2022

Chairman of the Doctoral Board: Prof. Dr. Steven Ongena

To my beloved family

Acknowledgements

I am thankful to many people for guiding me through my Ph.D. studies. First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Dr. Erich Walter Farkas, for his guidance, availability, and constant support in all matters. This dissertation as well as the industry experience that I gained during my time at UZH would not have been possible without his help and generosity. Moreover, I am grateful to Prof. Dr. Markus Leippold for providing valuable feedback on my dissertation and for being a part of my doctoral committee. I would also like to thank Prof. Dr. Damir Filipović for his fruitful comments on my work and for attending my defense as the SFI Senior Chair.

My gratitude is extended to my co-authors Dr. Nikola Vasiljević, Prof. Dr. Paweł Polak, Raphael Burkhardt, and David Anderson for their passion, curiosity, and positivism towards research. I have truly enjoyed our collaboration and learned a lot from you. Furthermore, I would like to thank Dr. Yves Allkofer for the opportunity to join his team at Credit Suisse at the end of my doctoral studies. My time at Quantitative Strategists has been very enriching, both on a professional and a personal level.

During my entire journey as Ph.D. student, I was fortunate to be surrounded by smart colleagues and good friends who greatly influenced my last four years. In particular, I would like to thank my colleagues at the university – Isabella, Alex, Patrick, Ludovic, Ming, Jan, Justin, Sean, Simon, Michal, Ana, Emanuela, Hanlin, Chantal, Michelle, and Sarah. I feel lucky to have you during this phase of my life.

Last and most important, I am indebted to my family for their unconditional support and love. There are no words that could express my thankfulness to my parents who have spared no effort to allow me to pursue my goals and dreams in life and who have always stood by my side on my journey.

Zurich, April 2022 *U.U.*

Contents

Part I

Introduction

Introduction and Summary of Research Results

The dissertation "Applications of Statistical Learning in Quantitative Finance" consists of four papers that utilize various methods arising from the broader area of statistical learning and apply them to solve problems from the field of quantitative finance. The applications of statistical learning methods in finance are continuously growing with theoretical and computational developments in both quantitative finance and computational statistics. This dissertation contributes to the existing literature by proposing novel solutions to practically relevant problems arising from quantitative risk management. Thereby, different methods from statistical learning are utilized in order to study and improve the risk-ambiguity-return spectrum of international investors and accelerate the pricing of American-style options. Namely, i) the optimal currency allocation for a risk-and-ambiguity averse investor is characterized as the solution to a generalized ridge regression, ii) a non-Gaussian returns model based on a continuous normal mean-variance mixture representation is utilized to develop a dynamic currency hedging strategy, iii) a sparse and stable optimization approach for a multi-currency asset allocation problem employing various regularization techniques is derived, and iv) a feed-forward neural network is applied for the acceleration of the pricing of American-style options. A brief summary of the four research articles is presented below.

In Chapter 1, the paper Ulrych and Vasiljević [2020] addresses the problem of determining an optimal currency allocation for a risk-and-ambiguity-averse international investor. A robust mean–variance model with smooth ambiguity preferences is used to derive the optimal currency exposure. The theoretical part of the paper shows that the sample-efficient currency demand can be calculated as the solution to a generalized ridge regression. Through the lens of these results, we demonstrate that our model offers a new explanation of the home currency bias as the optimal currency allocation under extreme ambiguity aversion. The investor's dislike for model uncertainty induces a disproportionately high currency hedging demand. The empirical analysis demonstrates how ambiguity leads to a larger estimation bias and simultaneously narrows the confidence interval of the optimal currency exposure. The out-of-sample backtest illustrates that accounting for ambiguity enhances the stability of optimal currency allocation and significantly improves the risk-adjusted portfolio performance net of transaction costs.

In Chapter 2, the paper Polak and Ulrych [2021] introduces a non-Gaussian dynamic currency hedging strategy for globally diversified investors with ambiguity. Assuming that ambiguity of a typical investor can be measured from market data, we associate it to non-Gaussianity of financial asset returns and compute an optimal ambiguity-adjusted mean-variance (dynamic) currency allocation. Next, we extend the filtered historical simulation method to numerically optimize an arbitrary risk measure, such as the expected shortfall. The out-of-sample backtest results show that the derived non-Gaussian dynamic currency hedging strategy outperforms the benchmarks of constant hedging and dynamic hedging with Gaussianity for all base currencies and net of transaction costs.

In Chapter 3, the paper Burkhardt and Ulrych [2022] introduces a sparse and stable optimization approach for a multi-currency asset allocation problem. We study the benefits of joint optimization of assets and currencies as opposed to the standard industry practice of managing currency risk via so-called currency overlay strategies. In our setting, a classical mean–variance problem in an international framework is augmented by several extensions that aim at reducing parameter uncertainty related to the input parameters and induce sparsity and stability of the asset and currency weights. These extensions integrate maximal net exposure to foreign currencies, shrinkage of the input parameters, and constraints on the norms of the asset- and currency-weight vectors. The empirical performance of the portfolio optimization strategies based on the proposed regularization techniques and the joint (i.e., asset and currency) optimization is tested out of sample. We demonstrate that the sparse and stable joint optimization approach consistently outperforms the standard currency overlay as well as the equally-weighted and the non-regularized global portfolio benchmarks net of transaction costs. This result shows that the common industry practice of employing currency overlay strategies is suboptimal and can be improved by a joint optimization over assets and currencies.

In Chapter 4, the paper Anderson and Ulrych [2022] proposes a novel method for accelerating the pricing of American options to near-instantaneous using a feed-forward neural network. Given the competitiveness of a market-making environment, the ability to speedily quote option prices consistent with an ever-changing market environment is essential. Thus, the smallest acceleration or improvement over traditional pricing methods is crucial to avoid arbitrage. The paper introduces an approach in which a neural network is trained over the chosen (e.g., Heston) stochastic volatility specification. Such an approach facilitates parameter interpretability, as generally required by the regulators, and establishes the presented method in the area of eXplainable Artificial Intelligence (XAI) for finance. We show that the proposed deep explainable pricer induces a speed-accuracy trade-off compared to the typical Monte Carlo or Partial Differential Equation-based pricing methods. Moreover, the proposed approach allows for pricing derivatives with path-dependent and more complex payoffs and is, given the sufficient accuracy of computation and its tractable nature, applicable in a market-making environment.

Methods from the field of computational statistics have been used in the financial industry for many decades, however, statistical learning in quantitative finance as a research area is still emergent and in many aspects incomplete. On the one hand, finance possesses a vast potential needed to effectively utilize methods from statistical learning, namely, the large amount of data, high-dimensional setting, sufficient computational resources, and direct profit and loss implications. On the other hand, financial institutions are continuously under close supervision by the regulators who are generally resistant to the so-called "black-box" algorithms but are reorienting more and more toward a data-driven regulation. This dissertation presents four papers that apply methods from statistical learning in an explainable and interpretable manner. Model explainability is one of the crucial aspects that would ensure investors and regulators with understanding and trust to use the proposed approaches in their decision-making. This is important since the topics investigated in this dissertation are practically relevant and widely discussed in the financial services industry, predominantly in the areas of strategic asset allocation and derivatives pricing.

Part II

Research Articles

Chapter 1

Ambiguity, Optimal Currency Overlay, and Home Currency Bias

A version of this paper has been submitted to the *Journal of Banking & Finance*.

Authors:

Urban Ulrych, Department of Banking and Finance, University of Zurich & Swiss Finance Institute

Nikola Vasiljević, Department of Banking and Finance, University of Zurich

Abstract

This paper addresses the problem of determining an optimal currency allocation for a risk-and-ambiguity-averse international investor. A robust mean–variance model with smooth ambiguity preferences is used to derive the optimal currency exposure. The theoretical part of the paper shows that the sample-efficient currency demand can be calculated as the solution to a generalized ridge regression. Through the lens of these results, we demonstrate that our model offers a new explanation of the home currency bias as the optimal currency allocation under extreme ambiguity aversion. The investor's dislike for model uncertainty induces a disproportionately high currency hedging demand. The empirical analysis demonstrates how ambiguity leads to a larger estimation bias and simultaneously narrows the confidence interval of the optimal currency exposure. The out-of-sample backtest illustrates that accounting for ambiguity enhances the stability of optimal currency allocation and significantly improves the risk-adjusted portfolio performance net of transaction costs.

Keywords: Ambiguity aversion, home currency bias, model uncertainty, optimal currency overlay, generalized ridge regression, international asset allocation.

JEL Classification: D81, D83, F31, G11, G15.

1.1 Introduction

Diversification is often said to be the only free lunch in finance. International asset allocation is a natural way to improve risk-adjusted portfolio performance—it opens access to a multitude of global investment opportunities and allows for diversification across asset classes, factors, styles, and geographies. A carefully tailored international portfolio can improve both growth potential and risk management. However, risks can only be reduced and not eliminated. One of the main challenges for global asset allocation is the currency risk. Fluctuations of exchange rates are influenced by macroeconomic, financial, and political factors. Shifts in currency markets can have a profound impact on portfolio performance. Therefore, investors have to decide on the hedging policy in their portfolios, i.e., the amount of foreign currency exposure that should be hedged.

Increasing complexity of financial markets and products, structural changes caused by secular trends (such as the fourth industrial revolution, aging population, and climate change) or exogenous shocks (such as the Covid-19 pandemic), as well as frequent short-term volatility bouts add to investors' ambiguity aversion (i.e., the aversion to difficult-to-quantify uncertainty) on top of the risk aversion. In this paper, we study optimal currency allocation in international portfolios under ambiguity. Building on the work of Maccheroni et al. [2013], we account for model uncertainty directly in the investor's ambiguity-adjusted preferences and solve a robust mean–variance optimization problem for currency overlay strategies. Our main theoretical contribution is that we provide a closed-form expression for optimal currency allocation under both risk and ambiguity aversion. In particular, we show that the optimal currency exposure can be interpreted as the solution of a generalized ridge regression, where ambiguity aversion is the key driver of the intensity of ridge regularization. Our model is tractable and offers an intuitive geometric interpretation of the optimal currency exposure shrinkage in the presence of ambiguity. This result represents a generalization of the solution obtained via ordinary least squares regression in the risk-only framework considered in Campbell et al. [2010], which builds on the work of Britten-Jones [1999]. The equivalence between the ambiguity-adjusted mean–variance optimization and the generalized ridge regression has broader implications. It establishes a link between two important and growing research fields—uncertainty in financial economics and statistical learning.

Considering ambiguity in the context of optimal currency allocation is relevant for at least two reasons. First, there is a general dissatisfaction with the empirical and predictive performance of the standard rational expectations paradigm and expected utility theory. Second, many empirical stylized facts cannot be explained using the classical, risk-only models. Embedding ambiguity aversion into financial modeling is currently perceived as one of the most promising avenues for future research.[1] Our paper contributes

[1] An excellent literature overview of the applications of ambiguity in asset pricing and portfolio selection is provided in

to the extant literature by accounting for the uncertainty associated with predictive models in an optimal currency allocation framework. Furthermore, we offer a novel explanation of the home currency bias, i.e., the empirical fact that many investors and firms avoid foreign currency exposure and therefore hold insufficiently diversified currency portfolios [e.g., Maggiori et al., 2020]. In our model, the optimal foreign currency exposure is inversely related to the ambiguity aversion. In the extreme case of an infinitely ambiguity-averse investor, we find that the optimal strategy is to fully hedge the currency risk.

In our empirical study, we consider currency overlay strategies in a portfolio context. Such strategies are widely applied in the asset and wealth management industry. They are designed for a specific purpose—to separate asset allocation and portfolio construction from currency hedging decisions. Arguably, this approach improves the risk management process. This is particularly important for active managers and their positioning over shorter, tactical investment horizons. Therefore, we follow the common practice and assume that portfolio weights are predefined. This allows us to focus on currencies, and isolate the impact of ambiguity aversion on the optimal foreign currency exposure and the home currency bias in particular.

In the first step, we demonstrate that ambiguity induces a statistical bias and shrinks the confidence interval of the proposed optimal currency exposure estimator. This result gives rise to a possible bias–variance trade-off. Second, in an out-of-sample backtest, we show that the efficiency of our estimator can be improved by embedding model uncertainty directly in the investor's utility function. Using a panel of seven developed market currencies from January 1999 until December 2019, we consider several popular predictive models presented to increase the predictive power of our optimal currency exposure model. Finally, we provide supporting evidence that hedging based on our ambiguity-adjusted model outperforms the minimum–variance and mean–variance approaches as well as the constant-hedging-ratio rules, net of transaction costs. The key takeaway is that accounting for ambiguity aversion stabilizes the optimal exposure estimates and reduces hedging turnover. Therefore, besides its theoretical appeal, our model is of particular interest for practical applications of currency hedging in the financial industry.

The paper is organized as follows. Section 1.2 reviews the existing literature. Section 1.3 introduces a theoretical framework for international asset allocation based on a robust mean–variance model under smooth ambiguity preferences. We provide a solution to the currency overlay optimization problem for a risk-and-ambiguity-averse investor, and a geometric interpretation of our results. Next, we show that the optimal currency allocation under extreme ambiguity aversion corresponds to the empirical puzzle of home currency bias in international portfolios. Section 1.4 presents an empirical study in which the impact of ambiguity aversion on the optimal currency exposure is investigated both in-sample and out-of-sample. Section 1.5 concludes. Additional proofs, explanations, and figures are delegated to the Appendix.

Guidolin and Rinaldi [2013].

1.2 Literature Review

Our paper is related to two strands of literature. This section provides an overview of the extant research that laid out the foundation for our study. The first group of papers is related to optimal currency hedging in the context of international asset allocation. The second strand addresses the estimation and model risk, specifically the impact of ambiguity on asset allocation.

1.2.1 Currency Hedging Policy

A number of hedging strategies have been presented in the literature over the past four decades. Some authors offered pragmatic but rather simplistic solutions to the optimal currency exposure problem. Others considered more nuanced approaches to currency hedging. Overall, there seems to be a broad consensus that (a) currency hedging tends to lower the portfolio volatility, and (b) conditional hedging outperforms strategies that employ fixed hedge ratios. When combined, these results indicate that one should consider both risk minimization and speculative side when deciding about the optimal hedging policy. The empirical results presented in the literature are varied across reference currencies, asset classes, asset allocation mixes, and investment horizons.

Perold and Schulman [1988] proposed full hedging of the face value of a foreign investment. They argued that currency trading is a zero-sum game on average; therefore, hedging represents a "free lunch" because it reduces volatility without a loss of expected return over the long haul.[2] Similar findings were presented in Madura and Reiff [1985] and Jorion [1989], among others.

These results stand in sharp contrast to Froot [1993], who argued that full hedging is beneficial only over short investment horizons. In the long run, full hedging can increase risk without an adequate return compensation. The main argument is that hedged returns are dominated by surprises in inflation and real interest rates. In particular, real interest rates are mean-reverting due to the purchasing power parity, and therefore provide a "natural hedge" over long periods. Since hedging currency exposure does not provide a protection against risk factors affecting long-term exchange rates, the optimal hedging ratio should be zero. However, these results are based on the currency regimes preceding the free-floating exchange rates. Using more recent data, Schmittmann [2010] found that the investment horizon is of limited importance for the hedging policy.

Black [1989, 1990] argued that all investors should apply a "universal hedging policy" irrespective of the portfolio composition and the reference currency. The hedging ratio is always less than one, as a consequence of Siegel [1972]'s paradox. Solnik [1993] challenged this view and showed that the equilibrium hedge ratio is specific for each investor, and it is a function of their risk preferences and relative wealth. On the other

[2] Naturally, active investors might decide to deviate from their currency hedging policy on a tactical basis.

hand, Gardner and Wuilloud [1995] and Gastineau [1995] advocated the use of 50% hedge as a baseline currency position. Their "middle-road approach" is an attempt to find a compromise between the universal hedging policy and various practical considerations (e.g., benchmarking, active management, hedging costs, and certain behavioral aspects). Additionally, we note that Ang and Bekaert [2002] obtained a similar hedge ratio using a regime-switching framework.

Glen and Jorion [1993] argued that currency hedging reduces portfolio risk; however it is beneficial only if it does not materially impact the portfolio return. Their analysis indicates that conditional hedging strategies are crucial for the performance of internationally diversified portfolios. Levy and Lim [1994] demonstrated that currency hedging often reduces risk at the expense of lower return. They concluded that this can be explained by the currency forward rate being a biased predictor of the future spot rate. De Roon et al. [2011] included currency positions as an additional asset class in internationally diversified portfolios. They showed that improvements in portfolio risk-return trade-off are primarily driven by speculative rather than risk hedging benefits. More recently, Boudoukh et al. [2019] formalized these ideas by introducing a modified portfolio mean–variance optimization approach and decomposing the optimal portfolio mix into three components: (a) an equity portfolio, (b) a risk-minimizing currency portfolio, and (c) an alpha-generating currency portfolio.

Jorion [1994] considered a global mean–variance optimization, where positions in assets and currencies can be determined simultaneously or separately. Either way, the optimal currency exposure strongly depends on the portfolio reference currency. Haefliger et al. [2002] proposed full hedging for fixed-income portfolios, while equity portfolios should be partially hedged or unhedged, depending on the correlations between equity and currency returns. Campbell et al. [2010] considered risk-minimizing currency overlay strategies for global equity and bond investors. Their analysis critically relies on the correlations between the currencies and core asset classes. For example, low currency–bond correlations indicate that international bond investors should fully hedge their currency exposure. This is consistent with the common practice of institutional investors. On the other hand, due to positive (negative) correlations between commodity (reserve) currencies and global equities, the optimal currency overlay for an international equity investor consists of a long position in reserve currencies and a short position in commodity currencies. A large body of literature on dynamic hedging concluded that currency strategies that are conditional on the interest rate spreads significantly improve portfolio performance compared to static hedging with forward contracts [e.g., Kroner and Sultan, 1993, Tong, 1996, De Roon et al., 2003, Brown et al., 2012, Caporin et al., 2014, Cho et al., 2020]. Additionally, several recent papers which explore the currency predictability using currency risk factors [e.g., Lustig et al., 2011, Verdelhan, 2018, Opie and Riddiough, 2020] demonstrated that dynamic hedging represents a promising route to achieve competitive out-of-sample performance.

Finally, within the realm of currency hedging, a special case of interest is the home currency bias, which represents investors' tendency to disproportionately hold assets denominated in their home currency [e.g., Solnik, 1974, Campbell et al., 2003]. Home currency bias is at least partially attributable to the home country bias. Burger et al. [2018] showed that this is particularly true in international bond portfolios. Maggiori et al. [2020] empirically show that the home currency bias has a substantial impact on global capital allocation. Finally, considering a more general setting, Bianchi and Tallon [2019] and Berger and Eeckhoudt [2021] argue that ambiguity aversion can result in a lower diversification, and an elevated portfolio risk.

1.2.2 Ambiguity and Asset Allocation

Most studies about optimal currency exposure rest on the assumption that investors know perfectly the true probability law governing the asset and currency return dynamics. However, investors are often uncertain about the model validity. The two primary concerns are estimation and model risk. For example, Eun and Resnick [1988] demonstrated that low accuracy of estimated input parameters—in particular the mean returns—is the main driver of a poor ex-ante performance of the joint optimization of asset allocations and currency hedge ratios.

A fundamental distinction between risk and uncertainty has been recognized and extensively studied since Knight [1921]. Generally, the risk can be defined if all relevant events are associated with a unique probability measure. The (Knightian) uncertainty refers to situations in which some events do not have an obvious probability assignment.[3] Ellsberg [1961] provided experimental evidence that agents are not always able to derive a unique probability distribution over the reference state space. This is also known as ambiguity, and a dislike of uncertainty is commonly referred to as ambiguity aversion.[4] The two most prominent approaches to incorporate uncertainty into investment decision making are Bayesian portfolio analysis and ambiguity-averse preferences.

Bayesian portfolio analysis allows for an inclusion of prior information about various quantities of interest (e.g., asset prices and macroeconomic variables) while accounting for the estimation risk and model uncertainty. An excellent literature review is presented in Avramov and Zhou [2010], who classified Bayesian portfolio studies into three categories based on the assumptions regarding asset return dynamics. The first strand of literature focuses on independently and identically distributed asset returns [e.g., among others, Klein and Bawa, 1976, Jorion, 1986, Black and Litterman, 1992, Pástor, 2000, Pástor and Stambaugh, 2000, Kan and Zhou, 2007, Tu and Zhou, 2010]. The priors used in these studies range from uninformative and data-driven to those based on asset pricing theories. The second group of papers considers the possibility that asset returns can be predicted by macroeconomic and fundamental financial variables such as growth,

[3] In other words, the Knightian uncertainty represents "non-probabilized" uncertainty, as opposed to the risk which corresponds to "probabilized" uncertainty.

[4] In accordance with the literature, we use uncertainty and ambiguity interchangeably.

inflation, dividend yield, term and credit spreads [e.g., Kandel and Stambaugh, 1996, Stambaugh, 1999, Barberis, 2000, Pástor and Stambaugh, 2002, Avramov, 2002, 2004, Avramov and Chordia, 2006, Wachter and Warusawitharana, 2009]. The third category comprises alternative, more complex models of asset returns such as stochastic volatility and regime switching [e.g., Han, 2006, Pástor and Stambaugh, 2009, Tu, 2010].

Ambiguity-averse preferences take a leap from the traditional use of rational expectations and maximization of (subjective) expected utility. Uncertainty regarding the asset return predictions is captured directly in the utility function. Intuitively, when a decision maker has too little information to form a single prior, she may consider a set of probability distributions instead. This notion was formalized in Schmeidler [1989], who stressed that the probability attached to an uncertain event may not reflect the heuristic amount of information that has led to that particular probability assignment. To enable encoding of information that additive probabilities cannot represent, non-additive probabilities (i.e., capacities) were proposed. Gilboa and Schmeidler [1989] extended this work by introducing multiple prior preferences. Building on these seminal papers, Anderson et al. [1998, 2003] and Hansen and Sargent [2001] extended the use of multi-prior criteria to the robust control theory. In that framework, a set of probabilities is generated by statistical perturbations of an approximating model. This approach corresponds to the situations in which agents have a specific model of reference and, acknowledging the possibility of errors, seek robustness against misspecifications.

Klibanoff et al. [2005] proposed that the ambiguity of a risky event can be characterized by a set of subjectively plausible cumulative probability distributions for this event. The decision maker subjectively weights these distributions, and the resulting preference relation describes the investor's attitude towards ambiguity. Building on this, Maccheroni et al. [2013] derived the analogue of the classic Arrow–Pratt approximation of the certainty equivalent under model uncertainty as described by the smooth model of decision making under ambiguity. They study its scope by deriving a tractable mean–variance model adjusted for ambiguity and solving the corresponding portfolio allocation problem. We use a similar approach in this work and connect this ambiguity aversion adjusted mean–variance preferences to the optimal currency exposure framework. The analytical tractability of the enhanced Arrow–Pratt approximation renders this model especially well suited for calibration exercises aimed at exploring the consequences of model uncertainty on the optimal currency allocations.

1.3 Theoretical Framework

In this section, we first set out the stage for our model by deriving the key preliminary results. In particular, we commence by considering currency hedging with forward contracts in an international portfolio context. Building on the work of Eun and Resnick [1988], closed-form expressions for hedged and unhedged portfolio returns are derived. These results are model-free in the sense that no specific dynamics for the asset and currency returns are assumed. Next, we recap the workhorse model in our framework— Maccheroni et al. [2013]'s robust mean–variance optimization, which is rooted in the Arrow-Pratt approximation of the certainty equivalent in the case of a risk-and-ambiguity-averse (RAA) agent who maximizes von Neumann–Morgenstern expected utility. Subsequently, we derive optimal currency overlay strategies for an RAA investor and discuss several special cases to provide economic background and intuition of our theoretical results. Finally, we show that a generalized ridge regression can be used to recover sample-efficient currency exposures, and offer a geometric interpretation of these findings.

1.3.1 Preliminaries

1.3.1.1 Foreign Currency Exposure and Portfolio Returns

Our starting point is a simple example of an unhedged foreign-currency-denominated asset. More specifically, we consider an asset i whose price in local currency at time t is denoted by $P_{i,t}$. The simple return on this asset from t to $t+1$ is given by $R_{i,t+1}$.[5] We introduce $S_{c,t}$ as the spot exchange rate in the reference/home currency (HC) per unit of foreign/local currency (LC) at time t, where c denotes the local currency of asset i. Alternatively, for financial assets that have embedded multi-currency exposure (e.g., global equity indices), c_i can be interpreted as a single currency in which prices of asset i are quoted on markets. The exchange rate return from t to $t+1$ is $e_{c,t+1}$. The unhedged return on asset i expressed in the home currency is

$$\tilde{R}^u_{i,t+1} = \frac{P_{i,t+1}S_{c,t+1}}{P_{i,t}S_{c,t}} - 1 = \underbrace{R_{i,t+1}}_{\text{LC asset return}} + \underbrace{e_{c,t+1}}_{\text{FX return}} + \underbrace{R_{i,t+1}e_{c,t+1}}_{\text{cross-product}}. \tag{1.3.1}$$

The above equation demonstrates that the unhedged asset return in home currency is driven by three components: (a) The asset return in local currency, (b) The exchange rate (FX) return, and (c) The second-order cross-product between the first two terms.

This result can be generalized from a single-asset case to a portfolio context. We consider an investor—with an arbitrary home currency—who holds a portfolio $\mathcal{P}$ which consists of N assets. The fraction of wealth invested in asset $i = 1, 2, \ldots, N$ is defined as $x_{i,t}$. Let us further assume that this portfolio has a direct exposure to K foreign currencies. For simplicity, we label the home currency by $c = 1$, whereas the foreign currencies are denoted by $c = 2, 3, \ldots, K+1$. The assets can be classified and grouped by their local currency.

[5] The following notation is applied throughout the paper. Local currency returns are denoted by $R_{\star,\star}$, where the first (second) subscript indicates the investment position (time). For example, $R_{i,t}$ is a return on asset i in period t in local currency. Home currency returns are denoted by $\tilde{R}^{\star}_{\star,\star}$. The subscripts bear the same meaning as for the local currency returns, whereas the superscript indicates currency hedging. For example, $\tilde{R}^h_{i,t}$ and $\tilde{R}^u_{i,t}$ represent hedged and unhedged home currency returns, respectively. To ease notation, we drop indexation of the home currency in mathematical expressions. All results derived in this paper hinge on the assumption that a home currency is prespecified.

The collection of all assets denominated in currency c held in a portfolio at time t is denominated by $\mathcal{A}_{c,t}$. The fraction of wealth directly exposed to currency c is defined as $w_{c,t} := \sum_{j \in \mathcal{A}_{c,t}} x_{j,t} \neq 0$. We introduce $R_{\mathcal{P}_c,t+1} := \sum_{j \in \mathcal{A}_{c,t}} \frac{x_{j,t}}{w_{c,t}} R_{j,t+1}$ as the return on sub-portfolio $\mathcal{P}_c$ that consists of all assets denominated in currency c. From Eq. (1.3.1) it follows that the unhedged return on sub-portfolio $\mathcal{P}_c$ expressed in an investor's home currency can be computed as

$$\tilde{R}^u_{\mathcal{P}_c,t+1} = R_{\mathcal{P}_c,t+1} + e_{c,t+1} + R_{\mathcal{P}_c,t+1} e_{c,t+1}. \tag{1.3.2}$$

Equations (1.3.1) and (1.3.2) usher in two representations for the unhedged return on portfolio $\mathcal{P}$:

$$\tilde{R}^u_{\mathcal{P},t+1} = \sum_{i=1}^{N} x_{i,t} \tilde{R}^u_{i,t+1} = \sum_{c=1}^{K+1} w_{c,t} \tilde{R}^u_{\mathcal{P}_c,t+1}, \tag{1.3.3}$$

with $\sum_{i=1}^{N} x_{i,t} = \sum_{c=1}^{K+1} w_{c,t} = 1$, for every t.[6] By substituting Eq. (1.3.2) in the second representation in Eq. (1.3.3) we obtain the following result:

$$\tilde{R}^u_{\mathcal{P},t+1} = \underbrace{\sum_{c=1}^{K+1} w_{c,t} R_{\mathcal{P}_c,t+1}}_{\text{LC sub-portfolio returns}} + \underbrace{\sum_{c=2}^{K+1} w_{c,t} e_{c,t+1}}_{\text{FX returns}} + \underbrace{\sum_{c=2}^{K+1} w_{c,t} R_{\mathcal{P}_c,t+1} e_{c,t+1}}_{\text{cross-products}}. \tag{1.3.4}$$

Having derived the expression for unhedged portfolio returns, we turn our attention to hedging via currency overlay strategies that can be implemented using forward exchange contracts. The forward exchange rate in home currency per unit of foreign currency c at time t is denoted by $F_{c,t}$.[7] We consider the forward contract with delivery date $t+1$. The forward premium, i.e., the return on the forward contract, is defined as $f_{c,t} := (F_{c,t} - S_{c,t})/S_{c,t}$. At time t, this quantity is known and represents the cost of carry.[8] We denote by $\phi_{c,t}$ the amount invested at time t in a forward exchange contract for currency c, expressed in the home currency as a fraction of total portfolio value. Similarly, $\phi_{c,t}/S_{c,t}$ represents the relative notional value of the forward contract in local currency c at time t. Hedging of foreign currency exposure can be achieved by selling a forward exchange contract (i.e., $\phi_{c,t} > 0$). By the no arbitrage principle this is analogous to shorting foreign bonds and holding domestic bonds.[9] The pay-off at time $t+1$ is $F_{c,t} - S_{c,t+1}$.

Furthermore, we extend the opportunity set by assuming that the total number of foreign currencies in which investors can trade is $M \geq K$. Therefore, an investor can enter positions in currencies to which her portfolio is not directly exposed. This optionality allows investors to expand the investment universe and implement various cross-hedging strategies. Anderson and Danthine [1981] and Eaker and Grant [1987]

[6] This setting permits short selling as the only condition is that the asset weights sum up to 1.
[7] The price of the forward contract is assumed to be zero at the inception.
[8] We note that $F_{1,t} = 1$ and $f_{1,t} = 0$ trivially hold for all t.
[9] Alternatively, this is equivalent to borrowing funds in foreign currency and lending the same amount in home currency.

showed that cross-hedging can reduce portfolio risk, albeit to a lesser extent than direct hedging. However, in the absence of direct hedging instruments, cross-hedging strategies provide a valuable alternative. Moreover, accounting for portfolio effects due to cross-hedging substantially improves efficiency and utility gains [e.g., see Gagnon et al., 1998].[10] For these reasons, cross-hedging is commonly applied in the wealth and asset management industry.

A hedged portfolio return is then equal to

$$\tilde{R}^{h}_{\mathcal{P},t+1} = \tilde{R}^{u}_{\mathcal{P},t+1} + \sum_{c=2}^{M+1} \phi_{c,t}(f_{c,t} - e_{c,t+1}), \tag{1.3.5}$$

where $f_{c,t} - e_{c,t+1} = (F_{c,t} - S_{c,t+1})/S_{c,t}$ represents the normalized payoff of a short (for $\phi_{c,t} > 0$) forward contract on currency c at time $t+1$. Since $S_{1,t} = F_{1,t} = 1$ for all t, the choice of $\phi_{1,t}$ is completely arbitrary. Nevertheless, we would like to keep the interpretation of $\phi_{c,t}$ as a fraction of total portfolio value corresponding to the notional of the forward contract on currency c. Therefore, we impose the following condition:

$$\phi_{1,t} = 1 - \sum_{c=2}^{M+1} \phi_{c,t}. \tag{1.3.6}$$

Consequently, all (net) currency exposures sum up to zero, i.e., the currency portfolio is a zero investment portfolio. For a portfolio that is directly exposed to currency c (i.e., meaning $w_{c,t} \neq 0$), the hedge ratio can be defined as $h_{c,t} \coloneqq \phi_{c,t}/w_{c,t}$. If $\phi_{c,t} = h_{c,t} = 0$, the assets denominated in currency c are unhedged. Conversely, if $\phi_{c,t} = w_{c,t}$ or $h_{c,t} = 1$, the assets are fully hedged.

Fully hedged returns can be computed by setting $\phi_{c,t} = w_{c,t}$ for $c = 1, 2, \ldots, K+1$, and $\phi_{c,t} = 0$ for $c = K+2, K+3, \ldots, M+1$ in Eq. (1.3.5). Cross-hedging with additional currencies is excluded in this case. Therefore, the fully hedged portfolio return can be expressed as

$$\tilde{R}^{fh}_{\mathcal{P},t+1} = \underbrace{\sum_{c=1}^{K+1} w_{c,t} R_{\mathcal{P}_c,t+1}}_{\text{LC returns}} + \underbrace{\sum_{c=2}^{K+1} w_{c,t} f_{c,t}}_{\text{FX forward premia}} + \underbrace{\sum_{c=2}^{K+1} w_{c,t} R_{\mathcal{P}_c,t+1} e_{c,t+1}}_{\text{cross-products}}. \tag{1.3.7}$$

Comparison of this result with the expression for unhedged portfolio return in Eq. (1.3.4) reveals that currency returns are replaced by forward premia in the case of a fully hedged portfolio. Therefore, currency hedging eliminates randomness stemming from the exchange rate fluctuations. This is due to the fact that—for any foreign currency c—the forward premium $f_{c,t}$ is known at time t. However, the exchange rate risk is not completely eliminated since the second-order cross-product term remains. The hedging positions are

[10] Cross-hedging portfolio effects hold even in the case when the hedging opportunity set is restricted to the currencies to which investor's portfolio is directly exposed.

entered at time t, and the exact currency exposure at time $t+1$ is unknown in advance.[11]

Unhedged and fully hedged portfolio returns represent two special cases of interest, which lie on the opposite sides of the currency hedging spectrum. To accommodate partial hedging of direct currency exposure and cross-hedging with additional currencies, we define the net exposure to currency c as

$$\psi_{c,t} := w_{c,t} - \phi_{c,t}. \tag{1.3.8}$$

Here, $w_{c,t}$ represents the direct currency exposure and $\phi_{c,t}$ reflects the position in a forward contract. We distinguish among four cases of currency hedging: (a) No hedging ($\psi_{c,t} = w_{c,t}$), (b) Full hedging ($\psi_{c,t} = 0$), (c) Partial hedging ($0 < \psi_{c,t} \leq w_{c,t}$), and ($d$) Over-hedging and under-hedging ($\psi_{c,t} < 0$ and $\psi_{c,t} > w_{c,t}$, respectively).

Cross-hedging strategies with the remaining $M - K$ currencies can be implemented through forward contracts as well. For $c = K+2, K+3, \ldots, M+1$, the direct currency exposure is zero (i.e., $w_{c,t} = 0$), and the net exposure is $\psi_{c,t} = -\phi_{c,t}$. Therefore, Eq. (1.3.5) can be rewritten in the following form:

$$\begin{aligned} \tilde{R}^{h}_{\mathcal{P},t+1} &= \tilde{R}^{fh}_{\mathcal{P},t+1} + \sum_{c=2}^{K+1} \psi_{c,t}(e_{c,t+1} - f_{c,t}) - \sum_{c=K+2}^{M+1} \phi_{c,t}(e_{c,t+1} - f_{c,t}) \\ &= \tilde{R}^{fh}_{\mathcal{P},t+1} + \sum_{c=2}^{M+1} \psi_{c,t}(e_{c,t+1} - f_{c,t}). \end{aligned} \tag{1.3.9}$$

Additionally, from Eq. (1.3.6) we compute the net home currency exposure:

$$\psi_{1,t} = - \sum_{c=2}^{M+1} \psi_{c,t}. \tag{1.3.10}$$

Eqs. (1.3.5) and (1.3.9) are mathematically equivalent, however they offer different economic interpretations. The former (latter) equation decomposes portfolio returns into unhedged (fully hedged) returns and a currency hedging (net currency exposure) component.

We reiterate that all results derived in this section are model-free. As such, they represent a cornerstone on top of which a modeling framework can be superposed. To illustrate this, we consider in Section 1.6.1 a particularly relevant case in point—the covered interest rate parity model.

[11] Over tactical investment horizons and in most market environments, the cross-product term is relatively small and can be neglected. Hence, nearly perfect currency hedging is typically achieved in practice. However, over longer investment horizons (e.g., beyond one year) or in the case of a sudden market crash, these terms could significantly impact the portfolio returns. To address this issue, one possible solution is to review currency hedging policies when market conditions change substantially.

1.3.1.2 Smooth Ambiguity Preferences and Robust Mean–Variance Optimization

Let us define a state space Ω that consists of all possible realizations of uncertainty. Sets of states of nature are called events ω, and the outcome space represents a σ-algebra $\mathcal{F}$, which contains random payoffs of an investor's decisions. A preference relation is defined over the mapping from Ω to $\mathcal{F}$. In a risk-only setting, all agents agree on the probability measure $\mathbb{P}$. To capture model uncertainty we introduce a space Δ of possible models $\mathcal{Q}$. We further assume that an agent's prior over all probability measures $\mathbb{Q}$ corresponding to the models in $\mathcal{Q}$ is given by μ. Then, we can compute the reduced probability $\bar{\mathbb{Q}} := \int_\Delta \mathbb{Q}\, d\mu(\mathbb{Q})$ induced by the prior μ—also called the barycenter of μ—which plays an important role in our setting.[12]

To account for model uncertainty, Maccheroni et al. [2013] (MMR) proposed a robust optimization that builds on the classical mean–variance expected utility framework and embeds smooth ambiguity model of Klibanoff et al. [2005] (KMM). The utility function is given by

$$\mathcal{U}(\ell) = \mathrm{E}_{\bar{\mathbb{Q}}}[\,\ell\,] - \frac{\lambda}{2}\mathrm{Var}_{\bar{\mathbb{Q}}}[\,\ell\,] - \frac{\theta}{2}\mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\ell\,]\,], \tag{1.3.11}$$

where ℓ is an uncertain prospect, and positive coefficients λ and θ represent the risk and ambiguity aversion, respectively. Therefore, the MMR model provides additional flexibility in a computationally tractable and economically meaningful way. Terms $\mathrm{E}_{\bar{\mathbb{Q}}}[\,\cdot\,]$ and $\mathrm{Var}_{\bar{\mathbb{Q}}}[\,\cdot\,]$ represent the reduced-probability estimators of the mean and variance obtained by combining predictions from different $\mathcal{Q}$-models. The underlying weighting scheme is captured by the agent's prior μ. The estimator $\mathrm{Var}_{\bar{\mathbb{Q}}}[\,\cdot\,]$ measures risk under the reduced probability measure, whereas the estimator $\mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\cdot\,]\,]$ quantifies the model uncertainty via the dispersion of different $\mathcal{Q}$-predictions with respect to μ. The latter can be computed as

$$\mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\ell\,]\,] = \int_\Delta \left(\int_\Omega \ell(\omega)\, d\mathbb{Q}(\omega) \right)^2 d\mu(\mathbb{Q}) - \left(\int_\Delta \left(\int_\Omega \ell(\omega)\, d\mathbb{Q}(\omega) \right) d\mu(\mathbb{Q}) \right)^2. \tag{1.3.12}$$

Higher estimates of the model uncertainty are indicative of investor's lower confidence in a single model. Conversely, if the investor's prior μ is a singleton (i.e., when represented by the Dirac function), the prospect is regarded as purely risky. In that case, all probability measures induced by the set of models $\mathcal{Q}$ are then mapped to the probability measure $\mathbb{P}$, which corresponds to the risk-only setting. Furthermore, the third term in Eq. (1.3.11) vanishes in the absence of ambiguity, and the model collapses into the classical mean–variance utility. Investors with exceptionally high or low confidence in a single model occupy the two ends of the ambiguity spectrum. Ambiguity-neutral investors fall somewhere in-between these two extremes, and their expectations are formed based on the reduced probability $\bar{\mathbb{Q}}$ instead of a single probabilistic model.

[12] Two remarks are due. First, the probability measure $\bar{\mathbb{Q}}$ is called the reduction of μ on Ω since it can be interpreted in terms of reduction of compound lotteries. Second, if $\mathrm{supp}(\mu) = \{\mathbb{Q}_1, \mathbb{Q}_2, \ldots, \mathbb{Q}_n\}$ is finite and $\mu(\mathbb{Q}_i) = \mu_i$ for $i = 1, 2, \ldots, n$, then $\bar{\mathbb{Q}}(A) = \mu_1\mathbb{Q}_1(A) + \mu_2\mathbb{Q}_2(A) + \cdots + \mu_n\mathbb{Q}_n(A)$, for any event (i.e., set of outcomes) $A \in \mathcal{F}$.

1.3.2 Optimal Currency Overlay with Ambiguity

We consider an RAA investor who wants to optimize currency exposure in her portfolio. The asset weights are assumed to be predefined and unaffected by the investor's decision regarding the currency overlay. Therefore, the investor's objective is to maximize her MMR utility function, i.e.,

$$\max_{\boldsymbol{\Psi}_t} \mathcal{U}(\tilde{R}^h_{\mathcal{P},t+1}) = \max_{\boldsymbol{\Psi}_t} \left\{ \mathrm{E}_{\bar{\mathbb{Q}}}[\,\tilde{R}^h_{\mathcal{P},t+1}\,] - \frac{\lambda}{2}\mathrm{Var}_{\bar{\mathbb{Q}}}[\,\tilde{R}^h_{\mathcal{P},t+1}\,] - \frac{\theta}{2}\mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^h_{\mathcal{P},t+1}\,]\,] \right\}. \tag{1.3.13}$$

Hedged portfolio returns $\tilde{R}^h_{\mathcal{P},t+1}$ are given in Eq. (1.3.9) and represent a risky and ambiguous prospect in our robust mean–variance optimization problem. The utility function is maximized with respect to the M-dimensional vector of net foreign currency exposures $\boldsymbol{\Psi}_t := (\psi_{2,t}, \psi_{3,t}, \ldots, \psi_{M+1,t})'$, effectively allowing for hedging of the direct currency exposures as well as cross-hedging with additional currencies.

The currency overlay optimization (1.3.13) can be cast into a quadratic programming format. Due to a linear relationship between portfolio returns and currency exposures—which is a consequence of hedging with forward contracts—the objective function amounts of a sum of linear and quadratic forms in $\boldsymbol{\Psi}_t$. By including an arbitrary set of linear constraints on the currency exposures, we obtain the following representation:

$$\begin{aligned} \boldsymbol{\Psi}^*_t = \quad & \underset{\boldsymbol{\Psi}_t}{\arg\min} \quad \left\{ \frac{1}{2}\boldsymbol{\Psi}'_t \mathbf{A} \boldsymbol{\Psi_t} + \mathbf{b}'\boldsymbol{\Psi_t} \right\} \\ & \text{subject to} \;\; \mathbf{C}\boldsymbol{\Psi_t} \preceq \mathbf{d}, \end{aligned} \tag{1.3.14}$$

where $\mathbf{A}$ is an $(M \times M)$-dimensional symmetric matrix and $\mathbf{b}$ is an $(M \times 1)$-dimensional vector given by

$$\mathbf{A} = \lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,] + \theta \mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,], \tag{1.3.15a}$$

$$\mathbf{b} = \lambda \mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1} - \mathbf{f}_t\,] + \theta \mathrm{Cov}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,], \mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,] - \mathrm{E}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]. \tag{1.3.15b}$$

An arbitrary set of L linear constrains on $\boldsymbol{\Psi}_t$ can be specified in an $(L \times M)$ matrix $\mathbf{C}$ and a $(L \times 1)$ vector $\mathbf{d}$. The operator $\preceq$ represents component-wise less-than-or-equal-to operator. Last but not least, the optimal home currency exposure can be computed using Eq. (1.3.10). The derivation of Eq. (1.3.14) and additional information regarding the quadratic program representation are provided in Appendix 1.6.2.

Linear constraints are highly relevant for practical applications because they are often based on regulatory requirements. For example, pension funds use such constraints to prevent excessive currency positions and therefore avoid unnecessary risks. However, adding restrictions comes at a cost because an analytical solution to a constrained optimization problem might not be attainable. Therefore, in a general case, the solution has to be obtained numerically through a quadratic programming procedure.

To demonstrate the inner workings of our model, we consider the unconstrained version of the optimization problem (1.3.14). The optimal net foreign currency exposure is then obtained in the closed form:

$$\begin{aligned}\boldsymbol{\Psi}_t^* = &-\Big(\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,] + \theta \mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1}-\mathbf{f}_t]\,]\Big)^{-1} \\ &\cdot\Big(\lambda \mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1}-\mathbf{f}_t\,] + \theta \mathrm{Cov}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,], \mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1}-\mathbf{f}_t]\,] - \mathrm{E}_{\bar{\mathbb{Q}}}[\mathbf{e}_{t+1}-\mathbf{f}_t]\Big),\end{aligned} \tag{1.3.16}$$

where the first bracket is an $(M \times M)$-dimensional positive definite and invertible matrix. Eq. (1.3.16) is the key theoretical results in our paper, and it represents a generalization of the optimal currency exposure derived in Campbell et al. [2010] in the absence of ambiguity. Furthermore, we emphasize that the optimal currency exposure can be decomposed into two components which have clear and precise economic meaning: (a) The ambiguity-adjusted hedging demand ($\boldsymbol{\Xi}_t^*$), and (b) The ambiguity-adjusted market price of currency risk ($\boldsymbol{\Lambda}_t^*$), which are defined as follows:

$$\begin{aligned}\boldsymbol{\Xi}_t^* := &-\Big(\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,] + \theta \mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1}-\mathbf{f}_t]\,]\Big)^{-1} \\ &\cdot\Big(\lambda \mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1}-\mathbf{f}_t\,] + \theta \mathrm{Cov}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,], \mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]\Big),\end{aligned} \tag{1.3.17a}$$

$$\boldsymbol{\Lambda}_t^* := \Big(\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,] + \theta \mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]\Big)^{-1} \cdot \mathrm{E}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]. \tag{1.3.17b}$$

To gain further insights we examine several special cases of interest and analyze the implications for the optimal currency hedging policy.

1.3.2.1 Infinitely Risk-Averse Investor

The objective of an infinitely risk-averse investor (i.e., $\lambda \to \infty$) is to minimize the variance (volatility) of portfolio returns. The optimal net foreign currency exposure for such an investor is

$$\boldsymbol{\Psi}_{t,\mathrm{risk}}^* := \lim_{\lambda\to\infty} \boldsymbol{\Psi}_t^* = -\mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]^{-1} \cdot \mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1}-\mathbf{f}_t\,]. \tag{1.3.18}$$

The optimal currency exposure is reduced to the hedging demand component $\boldsymbol{\Xi}_t^*$ defined in Eq. (1.3.17a), which is driven by the correlation between the fully hedged portfolio returns and excess currency returns. An infinitely risk-averse investor prefers long (short) positions in the currencies which tend to appreciate (depreciate) when the fully hedged portfolio loses value.[13] If the correlation is zero, it is optimal to fully hedge the currency risk. Intuitively, an infinitely risk-averse investor is concerned only with risk minimization and does not account for an expected compensation for the additional (currency) risk in her portfolio. Hence, she is better off if the currency exposure is fully hedged.

Although risk aversion dominates in this case, the above equation does not exclude ambiguity aversion,

[13] If the correlation is sufficiently negative (positive), the investor can reduce portfolio risk also by under-hedging (over-hedging). This can be accomplished by holding a long (short) position in excess of the portfolio weight $w_{c,t}$.

which is incorporated via the reduced probability $\bar{\mathbb{Q}}$. In the absence of model uncertainty, the reduced probability $\bar{\mathbb{Q}}$ collapses to the probability $\mathbb{P}$, and we obtain the classical minimum–variance result.

1.3.2.2 Infinitely Ambiguity-Averse Investor

The optimal net foreign currency exposure for an infinitely ambiguity-averse investor (i.e., $\theta \to \infty$) is given by

$$\mathbf{\Psi}^*_{t,\text{amb}} := \lim_{\theta\to\infty} \mathbf{\Psi}^*_t = -\text{Var}_\mu[\,\text{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,]^{-1} \cdot \text{Cov}_\mu[\,\text{E}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,], \text{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t]\,], \tag{1.3.19}$$

where matrix $\text{Var}_\mu(\text{E}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t])$ has to be positive definite to ensure existence of its inverse. The parameter and model uncertainty is captured by μ, the investor's prior probability over the space of possible models. An infinitely ambiguity-averse investor prefers long (short) positions in the currencies which exhibit a negative (positive) uncertainty-adjusted correlation with the fully hedged portfolio returns. The intuition is very similar to the case of an infinitely risk-averse investor, and the optimal currency exposure is again exclusively driven by the hedging demand. The only difference is that the model uncertainty (instead of the risk) is the first-order effect for an infinitely ambiguity-averse investor.

1.3.2.3 Ambiguity-Neutral Investor

We consider now an ambiguity-neutral agent (i.e., the case $\theta \to 0$). This is equivalent to the mean–variance setting where all uncertainty is captured by $\bar{\mathbb{Q}}$. The optimal net foreign currency exposure can be computed as

$$\begin{aligned}\mathbf{\Psi}^*_{t,\text{mv}} := \lim_{\theta\to 0} \mathbf{\Psi}^*_t &= -\text{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]^{-1} \cdot \Big(\text{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1} - \mathbf{f}_t\,] - \frac{1}{\lambda}\text{E}_{\bar{\mathbb{Q}}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\Big) \\ &= \mathbf{\Psi}^*_{t,\text{risk}} + \frac{1}{\lambda}\text{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]^{-1} \cdot \text{E}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,].\end{aligned} \tag{1.3.20}$$

The optimal mean–variance currency exposure can be expressed as a sum of the optimal minimum–variance currency portfolio Eq. (1.3.18) and the market price of currency risk, adjusted for the risk aversion λ. The market price of currency risk represents the trade-off between the expected excess return and risk associated with the respective currency exposure. Its effect on the optimal mean–variance currency exposure vanishes as $\lambda \to \infty$. Moreover, the optimal mean–variance currency exposure can be interpreted as an unambiguous currency overlay strategy. While the minimum variance strategy only seeks to reduce risk, the mean–variance currency exposure introduces an additional component that aims to generate an excess return (i.e., the "alpha"), hence representing a speculative currency demand.[14]

[14] We reiterate that, in our framework, the decision regarding the optimal currency risk-return trade-off is independent of the asset allocation policy for the underlying portfolio. Moreover, we stress that currency overlay strategies do not represent direct investments. They are merely risk management strategies—in our case implemented with forward contracts—that do not require any portfolio rebalancing or additional investments.

The general expression for the optimal currency exposure in Eq. (1.3.16) can be written in the form:

$$\mathbf{\Psi}_t^* = \mathbf{\Psi}_{t,\mathrm{mv}}^* + \big(\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,] + \theta \mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\,]\big)^{-1} \theta \mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\,] \cdot (\mathbf{\Psi}_{t,\mathrm{amb}}^* - \mathbf{\Psi}_{t,\mathrm{mv}}^*), \quad (1.3.21)$$

where we used the results derived in Eqs. (1.3.19) and (1.3.20), and assumed the positive definiteness of $\mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\,]$. The proof is provided in Appendix 1.6.3. This result represents an explicit correction of the optimal mean–variance currency exposure for an RAA investor, relative to a purely risk-averse investor. The term $(\mathbf{\Psi}_{t,\mathrm{amb}}^* - \mathbf{\Psi}_{t,\mathrm{mv}}^*)$ measures the difference between the optimal currency exposures for an infinitely ambiguity-averse and an unambiguous investor. The multiplier of this term accounts for the size and direction of the parameter and model uncertainty correction and vanishes for $\theta \to 0$. Therefore, this multiplier can be interpreted as the share of the total excess currency return variance which can be attributed to ambiguity (up to the scaling factors λ and θ).[15]

1.3.3 Ambiguity and the Home Currency Bias

In this section, we establish a link between the ambiguity and home currency bias. This is a novel result, and it represents one of the key theoretical contributions of our paper. It naturally arises in our robust mean–variance setting with parameter and model uncertainly. Following Maccheroni et al. [2013], we distinguish between purely risky and ambiguous assets.[16] International asset allocation problems with currency overlay strategies are particularly well suited for applications of our modeling framework. More specifically, in our framework, cash represents a risk-free asset, government bonds and equities viewed as purely risky, whereas currencies are ambiguous assets.

We consider an investor who is fully invested in a portfolio comprising only home (i.e., base-currency denominated) assets. In the literature, such a portfolio is commonly referred to as home-country-biased. Next, we assume that the investor wants to optimize her foreign currency exposure, given a set of feasible currency overlay strategies across the risk–ambiguity–return spectrum. Based on the asset classification introduced above, the underlying portfolio is treated as purely risky and the foreign currency positions are assumed to be ambiguous. Thus, $\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}_{\mathcal{P},t+1}^{fh}\,]$ is constant for all models $\mathbb{Q} \in \mathcal{Q}$ and $\mathrm{Cov}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}_{\mathcal{P},t+1}^{fh}\,], \mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\,] = 0$. Using Eq. (1.3.16), we compute the optimal currency exposure for a home-country-biased investor:

$$\begin{aligned} \mathbf{\Psi}_{t,\mathrm{hb}}^* = -\Big(&\mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,] + \frac{\theta}{\lambda}\mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\,]\Big)^{-1} \\ &\cdot\Big(\mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}_{\mathcal{P},t+1}^{fh}, \mathbf{e}_{t+1} - \mathbf{f}_t\,] - \frac{1}{\lambda}\mathrm{E}_{\bar{\mathbb{Q}}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\Big). \end{aligned} \quad (1.3.22)$$

To provide an economic intuition regarding the derived result, we differentiate between two cases. If an

[15] Alternatively, Eq. (1.3.21) can be interpreted as a convex combination of the two limiting cases $\mathbf{\Psi}_{t,\mathrm{mv}}^*$ and $\mathbf{\Psi}_{t,\mathrm{amb}}^*$.

[16] By definition, ambiguous assets are always both risky and ambiguous. Hence, we simply call them ambiguous assets.

investor is infinitely risk-averse (i.e., $\lambda \to \infty$), the optimal currency exposure converges to the minimum variance case given in Eq. (1.3.18). This case is closely examined in Section 1.3.2.1. In the case of an infinitely ambiguity-averse investor (i.e., $\theta \to \infty$), the optimal currency exposure converges to zero for all foreign currencies. Hence, the currency risk is fully hedged, and the investor is maximally biased towards the home/base currency. Using a security-level data set, Maggiori et al. [2020] provided empirical evidence that investors exhibit home currency bias in that they disproportionately hold securities denominated in their domestic currency. Our theoretical results show that high ambiguity aversion—in a setting where currencies are treated as ambiguous assets—can explain why investor holdings are biased toward their base currencies (i.e., the puzzle of insufficient currency diversification is driven by investor's elevated ambiguity aversion). Our findings are in line with Pflug et al. [2012], who theoretically—using a different modeling framework—showed that a uniform investment strategy is optimal when an investor exhibits high model uncertainty.

1.3.4 Generalized Ridge Regression Representation

1.3.4.1 Sample Efficiency

Portfolio optimization inputs are often estimated with errors. Our framework addresses this issue by accounting for the impact of parameter and model uncertainty on the estimation of expected returns. This can be of interest to asset managers and other practitioners seeking to determine a forward-looking optimal currency exposure in their portfolios. Furthermore, academic researchers often investigate the role of sampling error when constructing ex-post (i.e., in-sample) efficient portfolio weights or currency exposures. For example, Britten-Jones [1999] showed that the ordinary least squares (OLS) regression of a constant onto excess asset returns—without an intercept term—results in an estimated coefficient vector that represents a set of risky-asset-only portfolio weights for a sample-efficient tangency portfolio. A similar idea is applied in Campbell et al. [2010] to derive the optimal currency exposure. In this section, we demonstrate how our setting—which incorporates ambiguity—compares to the extant literature.

Our empirical loss function is defined as a negative of robust mean–variance utility function, i.e., $\mathcal{L}_{\mathbb{H}}(\tilde{R}^h_{\mathcal{P},t+1}) := -\mathcal{U}(\tilde{R}^h_{\mathcal{P},t+1})$. Therefore, the minimization of the loss function is equivalent to the maximization of the robust mean–variance utility with respect to the in-sample currency exposure

$$\widehat{\boldsymbol{\Psi}}^*_{t,\mathbb{H}} := \arg\min_{\boldsymbol{\Psi}_t} \mathcal{L}_{\mathbb{H}}(\tilde{R}^h_{\mathcal{P},t+1}) = \arg\min_{\boldsymbol{\Psi}_t} \left\{ \frac{\lambda}{2} \widehat{\mathrm{Var}}_{\mathbb{H}}[\, \tilde{R}^h_{\mathcal{P},t+1} \,] + \frac{\theta}{2} \mathrm{Var}_{\mu}[\, \mathrm{E}_{\mathbb{Q}}[\, \tilde{R}^h_{\mathcal{P},t+1} \,]\,] - \widehat{\mathrm{E}}_{\mathbb{H}}[\, \tilde{R}^h_{\mathcal{P},t+1} \,] \right\}, \quad (1.3.23)$$

where $\widehat{\mathrm{Var}}_{\mathbb{H}}[\, \tilde{R}^h_{\mathcal{P},t+1} \,]$ and $\widehat{\mathrm{E}}_{\mathbb{H}}[\, \tilde{R}^h_{\mathcal{P},t+1} \,]$ are the sample covariance matrix and the vector of sample averages of hedged portfolio returns, respectively. The index $\mathbb{Q}$ denotes probability measures in $\mathcal{Q}$, which capture the parameter and model uncertainty (as perceived by the investor during the sample period).

When studying the in-sample optimality of foreign currency exposures, the vector of average hedged portfolio returns is typically not included in calculations. There are, at least, two reasons for that. First, the sample averages are noisy estimators. Second, anchoring the sample-efficient currency overlay strategies to average realized returns (i.e., capturing the in-sample "alpha") has limited practical significance.[17] Therefore, in the remainder of this section, we assume $\widehat{\mathrm{E}}_{\mathbb{H}}[\, \tilde{R}^{h}_{\mathcal{P},t+1} \,] = 0$, and investigate the sample optimality from the (model) risk management perspective only. Under this assumption, the optimal in-sample currency overlay strategy is equal to the optimal in-sample hedging demand, i.e., $\widehat{\boldsymbol{\Psi}}^{*}_{t,\mathbb{H}} = \widehat{\boldsymbol{\Xi}}^{*}_{t,\mathbb{H}}$. For completeness, the in-sample ambiguity-adjusted market price of currency risk $\widehat{\boldsymbol{\Lambda}}^{*}_{t,\mathbb{H}}$ is studied in Appendix 1.6.4.

We consider the demeaned historical returns which come from the historical measure $\mathbb{H}$. We denote by $\mathbf{X}$ and $\mathbf{y}$ the $(T \times M)$ matrix of demeaned historical excess currency returns $\mathbf{e}_{t+1} - \mathbf{f}_t$, and the $(T \times 1)$ vector of demeaned historical fully hedged portfolio return $\tilde{R}^{fh}_{\mathcal{P},t+1}$, respectively, where T is the total number of observations. Thus, the sample covariance matrix of excess currency returns and the vector of covariances between the fully hedged portfolio and excess currency returns are given by

$$\widehat{\mathrm{Var}}_{\mathbb{H}}[\, \mathbf{e}_{t+1} - \mathbf{f}_t \,] = \frac{1}{T} \mathbf{X}' \, \mathbf{X},$$
$$\widehat{\mathrm{Cov}}_{\mathbb{H}}[\, \tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1} - \mathbf{f}_t \,] = \frac{1}{T} \mathbf{X}' \, \mathbf{y}.$$

Next, we define $\mathbf{W} := \frac{\lambda}{T}\mathbf{I}$, where $\mathbf{I}$ is a $(T \times T)$ identity matrix, $\mathbf{Z} := \theta \mathrm{Var}_{\mu}[\, \mathrm{E}_{\mathbb{Q}}[\tilde{R}^{h}_{\mathcal{P},t+1}] \,]$, and $\mathbf{z_0} := -\boldsymbol{\Psi}^{*}_{t,\mathrm{amb}}$.[18] Consequently, the loss function can be expressed as

$$\mathcal{L}_{\mathbb{H}}(\tilde{R}^{h}_{\mathcal{P},t+1}) = \frac{1}{2} \boldsymbol{\Psi}_t{}' \, \mathbf{X}' \, \mathbf{W} \mathbf{X} \boldsymbol{\Psi}_t + \boldsymbol{\Psi}_t{}' \, \mathbf{X}' \, \mathbf{W} \mathbf{y} + \frac{1}{2} \boldsymbol{\Psi}_t{}' \, \mathbf{Z} \boldsymbol{\Psi}_t + \boldsymbol{\Psi}_t{}' \, \mathbf{Z} \mathbf{z_0} + \text{“remaining terms”}, \tag{1.3.24}$$

where we explicitly write the terms which depend on $\boldsymbol{\Psi}_t$, and we capture the terms which do not affect the optimization under "remaining terms". Using Eq. (1.3.24), the minimization problem can be rewritten as

$$\begin{aligned}
\widehat{\boldsymbol{\Psi}}^{*}_{t,\mathbb{H}} &:= \arg\min_{\boldsymbol{\Psi}_t} \; \mathcal{L}_{\mathbb{H}}(\tilde{R}^{h}_{\mathcal{P},t+1}) \\
&= \arg\min_{\boldsymbol{\Psi}_t} \left\{ \frac{1}{2} \Big(\mathbf{y} + \mathbf{X} \boldsymbol{\Psi}_t \Big)' \, \mathbf{W} \Big(\mathbf{y} + \mathbf{X} \boldsymbol{\Psi}_t \Big) + \frac{1}{2} (\boldsymbol{\Psi}_t + \mathbf{z_0})' \, \mathbf{Z} (\boldsymbol{\Psi}_t + \mathbf{z_0}) + \text{“remaining terms”} \right\} \\
&= \arg\min_{\boldsymbol{\Psi}_t} \left\{ \big\| \mathbf{y} - \mathbf{X}(-\boldsymbol{\Psi}_t) \big\|^2_{\mathbf{W}} + \big\| (-\boldsymbol{\Psi}_t) - (-\boldsymbol{\Psi}^{*}_{t,\mathrm{amb}}) \big\|^2_{\mathbf{Z}} \right\},
\end{aligned} \tag{1.3.25}$$

where $\big\| \boldsymbol{\Psi}_t \big\|^2_{\mathbf{D}} = \boldsymbol{\Psi}_t' \, \mathbf{D} \boldsymbol{\Psi}_t$ stands for the weighted L^2-norm given a positive definite matrix $\mathbf{D}$.

This is an important theoretical result which shows that the optimal ex-post currency exposure for an

[17] Practitioners are mostly interested in forward-looking estimates, as opposed to the historical estimates. For this reason we conduct an out-of-sample backtest in Section 1.4.3.

[18] We assume that the inverse of $\mathrm{Var}_{\mu}[\, \mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t] \,]$ exists.

RAA agent—without the consideration of the in-sample currency "alpha"—can be found by a generalized ridge regression of the demeaned hedged portfolio returns on the demeaned excess currency returns, with shrinkage towards the infinitely ambiguity-averse optimal exposure driven by the degree and structure of ambiguity as captured in $\mathbf{Z}$. The weighting matrix is given by $\mathbf{W} = \frac{\lambda}{T}\mathbf{I}$ and accounts for agent's risk aversion λ. The positive-definite matrix driving the generalized penalization is given by $\mathbf{Z} = \theta \mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{h}_{\mathcal{P},t+1}\,]\,]$ and accounts for model uncertainty. The shrinkage target is given by $-\mathbf{\Psi}^{*}_{t,\mathrm{amb}}$, which is a negative of optimal currency exposure for an infinitely ambiguity-averse agent (1.3.19).

Two technical remarks are due. First, we note that if a constant term was included in the regression, i.e., an unambiguous prospect, the generalized penalty matrix $\mathbf{Z}$ would not have been positive definite. In fact, it would not generate a proper norm, and the shrinkage target would not exist. Hastie et al. [2009] argue that an intercept term should always be left out of the ridge penalty term. Otherwise, the penalization procedure would be dependent on the origin chosen for the predictor variable. This is precisely the reason why we use demeaned returns and exclude the intercept term from the regression. Second, we stress that the generalized ridge regression representation is a consequence of (a) The linear relationship between the portfolio returns and the underlying currency exposure, and (b) The choice of the robust mean–variance preferences. In (a), the linearity is inherited from forward contracts which are used for the hedging of currency risk in our framework. In (b), the ambiguity is measured in a quadratic manner, i.e., see Eq. (1.3.12), which opens the door to a generalized ridge regression representation. Hence, the differentiability and smoothness are sustained also in the presence of ambiguity, and a unique closed-form solution exists.

Last but not least, we note that in the limiting case of no model uncertainty (i.e., $\theta \to 0$), the penalty term vanishes. The optimal in-sample currency exposure can be estimated (up to the change of sign) via an OLS regression of the demeaned hedged portfolio returns on the demeaned excess currency returns, i.e.,

$$\widehat{\mathbf{\Psi}}^{*}_{t,\mathbb{H}}\Big|_{\theta \to 0} = \underset{\mathbf{\Psi}_t}{\arg\min} \left\{ \left\| \mathbf{y} - \mathbf{X}(-\mathbf{\Psi}_t) \right\|^{2}_{\mathbf{I}} \right\}. \tag{1.3.26}$$

This is in line with Campbell et al. [2010] and Schmittmann [2010] where the vector of optimal currency exposure is obtained as the negative of the slopes (without an intercept) of a multiple regression of the excess portfolio return on a constant and the vector of currency excess returns. Moreover, this is equivalent to demeaning the returns and regressing without a constant term. Hence, our paper nests the OLS approaches presented in the extant literature as a special case of no model uncertainty.

1.3.4.2 Geometric Interpretation

The regression (1.3.25) recovers the sample-efficient currency exposure. Furthermore, it provides a useful geometric interpretation of the optimal currency overlay strategies, which is discussed in this section.

We start with the risk-only case, which corresponds to the ordinary least squares linear regression. The dependent variable $\mathbf{y}$ represents the return on a fully hedged portfolio that has no exposure to currency risk. The independent variables $\mathbf{X}$ involve only the excess returns on currencies. The coefficients $\boldsymbol{\Psi}_t$ represent the currency exposures (i.e., the weights of a zero value currency portfolio). Finally, the regression residuals show the deviation of the pure currency portfolio $\mathbf{X}\boldsymbol{\Psi}_t$ from the fully hedged portfolio returns $\mathbf{y}$. The estimated optimal in-sample currency weights represent a currency overlay strategy that is the closest to the fully hedged portfolio returns in terms of the least-squares distance. The closer the two vectors, the larger amount of risk can be reduced via the currency position, where the optimal risk reductive currency exposure corresponds to the negative of the estimated regression coefficient $-\boldsymbol{\Psi}_t$.

In the presence of model uncertainty, the generalized penalty term corresponds to the utility loss arising from model uncertainty. The OLS regression has a geometric interpretation as an orthogonal projection of the hedged portfolio returns onto the space spanned by the excess currency returns. In the case of the generalized ridge regression, the estimated coefficients (i.e., the optimal exposures) still lie in the span of the predictors (i.e., the excess currency returns), while a shrinkage of the OLS-based solution is induced. The regularization magnitude θ and the structure of model uncertainty $\mathrm{Var}_\mu[\,\mathrm{E}_\mathbb{Q}[\,\tilde{R}^h_{\mathcal{P},t+1}\,]\,]$ control the distance and direction of the shrinkage towards the target $-\boldsymbol{\Psi}^*_{t,\mathrm{amb}}$. The generalized penalty is a quadratic form. Geometrically, this means that there is a non-zero-centered, ellipsoidal parameter constraint. Figure 1.1 provides an illustration of this effect. The red point corresponds to the two-dimensional vector of optimal currency exposure obtained via a non-regularized weighted least squares regression, whereas the blue point represents the shrinkage target. The black ellipsoid contains all sets of currency exposures that are feasible. For smaller (larger) regularization coefficient θ, these sets becomes larger (smaller) due to relaxation (tightening) of the constraint. The shape of the ellipsoid is determined by the model uncertainty matrix $\mathrm{Var}_\mu[\,\mathrm{E}_\mathbb{Q}[\,\tilde{R}^h_{\mathcal{P},t+1}\,]\,]$. The optimal solution is given as the tangent of the contour sets corresponding to the weighted least squares residual sums of squares and the set of feasible solutions.[19]

The sum of squared residuals in the weighted least squares regression and the ridge penalty are both convex in $\boldsymbol{\Psi}_t$. Therefore, a unique minimizer of the penalized sum of squares exists. The solution to Eq. (1.3.25) yields a sample-efficient optimal currency exposure for an RAA investor:

$$\begin{aligned}\widehat{\boldsymbol{\Psi}}^*_{t,\mathbb{H}} &= -\big(\mathbf{X}'\,\mathbf{W}\mathbf{X} + \mathbf{Z}\big)^{-1}\big(\mathbf{X}'\,\mathbf{W}\mathbf{y} + \mathbf{Z}\mathbf{z_0}\big)\\ &= \widehat{\boldsymbol{\Psi}}^*_{t,\mathbb{H},\mathrm{risk}} + \big(\mathbf{X}'\,\mathbf{W}\mathbf{X} + \mathbf{Z}\big)^{-1}\,\mathbf{Z}\,\big(\boldsymbol{\Psi}^*_{t,\mathrm{amb}} - \widehat{\boldsymbol{\Psi}}^*_{t,\mathbb{H},\mathrm{risk}}\big),\end{aligned} \tag{1.3.27}$$

where $\widehat{\boldsymbol{\Psi}}^*_{t,\mathbb{H},\mathrm{risk}} = -(\mathbf{X}'\,\mathbf{X})^{-1}\mathbf{X}'\,\mathbf{y}$ is the sample-efficient currency exposure in the absence of ambiguity. The

[19] In the classical ridge regression, the penalty is given by a scaled identity matrix, and the target is set at the origin. This implies that the feasible set is represented by a sphere centered at the origin. In the generalized ridge regression, the sphere becomes an ellipsoid centered at the particular shrinkage target.

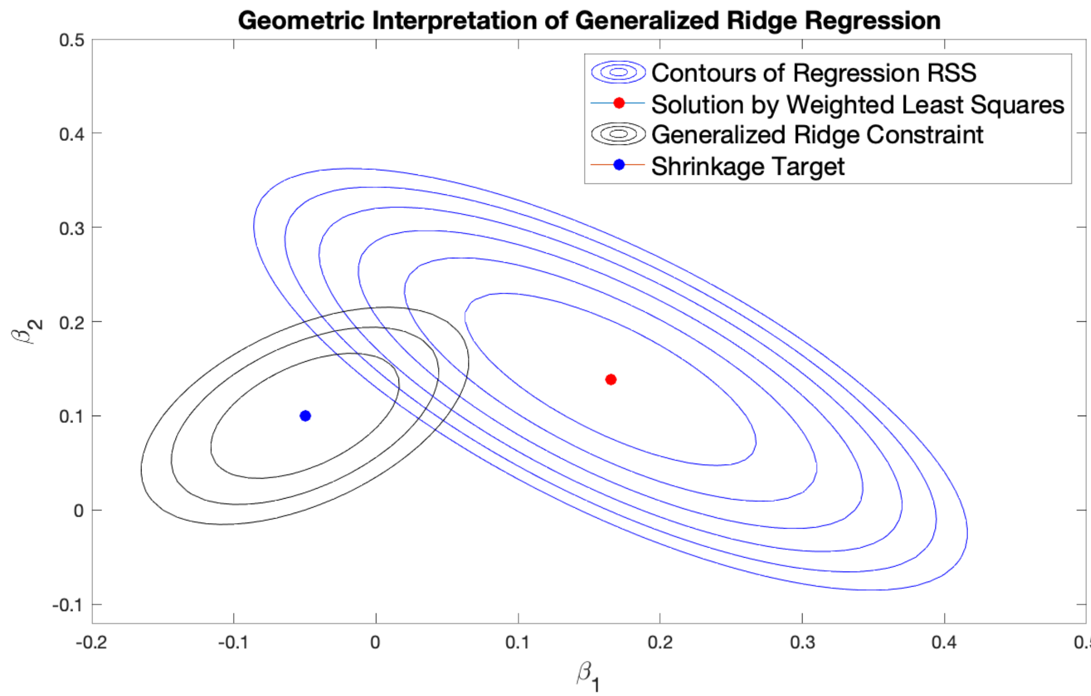

Figure 1.1: A geometric interpretation of a generalized ridge regression in two dimensions. The red point corresponds to the solution obtained by the weighted least squares regression. The contours centered at the red point represent the sums of squared residuals in the regression. The blue point is the shrinkage target and the ellipsoid around that point represents the contours of feasible points given a particular choice of the penalization parameter. The set of feasible points shrinks towards the target for larger values of the penalization parameter. The shape of the ellipsoid is determined by the positive definite matrix in the weighted L^2-norm. The optimal solution of this optimization problem is given as the tangent of the weighted least squares contour sets and the set of feasible points.

proof of this result is provided in Appendix 1.6.5. Eq. (1.3.27) represents the in-sample equivalent of the general solution given in Eqs. (1.3.16) and (1.3.21). The correction term arising from model uncertainty vanishes when $\theta \to 0$, or when $\mathbf{\Psi}^*_{t,\mathrm{amb}} = \widehat{\mathbf{\Psi}}^*_{t,\mathbb{H},\mathrm{risk}}$. In the latter case, the shrinkage target matches the OLS solution, which geometrically corresponds to an overlap of the red and the blue point. The risk-only sample-efficient solution is then attainable for any θ and the model uncertainty does not affect the OLS solution. Furthermore, assuming that the cash is risk-free, government bonds and equities are purely risky, and currencies are ambiguous assets, such as in Section 1.3.3, the ridge target (i.e., the blue point) becomes fixed in the origin. In the limit $\theta \to \infty$, the set of attainable currency exposures is shrunk to a single point in the origin, implying the optimality of home currency bias for the extreme ambiguity averse investor.

The penalty term in Eq. (1.3.25) represents the utility loss from the model uncertainty. Therefore, the risk-minimizing sample-efficient currency exposure $\widehat{\mathbf{\Psi}}^*_{t,\mathbb{H},\mathrm{risk}}$ can be thought of as the "first-best" solution in the absence of ambiguity. Once the model uncertainty is accounted for via the ambiguity aversion θ, the currency overlay strategy $\widehat{\mathbf{\Psi}}^*_{t,\mathbb{H}}$ becomes the "second-best" solution. The adjustment of the optimal currency exposure from the "first-best" to the "second-best" is illustrated in Fig. 1.1. We showed that there is a direct link, an equivalence, between a robust mean–variance utility representation and the solution to such utility maximization problem obtained with a generalized ridge regression. In such a way, we formally associated the areas of financial economics (i.e., a problem of asset/currency allocation) with statistical learning (i.e., arising from regularization).

1.4 Empirical Analysis

In this section, we turn our attention to the empirical analysis. First, we present the summary data statistics. Second, we study the effects of the ambiguity aversion on the optimal in-sample currency exposure. Moreover, we investigate if the ambiguity aversion can be a driver of currency under-diversification, known as the home currency bias. Finally, we design a currency hedging strategy and present the out-of-sample backtesting results. Our goal is to investigate the impact of accounting for the ambiguity aversion in currency hedging decisions on the performance of international portfolios out of sample.

1.4.1 Data

Following Campbell et al. [2010], we assembled a data set covering seven major developed economies: Australia, Canada, Switzerland, Eurozone, the United Kingdom, Japan, and the United States. We have collected daily time series of spot and forward exchange rates, short-term interest rates, and broad equity and government bonds total return indices from January 1999 (i.e., since the dawn of the euro) until December 2019. The data is sourced from Thomson Reuters Datastream and Bloomberg. This sample allows for an investigation of possible shifts in the optimal currency exposures during or after the major financial and economic crises, e.g., the Dot-Com Bubble 2000–2002, the Global Financial Crisis 2008–2009, and the European Sovereign Debt Crisis 2009–2011.

Table 1.1 reports the annualized mean return and volatility of major equity and government bond indices in local currencies, as well as currency returns for the Australian dollar (AUD), Canadian dollar (CAD), Swiss franc (CHF), euro (EUR), pound sterling (GBP), and Japanese yen (JPY) with respect to the US dollar (USD) from January 1999 until December 2019. Table 1.2 reports the correlations of equity returns (Panel A), government bond returns (Panel B), and currency returns averaged across all possible base currencies (Panel C). Across all three panels, we observe elevated correlations between the US and Canada, as well as among Swiss, Eurozone, and UK markets, reflecting their respective economic integration due to geographic proximity. All correlation coefficients are positive but relatively limited, hence indicating international diversification benefits during the observation period.

Summary Statistics

Panel A: Equities	Australia	Canada	Switzerland	Eurozone	UK	Japan	US
Return (%)	8.85	8.75	5.69	7.00	6.94	6.15	7.66
Volatility (%)	15.04	16.22	16.58	21.01	17.63	20.73	18.85
Panel B: Bonds	Australia	Canada	Switzerland	Eurozone	UK	Japan	US
Return (%)	4.93	3.78	2.05	3.55	4.04	1.07	3.79
Volatility (%)	3.26	2.57	1.73	2.44	2.55	1.30	3.47
Panel C: Currencies	AUD	CAD	CHF	EUR	GBP	JPY	USD
Return (%)	1.71	1.17	2.19	0.44	−0.72	0.61	–
Volatility (%)	12.53	9.00	10.79	9.71	9.26	10.23	–

Table 1.1: A summary table comprising the annualized means and the standard deviations of equity, bond, and currency returns for a fully hedged investor based in the US dollar over the period from January 1999 until December 2019.

1.4.2 In-Sample Analysis: The Impact of Ambiguity Aversion

We focus first on the in-sample analysis of the impact of the ambiguity aversion on the optimal currency exposure. Therefore, the reduced probability measure $\bar{\mathbb{Q}}$ is given by the historical measure $\mathbb{H}$. Moreover, we assume that the uncovered interest rate parity holds, i.e., the forward rate is an unbiased predictor of the future spot rate.[20] In connection to Section 1.3.3, we also assume that the cash is risk-free, government bonds and equities are purely risky, and currencies are ambiguous assets. This economic assumption centers the ridge target in the origin. Based on these assumptions, the optimal currency exposure can be derived from Eq. (1.3.22) as:[21]

$$\widehat{\mathbf{\Psi}}^{*}_{t,\mathbb{H}} = -\Big(\widehat{\mathrm{Var}}_{\mathbb{H}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,] + \frac{\theta}{\lambda}\widehat{\mathrm{Var}}_{\mu}[\,\widehat{\mathrm{E}}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\,]\Big)^{-1} \cdot \widehat{\mathrm{Cov}}_{\mathbb{H}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1} - \mathbf{f}_t\,]. \tag{1.4.1}$$

To focus on the impact of the ambiguity aversion on the sample-efficient currency exposure, we consider a setting which allows us to perform an in-sample analysis without explicitly specifying various predictive models. Without the loss of generality, we assume independence of investor's predictive models. In particular, we impose the condition $\widehat{\mathrm{Var}}_{\mu}[\,\widehat{\mathrm{E}}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,] = \frac{1}{T^2}\mathbf{I}$, where T is the number of observations in the sample and $\mathbf{I}$ is the identity matrix.[22] The scaling parameter $1/T^2$ is introduced so that the covariance matrices calculated under $\mathbb{H}$ and μ have the same order of magnitude. Cosequently, our model is transformed into an ordinary (i.e., non-generalized) ridge regression. Using the notation from previous section, the expression in

[20] We note that, in contrast to the covered interest rate parity, the uncovered interest rate parity is not an arbitrage relationship. It is an assumption based on equilibrium reasoning that may or may not hold.

[21] In effect, we have substituted $\bar{\mathbb{Q}}$ by $\mathbb{H}$, and the expectation of the normalized payoff of a forward contract has vanished because of the uncovered interest rate parity.

[22] This assumption is relaxed in the next section where we consider out-of-sample performance and introduce several popular foreign exchange forecasting models.

Cross-country Return Correlations

Panel A: Equities	Australia	Canada	Switzerland	Eurozone	UK	Japan	US
Australia	1.00						
Canada	0.22	1.00					
Switzerland	0.33	0.47	1.00				
Eurozone	0.33	0.53	0.73	1.00			
UK	0.35	0.54	0.81	0.81	1.00		
Japan	0.55	0.21	0.31	0.27	0.30	1.00	
US	0.14	0.73	0.50	0.54	0.53	0.12	1.00
Panel B: Bonds	Australia	Canada	Switzerland	Eurozone	UK	Japan	US
Australia	1.00						
Canada	0.15	1.00					
Switzerland	0.27	0.40	1.00				
Eurozone	0.12	0.40	0.49	1.00			
UK	0.25	0.49	0.53	0.55	1.00		
Japan	0.22	0.09	0.15	0.10	0.13	1.00	
US	0.08	0.77	0.34	0.38	0.45	0.08	1.00
Panel C: Currencies	AUD	CAD	CHF	EUR	GBP	JPY	USD
AUD	1.00						
CAD	0.53	1.00					
CHF	0.25	0.30	1.00				
EUR	0.43	0.44	0.68	1.00			
GBP	0.41	0.47	0.41	0.57	1.00		
JPY	0.09	0.24	0.41	0.35	0.28	1.00	
USD	0.29	0.54	0.37	0.44	0.50	0.53	1.00

Table 1.2: Cross-country correlations for equity, bond, and currency returns from January 1999 until December 2019. Panels A and B report the correlations of local-currency equity and bond returns, respectively. Panel C reports the correlations of currency returns, averaged across all possible base currencies.

Eq. (1.4.1) can be rewritten as:

$$\widehat{\boldsymbol{\Psi}}^{*}_{t,\mathbb{H}} = -(\mathbf{X}'\mathbf{X} + q\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}, \tag{1.4.2}$$

where the ridge regularization parameter is given by $q = \frac{\theta}{\lambda T}$.

Having set the stage for our in-sample analysis, we calculate the optimal currency exposure as a function of the risk and ambiguity aversion parameters λ and θ. To remain tractable, we present results for a portfolio exposed only to a single foreign currency. Similar conclusions can be drawn in a multi-currency setting. The upper (lower) plots in Figure 1.2 provide the optimal currency exposure in CHF and USD (CHF and EUR) for a EUR-based (USD-based) RAA investors.

First, we note that in the case of an extremely high risk aversion (i.e., when $\lambda \to \infty$), our optimization is reduced to the minimum variance problem, which can be solved via an ordinary least squares regression. This is a standard setting that is often considered in the literature. However, for finite values of the risk-aversion

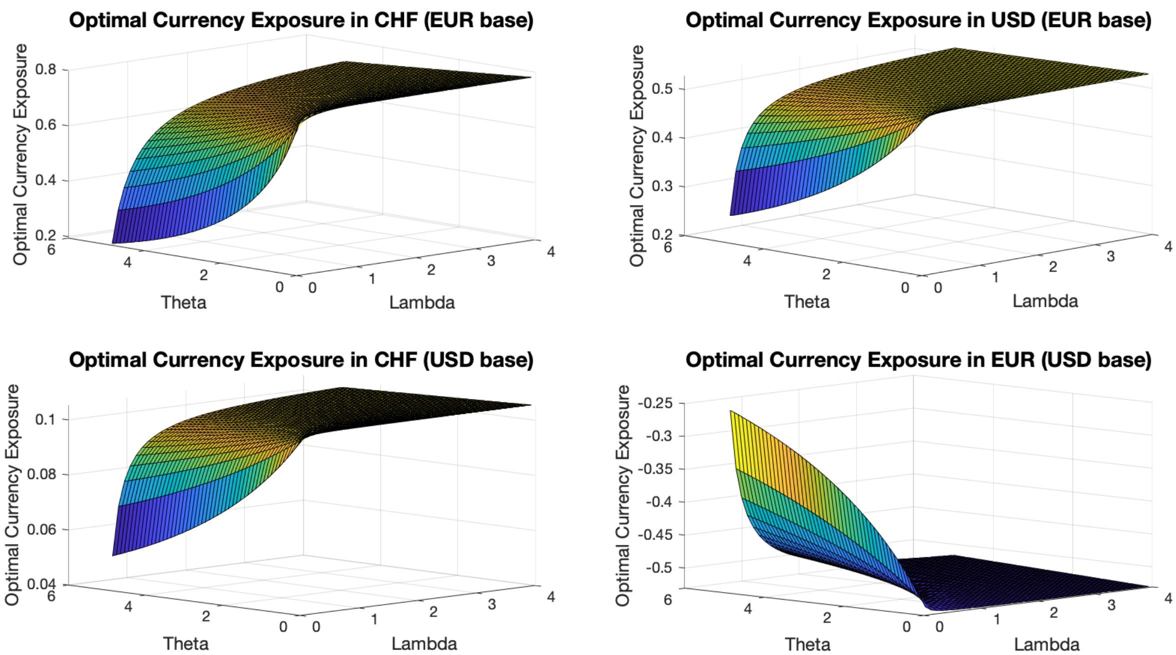

Figure 1.2: The optimal in-sample currency exposure in CHF, EUR, and USD for a EUR- and USD-based investor as a function of the risk and ambiguity aversion parameters λ and θ. We assume that (a) the uncovered interest rate parity condition is satisfied, and (b) the prediction models are independent, and it holds that $\widehat{\mathrm{Var}}_{\mu}[\,\widehat{\mathrm{E}}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,] = \frac{1}{T^2}\mathbf{I}$. The sample spans the period from January 1999 until December 2019.

parameter λ, the ambiguity aversion plays an important role. More specifically, the investor's confidence in her predictive models is getting weaker as the ambiguity aversion parameter θ increases. Therefore, an ambiguity-averse investor prefers to hold smaller exposure to foreign currencies. In the limiting case when $\theta \to \infty$, the optimal currency exposure converges to zero, which corresponds to full hedging of the currency risk. We also observe that the convergence rate is inversely related to the level of risk aversion due to the definition of the ridge penalization parameter q.

Second, we find that the optimal in-sample currency exposure is positive for those foreign currencies which are perceived as safe-haven currencies compared to the investor's reference currency, e.g., the Swiss franc and the US dollar relative to the euro. The ambiguity aversion has a particularly pronounced effect in the case when the optimal exposure in Swiss francs is estimated for a euro-based investor. For such an investor, the ambiguity-induced shrinkage in the ridge regression is stronger for the Swiss franc than for the US dollar because of a high correlation between the Swiss and Eurozone economies.

The optimal currency exposures presented in Fig. 1.2 are estimated from a finite sample. Being a function of the data, the optimal currency exposure estimator (1.4.2) is itself a random variable whose properties can be studied further. To address this question, we investigate how different values of λ and θ affect the confidence intervals of the optimal in-sample currency exposure. In the first step, we construct time series of

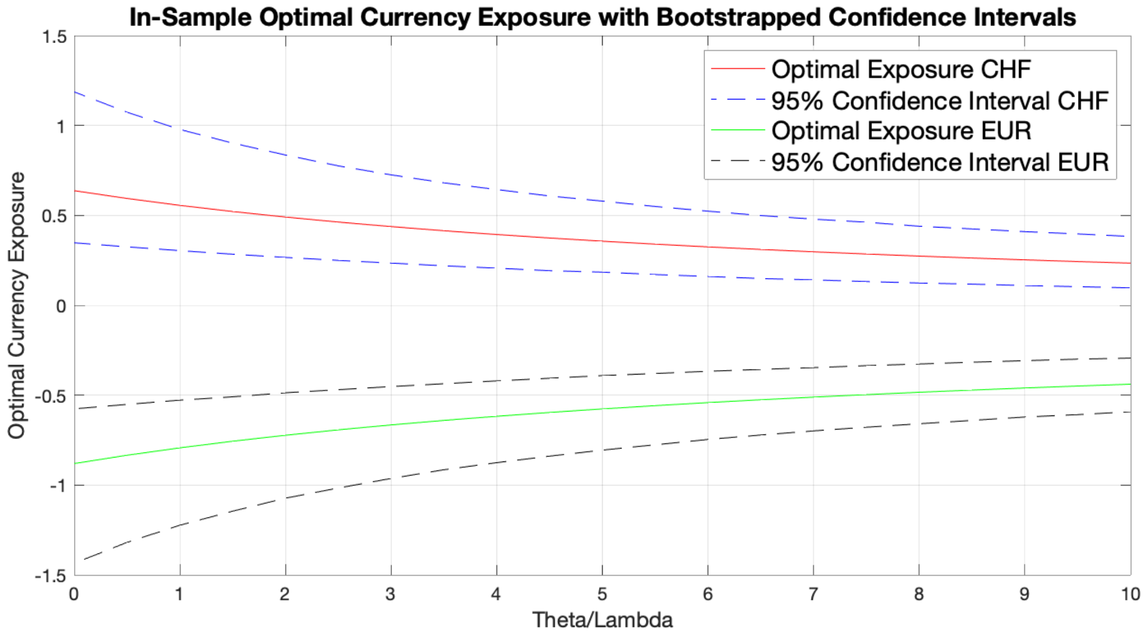

Figure 1.3: Optimal currency exposure in CHF and EUR and the corresponding 95% confidence intervals for a USD-based investor, as a function of the risk and ambiguity aversion parameters λ and θ. We assume that (a) the uncovered interest rate parity condition is satisfied, and (b) the prediction models are independent, and it holds that $\mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,] = \frac{1}{T^2}\mathbf{I}$. The sample spans the period from January 1999 until December 2019.

monthly non-overlapping returns. Then, assuming independent and identically distributed (i.i.d.) returns, we are able to estimate the confidence intervals of the optimal sample-efficient currency exposure by the means of non-parametric bootstrapping (i.e., a random sampling with replacement).

We consider a USD-based investor whose portfolio comprises Swiss, Eurozone, and US equities and bonds (equally-weighted). Figure 1.3 displays the estimated optimal exposure to the Swiss franc and the euro, jointly with the bootstrapped 95% confidence intervals as a function of the ratio of the ambiguity and risk-aversion parameters θ and λ.[23] As the ambiguity aversion parameter corresponds to the regularization parameter in the ridge regression, an increase in θ leads to a larger home currency bias. Additionally, the confidence intervals shrink with increasing ambiguity aversion. In the limiting case when $\theta \to \infty$, the optimal currency exposures and the confidence intervals for both CHF and EUR converge to zero.

The Swiss and Eurozone economies are highly integrated, which is reflected also in the large and positive correlation coefficient between the Swiss franc and the euro (see Panel C in Table 1.2). This result can impact the precision of the estimated optimal currency exposures, i.e., they can exhibit large standard errors. Indeed, we observe wide confidence intervals for $\theta = 0$ in Fig. 1.3. Moreover, a large positive coefficient for the Swiss franc is partially neutralized by a negative coefficient for the euro. This issue is

[23] Using the statistical learning terminology, the plotted functions can be interpreted as ridge regularization paths of optimal currency exposures.

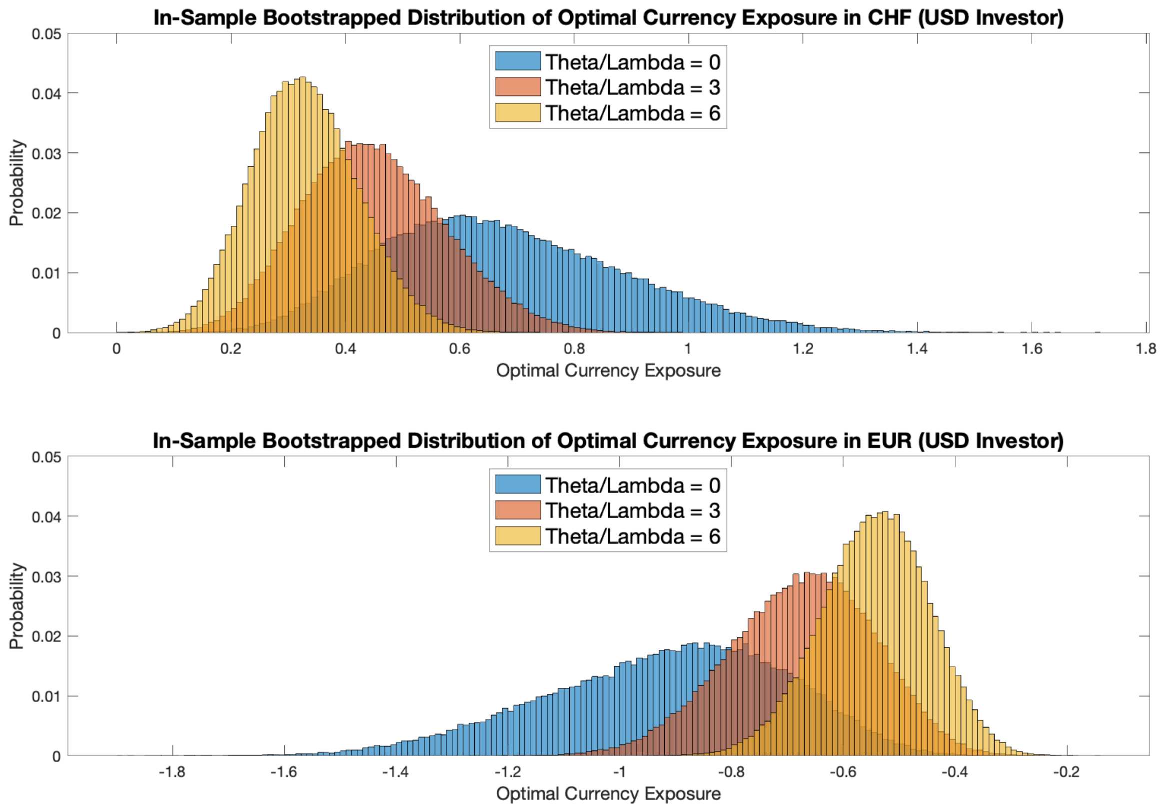

Figure 1.4: Bootstrapped distribution of optimal currency exposure in CHF and EUR for a USD-based investor, as a function of the risk and ambiguity aversion parameters λ and θ. We assume that (a) the uncovered interest rate parity condition is satisfied, and (b) the prediction models are independent, and it holds that $\mathrm{Var}_\mu[\,\mathrm{E}_\mathbb{Q}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,] = \frac{1}{T^2}\mathbf{I}$. The sample spans the period from January 1999 until December 2019.

mitigated by imposing a constraint on the coefficients in the ridge regression via the regularization term.[24] We provide an alternative illustration of these results in Fig. 1.4 by plotting the bootstrapped distributions of the optimal exposures to CHF and EUR for a USD-based investor. The distribution of the estimator for $\theta = 0$ (represented by the blue bars) corresponds to the risk-only case, which is typically considered in the literature. The distribution is markedly wide, demonstrating that the optimal currency exposure estimator exhibits large parameter uncertainty. For larger values of θ, the distribution shifts towards the ridge target—equal to zero in our example—and becomes significantly narrower.[25]

Finally, in Figure 1.5 we present an efficient surface—a generalization of the standard efficient frontier to a three-dimensional setting. To generate the efficient surface, we maximize the expected portfolio return

[24] In the presence of correlated regressors, ridge regression is a convenient alternative to OLS regression. Ridge regression projects the predictor variables onto the principal components of input data points and then shrinks the coefficients of the low–variance components more than the high-variance components. We refer an interested reader to Hastie et al. [2009] for additional information.

[25] In the limiting case when $\theta \to \infty$, the distribution converges to the Dirac delta function centered at the origin—implying the home currency bias.

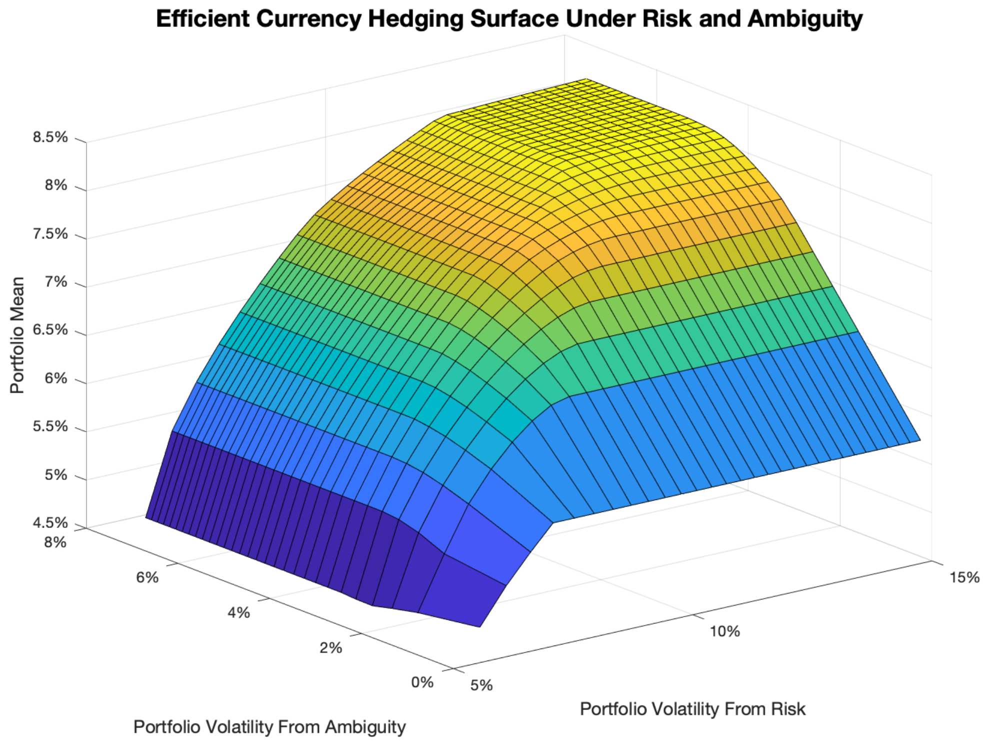

Figure 1.5: The efficient currency hedging surface for a USD-based investor who holds a global equally-weighted portfolio which consists of developed market equities and bonds. The efficient surface illustrates the optimal in-sample trade-off between the portfolio return and the two components of portfolio volatility driven by risk and ambiguity, respectively. We assume that (a) the uncovered interest rate parity condition is satisfied, and (b) the prediction models are independent, and it holds that $\mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,] = \frac{1}{T^2}\mathbf{I}$. The sample spans the period from January 1999 until December 2019.

on a pre-defined two-dimensional grid of the portfolio variance arising from risk and the portfolio variance arising from ambiguity. Mathematically, this optimization problem can be cast as a linear program with quadratic constraints. We consider a USD-based investor who holds an equally-weighted portfolio consisting of broad equity and bond indices representing the seven developed markets described in Section 1.4. The efficient surface results from an optimization over currencies, while the equity and bond portfolio weights are kept fixed. For practical reasons, we impose an upper bound on the leverage, which is equal to 100% of the currency position implied by the portfolio design, without any initial currency overlay.[26]

We note that, for a given level of ambiguity-induced portfolio volatility, one obtains a standard, two-dimensional efficient frontier (i.e., the expected return versus the volatility arising from risk). Moreover, the point on the efficient surface which corresponds to the minimum variance and the maximum ambiguity converges to the global minimum variance case in a standard mean–variance framework, which does not include ambiguity. The convergence is achieved only if the leverage constraint is sufficiently relaxed or removed, otherwise the standard minimum variance portfolio might not be feasible.

[26] Without any constraint on currency positions, the problem becomes unbounded as infinite leverage is possible.

1.4.3 Out-of-Sample Backtest

In this section, we consider an out-of-sample (OOS) backtest of different currency overlay strategies for a global multi-asset class equally-weighted portfolio consisting of cash, bonds, and equities. The analysis is performed for each of the seven possible reference/base currencies. We study three different model-based currency overlay strategies which are discussed throughout the paper: the risk-only minimum variance (RO-MinVar), the risk-only mean–variance (RO-MeanVar), and the ambiguity-adjusted mean–variance (AA-MeanVar). The OOS performance of these portfolios is compared to three benchmarks, which are specified in terms of rule-based, constant-hedging strategies: zero, half, and full hedging. Additionally, we include the leverage constraints described in the previous section.[27]

To simplify the analysis and better understand the impact of the ambiguity aversion on the optimal currency exposure, we assumed the independence of the investor's predictive models in Section 1.4.2. In contrast, we dispense with this assumption here and employ several popular exchange rate forecasting models to estimate the expected currency return. Our goal is to calculate the expected normalized payoff of the long forward contract $\mathrm{E}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]$, as well as the ambiguity covariance matrix $\mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,]$.[28] We follow the work of Cheung et al. [2019] who examine the predictive performance of various exchange rate models. More specifically, we consider seven forecasting models in total: (a) Historical average, (b) Uncovered interest rate parity, (c) Relative purchasing power parity, (d) Sticky price monetary model, (e) Behavioral equilibrium exchange rate (BEER) model, (f) Sticky price monetary model augmented by risk and liquidity factors, and (g) Yield curve slope model. Further information about the models and the macroeconomic data used for the exchange rate forecasting is provided in the Appendix.

We label the models by their respective probability measures $\mathbb{Q}_i$, where $i = 1, 2, \ldots, n$.[29] The set of all considered probability measures is denoted by $\mathcal{Q}$. In the first step, our goal is to estimate $\mathrm{E}_{\mathbb{Q}_i}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]$ for each model $\mathbb{Q}_i$. To compute $\mathrm{E}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]$ and $\mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,]$, we need to specify a prior μ, which governs the weighting of the models in $\mathcal{Q}$. A common approach is to use equal weighting with $\mu_i = 1/n$, where μ_i is the weight corresponding to the model $\mathbb{Q}_i$. However, the model performance can be potentially improved by calibrating the model weights μ to the data. Forecast combinations have been advocated to outperform individual forecasting models from both theoretical and empirical perspective, e.g., see Timmermann [2006].[30] We follow Conflitti et al. [2015] and compute the weights μ by minimizing the

---

[27] Generally, the unbounded versions of the models considered here correspond either to an OLS or a ridge regression, whereas the bounded models can be solved via quadratic programming.

[28] We note that the assumption $\mathrm{Cov}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,], \mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,] = \mathbf{0}$ remains intact. This economic interpretation of this condition is that fully hedged positions are treated as purely risky (i.e., non-ambiguous). Furthermore, this choice fixes the shrinkage target right at the origin.

[29] We keep the notation as general as possible. Specifically, in our applications $n = 7$.

[30] For example, Jorion [1986] introduced a simple Bayes estimator for a sample mean return and showed that it provides significant gains in portfolio selection problems.

mean square forecasting error for the ambiguity-adjusted mean–variance model under the constraints of $\sum_{i=1}^{n} \mu_i = 1$ and $\mu_i \geq 0$ (for $i = 1, 2, \ldots, n$).

In the backtest, currency hedging is implemented using forward contracts which are rolled over quarterly.[31] When a currency forward contract expires, the resulting P&L is recorded on the investor's cash account in the base currency. A new hedge is formed immediately thereafter and the process is repeated over time. The asset positions are not rebalanced during the backtesting period.[32] On the other hand, the positions in currency forward contracts and the cash balance in the investor's base currency change over time. This procedure allows us to isolate the effects of different currency overlay strategies on the portfolio performance. The results presented in this section are net of transaction costs, which are assumed to be two basis points relative to the notional of a currency forward contract.[33] Our estimate of the transaction costs is based on Ackermann et al. [2017].[34]

Table 1.3 reports the annualized average returns, volatilities, and performance metrics (i.e., Sharpe and Sortino ratios) of portfolio daily returns, as well as the average hedging turnover. The statistics are presented for all considered base currencies and across different hedging strategies described above. Our findings indicate that the out-of-sample benefit of currency hedging strongly depends on the investor's base currency and the model.

First, we compare the currency overlay strategies in terms of portfolio volatility. Full-hedging strategy—in comparison to the zero-hedging strategy—reduced volatility for all base currencies. This is particularly pronounced for Swiss, Japanese, and US investors, which is intuitive because CHF, JPY, and USD are considered to be safe-haven currencies (i.e., they tend to be less correlated with equities in down markets than other currencies). Next, we stress that all advanced currency overlay strategies outperform constant-hedging benchmarks in terms of the risk-reduction potential. The volatility reduction is largest for Japanese investors—it is about two thirds of the volatility generated by the zero-hedging strategy. After consolidating the results across all base currencies, we conclude that the best-performing model in terms of volatility reduction is the risk-only minimum variance model, closely followed by the ambiguity-adjusted mean–variance model.

Second, we compare the models in terms of the average portfolio returns. The full hedging tends to

31 Our results are robust to alternative specifications.

32 The relative portfolio weights change over time, but not the absolute weights that are set at the beginning of the backtest such that the portfolio is equally weighted.

33 Our findings are robust to alternative specifications of transaction costs as well.

34 In practice, currency forward contracts are typically implemented through a foreign exchange spot and swap transactions. For a trading strategy that requires currency forwards to be rolled over time, the spot transaction can be executed only at the beginning, and only swap contracts need to be rolled over. This is important because transaction costs for foreign exchange swaps are considerably lower than spot transactions (typically, by an order of magnitude). Ackermann et al. [2017] estimated the average cost of about two basis points for the combined effect of spot and swap transactions per roll-over. Therefore, this estimate could be seen as an upper bound for transaction costs since one could construct a cheaper currency hedging strategy by rolling over only swaps.

Out-of-Sample Backtest of Currency Overlay Strategies

	Constant Hedging			Optimized Currency Exposure		
Panel A: AUD	Zero	Half	Full	RO-MinVar	RO-MeanVar	AA-MeanVar
Return (%)	4.99	5.36	5.81	6.37	7.02	**7.45**
Volatility (%)	7.94	5.54	6.04	**4.50**	6.16	5.02
Sharpe Ratio	0.63	0.97	0.96	1.42	1.14	**1.49**
Sortino Ratio	0.93	1.36	1.32	2.04	1.60	**2.21**
Turnover	0	0.33	0.68	0.88	1.33	1.00
Panel B: CAD	Zero	Half	Full	RO-MinVar	RO-MeanVar	AA-MeanVar
Return (%)	5.59	5.23	4.90	5.37	5.89	**6.02**
Volatility (%)	7.92	6.87	7.46	**5.63**	6.84	5.70
Sharpe Ratio	0.71	0.76	0.66	0.95	0.86	**1.06**
Sortino Ratio	1.00	1.05	0.90	1.35	1.19	**1.49**
Turnover	0	0.45	0.91	1.22	1.41	1.16
Panel C: CHF	Zero	Half	Full	RO-MinVar	RO-MeanVar	AA-MeanVar
Return (%)	3.12	3.11	3.13	3.34	**3.98**	3.72
Volatility (%)	11.36	8.61	6.78	**5.13**	6.60	5.36
Sharpe Ratio	0.27	0.36	0.46	0.65	0.60	**0.69**
Sortino Ratio	0.38	0.49	0.63	0.91	0.83	**0.96**
Turnover	0	0.40	0.79	1.28	1.44	1.21
Panel D: EUR	Zero	Half	Full	RO-MinVar	RO-MeanVar	AA-MeanVar
Return (%)	5.43	4.86	4.32	4.68	**6.21**	5.04
Volatility (%)	8.53	7.48	7.30	5.59	7.23	**5.55**
Sharpe Ratio	0.64	0.65	0.59	0.84	0.86	**0.91**
Sortino Ratio	0.89	0.90	0.81	1.19	1.23	**1.27**
Turnover	0	0.46	0.92	1.21	1.33	1.14
Panel E: GBP	Zero	Half	Full	RO-MinVar	RO-MeanVar	AA-MeanVar
Return (%)	**7.35**	6.35	5.45	5.75	6.50	6.16
Volatility (%)	9.12	7.86	8.25	**6.20**	7.23	6.32
Sharpe Ratio	0.81	0.81	0.66	0.93	0.90	**0.98**
Sortino Ratio	1.17	1.14	0.91	1.32	1.26	**1.37**
Turnover	0	0.51	1.04	1.46	1.47	1.38
Panel F: JPY	Zero	Half	Full	RO-MinVar	RO-MeanVar	AA-MeanVar
Return (%)	**4.70**	4.03	3.46	3.70	4.50	4.28
Volatility (%)	15.23	10.88	7.48	**5.72**	6.98	5.91
Sharpe Ratio	0.31	0.37	0.46	0.65	0.65	**0.72**
Sortino Ratio	0.42	0.50	0.63	0.90	0.89	**1.00**
Turnover	0	0.45	0.92	1.36	1.45	1.28
Panel G: USD	Zero	Half	Full	RO-MinVar	RO-MeanVar	AA-MeanVar
Return (%)	4.58	4.69	4.82	5.09	**6.07**	5.44
Volatility (%)	10.99	8.87	7.49	**5.66**	7.05	5.77
Sharpe Ratio	0.42	0.53	0.64	0.90	0.86	**0.94**
Sortino Ratio	0.59	0.73	0.88	1.27	1.20	**1.33**
Turnover	0	0.44	0.88	1.31	1.49	1.26

Table 1.3: We report the out-of-sample annualized average return, volatility, Sharpe ratio, Sortino ratio, and turnover for six currency overlay strategies (the definitions are provided at the beginning of Section 1.4.3) and across seven base currencies. For RO-MinVar and RO-MeanVar we choose $\lambda = 3$, whereas for AA-MeanVar we set $\lambda = 3$ and $\theta = 4$. The underlying portfolio consists of equities and bonds (mixed in proportion 60/40 at the inception). Within each asset class, we assume an equally-weighted sub-portfolio that consists of the corresponding developed market indices.

lower portfolio return compared to the zero-hedging strategy. However, there are some exceptions, such as AUD, CHF, and USD. The advanced currency overlay strategies tend to outperform the constant-hedging strategies. The best-performing models which are based on the optimized currency exposure are the risk-only mean–variance model and the ambiguity-adjusted mean–variance model.

Third, we investigate the risk-adjusted portfolio performance as measured by the Sharpe and Sortino ratio. Among the constant-hedging strategies, the two performance metrics vary substantially across the base currencies. For example, if full hedging is applied, a USD-based (GBP-based) investor can substantially improve (lower) their Sharpe and Sortino ratios. Half-hedging strategies display relatively attractive risk-adjusted performance for AUD, CAD, and EUR. Most importantly, our out-of-sample backtest indicates that the advanced currency overlay strategies strongly outperform the constant-hedging strategies on a risk-adjusted basis and net of transaction costs. The largest relative increase in the Sharpe and Sortino ratio is observed for Australian, Swiss, and Japanese investors. We stress that the ambiguity-adjusted mean–variance model outperforms all other currency overlay strategies across the board. Therefore, using a battery of exchange rate forecasting models while accounting for the ambiguity aversion seems to add value and improve the traditional models which are based either on constant-hedging rules or on risk-only currency overlay strategies.

Finally, we study the turnover of our currency overlay strategies. We define the average hedging turnover of a strategy s as:

$$\overline{\mathrm{HT}}^s = \frac{1}{T} \sum_{t=1}^{T} \sum_{c=2}^{K+1} |\phi_{c,t}^s|, \tag{1.4.3}$$

where T is the number of trading days, K is the number of foreign currencies, and $\phi_{c,t}^s$ is the notional value of the hedging position for currency c at time t relative to the total portfolio value.[35] Based on our findings reported in Table 1.3, we note that—for a given currency overlay strategy—the average hedging turnover can vary substantially across different base currencies. For example, the turnover of the full-hedging strategy in AUD and GBP is 68% and 104%, respectively. The rationale is that AUD-based (GBP-based) investors, on average, receive positive (negative) pay-offs from their fully hedged currency forward positions, which in turn increase (decrease) the base-currency cash holdings. By construction, the advanced currency overlay strategies are more reactive to changing market conditions than the constant-hedging strategies, hence generating a higher turnover. Between the risk-only minimum variance and mean–variance currency hedging strategies, the former exhibits a lower turnover. This result can be explained by the fact that sample averages are noisy estimates, and their fluctuations over time can be substantial. This induces significant changes in the optimal mean–variance currency exposure in comparison to the minimum variance case, which

[35] The average cost of the strategy can be inferred from the turnover, portfolio value, and rebalancing frequency (quarterly in our applications).

in turn results in an elevated turnover. On the other hand, the ambiguity-adjusted mean–variance model typically exhibits lower turnover compared to the risk-only models. This occurs because ridge regression shrinks the estimated exposure towards zero, hence stabilizing the optimal exposure estimates over time, and consequently reducing the hedging turnover.

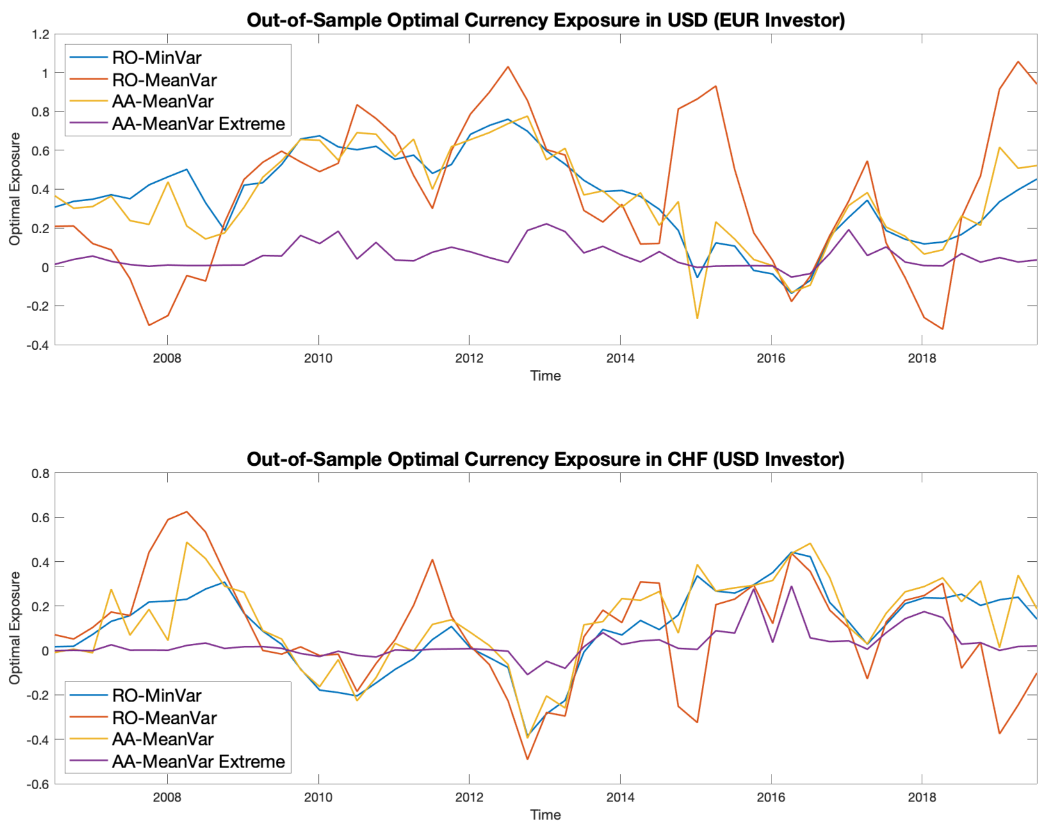

Figure 1.6: The out-of-sample optimal currency exposure in USD for a EUR-based investor and in CHF for a USD-based investor. We assume that exchange rate forecasting models are equally weighted. The risk-aversion parameter is set to $\lambda = 3$, and the optimal currency exposure is bounded between -1 and 1. The risk-only minimum variance (RO-MinVar) and risk-only mean–variance (RO-MeanVar) models do not depend on the ambiguity aversion. Models with ambiguity (AA-MeanVar) consider the following two cases of the parameter of ambiguity aversion θ: (a) “AA-MeanVar” with $\theta/\lambda = 2$ and (b) “AA-MeanVar Extreme” with $\theta/\lambda = 10000$. The latter demonstrates the out-of-sample optimality of home currency bias in the case of an extreme ambiguity aversion. The out-of-sample backtest spans the period from July 2006 until October 2019.

To conclude the out-of-sample backtest analysis and demonstrate empirically that ambiguity aversion can explain the home currency bias, we investigate how the estimated optimal currency exposure—in USD for a EUR-based investor and in CHF for a USD-based investor—varies over time. The optimal exposure is computed relative to the total portfolio value and is constrained between -1 and 1. Our findings are presented in Figure 1.6. Based on Eq. (1.3.20), the risk-only mean–variance strategy can be decomposed into a corresponding risk-only minimum variance strategy and the risk-adjusted currency return. The second term drives the difference between the two currency overlay strategies, and explains the elevated turnover of the risk-only mean–variance model compared to the risk-only minimum variance model. This effect can be observed both in Table 1.3 and Fig. 1.6. The ambiguity-adjusted mean–variance model is more robust because of the shrinkage towards the full-hedging strategy induced by the investor's ambiguity aversion. To gain further insights into the impact of the ambiguity on the dynamics of the optimal currency exposure, we consider two ambiguity-adjusted mean–variance strategies. Both models assume the risk-aversion parameter $\lambda = 3$, whereas the ambiguity aversion parameter θ is selected so that $\theta/\lambda = 2$ in one case (a standard ambiguity model), and $\theta/\lambda = 10000$ in the other case (an extreme ambiguity model). In the latter case, the optimal out-of-sample currency exposure is close to the full-hedging strategy. This empirical result shows that the home currency bias in international portfolios can be interpreted as the optimal currency exposure of an extremely ambiguity-averse investor.

1.5 Conclusions

The objective of this paper is to study the optimal currency allocation for a risk-and-ambiguity-averse investor and link the empirical puzzle of insufficient currency diversification—home currency bias—to extreme ambiguity aversion. We start by presenting a general framework for hedged portfolio returns in a model-free setting, without assumptions on the dynamics of asset or currency returns. These expressions for hedged portfolio returns are then combined with the robust mean–variance utility representation, and a closed-form solution that characterizes the optimal currency exposure for a risk-and-ambiguity-averse investor is derived. Using the derived expressions, we investigate the optimal currency allocation of an investor who regards local asset returns as purely risky and foreign currencies as ambiguous. We show that when the ambiguity aversion parameter $\theta \to \infty$, the optimal foreign currency exposure of the risk-and-ambiguity averse investor converges to 0, i.e., full hedging. This result shows that the home currency bias can be driven by investor's ambiguity aversion, emphasizing the importance of predictive modeling and currency risk management for international investors.

Furthermore, we show that the sample-efficient currency exposure for a risk-and-ambiguity-averse investor can be found by a generalized ridge regression. Thereby, the demeaned hedged portfolio returns are regressed on the demeaned currency excess returns and shrunk towards the infinitely ambiguity-averse optimal currency exposure. The generalized penalty term of the ridge regression corresponds to the utility loss arising from the model uncertainty and induces regularization. This regularization, biasing of the estimator that stabilizes the inference, is here not assumed a priori. It originates as a solution to the robust mean–variance maximization problem. This shows that accounting for ambiguity enables the formal relation between the areas of financial economics and statistical learning.

In the empirical part of the work, we explore the impact of ambiguity aversion on optimal currency allocations in and out of sample. We show that acknowledging ambiguity corresponds to an increase in bias and a simultaneous shrinkage of the confidence intervals of the optimal currency exposure estimator. This bias-variance trade-off gives rise to the possibility that the ambiguity-adjusted optimal currency exposure estimator could improve the portfolio performance out of sample. To examine this, we conduct an out-of-sample backtest examining the performance of the derived theoretical model on the historical market data. We implement various currency forecasting models presented in the existing literature and find that the proposed ambiguity-adjusted model outperforms constant hedging and risk-only based models net of transaction costs. This outperformance is achieved in terms of risk-adjusted returns measured by Sharpe and Sortino ratios. Ambiguity induces shrinkage, which stabilizes the optimal currency exposure estimations over the backtest. This yields a reduction in average hedging turnover and drives the outperformance of the ambiguity-adjusted model in comparison to the other currency hedging benchmarks.

This paper introduces a benchmark model that combines the risk and the ambiguity aversion of international investors. The model is intuitive since it allows for closed-form solutions and a geometric interpretation of obtained results but is involved enough to study the impact of investor's ambiguity on the optimal currency allocation in international portfolios. Given its underlying assumption of independent and identically distributed returns, our model does not capture empirical stylized facts present in the financial data. Therefore, this paper allows for possible generalizations in several research directions that can later be benchmarked to our model. For example, further implications of risk and ambiguity in higher moments can be pursued. Thereby, robust preferences capturing the coskewness and cokurtosis of asset and currency returns can be investigated. One can study their effect on the optimal currency exposure and home currency bias, potentially in closed-form. Another important consideration can also be the study of ambiguity around higher moments of the currency return distribution. Moreover, one can include the effect of transaction costs directly in the optimized utility function. This would correspond to an investor seeking to improve the risk-ambiguity-return characteristics in a cost-efficient manner. The transaction costs would, in this case, correspond to the L^1-normed penalty, such as in Brodie et al. [2009]. The generalized ridge regression arising in our model can then potentially be extended to the generalized elastic net regularized regression–a method that linearly combines the L^1 and L^2 penalties of the lasso and ridge methods.

From the empirical point of view, it might be valuable to investigate emerging market currencies and indices jointly with the already explored developed markets. It is also possible to extend an empirical analysis by using different asset classes, such as corporate bonds and commodities. For example, possible differences in optimal currency exposure of a portfolio consisting of investment grade vs. high yield bonds of developed and emerging markets can be investigated. Another potential prospect is studying specific sectors, such as energy, agriculture, precious metals, and observing if particular sectors in various market environments co-vary with currencies in specific ways. This relates to the growing academic interest in sector rotation investment strategies, where one can examine how currencies relate to the corresponding macro variables used for forecasting and sector rotation decisions. One can as well analyze specific investment styles, such as value, size, momentum, low volatility, and investigate the relation between currency exposure and different types of factor models.

1.6 Appendix

1.6.1 Embedding a Currency Model: The Covered Interest Rate Parity

Let us denote the nominal risk-free interest rate in currency c by $r_{c,t}$. Based on the notation introduced in Section 1.3, $r_{1,t}$ represents the risk-free interest rate in the home currency, whereas other risk-free interest rates correspond to the foreign currencies. The covered interest rate parity (CIRP) asserts that $F_{c,t}/S_{c,t} = (1+r_{1,t})/(1+r_{c,t})$.[36] Consequently, the forward premium can be computed as $f_{c,t} = (r_{1,t} - r_{c,t})/(1+r_{c,t}) \approx r_{1,t} - r_{c,t}$.[37] Plugging this approximation into Eq. (1.3.9) yields the following result for hedged excess portfolio returns:

$$\tilde{R}^{h}_{\mathcal{P},t+1} - r_{1,t} \approx \underbrace{\sum_{c=1}^{K+1} w_{c,t}(R_{\mathcal{P}_c,t+1} - r_{c,t})}_{\text{LC excess sub-portfolio returns}} + \underbrace{\sum_{c=2}^{M+1} \psi_{c,t}(e_{c,t+1} - r_{1,t} + r_{c,t})}_{\text{excess exchange rate return}} + \underbrace{\sum_{c=2}^{K+1} w_{c,t} R_{\mathcal{P}_c,t+1} e_{c,t+1}}_{\text{cross-products}} .$$

Therefore, the total return can be decomposed into three components: (a) Allocation-weighted average of the excess local-currency returns on sub-portfolios $\mathcal{P}_c$, (b) Net-exposure-weighted average of CIRP payoffs, and (c) Allocation-weighted average of cross-products between the local-currency returns on sub-portfolios $\mathcal{P}_c$ and the corresponding exchange rate returns. ♣

1.6.2 The Quadratic Programming Representation

To ease our notation, we introduce additional conventions. First, the vector of exchange rate returns is given by $\mathbf{e}_{t+1} = (e_{2,t+1}, e_{3,t+1}, \ldots, e_{M+1,t+1})'$. Second, the vector of forward exchange premia is $\mathbf{f}_t = (f_{2,t}, f_{3,t}, \ldots, f_{M+1,t})'$. Using the linearity of expectations and the variance sum law, we obtain the following results for the expected return, risk and model uncertainty for the investor's portfolio:

$$\mathrm{E}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{h}_{\mathcal{P},t+1}\,] = \mathrm{E}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,] + \mathbf{\Psi}'_t\,\mathrm{E}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,], \tag{1.6.1a}$$

$$\mathrm{Var}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{h}_{\mathcal{P},t+1}\,] = \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,] + \mathbf{\Psi}'_t\,\mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\mathbf{\Psi}_t + 2\mathbf{\Psi}'_t\,\mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1} - \mathbf{f}_t\,], \tag{1.6.1b}$$

$$\begin{aligned}\mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{h}_{\mathcal{P},t+1}\,]\,] &= \mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,]\,] + \mathbf{\Psi}'_t\,\mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,]\mathbf{\Psi}_t \\ &\quad + 2\mathbf{\Psi}'_t\,\mathrm{Cov}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,], \mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,].\end{aligned} \tag{1.6.1c}$$

We note that these equations[38] prominently feature the normalized forward payoff, i.e., the difference

[36] Among others, Du et al. [2018] showed that CIRP condition has been systematically violated after the Global Financial Crisis of 2008–2009. The CIRP deviations increase (a) at the quarter ends following the crisis, (b) when banks' balance sheet costs rise, (c) when other risk-free spreads increase, and (d) when nominal interest rates rise.

[37] We neglect the second-order cross-product term since it has a minuscule effect on the results.

[38] Observe that the term $\mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1} - \mathbf{f}_t\,]$ denotes the $(M \times 1)$-dimensional vector of covariances between $\tilde{R}^{fh}_{\mathcal{P},t+1}$ and $e_{c,t+1} - f_{c,t}$, for $c = 2, \ldots, M+1$, and equivalently for $\mathrm{Cov}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,], \mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,]$.

between $\mathbf{e}_{t+1}$ and $\mathbf{f}_t$.[39] This highlights the importance of hedging decision trade-off between entering a forward contract and retaining currency exposure.

By plugging the results from Eqs. (1.6.1a) to (1.6.1c) into Eq. (1.3.13), rearranging the terms, and dropping those which are not a function of $\mathbf{\Psi}_t$, we obtain the quadratic programming format (1.3.14).

The quadratic program (1.3.14) can be solved analytically when no linear constraints are imposed. Taking a derivative with respect to $\mathbf{\Psi}_t$ yields the first-order condition

$$-\mathbf{A}\mathbf{\Psi}_t - \mathbf{b} = \mathbf{0}, \tag{1.6.2}$$

where $\mathbf{0}$ is a $(M \times 1)$ vector of zeros. The Hessian matrix of second-order derivatives is $\mathrm{Hess}(\mathbf{\Psi}_t) = -A$. Since components of the random vector $(\mathbf{e}_{t+1} - \mathbf{f}_t)$ are linearly independent, the corresponding covariance matrix $\mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]$ is positive definite. Furthermore, risk and ambiguity aversion coefficients (i.e., λ and θ) are positive scalars, and $\mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,]$ is a positive semi-definite matrix. This implies that the Hessian matrix is negative definite.[40] We conclude that the first-order condition (1.6.2) is a necessary and sufficient condition to characterize the maximum of the currency overlay optimization problem (1.3.13) with respect to the vector of (unconstrained) net foreign currency exposures $\mathbf{\Psi}_t$. ■

1.6.3 The Optimal Currency Exposure: An Alternative Representation

The optimal currency exposure for an RAA investor (1.3.16) can be cast in the form (1.3.21) with the help of matrix algebra. Assuming positive definiteness of $\mathrm{Var}_{\mu}[\,\mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\,]$, Eq. (1.3.16) can be rearranged as

[39] Under the CIRP, the forward premium $f_{c,t}$ is the difference between the risk-free interest rates in foreign and home currency. Therefore, the random variable $e_{c,t+1} - f_{c,t}$ can be interpreted as the currency excess return.

[40] Equivalently, matrix $\mathbf{A}$ defined in Eq. (1.3.15a) is positive definite and therefore invertible.

follows:

$$
\begin{aligned}
\boldsymbol{\Psi}_t^* = \quad & -\Big(\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,] + \theta \mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]\Big)^{-1} \\
& \cdot \Big(\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,](\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,])^{-1}(\lambda \mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1}-\mathbf{f}_t\,] - \mathrm{E}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]) \\
& + \theta \mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,](\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,])^{-1}(\lambda \mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1}-\mathbf{f}_t\,] - \mathrm{E}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]) \\
& - \theta \mathrm{Var}_\mu(\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,])(\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,])^{-1}(\lambda \mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1}-\mathbf{f}_t\,] - \mathrm{E}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]) \\
& + \theta \mathrm{Cov}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,], \mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]\Big) \\
= \quad & -\Big(\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,] + \theta \mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]\Big)^{-1} \qquad (1.6.3) \\
& \cdot \Big((\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,] + \theta \mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]) \\
& \quad \cdot (\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,])^{-1}(\lambda \mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1}-\mathbf{f}_t\,] - \mathrm{E}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]) \\
& \quad - \theta \mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,](\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,])^{-1}(\lambda \mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1}-\mathbf{f}_t\,] - \mathrm{E}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]) \\
& \quad + \theta \mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]\mathrm{Var}_\mu(\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,])^{-1}\mathrm{Cov}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,], \mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]\Big) \\
= \quad & -\mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]^{-1}(\mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1}-\mathbf{f}_t\,] - \frac{1}{\lambda}\mathrm{E}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]) \\
& + \big(\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,] + \theta \mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]\big)^{-1}\theta \mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,] \\
& \quad \cdot \big(\mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]^{-1}(\mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1}-\mathbf{f}_t\,] - \frac{1}{\lambda}\mathrm{E}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]) \\
& \quad - \mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]^{-1}\mathrm{Cov}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,], \mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]\big) = \\
= \quad & \boldsymbol{\Psi}^*_{t,\mathrm{mv}} + \big(\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,] + \theta \mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]\big)^{-1}\theta \mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]\,(\boldsymbol{\Psi}^*_{t,\mathrm{amb}} - \boldsymbol{\Psi}^*_{t,\mathrm{mv}}),
\end{aligned}
$$

where we used the results derived in Eqs. (1.3.19) and (1.3.20). This concludes the derivation of Eq. (1.3.21). ■

1.6.4 The Ambiguity-Adjusted Market Price of Currency Risk

Starting from Eq. (1.3.16), the expression for the optimal in-sample currency exposure can be written as

$$
\begin{aligned}
\widehat{\boldsymbol{\Psi}}^*_{t,\mathbb{H}} = & -\Big(\lambda \widehat{\mathrm{Var}}_{\mathbb{H}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,] + \theta \widehat{\mathrm{Var}}_\mu[\,\widehat{\mathrm{E}}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]\Big)^{-1} \\
& \cdot \Big(\lambda \widehat{\mathrm{Cov}}_{\mathbb{H}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1}-\mathbf{f}_t\,] + \theta \widehat{\mathrm{Cov}}_\mu[\,\widehat{\mathrm{E}}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,], \widehat{\mathrm{E}}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]\Big) \qquad (1.6.4) \\
& + \Big(\lambda \widehat{\mathrm{Var}}_{\mathbb{H}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,] + \theta \widehat{\mathrm{Var}}_\mu[\,\widehat{\mathrm{E}}_{\mathbb{Q}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,]\,]\Big)^{-1}\widehat{\mathrm{E}}_{\mathbb{H}}[\,\mathbf{e}_{t+1}-\mathbf{f}_t\,].
\end{aligned}
$$

As discussed in Section 1.3.2, the first term in the final expression represents the ambiguity-adjusted hedging demand $\widehat{\boldsymbol{\Xi}}^*_{t,\mathbb{H}}$. The second term is the ambiguity-adjusted market price of currency risk $\widehat{\boldsymbol{\Lambda}}^*_{t,\mathbb{H}}$, which can be interpreted as a currency "alpha" demand.[41] We demonstrate below that the latter term can be calculated as the solution to a generalized ridge regression. More specifically, we consider the following problem:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \left\{ \left\|\mathbf{A}\mathbf{x} - \mathbf{b}\right\|_{\mathbf{P}}^2 + \left\|\mathbf{x}\right\|_{\mathbf{Q}}^2 \right\}.$$

This is a generalized ridge regression with the shrinkage target at the origin. Following van Wieringen [2021], it can be solved in a closed form:

$$\mathbf{x}^* = (\mathbf{A}'\mathbf{P}\mathbf{A} + \mathbf{Q})^{-1}\mathbf{A}'\mathbf{P}\mathbf{b}. \tag{1.6.5}$$

Using the notation introduced in Section 1.3.4.1, and defining $\mathbb{1}$ and $\mathbf{u}$ as the vector of ones and the average return vector of the columns in $\mathbf{X}$, respectively, we consider the following specification:

$$\begin{cases} \mathbf{A} := \mathbf{X}, \\ \mathbf{P} := \frac{\lambda}{T}\mathbf{I}, \\ \mathbf{b} := \frac{1}{T}\mathbb{1}, \\ \mathbf{Q} := \mathbf{Z} - \lambda\mathbf{u}\mathbf{u}', \\ \mathbf{x} := \mathbf{\Lambda}_t. \end{cases} \tag{1.6.6}$$

By combining Eqs. (1.6.5) and (1.6.6) we obtain the expression for the ambiguity-adjusted in-sample market price of currency risk:

$$\begin{aligned} \widehat{\mathbf{\Lambda}}_{t,\mathbb{H}}^* &= \left(\frac{\lambda}{T}\mathbf{X}'\mathbf{X} - \lambda\mathbf{u}\mathbf{u}' + \mathbf{Z}\right)^{-1}\frac{1}{T}\mathbf{X}'\mathbb{1} \\ &= \left(\lambda\widehat{\mathrm{Var}}_{\mathbb{H}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,] + \theta\widehat{\mathrm{Var}}_{\mu}[\,\widehat{\mathrm{E}}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,]\right)^{-1}\widehat{\mathrm{E}}_{\mathbb{H}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]. \end{aligned} \tag{1.6.7}$$

Therefore, this component of the the optimal in-sample currency exposure can be computed by regressing the vector of ones on the excess currency returns while the shrinkage is performed via a generalized L^2-penalty $\|\cdot\|_{\mathbf{Q}}^2$, where $\mathbf{Q}$ is the regularization matrix defined in Eq. (1.6.6). The positive definiteness of this matrix depends on the interaction between the ambiguity and pure alpha given by $\mathbf{u}$. In the absence of ambiguity (i.e., $\theta = 0$), the regularization matrix is $\mathbf{Q} = -\lambda\mathbf{u}\mathbf{u}'$, i.e., it is negative definite and its rank is equal to one. For positive and sufficiently large coefficient of ambiguity aversion θ, the regularization matrix $\mathbf{Q}$ becomes positive definite.

The regression formulation of the ambiguity-adjusted market price of currency risk represents the optimal trade-off between alpha, risk, and ambiguity. This currency sub-portfolio is then overlaid on top of the hedging demand derived in Eq. (1.3.25). ■

[41] The ambiguity-adjusted hedging demand and market price of currency risk are defined in Eqs. (1.3.17a) and (1.3.17b).

1.6.5 The Sample-Efficient Currency Exposure for an RAA Investor

To derive the expression (1.3.27), we have to solve the generalized ridge regression given in Eq. (1.3.25). Using the notation introduced in Section 1.3.4.1, and by taking a vector derivative with respect to $\mathbf{\Psi}_t$, we obtain the first-order optimality condition:

$$\mathbf{X}'\mathbf{W}\mathbf{X}\mathbf{\Psi}_t + \mathbf{X}'\mathbf{W}\mathbf{y} + \mathbf{Z}\mathbf{\Psi}_t + \mathbf{Z}\mathbf{z_0} \overset{!}{=} 0. \tag{1.6.8}$$

After some matrix algebra, the first equality in Eq. (1.3.27) can be written as

$$\widehat{\mathbf{\Psi}}^*_{t,\mathbb{H}} = -\big(\mathbf{X}'\mathbf{W}\mathbf{X} + \mathbf{Z}\big)^{-1}\big(\mathbf{X}'\mathbf{W}\mathbf{y} + \mathbf{Z}\mathbf{z_0}\big). \tag{1.6.9}$$

We note that the second-order vector derivative is positive definite, which ensures that the unique global minimum of the loss function is attained.

The above result can be further transformed into the following expression:

$$\begin{aligned}
\widehat{\mathbf{\Psi}}^*_{t,\mathbb{H}} &= -\big(\mathbf{X}'\mathbf{W}\mathbf{X} + \mathbf{Z}\big)^{-1}\big(\mathbf{X}'\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} + \\
&\quad + \mathbf{Z}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} - \mathbf{Z}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} + \mathbf{Z}\mathbf{z_0}\big) \\
&= -(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} + (\mathbf{X}'\mathbf{W}\mathbf{X} + \mathbf{Z})^{-1}\mathbf{Z}((\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} - \mathbf{z_0}) \\
&= \widehat{\mathbf{\Psi}}^*_{t,\mathbb{H},\text{risk}} + \big(\mathbf{X}'\mathbf{W}\mathbf{X} + \mathbf{Z}\big)^{-1}\mathbf{Z}\left(\mathbf{\Psi}^*_{t,\text{amb}} - \widehat{\mathbf{\Psi}}^*_{t,\mathbb{H},\text{risk}}\right),
\end{aligned} \tag{1.6.10}$$

where $\widehat{\mathbf{\Psi}}^*_{t,\mathbb{H},\text{risk}} = -(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} = -(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.
This concludes the derivation of Eq. (1.3.27). ∎

1.6.6 A Compendium of the Implemented Exchange Rate Forecasting Models

In Section 1.4.3 we utilize different exchange rate forecasting models to estimate the expected excess currency return $\mathrm{E}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]$ and the ambiguity covariance matrix $\mathrm{Var}_\mu[\,\mathrm{E}_{\mathbb{Q}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]\,]$. The choice of the models is based on the work of Cheung et al. [2019]. Here, we provide the specifications of the implemented models.

(1) Historical average: A sample average is calculated for a given set of historical spot and forward exchange rates.

(2) Uncovered interest rate parity (UIRP):

$$S_{c,t+1} = S_{c,t} + \widetilde{i}_t, \tag{1.6.11}$$

where $S_{c,t}$ is the exchange rate of currency c with respect to the home/base currency at time t, i_t is the interest rate at time t with maturity $t+1$, and tilde denotes the inter-country difference.[42]

(3) Relative purchasing power parity:

$$S_{c,t+1} = \beta_0 + \beta_1 \widetilde{p}_t + \epsilon_{t+1}, \tag{1.6.12}$$

where p_t is the logarithm of the consumer price index (CPI) at time t, and ϵ_{t+1} is the error term.

(4) Sticky price monetary model:

$$S_{c,t+1} = \beta_0 + \beta_1 \widetilde{m}_t + \beta_2 \widetilde{y}_t + \beta_3 \widetilde{i}_t + \beta_4 \widetilde{\pi}_t + \varepsilon_{t+1}, \tag{1.6.13}$$

where m_t is logarithm of the money supply, and y_t is logarithm of the real (i.e., inflation-adjusted) gross domestic product (GDP). The variables i_t and π_t represent the interest and inflation rate, respectively, and ε_{t+1} is the error term.

(5) Behavioral equilibrium exchange rate (BEER) model:

$$S_{c,t+1} = \beta_0 + \beta_1 \widetilde{p}_t + \beta_2 \widetilde{\omega}_t + \beta_3 \widetilde{r}_t + \beta_4 \widetilde{gdebt}_t + \beta_5 \widetilde{tot}_t + \beta_6 \widetilde{nfa}_t + \upsilon_{t+1}, \tag{1.6.14}$$

where p_t is the logarithm of the CPI, ω_t is the relative price of non-tradables, r_t is the real interest rate, $gdebt_t$ is the government-debt-to-GDP ratio, tot_t is the logarithm of the terms of trade, nfa_t is the net foreign assets, and υ_{t+1} is the error term.

(6) Sticky price monetary model augmented by risk and liquidity factors:

$$S_{c,t+1} = \beta_0 + \beta_1 \widetilde{m}_t + \beta_2 \widetilde{y}_t + \beta_3 \widetilde{i}_t + \beta_4 \widetilde{\pi}_t + \beta_5 VIX_t + \beta_6 TED_t + \xi_{t+1}, \tag{1.6.15}$$

which can be understood as the Sticky price monetary model augmented for the VIX and the three-month Treasury-Libor (TED) spread.

(7) Yield curve slope:

$$S_{c,t+1} - S_{c,t} = \beta_0 + \beta_1 \widetilde{i}_t + \beta_2 slope_t + \eta_{t+1}, \tag{1.6.16}$$

where the inter-country difference in the level of the three month interest rate is combined with the difference in the slope (ten-year minus three-month yields).

[42] The tilde-notation applies to all listed models.

The first specification is based on sample estimates. The second model does not require any estimation to generate a prediction. All remaining models are based on OLS regressions.

The financial markets variables (i.e., the interest rates, VIX, TED spread, spot and forward exchange rates) are obtained from Thomson Reuters Datastream and Bloomberg, as described in Section 1.4.1. The macroeconmic variables listed above are sourced from the International Monetary Fund (IMF), the Organization for Economic Co-operation and Development (OECD), Bloomberg, and the websites of the relevant central banks, i.e., the Reserve Bank of Australia (RBA), the Bank of Canada (BoC), the Swiss National Bank (SNB), the European Central Bank (ECB), the Bank of England (BoE), the Bank of Japan (BoJ), and the Federal Reserves (Fed). ♣

Chapter 2

Dynamic Currency Hedging with Non-Gaussianity and Ambiguity

A version of this paper has been submitted to the *Journal of Financial and Quantitative Analysis*.

Authors:

Paweł Polak, Department of Applied Mathematics and Statistics, Stony Brook University

Urban Ulrych, Department of Banking and Finance, University of Zurich & Swiss Finance Institute

Abstract

This paper introduces a non-Gaussian dynamic currency hedging strategy for globally diversified investors with ambiguity. Assuming that ambiguity of a typical investor can be measured from market data, we associate it to non-Gaussianity of financial asset returns and compute an optimal ambiguity-adjusted mean-variance (dynamic) currency allocation. Next, we extend the filtered historical simulation method to numerically optimize an arbitrary risk measure, such as the expected shortfall. The out-of-sample backtest results show that the derived non-Gaussian dynamic currency hedging strategy outperforms the benchmarks of constant hedging and dynamic hedging with Gaussianity for all base currencies and net of transaction costs.

Keywords: Currency Hedging, Ambiguity, Filtered Historical Simulation, Expected Shortfall, Non-Gaussianity, International Asset Allocation, Currency Risk Management.

JEL Classification: C53, C58, F31, G11, G15.

2.1 Introduction

Efficient international asset allocation lies at the core of quantitative risk management practices. To enhance diversification, investors hold global portfolios with exposure to different asset classes and underlying currencies. Investing internationally reduces the exposure to systematic domestic market risk and offers an opportunity for enhanced portfolio growth and improved risk-adjusted portfolio performance. On the other hand, investing in foreign markets creates an exposure to the currency exchange rate variation. Portfolio losses driven by adverse exchange rate movements represent one of the major risks for market participants with multi-currency portfolios, such as pension funds, insurance companies, banks, multinational firms, and other financial intermediaries. Consequently, understanding and managing the currency risk of international portfolios is essential from both theoretical and empirical perspectives.

The currency allocation decision is a special case of the general portfolio theory for international investors. As in any portfolio optimization problem, investors inevitably face parameter and model uncertainty. To integrate uncertainty into portfolio optimization process, methods based on Bayesian portfolio analysis (see Black and Litterman, 1992, Pástor, 2000, Tu and Zhou, 2010) and ambiguity-adjusted preferences (see Schmeidler, 1989, Hansen and Sargent, 2001, Klibanoff et al., 2005, Maccheroni et al., 2013) have been developed.[1] The former incorporate investor's prior information arising from news, macroeconomic data, and asset pricing models, which is otherwise ignored in the classical statistical analysis. The latter differentiate between the "non-probabilized" uncertainty, also known as ambiguity, contrary to "probabilized" uncertainty widely regarded as risk. They deviate from the traditional rational expectations paradigm and the corresponding maximization of expected utility by modeling ambiguity (i.e., uncertainty about the asset return) directly in the utility function. We build upon the literature on ambiguity-adjusted preferences by studying the relation between ambiguity and non-Gaussianity. We employ a flexible non-Gaussian returns model in the assumed dynamics of future asset and currency returns and show that ambiguity can be associated with a continuous normal mean-variance mixture representation of our non-Gaussian returns model. This approach allows us to account for ambiguity in the construction of optimal currency allocation. We show that the continuous mixing random variable of the assumed non-Gaussian return dynamics is analogous to employing a continuum of possible models in the ambiguity-adjusted preferences. Moreover, we apply the derived concept to solve the problem of optimal currency allocation in international portfolios and propose an ambiguity-adjusted dynamic currency hedging strategy.

Managing the currency risk of an international portfolio is performed via currency hedging. Different derivative instruments, such as forwards, futures, swaps, and options, can be utilized to perform hedging.

[1] More recently, Nguyen et al. [2021] introduce a universal framework for robustifying any distribution-based risk measurement against mean-covariance uncertainty.

The primary purpose of hedging is offsetting the change in the value of the target being hedged. In currency risk management, the hedging portfolio is set up such that any decrease (increase) in the value of the portfolio in domestic currency due to a change in a specific currency exchange rate is offset by the gain (loss) on the corresponding currency forward contract. By entering into a currency forward contract, an investor locks in an exchange rate between the base and foreign currencies. Although the spot exchange rates continue to fluctuate, the investor has fixed the exchange rate over the life of the entered currency forward contract. At maturity, another contract is entered, and the process continues over time. The investor's main objective is to determine the optimal notional value of the used hedging instrument to enter, or equivalently, the relative amount of implicit foreign currency exposure to hedge.

Many researchers have studied the problem of managing currency risk in international portfolios, and a number of currency hedging strategies have been presented in the existing literature. In the beginning, opinions among different authors were divided. Perold and Schulman [1988] proposed 100% hedging as the optimal strategy, whereas Froot [1993] concluded that full hedging could actually increase risk without an adequate return compensation in the long run. On the other hand, Black [1989] derived a model in which, under strong assumptions, all investors apply a universal hedging policy, irrespective of the portfolio composition and the reference currency. Glen and Jorion [1993] showed that international diversification decreases portfolio risk irrespective of assets being hedged or not. With time researchers started to agree on the optimal hedging practice. Especially after Solnik [1993] demonstrated that, in the short term, the optimal currency hedging is specific for each investor, and it depends on the portfolio structure.[2]

A broad consensus that currency hedging tends to lower portfolio volatility has been formed. Haefliger et al. [2002] proposed full hedging of fixed-income portfolios and partial hedging for the equity portfolios, depending on the underlying correlation structure between equity and currency returns.[3] Schmittmann [2010] analyzed the performance of static variance minimizing hedging ratios obtained with ordinary least squares in comparison to constant hedging. Campbell et al. [2010] proved that it is possible to find optimal hedging ratios that minimize volatility for arbitrary portfolios. Moreover, the authors showed that the US dollar, the euro, and the Swiss franc moved against world equity markets and should therefore be appealing to risk-minimizing international equity investors irrespective of their low average returns.[4] More recently, Boudoukh et al. [2019] provided a decomposition of a currency overlay portfolio in a mean-variance framework consisting of a hedge sub-portfolio and an alpha-seeking currency sub-portfolio. Ulrych and Vasiljević [2020]

[2] On the firm (and not portfolio) level, corporate hedging of foreign currency risk under uncertain cash flows is, among others, analyzed in Chowdhry [1995], and under ambiguous exchange rate risk in Wong [2017].

[3] Cho et al. [2016] show that capital tends to move out of emerging into developed countries in global down markets, leading to depreciation (appreciation) of emerging (developed) currencies. Moreover, Bruno and Shin [2020] investigate emerging market corporates during the periods of currency depreciation.

[4] de Boer et al. [2020] confirmed most of these results in the extended sample period and showed that the role of the euro as a reserve currency vanishes during the financial crisis.

generalized this concept to risk and ambiguity aversion and showed that ambiguity could be one of the drivers for the so-called home currency bias (i.e., investors' tendency to disproportionately hold assets denominated in their home currency).[5]

All of the works mentioned above study currency risk management from the perspective of minimizing the volatility of portfolio returns. However, there are different ways to measure the risk of a portfolio. Throughout the past 20 years, a dominant coherent[6] measure of downside risk has arguably become the expected shortfall, which is the expected loss given that a value-at-risk threshold is breached. One of the important features of the expected shortfall is that it takes into account the tail of the loss distribution, which is exactly what investors are predominantly concerned about. Harris and Shen [2006] show that although currency hedging based on minimum variance reduces the volatility of portfolio returns, it can increase both negative skewness and excess kurtosis. The authors study the hedging effectiveness in terms of value-at-risk and expected shortfall. On a similar note, Guo and Ryan [2018] use the expected shortfall optimization framework to more accurately account for tail risk with non-Gaussian returns. A similar analysis is also performed in Álvarez-Díez et al. [2016]. Simulation via principal component analysis and subsequent optimization of the expected shortfall is presented in Topaloglou et al. [2002], where the authors jointly determine the asset weights and currency allocations.[7]

Many efforts have also been performed on the econometric modeling of asset and currency returns that drive the optimal hedging decision. Another strand of literature studying optimal currency hedging moves away from the so-called static modeling where independent and identically distributed asset and currency returns are assumed. It employs multivariate GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) time-series models and, given the time-dependent properties of such approaches, calls the corresponding hedging strategies dynamic.[8] For example, De Roon et al. [2003] show that dynamic hedges conditional on the interest rate spread outperform static currency hedging approaches and provide significant improvements in portfolio performance. Similar findings are, among others, also presented in Tong [1996], Brown et al. [2012], Caporin et al. [2014], and Cho et al. [2020]. Different authors presented improvements to the dynamic currency hedging strategies by studying various generalizations of those models. For instance, regime-switching dynamic hedging approaches are investigated in Lee and Yoder [2007a,b].[9] Moreover, Hsu et al. [2008] show that copula-based GARCH models perform more effectively compared to other dynamic

[5] Moreover, Bianchi and Tallon [2019] empirically show that ambiguity averse investors exhibit equity home bias—increased exposure to the domestic relative to the international stock market.

[6] Portfolio construction based on a non-coherent risk measure can lead to counterintuitive solutions that are suboptimal in terms of coherent risk measures.

[7] On a similar note, Maurer et al. [2019] use the principal component analysis to identify important global risk sources in foreign currency markets.

[8] Note that such approaches still consider a single period optimization framework and should be differentiated from dynamic programming approaches solving recursive multi-period optimization problems.

[9] For a regime-switching model based on a Bayesian framework incorporating parameter and model uncertainty with an application in portfolio optimization see Tu [2010].

hedging models (see also Paolella and Polak [2015b] and Paolella and Polak [2018] for applications of copula models in portfolio optimization). Wang et al. [2015] comprehensively investigate the hedging performance of the constant (full) hedging strategy in comparison to the static as well as dynamic minimum variance-based strategies in different futures markets, including commodities, currencies, and stock indices. Recent literature focuses on exploring the time-series predictability of exchange rate returns via the use of currency risk factors in dynamic currency hedges. Opie and Riddiough [2020] show that exploiting a forecastable component in global factor returns leads to a hedging strategy that outperforms other dynamic hedging benchmarks. On a similar note, Filipozzi and Harkmann [2020] show that carry trades are a part of an optimal portfolio.

We extend the existing approaches by modeling a set of asset and currency returns via a COMFORT model, as introduced in Paolella and Polak [2015a]. By enhancing Bollerslev [1990] CCC-GARCH model in several ways, our approach allows for the incorporation of fundamental stylized facts of financial asset returns, including volatility clustering, non-normality (i.e., excess kurtosis and asymmetry), and dynamics in the dependency between assets and currencies over time. A fast expectation-maximization (EM) algorithm is employed for estimation. This makes our proposed dynamic non-Gaussian hedging strategy applicable in a multivariate setting for a potentially large number of foreign currency exposures. Moreover, we establish a general relation between investor's ambiguity and the employed non-Gaussian returns model. We show that the continuous normal mean-variance mixture representation of the COMFORT model parametrizes the ambiguity-adjusted preferences from Maccheroni et al. [2013].

The class of COMFORT models is a general modeling framework that incorporates a conditional multivariate continuous normal mean-variance mixture distribution with GARCH dynamics. What distinguishes these models from Gaussian-based multivariate GARCH models is the use of a mixing random variable. Paolella et al. [2021] show that Alexander and Chibumba [1996] O-GARCH model enriched with COMFORT structure provides smooth portfolios that perform well out of sample and after the transaction costs even with daily rebalancing. Paolella et al. [2019] incorporate regime-switching structure into the model and demonstrate that, compared with similar Gaussian constructions, more robust and systematic regimes can be captured with the COMFORT model.

The main contribution of this paper lies in studying the impact of non-Gaussianity and ambiguity (defined as investor's uncertainty about the assets portfolio and currencies return distribution) on the dynamic currency hedging strategy. For that purpose, we employ the COMFORT model that drives the non-Gaussianity of the joint distribution of portfolio and currency returns. Each element of the vector return at time t is endowed with a common univariate shock. This univariate shock is interpretable as a common market factor that arises from new information coming to the market. We show that the mixing random variable characterizing this shock is related to investor's ambiguity. Ambiguity can be proxied through the size of the market factor's conditional higher moments, i.e., tails or variance in the case of the mean-variance investor.

Using the proposed model, we present a novel dynamic non-Gaussian currency hedging strategy that is able to optimize a currency hedge with respect to an arbitrary, e.g. tail-based, risk measure. Barone-Adesi et al. [1999] introduced the filtered historical simulation of GARCH processes to model the future distribution of asset returns. The simulation is based on the combination of parametric GARCH modeling under the Gaussian distribution and non-parametric simulation of historical portfolio returns. We extend this approach to the aforementioned continuous normal mean-variance mixture setting and present a semi-parametric extended filtered historical simulation method to model the future distribution of asset and currency returns. Such a method is consistent with the non-Gaussian time-series returns model employed in our study. This tractable approach enables us to numerically optimize a dynamic currency hedge with respect to an arbitrary risk measure.

We empirically validate our extended filtered historical simulation and the proposed dynamic non-Gaussian currency hedging strategy via an extensive out-of-sample backtesting empirical exercise. The results show that our method yields a robust and highly risk reductive hedging strategy for both optimizing the ambiguity-adjusted mean-variance as well as the expected shortfall. The proposed hedging strategy outperforms the benchmarks of constant hedging as well as equivalent dynamic approaches based on Gaussian-GARCH modeling net of transaction costs. Furthermore, the empirical analysis indicates that the non-Gaussian model proxies the level of ambiguity present in currency exchange markets. Taking into account the non-Gaussianity and ambiguity when making currency hedging decisions appears to be the main driver of outperformance of the dynamic currency hedging strategy presented in this paper.

The remainder of the paper is organized as follows. The theoretical model and some of its characteristics are established in Section 2.2. Section 2.3 demonstrates the empirical performance of the model. Section 2.4 provides some concluding remarks, and the Appendix gathers additional technical results and figures.

2.2 Model

In this section, the theoretical model is presented. First, following Ulrych and Vasiljević [2020], a general framework for portfolio optimization in an international context is introduced. Next, we briefly recap the common market factor non-Gaussian returns (COMFORT) model, introduced by Paolella and Polak [2015a], which is used for modeling the asset and currency returns. We then perform an analysis of optimal currency exposure in the theoretically tractable robust mean-variance model, as presented by Maccheroni et al. [2013]. A closed-form expression for the optimal mean-variance currency exposure with ambiguity is derived, and an explicit link between the non-Gaussian returns model and the effect of ambiguity aversion is established. Last, we propose a generalized filtered historical simulation algorithm utilized to numerically optimize an arbitrary risk measure.

2.2.1 Portfolio Return with Currency Hedging

We start by presenting a framework for modeling portfolio returns in an international setting. Thereby, investment in assets denominated in arbitrary currencies is possible. We assume hedging is performed with currency forward contracts and provide expressions for hedged portfolio returns.

Denote with $P_{i,t}$ a price of an asset i at time t, expressed in a local[10] currency (LC). The corresponding simple return[11] of this asset over the time period from t to $t+1$ is given by $R_{i,t+1}$. Assume a domestic currency is prespecified and denote with $S_{c_i,t}$ the spot currency exchange rate in the domestic currency per unit of foreign currency at time t, where c_i denotes the local currency in which asset i is denominated. The corresponding currency exchange rate return from t to $t+1$ is denoted by $e_{c_i,t+1}$. Assume that an investor does not rebalance her position in asset i between t and $t+1$. Then, the unhedged return of an asset i expressed in domestic currency $\tilde{R}^u_{i,t+1}$ admits the following decomposition

$$\tilde{R}^u_{i,t+1} = \frac{P_{i,t+1}S_{c_i,t+1}}{P_{i,t}S_{c_i,t}} - 1 = \underbrace{R_{i,t+1}}_{LC\ asset\ return} + \underbrace{e_{c_i,t+1}}_{FX\ return} + \underbrace{R_{i,t+1}e_{c_i,t+1}}_{cross-product}.$$

The single asset return from above motivates an equivalent analysis in a portfolio context. Consider an international investor with an arbitrary domestic currency who is invested in a portfolio $\mathcal{P}$ consisting of N assets. The portfolio weights of assets $i = 1, 2, \ldots, N$ at time t are denoted by $x_{i,t}$. The return of a portfolio is then given by

$$\tilde{R}^u_{\mathcal{P},t+1} = \sum_{i=1}^{N} x_{i,t}\tilde{R}^u_{i,t+1},$$

with $\sum_{i=1}^{N} x_{i,t} = 1$, for every t. We allow the portfolio to have a direct exposure to $K \leq N$ foreign currencies. Take $c = 1$ for the domestic currency and denote by $c = 2, 3, \ldots, K+1$ the foreign currencies. Since multiple assets can be denominated in the same currency, we can simplify the return of the unhedged international portfolio by grouping the currency exchange rate returns. Denote the set of assets denominated in a particular currency c held in portfolio at time t as $\mathcal{A}_{c,t}$. Then, the fraction of portfolio wealth exposed to currency c at time t is characterized by $w_{c,t} := \sum_{j\in\mathcal{A}_{c,t}} x_{j,t}$. Using this notation we can express the

[10] Local can in our context corresponds to both a foreign or a domestic currency.

[11] As we are working with portfolio returns in a cross-section of underlying assets, we throughout the work use simple, and not logarithmic, returns.

unhedged return on portfolio $\mathcal{P}$ in domestic currency as

$$\tilde{R}^{u}_{\mathcal{P},t+1} = \underbrace{\sum_{i=1}^{N} x_{i,t} R_{i,t+1}}_{LC\ asset\ returns} + \underbrace{\sum_{c=2}^{K+1} w_{c,t} e_{c,t+1}}_{FX\ returns} + \underbrace{\sum_{i=1}^{N} x_{i,t} R_{i,t+1} e_{c_i,t+1}}_{cross-products}. \tag{2.2.1}$$

Next, we extend the current framework by introducing hedging with currency forward contracts. Assume that the price of a currency contract in currency c is zero at inception and denote by $F_{c,t}$ the forward exchange rate in domestic currency per unit of foreign currency c at time t. The maturity of the forward contract is at time $t+1$. The forward premium is defined as $f_{c,t} \coloneqq (F_{c,t} - S_{c,t})/S_{c,t}$ and is a deterministic quantity known at time t. Note that for the domestic currency $S_{1,t} = F_{1,t} = 1$ and $e_{1,t} = f_{1,t} = 0$ trivially hold for all t.

Denote by $\phi_{c,t}$ the relative notional value of a forward currency exchange contract for currency c at time t, denominated in domestic currency and expressed as a fraction of total portfolio value. Hedging can be realized by selling a forward contract (i.e., shorting foreign bonds and holding domestic bonds), which we indicate by $\phi_{c,t} > 0$. A hedged portfolio return[12] is then equal to

$$\tilde{R}^{h}_{\mathcal{P},t+1} = \tilde{R}^{u}_{\mathcal{P},t+1} + \sum_{c=2}^{K+1} \phi_{c,t}(f_{c,t} - e_{c,t+1}), \tag{2.2.2}$$

where $f_{c,t} - e_{c,t+1} = (F_{c,t} - S_{c,t+1})/S_{c,t}$ expresses the payoff of a short forward contract (for $\phi_{c,t} > 0$) on currency c at time $t+1$ denominated in domestic currency. The choice of $\phi_{1,t}$ is arbitrary and we set it to $\phi_{1,t} = 1 - \sum_{c=2}^{K+1} \phi_{c,t}$ such that the currency portfolio is a zero investment portfolio, also known as the hedging portfolio. Since $w_{c,t} \neq 0$, we can define the hedge ratio as $h_{c,t} \coloneqq \phi_{c,t}/w_{c,t}$. This implies that assets are unhedged if $\phi_{c,t} = h_{c,t} = 0$, and fully hedged if $\phi_{c,t} = w_{c,t}$ or $h_{c,t} = 1$. Using $\phi_{c,t} = w_{c,t}$, fully hedged portfolio return can be expressed as

$$\tilde{R}^{fh}_{\mathcal{P},t+1} = \underbrace{\sum_{i=1}^{N} x_{i,t} R_{i,t+1}}_{LC\ asset\ returns} + \underbrace{\sum_{c=2}^{K+1} w_{c,t} f_{c,t}}_{FX\ forward\ premia} + \underbrace{\sum_{i=1}^{N} x_{i,t} R_{i,t+1} e_{c_i,t+1}}_{cross-products}.$$

In comparison to the unhedged portfolio return from Eq. (2.2.1), observe that stochastic FX return is replaced by the deterministic forward premia in the case of full currency hedging. This is how hedging eliminates the risk arising from currency exchange rate fluctuations. Since the exact currency exposure at time $t+1$ is unknown at time t, perfect hedging is impossible, and the second-order cross-terms remain.[13]

[12] Note that the investment universe could be extended by assuming that the total number of foreign currencies available on the market is greater than K.

[13] Over shorter time horizons (e.g., monthly or quarterly), and in most market environments, the cross-terms are negligible in practice. Nevertheless, over longer investment horizons or sudden large market movements, these terms can exhibit a

To increase the intuition of what the magnitude of $\phi_{c,t}$ means, we study the currency exposures $\psi_{c,t}$. We define the net exposure to currency c as $\psi_{c,t} \coloneqq w_{c,t} - \phi_{c,t}$, where $w_{c,t}$ represents the direct currency exposure and $\phi_{c,t}$ reflects the currency hedging position. This notation allows us to characterize the cases of i) no hedging: $\psi_{c,t} = w_{c,t}$, ii) full hedging: $\psi_{c,t} = 0$, iii) partial hedging: $0 < \psi_{c,t} \leq w_{c,t}$, and iv) over and under hedging: $\psi_{c,t} < 0$ and $\psi_{c,t} > w_{c,t}$, respectively.

The expression for hedged portfolio return from Eq. (2.2.2) can now be interpreted also as

$$\tilde{R}^{h}_{\mathcal{P},t+1} = \tilde{R}^{fh}_{\mathcal{P},t+1} + \sum_{c=2}^{K+1} \psi_{c,t}(e_{c,t+1} - f_{c,t}), \tag{2.2.3}$$

where the exposure to the domestic currency is, equivalently to $\phi_{1,t}$, computed as $\psi_{1,t} = -\sum_{c=2}^{K+1} \psi_{c,t}$. This shows that the individual currency exposures add to zero, and the currency portfolio is indeed a zero investment portfolio. Equation (2.2.3) is useful as it provides a decomposition of the hedged portfolio return into a component of fully hedged asset returns, which are as close as possible to being orthogonal to currencies, and a component of net currency exposures.

2.2.2 Non-Gaussian Returns Model and Ambiguity

The presented hedged portfolio returns are computed in a model-free setting, using only the definitions of an asset return and a payoff of a currency forward contract. This enables us to specify an arbitrary econometric model for the asset and currency returns and utilize the expressions derived above. Our modeling choice is a non-Gaussian returns model that takes into account the primary stylized facts of financial asset returns, such as volatility clustering, non-normality, and dynamics in the dependency between assets and currencies over time, and presents a generalization to the classical Gaussian-GARCH models. Moreover, it enables us to establish a direct connection between non-Gaussianity and ambiguity about asset returns.

We denote the $(M \times 1)$ return vector at time t by $\mathbf{Y}_t$. Its conditional time-varying distribution is assumed to be multivariate asymmetric variance-gamma (MVG), which is a special case of the multivariate generalized hyperbolic (MGHyp). Using the mixture representation of the MGHyp, see for example McNeil et al. [2015] for details, we can express the return vector as

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\gamma} G_t + \boldsymbol{\epsilon}_t, \quad \text{with} \quad \boldsymbol{\epsilon}_t = \mathbf{H}_t^{\frac{1}{2}} \sqrt{G_t} \mathbf{Z}_t, \tag{2.2.4}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ are $(M \times 1)$ vectors, $\mathbf{H}_t$ is a positive definite, symmetric, conditional $(M \times M)$ dispersion matrix, $\mathbf{Z}_t$ is a sequence of independent and identically distributed standard normal random $(M \times 1)$ vectors, and $(G_t \mid \mathcal{F}_{t-1}) \sim GIG(a, b, p)$ is a scalar mixing random variable, independent of $\mathbf{Z}_t$, with the generalized

significant impact on portfolio returns.

inverse Gaussian (GIG) density, given by

$$f_{G_t|\mathcal{F}_{t-1}}(x) = \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})} x^{p-1} e^{-(ax+b/x)/2}, \quad x > 0, \tag{2.2.5}$$

with K_p denoting the modified Bessel function of the second kind, $a > 0$, $b > 0$, and p are real parameters; and where the information set a time t is given as the sigma algebra generated by the history of returns $\mathcal{F}_t = \sigma(\{\mathbf{Y}_1, \ldots, \mathbf{Y}_t\})$. MVG is a limiting case of MGHyp with two GIG parameters fixed at $a = 2$, $b = 0$, and a positive shape parameter $p > 0$ estimated from the data. The conditional $(M \times M)$ dispersion matrix H_t is decomposed as the product of scale terms and a conditional dependency matrix via

$$\mathbf{H}_t = \mathbf{S}_t \boldsymbol{\Gamma} \mathbf{S}_t, \tag{2.2.6}$$

where $\mathbf{S}_t$ is a $(M \times M)$ diagonal matrix with conditional scale terms $s_{m,t} > 0$, for $m = 1, 2, \ldots, M$, and $\boldsymbol{\Gamma}$ is a time-invariant and symmetric (with ones on the main diagonal) conditional $(M \times M)$ dependency[14] matrix, such that $\mathbf{H}_t$ is positive definite. The univariate scale terms $s_{m,t}$ are modeled by a GARCH(1,1) process

$$s_{m,t}^2 = \omega_m + \alpha_m \epsilon_{m,t-1}^2 + \beta_m s_{m,t-1}^2, \tag{2.2.7}$$

where $\epsilon_{m,t} = Y_{m,t} - \mu_m - \gamma_m G_t$ is the mth element of $\boldsymbol{\epsilon}_t$, and $\omega_m \geq 0$, $\alpha_m \geq 0$, $\beta_m \geq 0$, for $m = 1, 2, \ldots, M$.

In the model specified in Eq. (2.2.4), $\boldsymbol{\mu}$ is the location vector, $\boldsymbol{\gamma}$ is the asymmetry vector, and $\mathbf{H}_t$ is the dispersion matrix of the conditional distribution of $\mathbf{Y}_t$, while the conditional mean and covariance matrix are, by the law of iterated expectations, and the law of total variance, respectively, given by

$$\begin{aligned} \mathrm{E}[\mathbf{Y}_t \mid \mathcal{F}_{t-1}] &= \boldsymbol{\mu} + \mathrm{E}[G_t \mid \mathcal{F}_{t-1}]\boldsymbol{\gamma}, \quad \text{and} \\ \mathrm{Var}(\mathbf{Y}_t \mid \mathcal{F}_{t-1}) &= \mathrm{E}[G_t \mid \mathcal{F}_{t-1}]\mathbf{H}_t + \mathrm{Var}(G_t \mid \mathcal{F}_{t-1})\boldsymbol{\gamma}\boldsymbol{\gamma}', \end{aligned} \tag{2.2.8}$$

where $\mathrm{Var}(G_t \mid \mathcal{F}_{t-1}) = \mathrm{E}[G_t^2 \mid \mathcal{F}_{t-1}] - (\mathrm{E}[G_t \mid \mathcal{F}_{t-1}])^2$.

The generalization of the non-Gaussian COMFORT model compared to the Gaussian-GARCH approaches is achieved by introducing the mixing random variable G_t. The latter class of models can be thought of as the COMFORT model with constant G_t, for every t. The mixing random variable can be interpreted as a common market factor as it accounts for information arrivals and jumps in such a way that, conditional on it, the returns distribution is Gaussian. It can also be viewed as a source of ambiguity about the assets (and currencies) return distribution – see Section 2.2.3.1. Increased moments of G_t (i.e., the level

[14] Note that $\boldsymbol{\Gamma}$ is a correlation matrix only conditionally on the realization of the mixing process. Therefore, we call $\boldsymbol{\Gamma}$ the dependency matrix.

and uncertainty around the future shocks) correspond to greater ambiguity of the future return realizations[15] and vice versa. Note that this mixing random variable drives the dynamics of conditional (co-)moments of the returns and results in an enhanced ability for risk management and asset (currency) allocation. Because of the asset-specific conditional asymmetry coefficient $\boldsymbol{\gamma}$, the impact of G_t varies across assets. The maximum likelihood estimation of the COMFORT model is feasible and computationally inexpensive via the use of the EM algorithm; see Paolella and Polak [2015a] for details.

2.2.3 Optimizing Currency Exposure in an International Portfolio

2.2.3.1 Ambiguity-Adjusted Mean-Variance Currency Exposure

We start the analysis with an ambiguity-adjusted mean-variance investor who is optimizing the currency exposure of her international portfolio. We assume that the asset (portfolio) weights are predetermined, and the investor is optimizing only the currency exposure, which is altered by taking positions in currency forward contracts. Such hedging (or currency overlay) strategies are prevalent in the asset and wealth management industry, where management of currency risk is treated in isolation from the asset allocation, see Kim and Chance [2018]. To account for ambiguity, we employ the robust mean-variance utility function from Maccheroni et al. [2013]. The authors consider a space Δ of possible models (i.e., forecasts) $\mathcal{Q}$ that represent investor's ambiguity. An agent's prior over all probability measures $\mathbb{Q}$ corresponding to the models in $\mathcal{Q}$ is given by η. The reduced probability is then expressed as $\bar{\mathbb{Q}} := \int_\Delta \mathbb{Q}\, d\eta(\mathbb{Q})$.

On the other hand, note that in the COMFORT model, the GARCH dynamics are imposed on the dispersion matrix $\mathbf{H}_t$ that is proportional to the covariance of the returns only conditionally on the realization of the mixing random variable, i.e., $\mathrm{Cov}(\mathbf{Y}_t \mid G_t = g, \mathcal{F}_{t-1}) = g\mathbf{H}_t$. Therefore, the COMFORT model allows us to parameterize the ambiguity from Maccheroni et al. [2013] in the conditional dynamics. The probability density function of a normal variance-mean mixture $f_{\mathbf{Y}_t|\mathcal{F}_{t-1}}(x)$ with mixing probability density $f_{G_t}(g)$, as defined in Eq. (2.2.4), is given by

$$f_{\mathbf{Y}_t|\mathcal{F}_{t-1}}(x) = \int_0^\infty \mathcal{N}_{\mathbf{Y}_t|G_t,\mathcal{F}_{t-1}}(x \mid \boldsymbol{\mu} + \boldsymbol{\gamma} g, g\mathbf{H}_t) f_{G_t}(g) dg, \tag{2.2.9}$$

where $\mathcal{N}_{\mathbf{Y}_t|G_t,\mathcal{F}_{t-1}}(\cdot \mid \cdot, \cdot)$ is a probability density function of a multivariate Gaussian distribution and $f_{G_t}(g)$ is given in Eq. (2.2.5). Assuming that we can measure the ambiguity of a typical investor from the returns observed on the market, this representation displays a link between the Maccheroni et al. [2013] ambiguity model and the normal mean-variance mixture model employed in the COMFORT setting. Thereby, it holds that $\mathbb{Q} \equiv \mathcal{N}_{\mathbf{Y}_t|G_t,\mathcal{F}_{t-1}}(\cdot \mid \boldsymbol{\mu} + \boldsymbol{\gamma} g, g\mathbf{H}_t)$, $\eta \equiv G_t$, and $\bar{\mathbb{Q}} \equiv \mathbf{Y}_t \mid \mathcal{F}_{t-1}$. Probability measure $\mathbb{Q}$ that corresponds to a single model from $\mathcal{Q}$ is in the COMFORT setting given by a Gaussian distribution. These models

[15] This is statistically equivalent to heavier tails of the forecasted return distribution.

are mixed according to investor's prior η, which is parametrized by the mixing random variable G_t in the COMFORT model. Finally, the reduced probability $\bar{\mathbb{Q}}$ is the normal mean-variance mixture given by the MGHyp distribution, as shown in Eq. (2.2.9), and the corresponding continuous mixing random variable G_t enables us to introduce a market-based measure of investor's ambiguity as the expectation of G_t (see Section 2.3.2 for an empirical example).

Consider a hedged portfolio return $\tilde{R}^h_{\mathcal{P},t+1}$ from Eq. (2.2.3). As proposed in Maccheroni et al. [2013], the objective of the investor with ambiguity-adjusted mean-variances preferences, given the risk and ambiguity aversion coefficients $\lambda \geq 0$ and $\theta \geq 0$, respectively, is to maximize

$$\max_{\mathbf{\Psi}_t} U(R^h_{\mathcal{P},t+1}) = \max_{\mathbf{\Psi}_t} \left\{ \mathrm{E}_{\bar{\mathbb{Q}}}[R^h_{\mathcal{P},t+1}] - \frac{\lambda}{2}\mathrm{Var}_{\bar{\mathbb{Q}}}(R^h_{\mathcal{P},t+1}) - \frac{\theta}{2}\mathrm{Var}_{\eta}(\mathrm{E}_{\mathbb{Q}}[R^h_{\mathcal{P},t+1}]) \right\}, \tag{2.2.10}$$

where $\mathbf{\Psi}_t = (\psi_{2,t}, \ldots, \psi_{K+1,t})'$ denotes a $(K \times 1)$-dimensional vector of foreign currency exposures. The argument $\mathbf{\Psi}^*_t$ which maximizes the robust mean-variance utility from Eq. (2.2.10) is the optimal currency exposure for a risk-and-ambiguity-averse international investor. It is derived in Ulrych and Vasiljević [2020] and is, in closed-form, given by

$$\begin{aligned} \mathbf{\Psi}^*_t = &-\Big(\lambda \mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,] + \theta \mathrm{Var}_{\eta}[\,\mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\,]\Big)^{-1} \times \\ &\Big(\lambda \mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1} - \mathbf{f}_t\,] + \theta \mathrm{Cov}_{\eta}[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,], \mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\,] - \mathrm{E}_{\bar{\mathbb{Q}}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\Big), \end{aligned} \tag{2.2.11}$$

where $\mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,]$ and $\mathrm{Var}_{\eta}[\,\mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\,]$ denote the $(K \times K)$-dimensional covariance matrices of the random vector $(\mathbf{e}_{t+1} - \mathbf{f}_t)$ under $\bar{\mathbb{Q}}$ and η, respectively, $\mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1} - \mathbf{f}_t\,]$ and $\mathrm{Cov}_{\eta}[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,], \mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\,]$ denote the $(K \times 1)$-dimensional vectors of covariances between $\tilde{R}^{fh}_{\mathcal{P},t+1}$ and $e_{i,t+1} - f_{i,t}$, for $i = 1, \ldots, K$, under $\bar{\mathbb{Q}}$ and η, respectively, and $\mathrm{E}_{\bar{\mathbb{Q}}}[\mathbf{e}_{t+1} - \mathbf{f}_t]$ denotes the expectation of the random vector $(\mathbf{e}_{t+1} - \mathbf{f}_t)$ under $\bar{\mathbb{Q}}$.[16] The presented expression allows for i) over hedging: shorting foreign currency in excess of what would be required to fully hedge the implicit currency exposure, and ii) under hedging: holding foreign currency in excess to the current exposure of $w_{c,t}$.[17]

Next, we derive the optimal ambiguity-adjusted mean-variance currency exposure given the dynamics of the non-Gaussian returns model presented in Section 2.2.2. Using the notation from Eq. (2.2.4), we model the $(K + 1 \times 1)$-dimensional random vector as $\mathbf{Y}_{t+1} = (\tilde{R}^{fh}_{\mathcal{P},t+1}, e_{2,t+1} - f_{2,t}, \ldots, e_{K+1,t+1} - f_{K+1,t})$. Given the conditional dispersion matrix $\mathbf{H}_t$ defined in Eq. (2.2.6), denote with $\mathbf{H}^c_t$ the $(K \times K)$-dimensional matrix without the first row and the first column of $\mathbf{H}_t$, hence corresponding to the conditional dispersion

16 Since forward premium $f_{c,t}$ can be interpreted as a riskless return attained by entering a forward contract on currency c, we refer to $e_{c,t+1} - f_{c,t}$ as the currency excess return on currency c.

17 To avoid over or under hedging, for example, because of regulatory constraints, the same optimization problem can be cast in a quadratic programming format. There, arbitrary linear constraints on optimal exposures can be prescribed, and the solution is then obtained numerically; for more details, see Ulrych and Vasiljević [2020].

only among the currency excess returns. In a similar fashion, denote with $\boldsymbol{h}_t^c$ the $(K \times 1)$-dimensional vector of the first column - without the first element - of the matrix $\mathbf{H}_t$. Hence, $\boldsymbol{h}_t^c$ corresponds to the conditional dispersion between the fully hedged portfolio return and the currency excess returns. Similarly, define the $(K \times 1)$-dimensional vectors $\boldsymbol{\mu_c}$ and $\boldsymbol{\gamma_c}$, and a scalar γ_1, where $\boldsymbol{\gamma} = (\gamma_1, \boldsymbol{\gamma}_c')'$. Taking into account $\mathbb{Q} \equiv \mathcal{N}_{\mathbf{Y}_t|G_t,\mathcal{F}_{t-1}}(\cdot \mid \boldsymbol{\mu} + \boldsymbol{\gamma} g, g\mathbf{H}_t)$, $\eta \equiv G_t$, and $\bar{\mathbb{Q}} \equiv \mathbf{Y}_t \mid \mathcal{F}_{t-1}$ and utilizing the expressions from Eq. (2.2.8), we have

$$\begin{aligned}
\mathrm{Var}_{\bar{\mathbb{Q}}}[\,\mathbf{e}_{t+1} - \mathbf{f}_t\,] &= \mathrm{E}[G_{t+1} \mid \mathcal{F}_t]\,\mathbf{H}^c_{t+1|\mathcal{F}_t} + \mathrm{Var}(G_{t+1} \mid \mathcal{F}_t)\,\boldsymbol{\gamma_c}\boldsymbol{\gamma_c}', \\
\mathrm{Cov}_{\bar{\mathbb{Q}}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}, \mathbf{e}_{t+1} - \mathbf{f}_t\,] &= \mathrm{E}[G_{t+1} \mid \mathcal{F}_t]\,\mathbf{h}^c_{t+1|\mathcal{F}_t} + \mathrm{Var}(G_{t+1} \mid \mathcal{F}_t)\gamma_1\boldsymbol{\gamma_c}, \\
\mathrm{Var}_\eta[\,\mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\,] &= \mathrm{Var}(G_{t+1} \mid \mathcal{F}_t)\,\boldsymbol{\gamma_c}\boldsymbol{\gamma_c}', \\
\mathrm{Cov}_\eta[\,\mathrm{E}_{\mathbb{Q}}[\,\tilde{R}^{fh}_{\mathcal{P},t+1}\,], \mathrm{E}_{\mathbb{Q}}[\mathbf{e}_{t+1} - \mathbf{f}_t]\,] &= \mathrm{Var}(G_{t+1} \mid \mathcal{F}_t)\gamma_1\boldsymbol{\gamma_c}, \quad \text{and} \\
\mathrm{E}_{\bar{\mathbb{Q}}}[\mathbf{e}_{t+1} - \mathbf{f}_t] &= \boldsymbol{\mu_c} + \mathrm{E}[G_{t+1} \mid \mathcal{F}_t]\,\boldsymbol{\gamma_c}.
\end{aligned} \tag{2.2.12}$$

Plugging the terms from Eq. (2.2.12) into the model-free exposure from Eq. (2.2.11), we obtain the optimal COMFORT-based ambiguity-adjusted mean-variance currency exposure as

$$\begin{aligned}
\boldsymbol{\Psi}^*_{t,COM} = -&\Big[\lambda\mathrm{E}[G_{t+1} \mid \mathcal{F}_t]\,\mathbf{H}^c_{t+1|\mathcal{F}_t} + (\lambda+\theta)\mathrm{Var}(G_{t+1} \mid \mathcal{F}_t)\,\boldsymbol{\gamma_c}\boldsymbol{\gamma_c}'\Big]^{-1} \times \\
&\Big[\lambda\mathrm{E}[G_{t+1} \mid \mathcal{F}_t]\,\mathbf{h}^c_{t+1|\mathcal{F}_t} + (\lambda+\theta)\mathrm{Var}(G_{t+1} \mid \mathcal{F}_t)\gamma_1\boldsymbol{\gamma_c} - (\boldsymbol{\mu_c} + \mathrm{E}[G_{t+1} \mid \mathcal{F}_t]\,\boldsymbol{\gamma_c})\Big].
\end{aligned} \tag{2.2.13}$$

Even though there is a distinct term characterizing ambiguity in the robust utility function from Eq. (2.2.10), we see that ambiguity-like behavior in the COMFORT-based optimal currency exposure from Eq. (2.2.13) persists even in the case when $\theta \to 0$. The reason for this lies in the presence of the mixing random variable G_t. Only when $\mathrm{Var}(G_{t+1} \mid \mathcal{F}_t) \to 0$, the modeled returns become conditionally Gaussian, and the ambiguity about the returns distribution vanishes (i.e., there is no uncertainty about the future market shocks). Simultaneously, the COMFORT model converges to the standard constant conditional correlation (CCC) Gaussian-GARCH type of dynamics, which is the prevailing modeling choice present in the existing dynamic currency hedging literature. However, when $\mathrm{Var}(G_{t+1} \mid \mathcal{F}_t) > 0$, the ambiguity is present and its effect on the optimal currency allocation gets amplified via the parameter of ambiguity aversion θ, given the optimized robust mean-variance utility. On the other hand, the magnitude of ambiguity is expressed through the size of the conditional variance of the mixing random variable.[18] The larger the uncertainty around the possible market shock, the larger the effect on the optimal ambiguity-adjusted mean-variance currency exposure, in comparison to the non-ambiguous Gaussian-GARCH case. The direction of the effect of ambiguity on the optimal currency allocation is driven by the asset-specific conditional asymmetry vector

[18] Since we are, in this section, studying the optimal robust mean-variance currency exposure, only the first two moments affect the corresponding return ambiguity.

γ and hence varies across currencies. It can also be understood as the ambiguity-induced shrinkage of the optimal currency allocation. The robust mean-variance portfolio analysis based on Eq. (2.2.10) shows, in a closed form, the link between investor's ambiguity and non-Gaussianity of financial returns. In general, using our framework, an ambiguity averse investor can also use an arbitrary risk measure (e.g., expected shortfall) to optimize her portfolio.

Analogously to Eq. (2.2.10), one can introduce the general ambiguity-adjusted mean-expected shortfall portfolio objective function. Given the risk and ambiguity aversion coefficients $\lambda \geq 0$ and $\theta \geq 0$, respectively, the investor seeks to maximize

$$\max_{\boldsymbol{\Psi}_t} \left\{ \mathrm{E}_{\bar{\mathbb{Q}}}[R^h_{\mathcal{P},t+1}] - \frac{\lambda}{2}\mathrm{ES}_{\bar{\mathbb{Q}}}(R^h_{\mathcal{P},t+1}) - \frac{\theta}{2}\mathrm{Var}_{\eta}(\mathrm{E}_{\mathbb{Q}}[R^h_{\mathcal{P},t+1}]) \right\}, \tag{2.2.14}$$

where $\boldsymbol{\Psi}_t = (\psi_{2,t}, \ldots, \psi_{K+1,t})'$ denotes a $(K \times 1)$-dimensional vector of foreign currency exposures. Eq. (2.2.14) differs from Eq. (2.2.10) only in the middle term that corresponds to the portfolio tail-based risk measure. The last term has to be symmetric because investors are model uncertain for arbitrary changes in the parameters. In the next section, we derive numerical methods for optimizing objective functions like Eq. (2.2.14) under COMFORT dynamics.

2.2.3.2 General Utility Optimal Currency Exposure

Working with the ambiguity-adjusted mean-variance utility is useful as it yields a closed-form solution and thereby enhances the understanding of the estimation of the optimal currency exposure. In this section, we propose a numerical algorithm that extends the computation of the optimal currency exposure to a general, e.g., tail-based, risk measure. Furthermore, all equations introduced so far dealt with returns and the corresponding currency exposure measured over a single time period starting at time t and ending at $t+1$, where $t+1$ denotes the maturity of the hedge. A usual practical choice for a maturity of a currency hedge is a quarter (or a month, a half-year). However, the time-series models, such as Gaussian-GARCH and COMFORT, are commonly calibrated using daily returns. In such a case, one needs a conditional distribution of the multi-period returns, where the period length depends on the hedge maturity. To circumvent multi-period modeling, one could employ coarser sampling; however, such a method is inefficient since it discards a lot of data and information. Alternatively, computing with overlapping quarterly (or monthly) returns is unsuitable either since it induces an artificial serial dependence leading to inefficient and biased estimates. Another possibility would be utilizing the daily time-frequency and scaling the one-step-ahead conditional distribution estimates. This method exaggerates volatility of volatility and is, therefore, also not appropriate. In this section, we present an approach that circumvents such modeling issues. It uses daily modeling frequency and applies either a parametric bootstrap or its non-parametric equivalent called filtered historical simulation in order to build up a cumulative (i.e., multi-period) conditional distribution of the random vector of returns $\mathbf{Y}_{t+1}$. Additionally, this approach enables us to numerically optimize a general risk measure—for which closed-form solutions necessarily do not exist—such as expected shortfall.

We start with the parametric version of the return accumulation algorithm. Consider the COMFORT dynamics with a daily modeling frequency. We aim to build up a conditional distribution of h-step ahead cumulative returns $\mathbf{Y}_{T+h|\mathcal{F}_T}(h)$ for $h > 1$ corresponding to the hedge maturity (i.e., expressed in terms of days). To model the conditional dispersion matrix $\mathbf{H}_{T+h|\mathcal{F}_T}$ we need an expression for the univariate scale terms given in Eq. (2.2.7). Assuming a parametric distributional structure of the COMFORT model, we prove, in the Appendix, that the forecasted scale terms $\hat{s}^2_{k,T+h|\mathcal{F}_T} := \mathrm{E}\left[s^2_{k,T+h} \mid \mathcal{F}_T\right]$ are given by

$$\hat{s}^2_{k,T+h|\mathcal{F}_T} = \begin{cases} \hat{\omega}_k + \hat{\alpha}_k \left(Y_{k,T} - \hat{\mu}_k - \hat{\gamma}_k \mathrm{E}[G_T \mid \mathcal{F}_T]\right)^2 + \hat{\beta}_k \hat{s}^2_{k,T}, & \text{for } h = 1, \\ \hat{\omega}_k + \hat{s}^2_{k,T+h-1|\mathcal{F}_T} \left(\hat{\alpha}_k \mathrm{E}[G_{T+h-1} \mid \mathcal{F}_T] + \hat{\beta}_k\right), & \text{for } h > 1, \end{cases} \tag{2.2.15}$$

where $\mathrm{E}[G_T \mid \mathcal{F}_T]$ is imputed from the E-step of the EM-algorithm; and $\mathrm{E}[G_{T+h} \mid \mathcal{F}_T] = \mathrm{E}[G_T]$ for any $h > 1$ because of the assumed independence from previous data and iid structure of the mixing random variable G_t. The optimal prediction is obtained in the L^2 sense. It differs from a standard Gaussian-GARCH case as in Francq and Zakoian [2019], because of the additional mixing random variable factor for $h > 1$.

Equation (2.2.15) enables us to construct the daily conditional dispersion matrix $\mathbf{H}_{T+h|\mathcal{F}_T}$ for any $h > 1$. Employing Monte-Carlo sampling, we can repeatedly draw sequences of consecutive returns

$$\mathbf{Y}^{(i)}_{T+1|\mathcal{F}_T} \to \mathbf{Y}^{(i)}_{T+2|\mathcal{F}_T} \to \cdots \to \mathbf{Y}^{(i)}_{T+h|\mathcal{F}_T}, \quad i = 1, 2, \ldots B, \tag{2.2.16}$$

where B denotes the number of simulations (i.e., bootstraps). To simulate these single-period returns, one uses the representation of the model as given in Eq. (2.2.4) and repeatedly samples from the corresponding multivariate standard normal and generalized inverse Gaussian distributions, as estimated at time T. Utilizing these distributions is the reason that the presented approach is called parametric. After simulating the single-period returns, one simply needs to aggregate them into the h-step conditional cumulative return distribution $Y^{(i)}_{T+h|\mathcal{F}_T}(h)$ via

$$\mathbf{Y}^{(i)}_{T+h|\mathcal{F}_T}(h) = (1 + \mathbf{Y}^{(i)}_{T+1|\mathcal{F}_T})\,(1 + \mathbf{Y}^{(i)}_{T+2|\mathcal{F}_T}) \ldots (1 + \mathbf{Y}^{(i)}_{T+h|\mathcal{F}_T}) - 1, \quad i = 1, 2, \ldots B, \tag{2.2.17}$$

where the multiplication is performed on a component-by-component basis.

Now we present the non-parametric version of the described algorithm. The foundation of our procedure is the well-known filtered historical simulation (FHS) approach, as presented in Barone-Adesi et al. [1999]. The authors introduced a simulation model that does not impose any theoretical distribution on the data.

It uses the historical distribution of the return series solely. The procedure is utilized via a GARCH filter that aims to remove the serial correlation and volatility clusters present in the data set. The filtered returns are rendered identically and independently distributed, and the non-parametric bootstrapping can hence be applied. We extend this approach to the COMFORT model that, in addition to the GARCH component, also employs a mixing random variable.

Consider the COMFORT model specification as given in Eq. (2.2.4). Denote an arbitrary asset (currency) in the model with k and assume we work with the data series originating at times $t = 1, \dots T$. The standardized residuals of the asset k at times $t = 1, \dots T$ are given by

$$\tilde{e}_k^{(t)} = \frac{Y_{k,t} - \hat{\mu}_k - \hat{\gamma}_k \mathrm{E}[G_t \mid \mathcal{F}_t]}{\hat{s}_{k,t}\sqrt{\mathrm{E}[G_t \mid \mathcal{F}_t]}},$$

where the term $\mathrm{E}[G_t \mid \mathcal{F}_t]$ corresponds to the imputed mixing random variable from the expectation step of the EM-algorithm used for the estimation of the model parameters.

Under the COMFORT model specification, the standardized residuals are independent and identically distributed by definition and hence suitable for historical simulation – parametric approach. Empirical observations might not exactly satisfy this. However, filtered historical innovations can be drawn randomly with replacement and used as innovations to generate pathways of future returns consistent with the estimated time-series model.[19] We denote with $\tilde{\mathbf{Z}}$ the $(K \times T)$ matrix of standardized residuals, where $[\tilde{\mathbf{Z}}]_{k,t} = \tilde{e}_k^{(t)}$.

In addition to the standardized residuals, we also need to standardize the corresponding filtered mixing random variables that are utilized in the filtered historical simulation. Consider a sample of $t = 1, \dots, T$ filtered variables $\mathrm{E}[G_t \mid \mathcal{F}_t]$. We employ a kernel cumulative distribution function (cdf) estimation. This is a non-parametric way to estimate the cdf of a random variable. Kernel cdf estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample. For us, the sample consists of the imputed mixing random variables $\mathrm{E}[G_t \mid \mathcal{F}_t]$, for $t = 1, \dots, T$. A kernel cdf over this sample is estimated and denoted by $K_{cdf}(\cdot)$. Then, $K_{cdf}(\mathrm{E}[G_t \mid \mathcal{F}_t])$ represents a realization of an empirical kernel cumulative distribution function of the filtered mixing random variable, for each $t = 1, \dots, T$.

By construction of the COMFORT model, we have $(G_{T+1} \mid \mathcal{F}_T) \sim GIG(a, b, p)$. Denote with $\mathrm{GIG}_{\mathcal{F}_T}(\cdot)$ the corresponding cumulative distribution function. One then, for each $t = 1, \dots T$, obtains standardized conditional (on the information $\mathcal{F}_T$ available at time T) mixing random variables as

$$\tilde{G}^{(t)} = \mathrm{GIG}_{\mathcal{F}_T}^{-1}\left(K_{cdf}\left(\mathrm{E}[G_t \mid \mathcal{F}_t]\right)\right),$$

[19] Filtered historical simulation can be seen as a non-parametric bootstrap for time-series models.

where $\mathrm{GIG}^{-1}_{\mathcal{F}_T}(\cdot)$ denotes the (generalized) inverse of $\mathrm{GIG}_{\mathcal{F}_T}(\cdot)$ with the parameters estimated from the data. A $(T \times 1)$-dimensional vector $\tilde{\mathbf{G}}$ is a vector of standardized mixing random variables, where each vector component t, for $t = 1, \ldots, T$, corresponds to the equivalent entry in the matrix of standardized residuals $\tilde{\mathbf{Z}}$.[20] For $T+1$, the filtered historical simulation of $\mathbf{Y}^{FHS,(i)}_{T+1|\mathcal{F}_T}$ is carried out by

$$\mathbf{Y}^{FHS,(i)}_{T+1|\mathcal{F}_T} = \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\gamma}}\tilde{G}_{T+1|\mathcal{F}_T} + \sqrt{\tilde{G}_{T+1|\mathcal{F}_T}}\hat{\mathbf{S}}_{T+1|\mathcal{F}_T}\tilde{\mathbf{Z}}_{T+1|\mathcal{F}_T}, \quad i = 1, 2, \ldots B,$$

where a scalar $\tilde{G}_{T+1|\mathcal{F}_T}$ and a $(K \times 1)$-dimensional vector $\tilde{\mathbf{Z}}_{T+1|\mathcal{F}_T}$ are realizations of $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{Z}}$, for a randomly chosen t, consistent among both variables. Moreover, $\hat{\mathbf{S}}_{T+1|\mathcal{F}_T}$ is a $(K \times K)$-dimensional diagonal matrix of strictly positive conditional scale terms $s_{k,T+1|\mathcal{F}_T}$, where

$$\hat{s}^2_{k,T+1|\mathcal{F}_T} = \hat{\omega}_k + \hat{\alpha}_k(Y_{k,T} - \hat{\mu}_k - \hat{\gamma}_k \mathrm{E}[G_T \mid \mathcal{F}_T])^2 + \hat{\beta}_k \hat{s}^2_{k,T}.$$

The h-step prediction, where $h > 1$, is recursively obtained through

$$\mathbf{Y}^{FHS,(i)}_{T+h|\mathcal{F}_T} = \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\gamma}}\tilde{G}_{T+h|\mathcal{F}_T} + \sqrt{\tilde{G}_{T+h|\mathcal{F}_T}}\hat{\mathbf{S}}_{T+h|\mathcal{F}_T}\tilde{\mathbf{Z}}_{T+h|\mathcal{F}_T}, \quad i = 1, 2, \ldots B,$$

where $\tilde{G}_{T+h|\mathcal{F}_T}$ and $\tilde{\mathbf{Z}}_{T+h|\mathcal{F}_T}$ are randomly and consistently sampled from $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{Z}}$, and $\hat{\mathbf{S}}_{T+h|\mathcal{F}_T}$ is constructed via

$$\hat{s}^{2,(i)}_{k,T+h|\mathcal{F}_T} = \hat{\omega}_k + \hat{\alpha}_k(Y^{FHS,(i)}_{k,T+h-1|\mathcal{F}_T} - \hat{\mu}_k - \hat{\gamma}_k\tilde{G}_{T+h-1|\mathcal{F}_T})^2 + \hat{\beta}_k \hat{s}^{2,(i)}_{k,T+h-1|\mathcal{F}_T}, \quad i = 1, 2, \ldots B.$$

On every step of the described process, a random date t (among $1, 2, \ldots, T$) is selected and applied to the associated set of standardized residuals and corresponding standardized mixing random variables. In such a way, the dependence structure between the modeled multivariate return series is preserved. Note that the matrix $\tilde{\mathbf{Z}}$ is constructed in a non-parametric manner, whereas the vector $\tilde{\mathbf{G}}$ is established via the parametric assumption of the $GIG(\cdot)$ distribution. This combination of a parametric and a non-parametric part of the procedure renders our approach semi-parametric, conditional on the COMFORT dynamics.

The construction of the h-period cumulative filtered return $\mathbf{Y}^{FHS,(i)}_{T+h|\mathcal{F}_T}(h)$ is exactly the same as in the parametric case described in Eqs. (2.2.16) and (2.2.17). At first, the single-period returns are obtained recursively. Next, they are aggregated to cumulative h-period returns. The procedure is repeated B times to obtain the predicted FHS distribution of the multivariate return vector. This computationally cheap numerical algorithm renders our model general and tractable.

For a large enough number of simulations B, it is now possible to numerically optimize an arbitrary risk

[20] Consistent sampling of $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{Z}}$ is essential in order to preserve the implicit conditional dependence structure and the stability of the simulated dynamics.

measure/utility function. For example, the extended filtered historical simulation can be used to generate a forward-looking sample of multivariate returns. Then, the currency exposure can be optimized with respect to the expected shortfall by formulating the optimization problem in the linear programming format, such as presented in Rockafellar and Uryasev [2000].

2.2.4 Currency Hedging Strategy

Here, we propose a dynamic currency hedging strategy. This is a fixed plan which performs trading (i.e., hedging) decisions through time, applying the previously discussed estimation, simulation, and optimization approaches. The main idea is to run the currency exposure optimization procedure for every pre-specified rolling window[21] (e.g., quarterly or monthly) and enter the computed positions in the currency forward contracts for each currency in the portfolio over time. Below, the currency hedging strategy algorithm is specified.

Algorithm 1 Dynamic currency hedging strategy algorithm

Input: Initial portfolio (number of shares of each stock, or an equivalent number of fixed-income instruments, commodity contracts, or other investment positions), a base currency, a risk measure to optimize, constraints on the currency exposure, and a maturity of the hedge.
Output: Optimal (constrained) currency exposure given a chosen risk measure over time.

for `each rolling window` **do**

1) ESTIMATION: Calculate the current asset weights and estimate the chosen multivariate non-Gaussian time-series model.
2) FILTERED HISTORICAL SIMULATION: Perform the extended filtered historical simulation to construct forward-looking cumulative returns.
3) OPTIMIZATION: Given the chosen risk measure and constructed cumulative returns run the numerical optimization of the (possibly constrained) optimal currency exposure.
4) HEDGE: Enter into the computed amount of currency forward contracts for each currency in the portfolio. At maturity, the payoffs of the held currency forward contracts occur and are kept as domestic cash.

end for

The estimation part of the algorithm depends on the chosen underlying model, e.g. the Gaussian-GARCH or the non-Gaussian returns model we propose. Then, the (extended) filtered historical simulation is performed in order to construct the forward-looking cumulative returns consistent with the chosen time-series model. The optimization part depends on the chosen risk measure. The unconstrained robust mean-variance optimization can be performed with the analytically tractable procedure, while for general risk measures, such as the constrained expected shortfall, numerical optimization (e.g., linear programming) techniques are employed. Given the optimized values, the computed amount of currency forward contracts are entered.

[21] Note that the frequency of hedging should practically not be too high because of the induced hedging transaction costs.

At maturity, the forward payoffs occur[22] and the whole process is repeated over time (i.e., for each rolling window). We test the out-of-sample performance of the presented currency hedging strategy in the next section.

2.3 Empirical Analysis

In this section, we empirically investigate the dynamic currency hedging strategy presented above. An out-of-sample backtest is conducted and the performance of the proposed non-Gaussian hedging strategy is compared to the benchmarks of constant and Gaussian-GARCH-induced hedging, net of transaction costs.

2.3.1 Data

The empirical analysis covers 10 developed economies: Australia, Canada, Switzerland, Denmark, Eurozone, United Kingdom, Japan, New Zealand, Sweden, and United States. The time-series data of spot and forward currency exchange rates, short-term interest rates, equity market total return indices, and government bond total return indices are employed. The data series are available at a daily frequency and are obtained from Thomson Reuters Datastream and Bloomberg. The sample period starts on 01/01/1999, when the euro was introduced to the world financial markets, and ends on 12/31/2019.

The data enables us to test the out-of-sample performance of the hedging strategy over some major events that happened in the world financial markets and had a profound effect on the currency exchange rates, such as the global financial crisis of 2008–2009, the European sovereign debt crisis of 2009-2011, the Swiss franc unpeg of 2015, the Brexit referendum of 2016, and others. We investigate the performance of the dynamic non-Gaussian currency hedging strategy and compare it to constant hedging benchmarks, and Gaussian (i.e., non-ambiguous) approaches presented in the existing literature.

Table 2.1 reports the full sample annualized mean, annualized volatility, skewness, and kurtosis of local equity, local bond, and currency excess (i.e., for the USD base currency) returns. Equity and bond mean returns are consistently positive among all investigated countries, whereas around half of the countries in our sample exhibit positive mean currency excess returns. Volatilities are the lowest for bonds, followed by currencies, and are the largest for equities. While equity and bond returns generally exhibit negative skewness, currency return skewness changes sign across different countries. Bond returns display a mild excess kurtosis, whereas equity and currency returns exhibit large excess kurtosis, consistently among all countries. Those summary statistics are in line with the commonly known empirical stylized facts of financial asset returns.

[22] In the out-of-sample backtest analyzed in the next section, we assume that the payoffs from forward currency contracts are kept as domestic cash.

Summary Statistics

	AUD	CAD	CHF	DKK	EUR	GBP	JPY	NZD	SEK	USD
Equities										
Mean	8.88%	8.56%	6.34%	11.07%	6.82%	6.75%	5.95%	9.93%	11.09%	8.11%
Vol	14.80%	15.80%	16.25%	17.93%	19.86%	17.26%	20.35%	10.16%	22.33%	18.00%
Skew	−0.40	−0.45	−0.18	−0.24	−0.10	−0.09	−0.23	−0.45	0.08	−0.08
Kurt	8.77	12.96	10.26	9.63	10.34	9.54	9.65	7.81	7.24	11.88
Bonds										
Mean	4.96%	3.75%	1.97%	3.27%	3.14%	3.90%	1.02%	4.92%	3.44%	3.92%
Vol	3.12%	2.61%	1.64%	2.43%	2.36%	2.54%	1.22%	2.19%	2.40%	3.21%
Skew	0.09	−0.06	0.07	−0.19	−0.10	−0.08	−0.09	0.06	−0.03	−0.01
Kurt	5.26	5.41	7.11	5.92	5.48	5.38	8.44	6.70	5.80	5.68
Currencies										
Mean	3.36%	1.27%	0.43%	−0.37%	−0.41%	−0.22%	−1.55%	4.24%	−0.46%	·
Vol	12.21%	8.77%	10.48%	9.42%	9.45%	9.20%	9.97%	12.74%	11.36%	·
Skew	−0.62	0.16	1.15	0.22	0.21	−0.42	0.32	−0.29	0.23	·
Kurt	14.89	8.65	31.72	5.85	5.81	13.21	6.93	8.17	6.72	·

Table 2.1: This table presents the full sample unconditional annualized mean, annualized volatility, skewness, and kurtosis of local equity, local bond, and currency (against the USD) returns for 10 different economies.

Figure 2.1 depicts the quantile-quantile plot of the normalized empirical sample of the currency exchange rate (with USD base currency) and the equally-weighted fully hedged portfolio returns plotted versus the standard normal distribution. Again, one can observe the well-known stylized fact of the heavy-tailed return distribution present in all currency pairs as well as the fully hedged portfolio returns. These observations motivate the use of a non-Gaussian hedging strategy presented in the theoretical part of this paper.

Figure 2.2 compares a single asset cumulative return distribution constructed from both the parametric simulation and the filtered historical simulation (FHS) given the assumed COMFORT dynamics. Both approaches yield similar distributions showing that the model is well parameterized. However, one can observe that the forecasted distribution arising from FHS is more negatively skewed and heavy-tailed. This is more in line with what can be observed empirically; hence, we decided to use the FHS approach in the out-of-sample backtest analysis.

2.3.2 Out-of-Sample Backtest

The backtest analysis presented in this section is performed with hedging on a quarterly[23] time horizon. It is conducted for international portfolios comprised of three asset classes: equity, bonds, and cash. Moreover, we investigate various choices of investor's base currency. All results analyzed in this section are presented net of transaction costs, which are assumed to be two basis points relative to the notional of the entered

[23] Note that the results are robust also to alternative specifications.

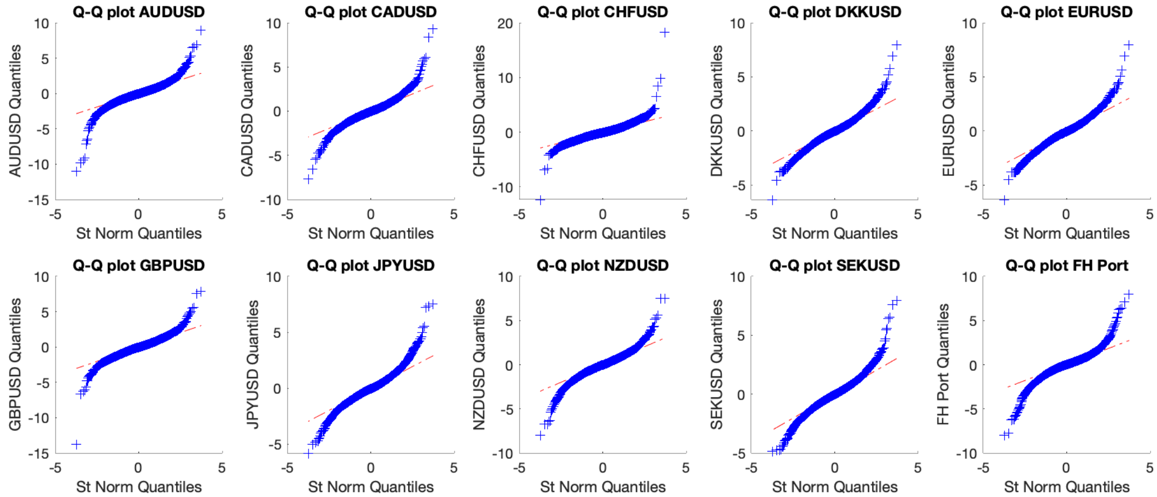

Figure 2.1: Quantile-quantile plot of the normalized excess currency returns (with respect to the USD) and the fully hedged (FH) equally-weighted 75/25 portfolio returns versus the standard normal distribution. Observe that the empirical sample of excess currency and the fully hedged portfolio returns exhibit heavier tails compared to the Gaussian distribution.

currency forward contract for each foreign currency present in the portfolio.

In practice, a currency forward contract is constructed with two different types of transactions: an FX spot and an FX swap. The spot transaction offsets the first leg of the FX swap, and what one is left with is equivalent to the currency forward. Note that transaction costs are paid for both spot and swap transactions. However, when an investor is rolling over currency forward contracts over time (e.g., as in our dynamic currency hedging setting), the initial spot transaction can be avoided, and only a new FX swap can be entered at maturity of the old one. This is important because transaction costs on FX swaps are considerably lower compared to a forex spot transaction, see **?**. In the light of falling transaction costs in the last years, Jahan-Parvar and Zikes [2019] show that the effective spread on the most traded foreign exchange rates is below one basis point, meaning that the transaction costs on the FX swaps are even lower. Therefore, our choice of transaction costs of two basis points can be interpreted as very conservative.

We analyze a buy-and-hold portfolio specified at the beginning of the backtest such that the portfolio is equally weighted. The portfolio is specified in terms of the number of shares of each stock or the equivalent for the fixed income or cash positions. Throughout the backtest, we do not rebalance these asset positions. Their relative portfolio weights change over time, but not the absolute weights set at the beginning of the backtest. Hence, we work with a so-called buy-and-hold portfolio. When a currency forward contract used for hedging expires, its payoff is kept as domestic cash. Hence, an investor changes only positions in currency forward contracts (i.e., as determined by different estimation models or kept constant in the case of constant hedging) and domestic cash (i.e., depending on the currency forward payoffs). We apply such a procedure since we aim to test the performance of different currency hedging strategies in isolation from any other

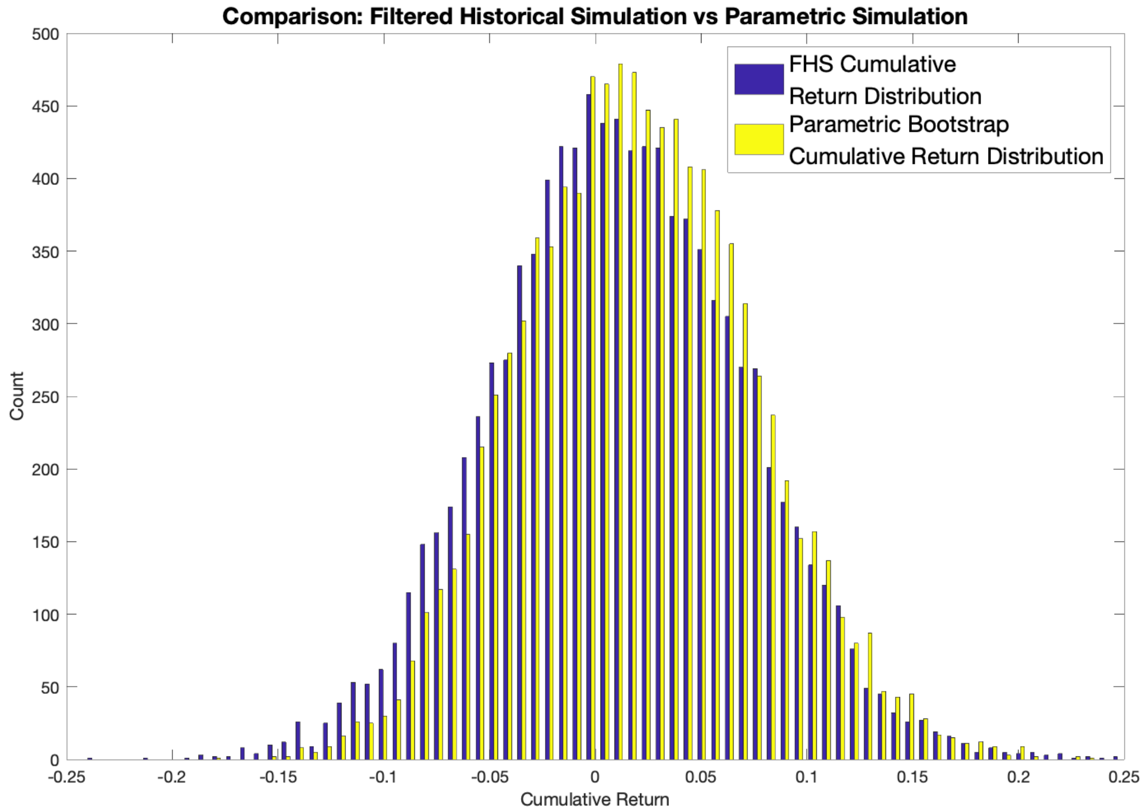

Figure 2.2: Comparison between the cumulative return distribution constructed from parametric simulation and filtered historical simulation, given the COMFORT model dynamics.

effects on portfolio performance—it is a standard approach in this type of analysis; among others, see also De Roon et al. [2003] and Cho et al. [2020].

We start the empirical analysis by investigating the dynamic optimal currency exposure obtained from different models and for different risk measures. We compare investors who manage their currency exposure with respect to (i) the robust mean-variance (MV), and (ii) the minimum expected shortfall (ES) at a confidence level $\alpha = 0.85$. To illustrate the performance of the FHS algorithm from Section 2.2.3.2, we work with the minimum expected shortfall portfolio optimization, which is a limiting case of Eq. (2.2.14) with $\lambda \to \infty$. The general case of Eq. (2.2.14) is a simple extension using the moments given in Eq. (2.2.12). Note that, when $\lambda \to \infty$, at least for the mean-variance portfolio and for all levels of ambiguity aversion $\theta \geq 0$, we can see from Eq. (2.2.13), that our non-Gaussian returns model still induces optimal currency allocations adjusted for non-Gaussianity, through the first two moments of the mixing random variable.

The hedging decisions are driven by a CCC-GARCH-based multivariate normal (MN) model and by the non-Gaussian (COMFORT) time-series model, as presented in previous sections. The rolling calibration window consists of one year of historical data. The analyzed portfolio is set up at the beginning of the backtest as equally-weighted among 10 different economies (meaning $w_{c,0} = 0.1$ for each currency c) where

75% of wealth is put into equities and 25% into government bonds. For practical reasons, such as regulatory constraints, we constrain the optimal currency exposure to lie on the interval $[-2w_{c,0}, 3w_{c,0}]$ for each currency c over the whole backtest.[24] This constraint can be interpreted as a form of shrinkage; see for example Jagannathan and Ma [2003].

Optimizing the robust mean-variance utility only depends on the first two moments of the forecasted cumulative return distribution. Therefore, we expect that hedging decisions arising from MN and COMFORT models, in this case, do not differ much. Subplot (a) of Fig. 2.3 depicts the optimal MV exposure in GBP for a EUR-based investor.[25] In this, as well as other subsequent plots, the optimal exposure is expressed in relative terms compared to the total portfolio value. Observe that both models generally agree on the optimal amount of exposure in GBP, as expected by the dependence on only the first two moments, whilst the COMFORT model yields more stable hedging decisions over time.

The difference in estimated optimal exposure arising from the Gaussian and non-Gaussian models becomes more distinct when optimizing the ES. This occurs because ES depends on higher moments (i.e., the tail of the return distribution), which is not the case for MV utility. One can observe the optimal ES exposure in DKK for a USD-based investor in subplot (b) of Fig. 2.3. The asymmetries captured in the COMFORT model play a large role in the computed optimal currency exposure. Notice that the optimal ES exposure is generally larger in absolute terms in the non-Gaussian case. There, the changing market conditions and ambiguity are captured through mixing in the non-Gaussian model. It is intuitively clear that employing a non-Gaussian model yields more distinct optimized currency exposures than the Gaussian model when a tail-based risk measure is utilized for the optimization.

To end the analysis of optimal currency exposures, let us focus on the non-Gaussian hedging strategy and investigate the difference between the corresponding optimal MV and ES currency exposures. Subplot (c) of Fig. 2.3 shows optimal minimum variance and ES exposures in JPY for a EUR-based investor. Since hedging is used to manage risk from either variance or ES perspective, one can observe that both objectives produce similar patterns of time-varying optimal currency exposure. Nonetheless, slightly more excessive positions are computed when optimizing ES in comparison to variance.

In Table 2.2, observe the results of the out-of-sample backtest comparing the performance of model-based currency hedging to the benchmarks of zero and full hedging. We investigate the performance of MN and COMFORT-based hedging. The optimization is conducted for MV and ES risk measures. The 75/25 portfolio comprised of 10 different countries is analyzed for different choices of the base currency, and all results are presented net of transaction costs. Moreover, the results in Table 2.2 are presented for the risk and ambiguity aversion parameters $\lambda = 4$ and $\theta = 5$, respectively. Further analysis around the choice of λ

[24] Let us note that other equivalent constraint specifications do not affect the validity of results.

[25] Note that the portfolio consists of 10 different currency exposures, whereas we only plot a single one here for the sake of clarity.

Hedging strategy out-of-sample backtest results

75/25 Portfolio	No Hedge	Full Hedge	MV-MN	ES-MN	MV-COM	ES-COM
Base: CAD						
Volatility	10.07%	9.03%	7.46%	7.51%	**7.08%**	7.43%
Mean	**7.01%**	6.53%	6.06%	6.32%	6.52%	6.75%
Sharpe Ratio	0.70	0.72	0.81	0.84	**0.92**	0.91
Sortino Ratio	0.98	1.00	1.15	1.20	**1.32**	1.29
CEQ	4.98%	4.90%	4.94%	5.19%	5.52%	**5.65%**
Max Drawdown	37.84%	41.34%	33.49%	33.59%	28.17%	**27.79%**
Turnover	0	0.96	1.69	1.41	1.60	1.66
Base: CHF						
Volatility	13.34%	8.91%	7.47%	8.37%	**6.95%**	7.48%
Mean	**5.30%**	4.69%	4.16%	4.01%	4.49%	4.64%
Sharpe Ratio	0.40	0.53	0.56	0.48	**0.65**	0.62
Sortino Ratio	0.55%	0.72%	0.79%	0.68%	**0.92%**	0.88%
CEQ	1.74%	3.10%	3.04%	2.60%	**3.53%**	3.52%
Max Drawdown	51.75%	43.82%	36.70%	39.20%	**30.43%**	31.75%
Turnover	0	0.93	1.83	1.61	1.74	1.84
Base: EUR						
Volatility	10.69%	8.49%	6.99%	7.56%	**6.68%**	7.14%
Mean	**6.62%**	5.85%	5.53%	5.73%	5.79%	5.86%
Sharpe Ratio	0.62	0.69	0.79	0.76	**0.87**	0.82
Sortino Ratio	0.86	0.95	1.12	1.08	**1.24**	1.16
CEQ	4.33%	4.41%	4.55%	4.58%	**4.89%**	4.84%
Max Drawdown	44.92%	36.67%	29.34%	31.71%	24.22%	**24.09%**
Turnover	0	0.91	1.69	1.44	1.64	1.80
Base: GBP						
Volatility	11.18%	9.73%	8.10%	9.06%	**7.43%**	7.60%
Mean	**8.27%**	7.47%	7.00%	7.17%	7.28%	7.58%
Sharpe Ratio	0.74	0.77	0.86	0.79	0.98	**1.00**
Sortino Ratio	1.05	1.07	1.23	1.11	1.40	**1.42**
CEQ	5.77%	5.58%	5.69%	5.53%	6.17%	**6.43%**
Max Drawdown	31.27%	41.43%	35.58%	37.87%	26.04%	**25.56%**
Turnover	0	1.05	1.96	1.70	1.82	1.82
Base: USD						
Volatility	13.40%	11.75%	9.88%	10.56%	**9.06%**	9.09%
Mean	**8.17%**	7.24%	6.86%	7.27%	7.10%	7.33%
Sharpe Ratio	0.61	0.62	0.69	0.69	0.78	**0.81**
Sortino Ratio	0.86	0.85	0.98	0.97	1.11	**1.14**
CEQ	4.57%	4.48%	4.91%	5.04%	5.46%	**5.67%**
Max Drawdown	51.44%	53.28%	48.56%	49.07%	42.51%	**38.86%**
Turnover	0	1.17	2.02	1.67	1.89	1.91

Table 2.2: Observe the realized annualized portfolio return volatility, mean, Sharpe and Sortino ratios, certainty equivalent (CEQ) for the risk aversion parameter of $\lambda = 4$, maximum drawdown, and turnover of different model-based and constant hedging strategies obtained in the out-of-sample backtest and corrected for transaction costs. We present the results for the buy-and-hold portfolio initially comprised of 75% equity and 25% bonds, equally-weighted between 10 different economies. We analyze the multivariate normal (MN) model based on the CCC-GARCH time-series structure and the proposed non-Gaussian COMFORT (COM) time-series model. Currency hedging based on optimizing the robust mean-variance (MV) and expected shortfall (ES) risk measures are investigated for different choices of the base currency.

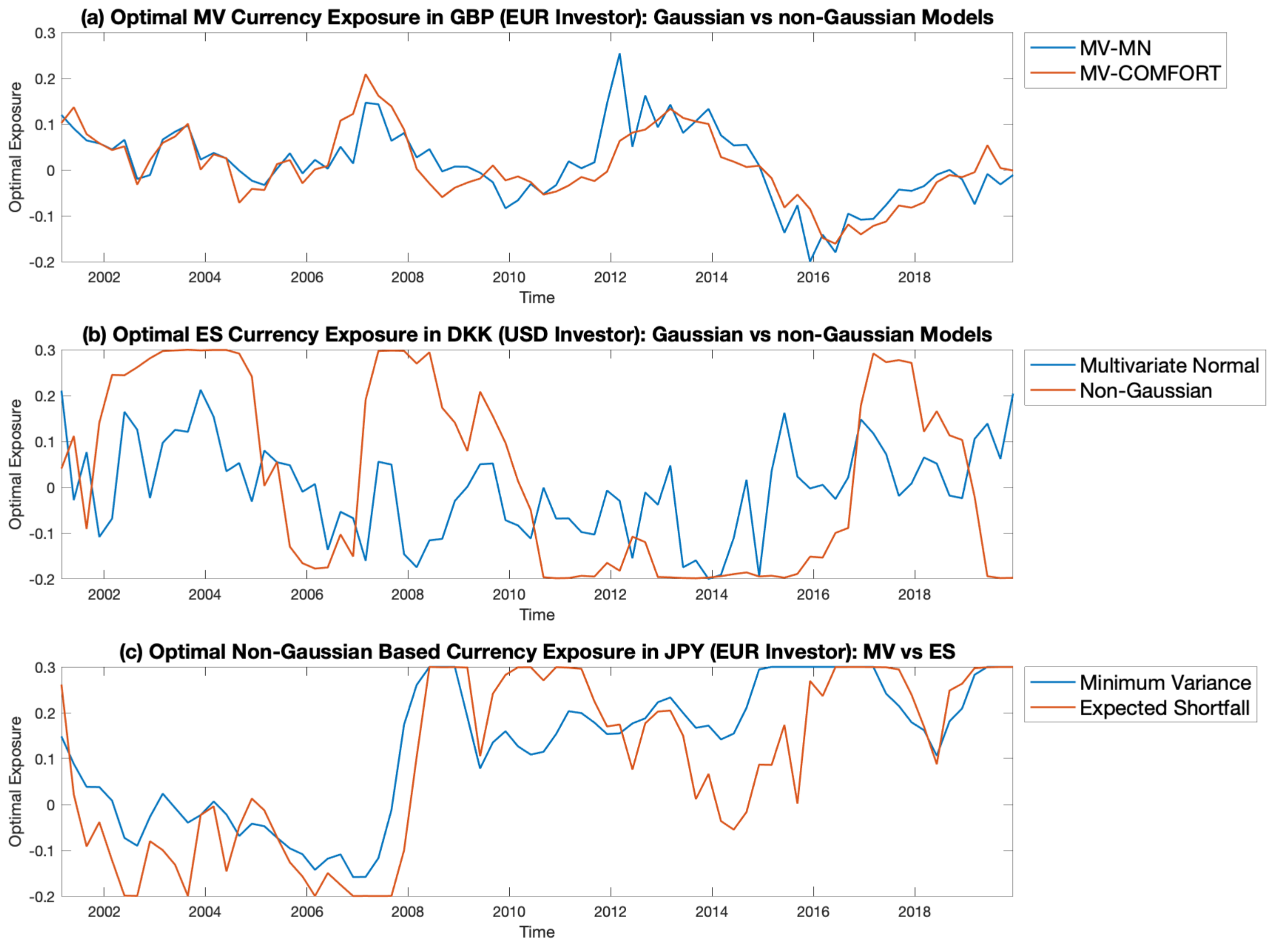

Figure 2.3: Subplot (a) depicts the optimal robust mean-variance (MV) exposure in GBP for a EUR-based investor as computed by multivariate normal (MN) and non-Gaussian (COMFORT) models. Subplot (b) shows optimal expected shortfall (ES) exposure in DKK for a USD-based investor as computed by multivariate normal and non-Gaussian models. Subplot (c) illustrates the optimal minimum variance and minimum expected shortfall exposures in JPY for a EUR-based investor as computed by the non-Gaussian model.

and θ is presented in Fig. 2.4.

Observe that all hedging strategies, in general, reduce the risk of the portfolio compared to constant (i.e., zero and full) hedging. This is expressed through the reduction of volatility and maximum drawdown. The annualized mean of the portfolio returns is the highest in the case of no hedging. One factor driving this result is that no hedging is the only hedging strategy unaffected by transaction costs. Importantly, model-induced hedging consistently outperforms the constant hedging benchmarks in terms of risk-adjusted returns measured by Sharpe and Sortino[26] ratios as well as the certainty equivalent[27] (CEQ) return.

[26] Sortino ratio penalizes negatively skewed portfolio return distributions by computing the volatility of returns only below the risk-free rate, hence accounting for the downside risk.

[27] The CEQ return indicates the certain return a risk-averse investor would accept rather than investing in a strategy with the same but uncertain return. The measure directly captures the standard mean-variance utility function with a predetermined

We express the average (i.e., quarterly) hedging turnover $\overline{\mathrm{HT}^s}$ of a currency hedging strategy s as

$$\overline{\mathrm{HT}^s} = \frac{1}{T}\sum_{t=1}^{T}\sum_{c=2}^{K+1}|\phi^s_{c,t}|, \tag{2.3.1}$$

where T is the number of trading instances, K is the number of foreign currency exposures, and $\phi^s_{c,t}$ is the notional of a currency forward contract used for hedging currency c at time t. The notional is expressed in relative terms compared to the total portfolio value. An average amount of trading for a hedging strategy s computed over the whole backtest is indicated by $\overline{\mathrm{HT}^s}$. Table 2.2 shows that the Gaussian-based MV hedging strategy on average yields higher turnover compared to its non-Gaussian analog. This is a result of increased stability of the non-Gaussian hedging, as also observed in Fig. 2.3. On the other hand, the non-Gaussian currency hedging strategy generates, on average, a larger trading turnover compared to its Gaussian equivalent in the case of ES optimization. The optimization of ES is more stable in the case of the non-Gaussian model. However, it produces more excessive positions, as observed in Fig. 2.3, which generate on average larger turnover. Moreover, given the regulatory constraints or the investor preferences, the turnover of the hedging strategy can be additionally controlled by specifying the currency exposure constraints in the optimization process.

The out-of-sample backtest demonstrates that the non-Gaussian currency hedging strategy consistently outperforms its Gaussian counterpart net of transaction costs. Given any possible portfolio base currency, the proposed non-Gaussian returns model constantly achieves the highest Sharpe/Sortino ratio and the CEQ return, and at the same time, the lowest maximum drawdown. In the case of risk-adjusted returns, the outperformance of MV-COMFORT versus ES-COMFORT depends on the chosen base currency, while in four out of five cases, the ES-COMFORT model outperforms in terms of maximum drawdown. This outperformance is driven by the non-Gaussianity (expressed through the asymmetry and heavy-tails) of the forecasted cumulative returns and ambiguity as captured by the non-Gaussian time-series model. When additional information provided by the asymmetry parameters of the model is utilized, the forecasting power of the cumulative return distribution increases, and the outperformance in comparison to the (non-ambiguous) Gaussian CCC-GARCH model results.

Figure 2.4 depicts the out-of-sample realized Sharpe ratio for the ambiguity-adjusted mean-variance model given various base currencies in dependence of risk and ambiguity aversion parameters λ and θ, respectively. Observe that the parameter of risk aversion has a larger impact on the realized Sharpe ratio compared to the parameter of ambiguity aversion. This is also expected given the derived optimal exposure from Eq. (2.2.13). Moreover, ambiguity aversion is more relevant when an investor exhibits low risk aversion—the slope in the ambiguity aversion direction of the Sharpe ratio surface is larger for small risk

level of the investor's risk aversion.

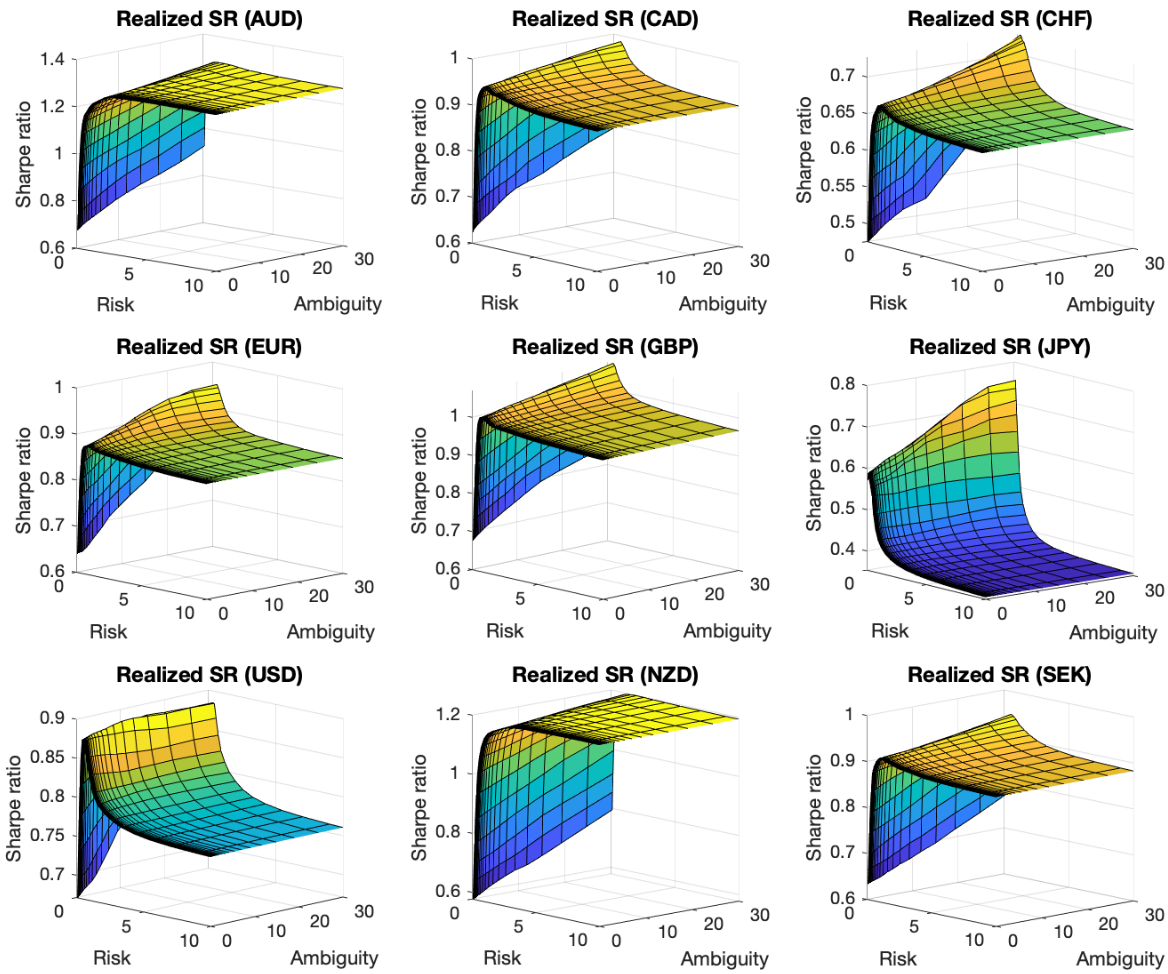

Figure 2.4: Out-of-sample realized Sharpe ratio surfaces from a rolling window exercise for different risk and ambiguity aversion parameters λ and θ, respectively, for the ambiguity-adjusted mean-variance model from Eq. (2.2.10). The subplots represent different base currencies given the analyzed 75/25 portfolio.

aversion than for large risk aversion. When investors admit low risk aversion, it is beneficial to be more ambiguity averse, as illustrated in the increased realized Sharpe ratio. This shows that ambiguity becomes an essential component in portfolio optimization when investors are taking more risks. On the other hand, for larger values of risk aversion, the portfolio performance stabilizes and converges towards the minimum variance-based Sharpe ratio.

Figure 2.5 shows the portfolio drawdown of the USD-based portfolio from Table 2.2. One can observe that model-induced hedging substantially reduces the portfolio drawdown over the whole period of the backtest. This is the most prominent in the time of the global financial crisis of 2008–2009. Many investors who are, for example, withdrawing retirement money from their pension funds are mostly concerned about drawdowns. Consequently, large drawdowns can be extremely problematic not only for retirees but also

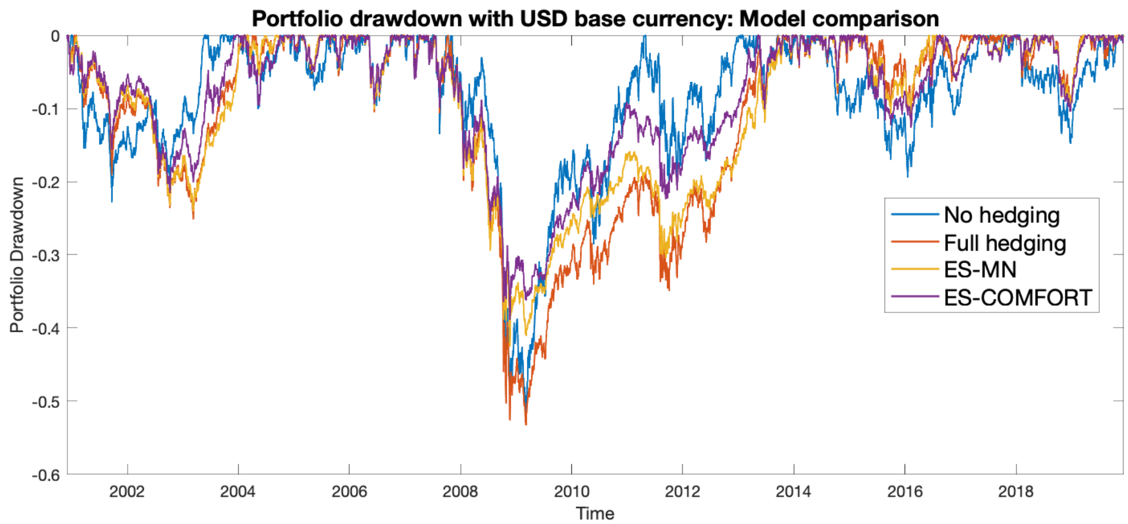

Figure 2.5: This plot shows a drawdown of the analyzed 75/25 portfolio for a USD-based investor. Zero hedging, full hedging, and the expected shortfall-based hedging as computed by multivariate normal (MN) and non-Gaussian (COMFORT) based currency hedging strategies are presented.

for asset managers. We show that when investors use the non-Gaussian returns model and the expected shortfall as a risk measure, they can significantly reduce portfolio drawdowns and achieve higher risk-adjusted portfolio returns net of transaction costs.

Portfolio drawdowns can also be driven by specific events occurring in financial markets that have profound effects on currency exchange rates. A prominent example of such an event is the decision of the Swiss National Bank to scrap its currency peg of 1.20 to the euro on January 15, 2015. The Swiss franc immediately appreciated against almost all other currencies, most notably around 20% against the euro. Another example of an event that has a large effect on currency markets is the Brexit referendum where voters in the United Kingdom, on June 23, 2016, decided to withdraw from the European Union. Consequently, the British pound depreciated against almost all major currencies.

In Fig. 2.6, we study the effect of these two events on the portfolio performance. We make use of the portfolio return decomposition as given in Eq. (2.2.3), where hedged portfolio return is presented as a sum of two components: i) fully hedged asset return, and ii) net currency exposure return. We observe the effect of the proposed dynamic currency hedging strategy on the currency component of the hedged portfolio return. Subplot (a) presents the decomposed cumulative return plot around the period of the Swiss franc unpeg for a CHF-based investor. Since other currencies depreciated against the CHF, one can observe a large drop in the currency component (i.e., summed over all foreign currencies present in the portfolio) for the no hedging benchmark. Full hedging straightforwardly represents a currency component that is through the whole backtest close to zero. Minuscule deviations are present because of changing interest rates and the fact that perfect hedging is impossible in practice. More importantly, observe the large positive outperformance of the

Figure 2.6: Subplot (a) illustrates the cumulative performance decomposition of fully hedged assets and net currency exposures for different currency hedging strategies and a CHF-based investor around the Swiss franc unpeg. Subplot (b) depicts the cumulative performance of the GBP net currency exposure for different currency hedging strategies and a USD-based investor around the Brexit referendum.

currency component of the portfolio managed using the proposed non-Gaussian currency hedging strategy for both MV and ES-based optimization. A similar pattern can also be recognized in subplot (b). There, a net currency exposure in GBP (i.e., in isolation from other currency exposures) is plotted for a USD-based investor. One can study the effect of the Brexit referendum on the unfavorable performance of the GBP currency component in the portfolio without currency hedging. On the other hand, the non-Gaussian currency hedging strategy again prevents the portfolio drawdown and at the same time also provides an outperformance in comparison to full hedging.

The empirical analysis shows that the non-Gaussian currency hedging strategy outperforms constant and Gaussian-based hedging strategies. This gives rise to the study of the mixing random variable that drives the non-Gaussianity and ambiguity. Recall that the mixing random variable is given by $G_t \sim GIG(a, b, p)$, with the generalized inverse Gaussian distribution, as described in Section 2.2.2. To simplify the analysis we fix parameters $a = 2$ and $b = 0$. This transforms the GIG distribution of the mixing random variable into a particular case of the gamma distribution. A normal mean-variance mixture with gamma distribution is known as the variance-gamma distribution.[28] Given its ability to model asymmetry and heavy tails, the variance-gamma distribution is well known in financial literature, see Madan and Seneta [1990].

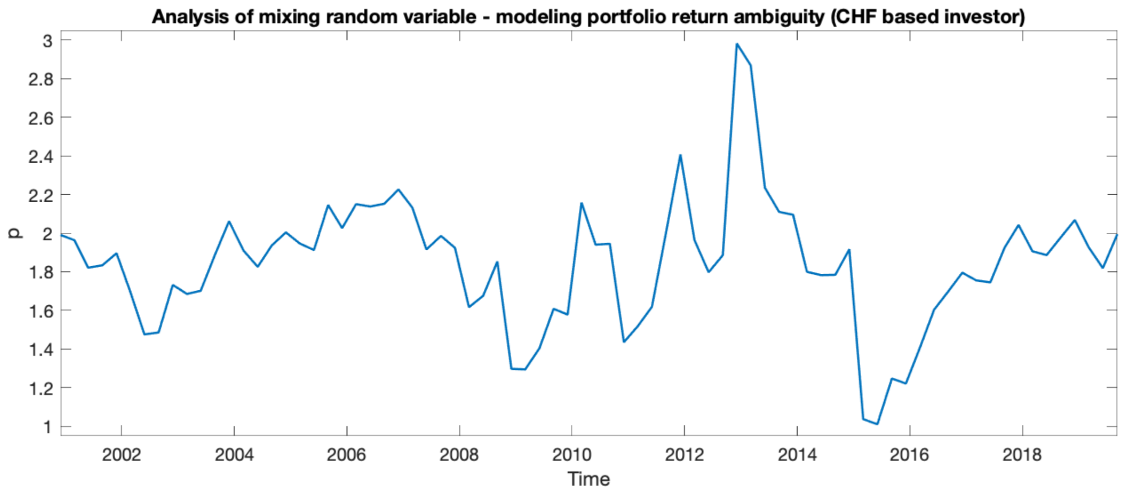

Figure 2.7: This figure shows the rolling window forecasted values of parameter p for a CHF-based ES minimizing investor, for $a = 2$ and $b = 0$ in $G_t \sim GIG(a, b, p)$, where GIG denotes the generalized inverse Gaussian distribution. Since the portfolio return tail probability and kurtosis are decreasing functions of p, lower values of parameter p proxy for a larger international portfolio return ambiguity.

In Section 2.2.3, we showed that the mixing random variable can be interpreted as a proxy of ambiguity about the portfolio and currencies return distribution. Here, we study this mixing random variable in a special case of the variance-gamma distribution. Given the fixed values of parameters a and b, we plot the forecasted values of parameter $p = \mathrm{E}[G_t]$ for G_t as returned by the EM algorithm throughout the rolling window backtest for the ES minimizing CHF-based investor in Fig. 2.7. Note that the tail probability is a decreasing function of parameter p, see Finlay [2009]. Furthermore, also the kurtosis of forecasted portfolio returns is a decreasing function of p, see Nestler and Hall [2019]. Therefore, smaller values of p proxy for a larger ambiguity expressed through heavier tails (i.e., increased kurtosis) of the forecasted portfolio return distribution. Observe how model ambiguity is increased during the period of the global financial crisis of 2008–2009 and the Swiss franc unpeg of 2015. Figure 2.7 shows that the non-Gaussian model is able to capture the increased ambiguity present in financial markets via a purely data-driven approach that utilizes only the historical returns data. The mixing random variable is a proxy of model ambiguity which in turn influences the currency allocation decision. Modeling and exploiting the non-Gaussianity and ambiguity are the key drivers of outperformance of the dynamic non-Gaussian currency hedging strategy proposed in this

[28] Note that the variance-gamma distribution is a special case of the generalized hyperbolic distribution.

paper.

2.4 Conclusions

The main goal of this paper is to develop a flexible dynamic currency hedging strategy that accounts for the stylized facts of financial returns and allows for a linkage to investor's ambiguity. The proposed hedging strategy enhances the existing approaches by modeling a set of asset and currency returns via a conditional multivariate continuous normal mean-variance mixture distribution with GARCH dynamics in the scale term. We show that the mixing random variable can be understood as a market-based measure of ambiguity from Maccheroni et al. [2013].

Moreover, we derive a semi-parametric extended filtered historical simulation method to model the future distribution of asset and currency returns. This method combines the parametric non-Gaussian time-series model employed in our study and a non-parametric simulation of historical portfolio returns. With the use of this tractable method, we can simulate a forecasted conditional cumulative distribution of portfolio returns and numerically optimize a currency hedge with respect to an arbitrary risk measure. Utilizing this procedure, we propose an algorithm for a dynamic non-Gaussian currency hedging strategy that can optimize a general risk measure in a multi-step-ahead forecast.

In the empirical part of the paper, we demonstrate the performance of the proposed non-Gaussian currency hedging strategy. An out-of-sample backtest on historical market data of 10 developed economies over the period from 1999 to 2019 is performed. The empirical results reveal that our non-Gaussian-based method yields a robust, stable, and highly risk reductive hedging strategy. It outperforms the benchmarks of constant hedging as well as equivalent approaches based on Gaussian-GARCH modeling in terms of risk-adjusted returns and portfolio drawdown net of transaction costs. This outperformance is driven by the asymmetries and heavy tails captured by the non-Gaussian returns model. Moreover, we show that the employed mixing random variable is able to proxy for ambiguity about future portfolio returns and incorporate this additional information in the currency allocation decision.

Our work allows for several theoretical and empirical extensions. Since currency forward contracts employed for hedging in the current setup are linear instruments, one could study hedging with currency options, especially to investigate the effect of mitigating the currency downside risk. In such a case, the setup would become non-linear, which would require an enhanced theoretical framework. Moreover, one could extend the dynamic currency hedging approach by allowing for currency re-hedging. Thereby, the optimization algorithm is run every day (or week), and an investor compares the currently hedged amount of foreign currency exposure with the newly computed optimal currency allocation taking into account the most recent market data. Then, depending on a prespecified re-hedging threshold governing the frequency of re-hedging, an investor decides whether a re-hedge of the foreign currency exposure is required. Furthermore, one could investigate the currency hedging of portfolios exposed to emerging market economies. Trading emerging market currencies exhibits considerably higher transaction costs compared to trading the developed market currencies. Hedging currencies with higher transaction costs could potentially be directly modeled by including an L^1 penalty term[29] to emerging market currencies in the optimization problem, yielding a potentially sparse currency hedge.

To summarize, an important result that emerges from our analysis is that accounting for non-Gaussianity in the econometric modeling of portfolio return dynamics appears to significantly improve the performance of the dynamic currency hedging strategy net of transaction costs. Moreover, we show that non-Gaussianity can be associated with ambiguity around the distribution of international portfolio returns. In addition to an academic interest in studying the problem of optimal currency hedging, this research topic is also practically relevant and widely discussed in the financial services industry, especially in the areas of strategic asset allocation and wealth management (see Chang, 2009, Bender et al., 2012, and de Boer, 2016).

[29] An L^1 constraint can be seen as a proxy for transaction costs, see Brodie et al. [2009].

2.5 Appendix

2.5.1 Derivation of Eq. (2.2.15)

This derivation applies the econometric model presented in Section 2.2.2. Using Eq. (2.2.4) we see that $\mathbf{Y}_t - \boldsymbol{\mu} - \boldsymbol{\gamma} G_t = \mathbf{H}_t^{\frac{1}{2}} \sqrt{G_t}\mathbf{Z}_t$, and by Eq. (2.2.6) it holds that

$$(Y_{m,t} - \mu_m - \gamma_m G_t)^2 = G_t s_{m,t}^2 \tilde{Z}_{m,t}^2,$$

for $m = 1, 2, \ldots, M$, where we define $\tilde{\mathbf{Z}}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma})$. Utilizing Eq. (2.2.7) we can express

$$\begin{aligned}
s_{m,t}^2 &= \omega_m + \alpha_m G_{t-1} s_{m,t-1}^2 \tilde{Z}_{m,t-1}^2 + \beta_m s_{m,t-1}^2 \\
&= \omega_m + \alpha_m s_{m,t-1}^2 \left(G_{t-1} \tilde{Z}_{m,t-1}^2 - G_{t-1} \right) + \beta_m s_{m,t-1}^2 + \alpha_m s_{m,t-1}^2 G_{t-1} \\
&= \omega_m + \alpha_m s_{m,t-1}^2 G_{t-1} \left(\tilde{Z}_{m,t-1}^2 - 1 \right) + \alpha_m s_{m,t-1}^2 \left(G_{t-1} - \mathrm{E}[G_{t-1}] \right) + \beta_m s_{m,t-1}^2 + \alpha_m s_{m,t-1}^2 \mathrm{E}[G_{t-1}].
\end{aligned}$$

Define $\hat{s}_{m,t+h|\mathcal{F}_t}^2 := \mathrm{E}\left[s_{m,t+h}^2 \mid \mathcal{F}_t \right]$ as the optimal prediction in the L^2 sense of $s_{m,t+h|\mathcal{F}_t}^2$. Since $\mathrm{E}\left[\tilde{Z}_{m,t+h}^2 - 1 \,|\, \mathcal{F}_t \right] = 0$ and $\mathrm{E}\left[G_{t+h} - \mathrm{E}[G_{t+h}] \,|\, \mathcal{F}_t \right] = 0$ for all $h > 0$, then

$$\hat{s}_{m,t+h|\mathcal{F}_t}^2 = \begin{cases} \hat{\omega}_m + \hat{\alpha}_m \left(Y_{m,t} - \hat{\mu}_m - \hat{\gamma}_m \mathrm{E}[G_t \mid \mathcal{F}_t] \right)^2 + \hat{\beta}_m \hat{s}_{m,t}^2, & \text{for } h = 1, \\ \hat{\omega}_m + \hat{s}_{m,t+h-1|\mathcal{F}_t}^2 \left(\hat{\alpha}_m \mathrm{E}[G_{t+h-1} \mid \mathcal{F}_t] + \hat{\beta}_m \right), & \text{for } h > 1, \end{cases}$$

which concludes the derivation of Eq. (2.2.15).

2.5.2 Pairwise Correlations

Figure 2.8 shows a heat plot of country pairwise unconditional (i.e., full sample) correlations between equity index, fixed income index, and currency exchange rate returns, considering the US dollar as the base currency. Different asset classes can be recognized by larger pairwise correlations. It is worth noting that the correlations of Japanese assets, especially bonds, are very low compared to other countries. Moreover, let us comment on the cross-correlations among different asset classes. The unconditional correlations between equities and bonds are stable across different countries and generally slightly negative. Also, the correlations between bonds and currencies are stable across different countries and take values around zero–this is the rationale for the optimality of full currency hedging in bond-only portfolios. The median correlation between bonds and currencies is -0.05. The largest variation in pairwise correlations among different countries is exhibited among equities and currencies; the median correlation here is 0.11. Note that such correlations change over time, whereas here, only the sample unconditional correlation is displayed. This motivates the use of the multivariate non-Gaussian returns model that captures some of the dynamics in the dependency between assets and currencies over time.

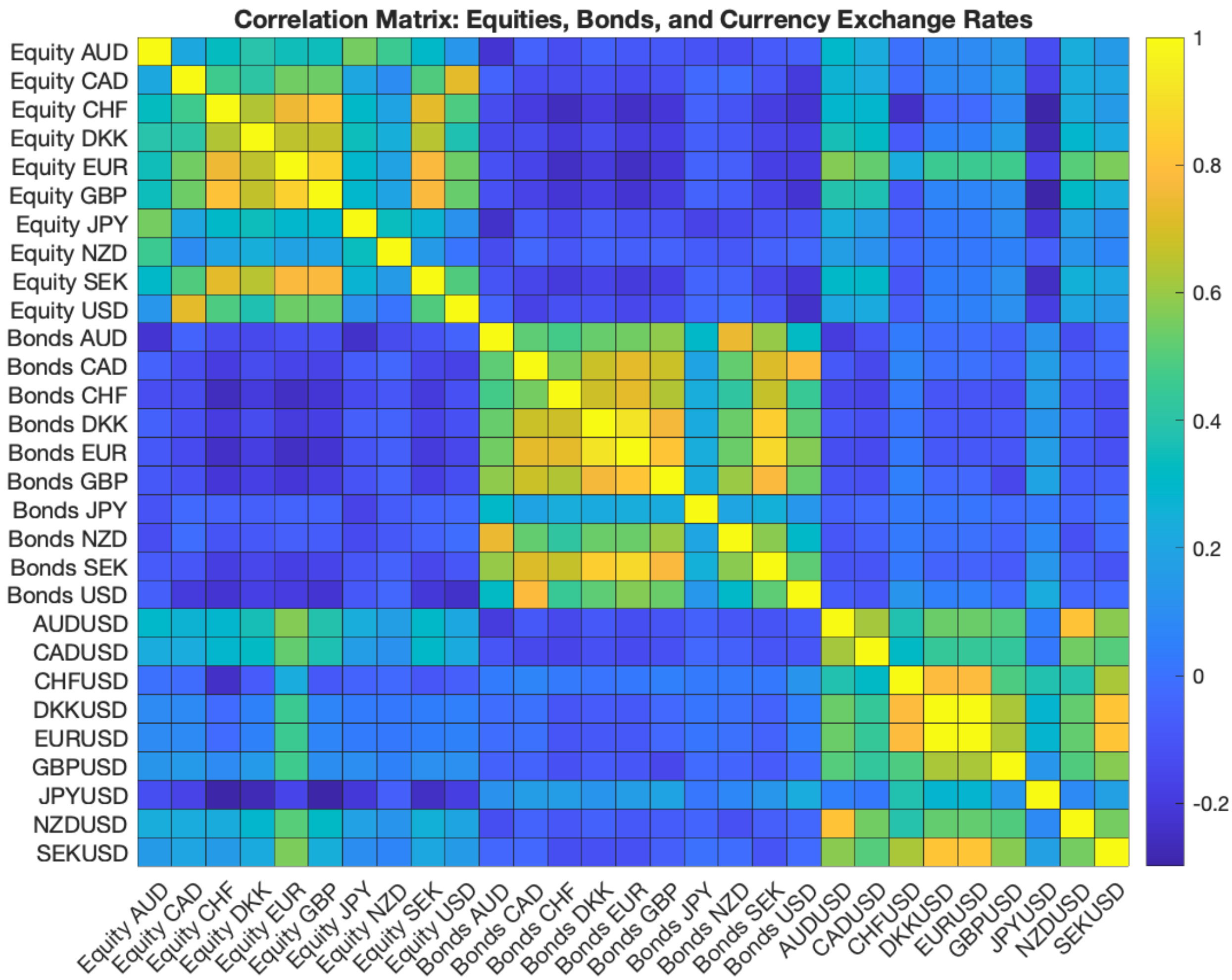

Figure 2.8: A heat map of the unconditional pairwise correlations between equity, government bond, and currency exchange rate returns (for USD as the base currency) for the 10 economies analyzed is plotted here. Notice that the correlations within each asset class are generally larger in comparison to the cross-correlations between assets from different asset classes. One can also observe larger correlations between countries located geographically close to one another, reflecting larger integration of such markets.

Chapter 3

Sparse and Stable International Portfolio Optimization and Currency Risk Management

A version of this paper has been submitted to the *Journal of International Money and Finance*.

Authors:

Raphael Burkhardt, Department of Banking and Finance, University of Zurich

Urban Ulrych, Department of Banking and Finance, University of Zurich & Swiss Finance Institute

Abstract

This paper introduces a sparse and stable optimization approach for a multi-currency asset allocation problem. We study the benefits of joint optimization of assets and currencies as opposed to the standard industry practice of managing currency risk via so-called currency overlay strategies. In our setting, a classical mean–variance problem in an international framework is augmented by several extensions that aim at reducing parameter uncertainty related to the input parameters and induce sparsity and stability of the asset and currency weights. These extensions integrate maximal net exposure to foreign currencies, shrinkage of the input parameters, and constraints on the norms of the asset- and currency-weight vectors. The empirical performance of the portfolio optimization strategies based on the proposed regularization techniques and the joint (i.e., asset and currency) optimization is tested out of sample. We demonstrate that the sparse and stable joint optimization approach consistently outperforms the standard currency overlay as well as the equally-weighted and the non-regularized global portfolio benchmarks net of transaction costs. This result shows that the common industry practice of employing currency overlay strategies is suboptimal and can be improved by a joint optimization over assets and currencies.

Keywords: International Asset Allocation, Currency Risk Management, Currency Overlay, Shrinkage Estimation, Regularization, Mean–Variance Optimization.

JEL Classification: C61, F31, G11, G15.

3.1 Introduction

International diversification is a common practice among asset managers aimed at improving the portfolio risk-return profile. This improvement can be achieved by the outperformance of foreign markets or by the mitigation of the (otherwise domestic) systematic market risk. However, global portfolios are prone to a new aspect of risk stemming from fluctuating currency exchange rates. Consequently, managing currency exposure lies at the core of the risk management practice of international investors. In practice, a standard approach for hedging currency risk is performed via a currency overlay, see Kim and Chance [2018]. In a currency overlay, an agent first optimizes the asset weights and then subsequently determines the currency exposures by taking positions in currency forward contracts. This paper investigates the potential benefits of a joint optimization approach to international asset allocation where asset weights and currency exposures are computed in a single (i.e., joint) optimization. While the potential of a joint optimization approach has been explored before, this paper proposes a novel framework that integrates optimization techniques inducing sparsity and stability of the asset and currency weights. Such an approach provides a generalization to the standard Markowitz setting and is aimed at improving the out-of-sample portfolio performance.

The joint optimization approach considered in this paper has the potential to improve the portfolio performance in comparison to a separate optimization routine (i.e., a currency overlay) as it takes into account the dependencies between single assets and currencies when determining the optimal portfolio, see Jorion [1994]. The improvement of portfolio performance crucially depends on the estimation of the optimization parameters. To deal with the problems related to parameter uncertainty in the mean–variance framework, we enhance the multi-currency asset allocation problem with regularization methods arising from the area of statistical learning. The main objective of this paper is to study a general approach that produces stable and robust multi-currency portfolios. We contribute to the existing literature by presenting a novel regularization-based approach for international portfolio optimization and testing its out-of-sample performance net of transaction costs. Our analysis shows that the widespread industry practice of utilizing currency overlay strategies is suboptimal and can be improved by the proposed regularized joint optimization over assets and currencies.

In the methodological part of the paper, we introduce a model that describes a return of an international portfolio that includes domestic and foreign assets together with the currency forward contracts employed to hedge the portfolio currency risk. Building on this formulation, we characterize the separate and joint optimization problems that investors solve when investing internationally. Then, we introduce different portfolio rules (i.e., optimization approaches) that represent extensions of the classical mean–variance framework directed at potentially improving the out-of-sample portfolio performance, see DeMiguel et al. [2009a]. These optimization procedures aim at reducing the inherent parameter uncertainty problem by either constraining the norms of the asset- and currency-weight vectors or by shrinking the optimization parameters, hence inducing sparsity and stability (i.e., robust portfolios with potentially only a few active positions). Additionally, we introduce practically inspired constraints that govern the maximal net exposure to foreign currencies. The computational solution to the sparse and stable (SAS) international asset and currency allocation problem is provided in terms of quadratic programming and cross-validation. The introduction of the SAS multi-currency asset allocation framework presents the methodological contribution of this paper.

The empirical analysis compares the joint and separate versions of the international portfolio optimization problem. A data sample of stocks, spot, and forward currency exchange rates for ten developed economies ranging from January 1999 to December 2019 is utilized. The out-of-sample performance of the proposed trading strategies is investigated and compared to the benchmark of the fully hedged equally-weighted and the classical mean–variance portfolios. The out-of-sample backtest demonstrates that the SAS joint optimization consistently outperforms the equivalent approaches based on separate optimization as well as the equally-weighted and the non-regularized portfolio benchmarks net of transaction costs. This finding represents the empirical contribution of this paper. We show that the widespread approach of first constructing an asset portfolio and then employing currency overlay managers for the subsequent management of currency risk is suboptimal and can be improved by a joint optimization over assets and currencies. This result is driven by the proposed regularization approaches that tackle the problem of parameter uncertainty by inducing a bias-variance trade-off that leads to improved out-of-sample portfolio performance.

The rest of the paper is organized as follows. Section 3.2 reviews the existing literature. Section 3.3 introduces a methodological framework for multi-currency asset allocation, extends this framework by regularization techniques inducing sparsity and stability, and provides a corresponding computational solution via quadratic programming and cross-validation. Section 3.4 investigates the empirical performance of the proposed model. Section 3.5 concludes. Additional derivations are presented in the Appendix.

3.2 Literature Review

International diversification is a standard practice of institutional investors such as mutual funds, pension funds, insurers, banks, corporate entities, and other financial intermediaries. While investing internationally offers an opportunity of improving the risk-return profile of a portfolio, it also exposes an investor to additional risk stemming from the currency exchange rates. The asset allocation and currency hedging decisions are often considered separately. There are numerous approaches in the literature exploring the benefits of hedging the implicit currency exposure of global portfolios.

At an early stage, researchers focused on investigating the potential benefits of fully hedging the implicit currency exposure. Solnik [1974] showed that when assets and currencies are uncorrelated and there is no speculative demand for currencies, it is optimal for an investor to fully hedge the implicit currency exposure. Perold and Schulman [1988] further investigated the benefits of a fully hedged portfolio empirically and found that US investors can reduce the risk of an internationally diversified portfolio of bonds and equities by fully hedging the implicit foreign currency exposure.

Moving away from the full hedging rule, another strand of the literature is positioned around investigating the so-called currency overlay strategies. The overlay strategies aim to reduce the implicit currency risk of a global portfolio by creating a currency portfolio that is laid over the asset portfolio, thereby improving the risk-return profile of the portfolio. A relevant characteristic of a currency overlay is that the asset allocation is separated from the currency allocation, which is determined subsequently after the asset weights have already been fixed. Glen and Jorion [1993] added forward contracts to an internationally diversified portfolio and showed that their inclusion results in a statistically significant increase in the portfolio performance. Campbell et al. [2010] found that some currencies move against world equity markets and thus have, despite their low average returns, attractive properties for risk-minimizing international investors. Their model takes into account the correlations between currencies and fully hedged asset returns. Dynamic currency hedging strategies have been shown to outperform the static approaches by providing significant improvements in portfolio performance.[1] On a more recent note, Boudoukh et al. [2019] derive a decomposition of a currency overlay to a risk-minimizing hedge part and an alpha-generating part. Ulrych and Vasiljević [2020] extend this approach by deriving closed-form expressions for optimal currency allocation under risk and ambiguity aversion. The authors show that ambiguity can be one of the drivers of the insufficient currency diversification puzzle known as the home currency bias. Opie and Riddiough [2020] present a method that exploits the time-series predictability of currency returns by exploiting a forecastable component in global factor returns. Their currency overlay strategy outperforms alternative dynamic hedging benchmarks. Polak and Ulrych [2021] propose a non-Gaussian approach that relies on the extended filtered historical simulation and numerical optimization of a currency overlay with respect to an arbitrary risk measure.

All of the above-mentioned papers largely contribute to the literature on managing currency risks in international portfolios. However, they assume that the portfolio is first constructed without taking into account foreign currency exposures and then the optimal hedging decision is conducted subsequently. Therefore, these overlay strategies are inherently suboptimal since the interactions between the individual assets and the exchange rates are ignored. Jorion [1994] investigated the potential benefits of a joint optimization approach versus employing specialized currency overlay managers. He found that while overlay management can add value to global equity and bond portfolios, it cannot improve portfolio performance as much as a joint approach can. Following his work, Suzuki et al. [1997] propose an efficient algorithm that obtains the globally optimal solution by employing both a mean-absolute deviation and a mean–variance model. Using historical

[1] See, among others, De Roon et al. [2003], Brown et al. [2012], Caporin et al. [2014], and Cho et al. [2020].

data, they show that this approach is more efficient than a traditional two-step (i.e., separate) optimization. In a subsequent paper, Konno and Li [2000] improve the mean-absolute deviation model in international markets by using a beta pricing method for the expected asset returns and show that this method improves the previous results and leads to an improved and more stable performance. An integrated approach of multi-currency asset allocation, where simulation is performed via the principal component analysis and a subsequent optimization is executed with respect to the expected shortfall risk measure, is presented in Topaloglou et al. [2002]. On a similar note, Mhiri and Prigent [2010] extend the mean–variance optimization paradigm in international portfolio optimization to higher moments by adding skewness and kurtosis to the utility function. More recently, Chatsanga and Parkes [2017] include the cost of carry, transaction costs, and margin requirements of forward contracts used for hedging directly in the portfolio return calculation. They show that the inclusion of such costs significantly changes the optimal decisions and improves portfolio performance. Huang and Wang [2021] extend the risk-based (i.e., mean–variance) approach of international portfolio construction to uncertainty. They propose an ambiguity-adjusted mean–chance model for international portfolio selection and derive closed-form solutions.

This paper builds on the existing work on optimal asset and currency allocation in international markets by combining the idea of the joint optimization approach with more recent advances in the portfolio optimization literature. The main objective of this work is to develop an approach that produces stable and robust portfolios consisting of multi-currency assets. While the existing literature already shows that the joint optimization approach has the potential to produce portfolios with improved risk-return properties, these results were principally established in sample as the same period was used for the optimization and the measurement of the portfolio performance. In-sample optimality implies that the joint approach can only improve portfolio performance. However, the investors are, in practice, concerned with the out-of-sample portfolio performance that can be worsened if in-sample over-fitting takes place.

The classical mean–variance optimization, as introduced by Markowitz [1952], is guaranteed to find the best portfolio in terms of the risk-return trade-off when the true asset means and covariances are known. However, this is practically impossible. Mean–variance portfolios often produce very unstable weights that substantially change for small adjustments in the input parameters. Moreover, the optimal portfolio weights are prone to attaining large positive/negative values. Portfolios constructed via the use of the sample mean and the sample covariance matrix generally perform very poorly out-of-sample, see e.g. Michaud [1989] and Jorion [1986].

There exist multiple approaches that extend the Markowitz framework by tackling the inherent parameter uncertainty of the optimization problem. One approach widely used in the literature has been the shrinkage of the input parameters. Jorion [1986] for example uses a Bayes-Stein estimator for the mean to reduce parameter uncertainty. Ledoit and Wolf [2003, 2004] propose shrinking the sample covariance matrix towards a shrinkage target with a lower variance. To deal with the large negative weights of the optimal mean–variance portfolio, as well as potential regulatory requirements, no short-sale constraints are often imposed on the portfolio weights. Jagannathan and Ma [2003] show that imposing lower and upper constraints on the portfolio weights can also be interpreted as shrinkage of the covariance matrix. These approaches have been shown to improve the out-of-sample performance of the optimal mean–variance portfolios.

DeMiguel et al. [2009b] showed that many portfolio strategies, including those mentioned above, fail to consistently outperform the naive diversification of the $1/N$ portfolio that assigns equal weights to all assets in the portfolio. In a subsequent paper, DeMiguel et al. [2009a] propose a method that does not focus on the input parameters of the optimization problem, but directly on the portfolio weights. The authors add $\mathcal{L}^1$ or $\mathcal{L}^2$ penalty terms to a minimum variance utility function, thereby penalizing large parameter values. These approaches, also known as the lasso and ridge regression in the statistical learning literature, are commonly utilized not only to deal with parameter uncertainty but also to prevent over-fitting, see e.g Friedman et al. [2001]. DeMiguel et al. [2009a] find that using these regularization approaches in the portfolio optimization context allows for building portfolios that produce improved risk-return performance compared to the naive $1/N$ portfolio. Furthermore, Brodie et al. [2009] also use a $\mathcal{L}^1$ penalty term in the utility function and show that the mean–variance portfolio constructed using this rule consistently outperforms the $1/N$ portfolio. The authors provide additional interpretations to the penalization term, interpreting it as a proxy for trading costs and showing that for a particular value of the penalization parameter, the lasso constraint is equivalent to a no short-sale constraint. Finally, Li [2015] employs an elastic net regularization (i.e., both a lasso and a ridge penalty term) for the mean–variance optimization problem and shows that the resulting portfolios manage to outperform other well-known portfolio optimization rules previously proposed in the literature, as well as the $1/N$ portfolio.

All of the shrinkage and penalization approaches mentioned above extend the basic mean–variance portfolio optimization framework but only deal with assets denominated in a single currency. Our paper contributes to the literature on international portfolio optimization by incorporating the regularization techniques aimed at improving the out-of-sample portfolio performance to the framework of the multi-currency asset allocation. Moreover, we in particular study the essential treatment of currency risk management and compare the currency overlay approach to the regularized joint optimization of assets and currencies. In addition to a theoretical interest in studying the problem of efficient multi-currency asset allocation, this is a practically oriented issue that is extensively debated in the asset and wealth management industry.

3.3 Model

In this section, a theoretical framework is presented. We consider a multi-currency-based asset allocation framework as derived in Ulrych and Vasiljević [2020]. Building on this model, we formulate the separate and joint optimization problems for the optimization of asset and currency weights, introduce the regularized versions of the international asset allocation problems, and provide numerical solutions in terms of quadratic programming algorithms.

3.3.1 Hedged Portfolio Return

Let $P_{i,t}$ represent a price of an asset i at time t, expressed in a local[2] currency, with $R_{i,t+1}$ denoting the simple return of this asset over the period t to $t+1$. Given an arbitrary domestic currency, $S_{c_i,t}$ denotes a spot currency exchange rate in domestic currency per unit of foreign currency c_i at time t, where c_i denotes a local currency asset i is denoted in. Let $e_{c_i,t+1}$ denote a currency exchange return from t to $t+1$ of a currency c_i with respect to the domestic currency. An international investor is concerned with returns expressed in her domestic currency. Thereby, an unhedged return $\tilde{R}^u_{i,t+1}$ of an asset i is given by

$$\tilde{R}^u_{i,t+1} = \frac{P_{i,t+1}S_{c_i,t+1}}{P_{i,t}S_{c_i,t}} - 1 = R_{i,t+1} + e_{c_i,t+1} + R_{i,t+1}e_{c_i,t+1}, \tag{3.3.1}$$

where, in line with with the existing literature, we assume that no rebalancing takes place between t and $t+1$. Observe that the unhedged asset return in domestic currency is expressed as a sum of the asset return in the local currency (i.e., $R_{i,t+1}$), the return of the currency exchange rate (i.e., $e_{c_i,t+1}$), and the (second-order) cross-term between the two (i.e., $R_{i,t+1}e_{c_i,t+1}$).

Now, consider an investor with a portfolio $\mathcal{P}$ consisting of $N \in \mathbb{N}$ assets. These assets can be denominated in different currencies $c = 1, 2, \ldots, K+1$, where $c = 1$ denotes the domestic currency. This implies that by investing in these assets, the investor can, next to the domestic currency, acquire an implicit exposure to up to $K \leq N$ foreign currencies. The investor's portfolio return in domestic currency is given by the weighted average of the single asset returns. Denote with $x_{i,t}$ the relative weight of the wealth invested in asset $i = 1, 2, \ldots, N$ and held from t to $t+1$. Then, the unhedged portfolio return $\tilde{R}^u_{\mathcal{P},t+1}$ is expressed by

$$\tilde{R}^u_{\mathcal{P},t+1} = \sum_{i=1}^{N} x_{i,t}\tilde{R}^u_{i,t+1},$$

where, since the asset weights are given as a fraction of the total portfolio value, $\sum_{i=1}^{N} x_{i,t} = 1$ holds for

[2] Local currency is simply the currency in which a certain asset is denominated. Given an investor's base currency, a local currency can refer to either a domestic or a foreign currency.

every t.

Next, we assume that for every time t and every currency c there exists a currency forward contract with a forward exchange rate $F_{c,t}$ expressed in domestic currency per unit of foreign currency c and delivery at $t+1$. Furthermore, we assume that $F_{c,t}$ is set in a way that the price of the corresponding forward contract equals to zero at inception. The forward premium (i.e., a rate of return corresponding to the forward exchange rate) is defined by $f_{c,t} \coloneqq (F_{c,t} - S_{c,t})/S_{c,t}$. Using this expression, the return of a short position in a currency forward contract for currency c is given by $(F_{c,t} - S_{c,t+1})/S_{c,t} = f_{c,t} - e_{c,t+1}$. Because an investor wishes to hedge her implicit foreign currency exposures, she sells these currencies forward. We denote with $\phi_{c,t}$ the relative notional value of a forward exchange rate contract for currency c at time t, expressed as a fraction of total portfolio value. A short position in the forward contract is indicated by $\phi_{c,t} > 0$.

An investor now has a possibility to not only hedge her implicit currency exposure to the currencies $c = 2, \ldots, K+1$, but also to expand her portfolio with forward contracts in currencies $c = K+2, \ldots, M+1$, to which she does not yet have an implicit exposure. The return of a hedged portfolio is then expressed by

$$\tilde{R}^{h}_{\mathcal{P},t+1} = \tilde{R}^{u}_{\mathcal{P},t+1} + \sum_{c=2}^{M+1} \phi_{c,t}(f_{c,t} - e_{c,t+1}). \tag{3.3.2}$$

For the domestic currency $c = 1$, $S_{1,t} = F_{1,t} = 1$ and $e_{1,t} = f_{1,t} = 0$ trivially hold for all t. This implies that the choice of $\phi_{1,t}$ is arbitrary.

The goal of the inclusion of currency forward contracts into the portfolio is to reduce the currency risk stemming from the investment in foreign assets. In other words, an investor is interested in optimally managing the exposure to foreign currencies. We define the (net) currency exposure to currency c at time t by $\psi_{c,t} \coloneqq w_{c,t} - \phi_{c,t}$, where $w_{c,t}$ denotes the implicit exposure to currency c. This implicit currency exposure can be expressed as $w_{c,t} = \sum_{i \in \mathcal{A}_c} x_{i,t}$, with $\mathcal{A}_c$ representing the set of all assets denoted in currency c. The case of zero exposure to foreign currencies, $\psi_{c,t} = 0$ for all $c = 2, \ldots, M+1$, corresponds to the fully hedged portfolio, which is equivalent to $\phi_{c,t} = w_{c,t}$. Moreover, we assume that $\sum_{c=1}^{M+1} \phi_{c,t} = 1$, which implies

$$\phi_{1,t} = 1 - \sum_{c=2}^{M+1} \phi_{c,t}, \tag{3.3.3}$$

showing that the domestic weight $\phi_{1,t}$ is trivially determined after the foreign currency weights have been set. We impose the summation to one such that, using Eq. (3.3.3) and the definition of $\psi_{c,t}$, one obtains $\psi_{1,t} = -\sum_{c=2}^{M+1} \psi_{c,t}$. Therefore, the sum of all net currency exposures is equal to zero, which implies that the currency portfolio is a zero investment portfolio (i.e., a hedging portfolio).

Building on the definition of $\mathcal{A}_c$, we define a matrix $\boldsymbol{C} \in \mathbb{R}^{N \times M}$, whose rows indicate a foreign currency denomination of an asset i by setting $\boldsymbol{C}_{i,c} = 1$ if asset i is traded in currency c and zero otherwise. This can then be written as

$$\boldsymbol{C}_{i,c} = \begin{cases} 1, & \text{if } i \in \mathcal{A}_{c+1}, \\ 0, & \text{otherwise}, \end{cases} \tag{3.3.4}$$

where the notation $c+1$ is used so that the first column indicates all the assets belonging to the first foreign currency and the domestic currency is skipped. This notation proves convenient in the equations below and arises from the above statement that the choice of the domestic currency weight $\phi_{1,t}$ (and equivalently $\psi_{1,t}$) is implied by the choice of the other M currency weights. This notation allows us to express a vector of implied foreign currency exposures $\boldsymbol{w}_t = (w_{2,t}, ..., w_{M+1,t})^T \in \mathbb{R}^{M\times 1}$ at time t as

$$\boldsymbol{w}_t = \boldsymbol{C}^T \boldsymbol{x}_t, \tag{3.3.5}$$

where $\boldsymbol{x}_t \in \mathbb{R}^{N\times 1}$ represents the vector of asset weights at time t and $\boldsymbol{C}^T$ is the transpose of the matrix $\boldsymbol{C}$. The domestic implicit exposure is trivially computed as $w_{1,t} = 1 - \boldsymbol{1}_M^T \boldsymbol{w}_t$, where $\boldsymbol{1}_M \in \mathbb{R}^{M\times 1}$ denotes a vector of ones. Note that the vector $\boldsymbol{w}_t$ is of dimension $(M \times 1)$, even though the assets in the portfolio are implicitly exposed to K foreign currencies. This implies that the elements $w_{K+2,t} = w_{K+3,t} = \ldots = w_{M+1,t} = 0$ are all equal to zero and the same holds for the corresponding rows of the matrix $\boldsymbol{C}$. The matrix $\boldsymbol{C}$ allows us to express the unhedged return in domestic currency from Eq. (3.3.1) in a vectorized form as

$$\tilde{\boldsymbol{R}}^u_{t+1} = \boldsymbol{R}_{t+1} + \boldsymbol{C}\boldsymbol{e}_{t+1} + (\boldsymbol{R}_{t+1} \odot \boldsymbol{C}\boldsymbol{e}_{t+1}), \tag{3.3.6}$$

where $\odot$ denotes the Hadamard product[3] (i.e., the element-wise product) and $\tilde{\boldsymbol{R}}^u_{t+1} \in \mathbb{R}^{N\times 1}$, $\boldsymbol{R}_{t+1} \in \mathbb{R}^{N\times 1}$, and $\boldsymbol{e}_{t+1} = (e_{2,t+1}, ..., e_{M+1,t+1})^T \in \mathbb{R}^{M\times 1}$ are the unhedged return vector in domestic currency, the vector of simple returns in local currencies, and the vector of simple currency returns, respectively. Taking $\boldsymbol{\phi}_t = (\phi_{2,t}, ..., \phi_{M+1,t}) \in \mathbb{R}^{M\times 1}$ and $\boldsymbol{f}_t = (f_{2,t}, ..., f_{M+1,t})^T \in \mathbb{R}^{M\times 1}$ as the vectors of currency forward weights and forward premiums, respectively, Eq. (3.3.2) can be expressed in a vectorized form as

$$\tilde{R}^h_{\mathcal{P},t+1} = \boldsymbol{x}_t^T \tilde{\boldsymbol{R}}^u_{t+1} + \boldsymbol{\phi}_t^T (\boldsymbol{f}_t - \boldsymbol{e}_{t+1}). \tag{3.3.7}$$

As mentioned above, the fully hedged portfolio is obtained by $\boldsymbol{\phi}_t = \boldsymbol{w}_t$. Using Eq. (3.3.5), this can be expressed as $\boldsymbol{\phi}_t = \boldsymbol{C}^T \boldsymbol{x}_t$. Plugging this into Eq. (3.3.7) yields the expression for the return of the fully hedged portfolio $\tilde{R}^{fh}_{\mathcal{P},t+1}$ given by

$$\tilde{R}^{fh}_{\mathcal{P},t+1} = \boldsymbol{x}_t^T \tilde{\boldsymbol{R}}^u_{t+1} + \boldsymbol{x}_t^T \boldsymbol{C}(\boldsymbol{f}_t - \boldsymbol{e}_{t+1}) = \boldsymbol{x}_t^T (\boldsymbol{R}_{t+1} + \boldsymbol{C}\boldsymbol{f}_t + (\boldsymbol{R}_{t+1} \odot \boldsymbol{C}\boldsymbol{e}_{t+1})) = \boldsymbol{x}_t^T \tilde{\boldsymbol{R}}^{fh}_{t+1},$$

[3] The Hadamard product, also known as the element-wise product, is an operation that takes two matrices/vectors of the same dimensions and produces another matrix, where each element (i, j) is the product of elements (i, j) of the original two matrices.

where $\tilde{\boldsymbol{R}}^{fh}_{t+1} = \boldsymbol{R}_{t+1} + \boldsymbol{C}\boldsymbol{f}_t + (\boldsymbol{R}_{t+1} \odot \boldsymbol{C}\boldsymbol{e}_{t+1})$ denotes the vector of fully hedged asset returns. This is a return that can be achieved by matching a long position in asset i with a short position in a forward contract for a currency c_i at time t, and vice versa. Compared to the unhedged returns from Eq. (3.3.6), the term $\boldsymbol{C}\boldsymbol{e}_{t+1}$ is replaced by $\boldsymbol{C}\boldsymbol{f}_t$. This shows that instead of the random currency return $\boldsymbol{e}_{t+1}$ the fully hedged investor secures the safe forward premium $\boldsymbol{f}_t$, thereby reducing the uncertainty arising from fluctuating currency exchange rates. The investor is, however, not able to eliminate all currency risk in the portfolio, as the cross-product term $\boldsymbol{R}_{t+1} \odot \boldsymbol{C}\boldsymbol{e}_{t+1}$ remains. This is also the reason why a perfect currency hedge is impossible to achieve. It would require an investor to know the value of her foreign assets at the maturity of the hedge $t+1$ already when choosing the forward weights at the inception of the currency forward contract at time t.

3.3.2 International Portfolio Optimization Problem

The international asset allocation problem involves a combined choice of investing into assets and currencies. The corresponding weights $\boldsymbol{x}_t$ and $\boldsymbol{\phi}_t$, as denoted in this paper, are determined such that they maximize a given utility function. We employ a classical mean–variance utility function and formulate two possible optimization problems that arise in the international portfolio optimization setting: i) the currency overlay, where first the asset positions are determined, and subsequently, given the chosen asset positions, the currency weights are optimized, and ii) the joint optimization setting where assets and currencies are determined in a single optimization.[4]

We start with the specification of a separate portfolio optimization. By $\boldsymbol{\mu}^a = \mathbb{E}(\tilde{\boldsymbol{R}}^{fh}_{t+1})$ and $\boldsymbol{\mu}^c = \mathbb{E}(\boldsymbol{f}_t - \boldsymbol{e}_{t+1})$ denote the expectations of the fully hedged asset returns and the currency forward excess returns (i.e., arising from the pay-offs of the entered currency forward contracts), respectively. The terms $\boldsymbol{\Sigma}^a = \mathrm{Var}(\tilde{\boldsymbol{R}}^{fh}_{t+1})$ and $\boldsymbol{\Sigma}^c = \mathrm{Var}(\boldsymbol{f}_t - \boldsymbol{e}_{t+1})$ denote the $(N \times N)$ covariance matrix of the fully hedged asset returns and the $(M \times M)$ covariance matrix of the currency forward excess returns, respectively. Furthermore, we define $\boldsymbol{\Sigma}^{ac} = \mathrm{Cov}(\tilde{\boldsymbol{R}}^{u}_{t+1}, \boldsymbol{f}_t - \boldsymbol{e}_{t+1})$ as the $(N \times M)$ cross-covariance matrix between asset and currency returns. The term $\gamma > 0$ represents the coefficient of relative risk aversion. The separate international

[4] Optimality in our paper should be understood in the context of risk-return trade-off in a mean–variance sense.

portfolio optimization problem is then given by

$$\begin{aligned}
\boldsymbol{x}_t^* &= \arg\max_{\boldsymbol{x}_t} \Big\{ \ \mathbb{E}(\tilde{R}^{fh}_{\mathcal{P},t+1}) - \frac{\gamma}{2}\mathrm{Var}(\tilde{R}^{fh}_{\mathcal{P},t+1}) \ \Big\} \\
&= \arg\max_{\boldsymbol{x}_t} \Big\{ \ \boldsymbol{x}_t^T \boldsymbol{\mu}^a - \frac{\gamma}{2}\boldsymbol{x}_t^T \boldsymbol{\Sigma}^a \boldsymbol{x}_t \ \Big\} \\
&\qquad \text{subject to} \qquad \boldsymbol{x}_t^T \mathbf{1}_N = 1, \\
\boldsymbol{\phi}_t^* \mid \boldsymbol{x}_t^* &= \arg\max_{\boldsymbol{\phi}_t} \Big\{ \ \mathbb{E}(\tilde{R}^{h}_{\mathcal{P},t+1}) - \frac{\gamma}{2}\mathrm{Var}(\tilde{R}^{h}_{\mathcal{P},t+1}) \mid \boldsymbol{x}_t^* \ \Big\} \\
&= \arg\max_{\boldsymbol{\phi}_t} \Big\{ \ \boldsymbol{\phi}_t^T \boldsymbol{\mu}^c - \frac{\gamma}{2}\boldsymbol{\phi}_t^T \boldsymbol{\Sigma}^c \boldsymbol{\phi}_t - \gamma \boldsymbol{x}_t^{*T} \boldsymbol{\Sigma}^{ac} \boldsymbol{\phi}_t \ \mid \boldsymbol{x}_t^* \Big\}.
\end{aligned} \tag{3.3.8}$$

The first optimization step is a standard mean–variance optimization of the asset portfolio. Here, the currencies are not considered and the goal is to optimize the mean–variance trade-off for the fully hedged portfolio.[5] In the second optimization step, the currency overlay step, an investor maximizes the mean–variance utility conditionally on the already determined asset weights (i.e., as obtained in the first optimization step), with the hedged return as the object of interest, focusing on the optimization of the currency exposure. Note that only the asset weights need to add up to one. For the currency weights, the weight $\phi_{1,t}$, which is not a part of the optimization, can simply be chosen such that the currency weights sum up to one, see Eq. (3.3.3).

Because of the conditioning on $\boldsymbol{x}_t^*$, terms that do not depend on $\boldsymbol{\phi}_t$ can be dropped out of the formulation of the second optimization step. Notice that the choice of the optimal currency weights is only influenced by the previously determined asset weights through the term $\boldsymbol{x}_t^{*T}\boldsymbol{\Sigma}^{ac} = \mathrm{Cov}(\tilde{R}^u_{\mathcal{P},t+1}, \boldsymbol{f}_t - \boldsymbol{e}_{t+1})$, which is a $(1 \times M)$ vector representing the covariances between the unhedged portfolio return and the currency forward returns. Clearly, the correlations between the single assets and currencies are not considered. Thus, the separate optimization approach can be understood as a form of dimensionality reduction. Moreover, one can think of specialized equity (bond) asset managers being employed in the first optimization step. Thereby, asset positions are determined and a specialized equity (bond) asset manager does not speculate on currencies when deciding on the asset allocations.[6] Subsequently, a specialized currency overlay manager determines the optimal currency allocations, given the already determined asset allocation. As discussed in Kim and Chance [2018], the described separate optimization is a standard approach of optimizing international portfolios in practice.

Next, we formulate a joint international asset allocation optimization problem. Consider the vector $\boldsymbol{\theta}_t = (x_{1,t}, ..., x_{N,t}, \phi_{2,t}, ..., \phi_{M+1,t})^T \in \mathbb{R}^{(N+M)\times 1}$ which is a combined vector of the asset weights $\boldsymbol{x}_t$ and the currency weights $\boldsymbol{\phi}_t$. We denote by $\boldsymbol{r}_{t+1} = (\tilde{R}^u_{1,t+1}, ..., \tilde{R}^u_{N,t+1}, (f_{2,t} - e_{2,t+1}), ..., (f_{M+1,t} - e_{M+1,t+1}))$

[5] In the separate optimization approach, an investor optimizes the asset allocation without the consideration of the underlying implicit currency exposure. This is the reason for the fully hedged, as opposed to the unhedged, formulation of the first optimization step. The currency allocation is then performed in the second (i.e., currency overlay) optimization step.

[6] This is a reason for the fully hedged return specification in the first optimization step.

the combined vector of asset returns expressed in a domestic currency and currency forward excess returns. Furthermore, the term $\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{r}_{t+1})$ denotes the $((N+M) \times 1)$ vector of its expected returns and $\boldsymbol{\Sigma} = \mathrm{Var}(\boldsymbol{r}_{t+1})$ the corresponding $((N+M) \times (N+M))$ covariance matrix. The joint international portfolio optimization problem is given by

$$\begin{aligned} \boldsymbol{\theta}_t^* &= \arg\max_{\boldsymbol{\theta}_t} \left\{ \ \mathbb{E}(\tilde{R}^h_{\mathcal{P},t+1}) - \frac{\gamma}{2}\mathrm{Var}(\tilde{R}^h_{\mathcal{P},t+1}) \ \right\} \\ &= \arg\max_{\boldsymbol{\theta}_t} \left\{ \ \boldsymbol{\theta}_t^T \boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{\theta}_t^T \boldsymbol{\Sigma} \boldsymbol{\theta}_t \ \right\} \\ &\qquad \text{subject to} \qquad \boldsymbol{\theta}_t^T \mathbf{q}_{N,M} = 1, \end{aligned} \tag{3.3.9}$$

where $\mathbf{q}_{N,M}$ denotes a $((N+M) \times 1)$ vector with the first N elements equal to one and the remaining M elements equal to zero. Therefore, the constraint is equivalent to $\boldsymbol{x}^T \mathbf{1}_N = 1$ in the separate optimization problem.

This formulation of the optimization problem determines the mean–variance efficient vector $\boldsymbol{\theta}_t$ by jointly optimizing over the asset weights and currency exposures.[7] Here, also the correlations between the single assets and currencies are considered. The optimization problem is, correspondingly, of higher dimension compared to the separate optimization. As such, it is more prone to parameter uncertainty and over-fitting that can potentially lead to a worsened out-of-sample portfolio performance. To tackle this problem, we introduce regularization to the international asset allocation problem in the next section.

3.3.3 Sparse and Stable International Portfolios

The solutions to both the separate and the joint optimization problem depend on the true expected returns and variances that are generally unknown. One commonly used approach is based on using the sample counterparts $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ in place of the true parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. This plug-in approach, however, tends to perform poorly out of sample, see DeMiguel et al. [2009b]. This is the reason for the introduction of alternative methods that deal with the problem of parameter uncertainty in international portfolio optimization.

This section introduces some extensions to the standard mean–variance framework and translates them into the context of the multi-currency portfolio optimization problem. We start with focusing on the sparsity of asset and currency weights, introduce stable international portfolios, and subsequently combine the two into a SAS international asset allocation approach. We finish by adjusting the optimization approach by adding currency constraints that are specific to the international optimization problem and are motivated by practical considerations.

[7] Note that maximizing the mean–variance utility function introduced above can also be interpreted as solving a linear regression problem for a suitable choice of regression parameters, see Li [2015] for more details.

3.3.3.1 Sparse Multi-Currency Portfolio

A sparse portfolio is a portfolio with some (i.e., possibly many) of the portfolio weights set to zero. This is a desirable property especially when the asset universe is large. Furthermore, setting some weights to zero has the potential benefit of lowering the transaction costs. The idea of sparse portfolios has been studied in the finance literature, Britten-Jones [1999], for example, introduce an F-test to test whether multiple asset weights are statistically different from zero.[8] Another approach was presented by Garlappi et al. [2007], who propose setting all portfolio weights to zero (i.e., only investing in risk-free assets) when the squared Sharpe ratio is not statistically different from zero.

Sparsity is introduced into the portfolio optimization problem by constraining the $\mathcal{L}^1$-norm of the portfolio weights. The $\mathcal{L}^1$-norm is defined by $\|\boldsymbol{x}\|_1 = \sum_{i=1}^N |x_i|$. Sparsity arises from the fact that the optimal portfolio is usually located at one of the corners of the ($\mathcal{L}^1$ induced) constraint region, where one or more weights are equal to zero (e.g., see Figure 3.12 in Hastie et al. [2009]). Next to sparsity, there are multiple potential benefits of introducing a constraint on the $\mathcal{L}^1$-norm. Such a constraint can induce stability into a portfolio by reducing the sensitivity of the optimization to potential collinearities between the assets, see Daubechies et al. [2004]. Through the introduction of the constraint, the estimation becomes biased; however, it has the potential to improve the out-of-sample portfolio performance through a bias-variance trade-off, see Tibshirani [1996]. Furthermore, the constraint also limits the amount of short-sale permitted in the portfolio, see DeMiguel et al. [2009a], which is a desirable property from a practical point of view. Another advantage is that the constraint can be seen as a proxy for transaction costs, see Brodie et al. [2009].[9]

Next, we specify a sparse international portfolio optimization problem. Consider a separate optimization problem as defined in Eq. (3.3.8). Consistent with the literature on regularization, we cast the optimization problem in terms of minimization and introduce the $\mathcal{L}^1$-normed penalty directly in the objective function.[10] The sparse separate optimization is given by

$$\begin{aligned} \boldsymbol{x}_t^* &= \arg\min_{\boldsymbol{x}_t} \Big\{ \frac{\gamma}{2} \boldsymbol{x}_t^T \hat{\boldsymbol{\Sigma}}^a \boldsymbol{x}_t - \boldsymbol{x}_t^T \hat{\boldsymbol{\mu}}^a + \lambda_1^a \|\boldsymbol{x}_t\|_1 \Big\} \\ &\qquad \text{subject to} \qquad \boldsymbol{x}_t^T \mathbf{1}_N = 1, \\ \boldsymbol{\phi}_t^* \mid \boldsymbol{x}_t^* &= \arg\min_{\boldsymbol{\phi}_t} \Big\{ \frac{\gamma}{2} \boldsymbol{\phi}_t^T \hat{\boldsymbol{\Sigma}}^c \boldsymbol{\phi}_t - \boldsymbol{\phi}_t^T \hat{\boldsymbol{\mu}}^c + \gamma \boldsymbol{x}_t^{*T} \hat{\boldsymbol{\Sigma}}^{ac} \boldsymbol{\phi}_t + \lambda_1^c \|\boldsymbol{\phi}_t\|_1 \ \mid \boldsymbol{x}_t^* \Big\}, \end{aligned} \tag{3.3.10}$$

8 An equivalent approach for testing a single asset is performed via a t-test.

9 Transaction costs normally consist of two components: a fixed “overhead” cost for every transaction, which is usually comparably small for large investments, and a proportional cost that results from the bid-ask spread of the market. As the $\mathcal{L}^1$ constraint limits the absolute size of the portfolio weights, this can also be seen as imposing a limit on the proportional transaction costs.

10 This is, by the method of Lagrange multipliers, equivalent to specifying the penalty term as a constraint of the given optimization problem.

where with ˆ we denote a sample estimate and the parameters λ_1^a and λ_1^c regulate the penalization of the assets and currencies, respectively.[11] Employing two separate parameters for assets and currencies allows for a more precise calibration of the model to the international portfolio optimization problem. A single parameter for both would indicate that both should be treated in the same way, i.e., currencies are just other assets. However, as explained above, only the assets are constrained to add up to one. Moreover, the transaction costs for currencies are generally lower than those of equities. The interpretation of the lasso constraint as a proxy for transaction costs also provides motivation for a separate penalization structure. The sparsity property induced by the lasso constraint could also imply that possibly many of the currency weights can become equal to zero if there is a single penalization parameter for assets and currencies. However, this is not a desirable property from the currency risk management perspective.

Equivalently to the separate optimization case, the sparse joint optimization problem is given by

$$\begin{aligned} \boldsymbol{\theta}_t^* &= \arg\min_{\boldsymbol{\theta}_t} \left\{ \frac{\gamma}{2} \boldsymbol{\theta}_t^T \hat{\boldsymbol{\Sigma}} \boldsymbol{\theta}_t - \boldsymbol{\theta}_t^T \hat{\boldsymbol{\mu}} + \|\boldsymbol{S}\boldsymbol{\theta}_t\|_1 \right\} \\ &\text{subject to} \qquad \boldsymbol{\theta}_t^T \mathbf{q}_{N,M} = 1, \end{aligned} \tag{3.3.11}$$

where $\boldsymbol{S}$ is an $((N+M)\times(N+M))$ matrix with all off-diagonal elements equal to zero, the first N diagonal elements equal to λ_1^a and the last M diagonal elements equal to λ_1^c. Such specification again allows for different optimal levels of penalization for assets and currencies.

The $\mathcal{L}^1$-norm proposed above is non-differentiable with respect to the weight vector, which is why the lasso problem does not yield an analytical solution. There are different techniques to solve the lasso problem numerically, such as the conjugate-gradient method or the LARS algorithm, see Efron et al. [2004]. In this paper, we introduce various types of constraints on the asset and currency weights and consequently require a general algorithm that is able to handle such constraints. We follow the approach described by Roncalli [2013], where auxiliary variables are introduced and the problem is reformulated as a quadratic programming problem in a higher dimension. The auxiliary variables $\Delta_i^+ \geq 0$ and $\Delta_i^- \geq 0$ are defined by

$$x_i = \Delta_i^+ - \Delta_i^-. \tag{3.3.12}$$

They represent the positive and the negative part of the weight x_i. Using the decomposition from Eq. (3.3.12), we can express the absolute value of the asset weight as

$$|x_i| = |\Delta_i^+ - \Delta_i^-| = \Delta_i^+ + \Delta_i^-, \tag{3.3.13}$$

11 For $\lambda_1^a = \lambda_1^c = 0$, the penalization is absent and the optimization problem corresponds to the one from Eq. (3.3.8). Moreover, larger choices of λ_1^a and λ_1^c imply larger penalization of the absolute portfolio weights, which can be interpreted as shrinking the weights towards zero.

which follows since only one of the values Δ_i^+ and Δ_i^- can be positive at a time. This decomposition enables us to cast the optimization problems given in Eqs. (3.3.10) and (3.3.11) as quadratic programs.

Consider a vector $\tilde{\boldsymbol{x}}_t = (x_{1,t}, \ldots, x_{N,t}, \Delta_1^-, \ldots, \Delta_N^-, \Delta_1^+, \ldots, \Delta_N^+) \in \mathbb{R}^{3N\times 1}$ and equivalently a vector $\tilde{\boldsymbol{\phi}}_t \in \mathbb{R}^{3M\times 1}$, where $\tilde{}$ denotes an augmented variable. These vectors allow us to formulate augmented quadratic programs of dimensions $3N$ and $3M$ that solve the sparse separate international portfolio optimization problem (with a constraint of the asset weights summing up to one) via

$$\begin{aligned} \tilde{\boldsymbol{x}}_t^* = \arg\min_{\tilde{\boldsymbol{x}}_t} & \left\{ \frac{\gamma}{2} \tilde{\boldsymbol{x}}_t^T \tilde{\boldsymbol{H}} \tilde{\boldsymbol{x}}_t - \tilde{\boldsymbol{x}}_t^T \tilde{\boldsymbol{m}} \right\}, \\ \text{subject to} \qquad & \tilde{\boldsymbol{A}} \tilde{\boldsymbol{x}}_t = \tilde{\boldsymbol{B}}, \end{aligned} \tag{3.3.14}$$

with

$$\tilde{\boldsymbol{H}} = \begin{pmatrix} \hat{\boldsymbol{\Sigma}}^a & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \tilde{\boldsymbol{m}} = \begin{pmatrix} \hat{\boldsymbol{\mu}}^a \\ -\lambda_1^a \mathbf{1}_N \\ -\lambda_1^a \mathbf{1}_N \end{pmatrix}, \quad \tilde{\boldsymbol{A}} = \begin{pmatrix} \mathbf{1}_N^T & \mathbf{0} & \mathbf{0} \\ \boldsymbol{I}_N & \boldsymbol{I}_N & -\boldsymbol{I}_N \end{pmatrix}, \quad \tilde{\boldsymbol{B}} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix},$$

and

$$\begin{aligned} \tilde{\boldsymbol{\phi}}_t^* \mid \tilde{\boldsymbol{x}}_t^* = \arg\min_{\tilde{\boldsymbol{\phi}}_t} & \left\{ \frac{\gamma}{2} \tilde{\boldsymbol{\phi}}_t^T \tilde{\boldsymbol{H}} \tilde{\boldsymbol{\phi}}_t - \tilde{\boldsymbol{\phi}}_t^T \tilde{\boldsymbol{m}} \mid \boldsymbol{x}_t^* \right\}, \\ \text{subject to} \qquad & \tilde{\boldsymbol{A}} \tilde{\boldsymbol{\phi}}_t = \tilde{\boldsymbol{B}}, \end{aligned}$$

with

$$\tilde{\boldsymbol{H}} = \begin{pmatrix} \hat{\boldsymbol{\Sigma}}^c & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \tilde{\boldsymbol{m}} = \begin{pmatrix} \hat{\boldsymbol{\mu}}^c - \gamma \hat{\boldsymbol{\Sigma}}^{ac^T} \boldsymbol{x}_t^* \\ -\lambda_1^c \mathbf{1}_M \\ -\lambda_1^c \mathbf{1}_M \end{pmatrix}, \quad \tilde{\boldsymbol{A}} = \begin{pmatrix} \boldsymbol{I}_M & \boldsymbol{I}_M & -\boldsymbol{I}_M \end{pmatrix}, \text{ and } \quad \tilde{\boldsymbol{B}} = \mathbf{0}.$$

As before, $\boldsymbol{I}_N$ denotes an $(N \times N)$ identity matrix and $\mathbf{1}_N$ an $(N \times 1)$ vector of ones. With the term $\mathbf{0}$ we denote matrices and vectors of zeros that have a dimension that ensures the correct overall dimension of the objective and the constraints. The hyperparameters λ_1^a and λ_1^c represent the penalization intensity of the $\mathcal{L}^1$-norm constraint. The second row of the equality constraint in the first part of the optimization (i.e., optimization over the assets) enforces the auxiliary variables decomposition from Eq. (3.3.12), while the first row represents the equality constraint of the asset weights summing up to one. The second part of the optimization (i.e., the currency overlay) requires only the constraint arising from the auxiliary problem.

Equivalently, the sparse joint international portfolio optimization problem is formulated as a quadratic program of dimension $3(N + M)$ via

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_t^* = \arg\min_{\tilde{\boldsymbol{\theta}}_t} & \left\{ \frac{\gamma}{2} \tilde{\boldsymbol{\theta}}_t^T \tilde{\boldsymbol{H}} \tilde{\boldsymbol{\theta}}_t - \tilde{\boldsymbol{\theta}}_t^T \tilde{\boldsymbol{m}} \right\}, \\ \text{subject to} \qquad & \tilde{\boldsymbol{A}} \tilde{\boldsymbol{\theta}}_t = \tilde{\boldsymbol{B}}, \end{aligned} \tag{3.3.15}$$

with

$$\tilde{\boldsymbol{H}} = \begin{pmatrix} \hat{\boldsymbol{\Sigma}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \tilde{\boldsymbol{m}} = \begin{pmatrix} \hat{\boldsymbol{\mu}} \\ -\lambda_1^a \mathbf{1}_N \\ -\lambda_1^c \mathbf{1}_M \\ -\lambda_1^a \mathbf{1}_N \\ -\lambda_1^c \mathbf{1}_M \end{pmatrix}, \quad \tilde{\boldsymbol{A}} = \begin{pmatrix} \mathbf{q}_{N,M}^T & \mathbf{0} & \mathbf{0} \\ \boldsymbol{I}_{N+M} & \boldsymbol{I}_{N+M} & -\boldsymbol{I}_{N+M} \end{pmatrix}, \quad \tilde{\boldsymbol{B}} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix}.$$

3.3.3.2 Stable Multi-Currency Portfolio

The stability of portfolio weights is another property that is desirable for the investor. The standard mean–variance efficient portfolio is sensitive to changes or errors in the input and is prone to producing extreme weights. This is problematic in practice since unstable weights have the potential to cause high transaction costs over time that can significantly diminish realized portfolio performance. Ledoit and Wolf [2003] show that instability is related to the estimation errors of the sample covariance matrix and its inverse needed for computing the mean–variance efficient portfolio. The inverse becomes especially problematic in the presence of multicollinearity. These instabilities motivate several approaches in the literature that aim at constraining the portfolio in a way that induces stability and thereby reduces the problem of parameter uncertainty.

A well-known approach of shrinking the covariance matrix is proposed in Ledoit and Wolf [2004]. The authors propose a shrinkage estimator for a covariance matrix given by

$$\hat{\boldsymbol{\Sigma}}_s = v\hat{\boldsymbol{\Sigma}}_g + (1-v)\hat{\boldsymbol{\Sigma}}, \tag{3.3.16}$$

where $\hat{\boldsymbol{\Sigma}}$ is a sample covariance matrix, $\hat{\boldsymbol{\Sigma}}_g$ is a shrinkage target with a lower variance[12], and v is a shrinkage intensity. Furthermore, Ledoit and Wolf demonstrate that optimized portfolios resulting from the covariance matrix shrinkage improve out-of-sample performance. Moreover, Li [2015] shows that using a shrinkage estimator of the form of Eq. (3.3.16) is equal to a linear transformation of the portfolio weights that depends on both the shrinkage target $\hat{\boldsymbol{\Sigma}}_g$ and the shrinkage intensity v. Another interpretation offered by the author is that the shrinkage estimator can be seen as a constraint on the $\mathcal{L}^2$-norm, weighted by the target covariance matrix $\hat{\boldsymbol{\Sigma}}_g$.

While the implementation of this estimator is straightforward for the joint international portfolio optimization approach, where the matrix $\hat{\boldsymbol{\Sigma}}$ can simply be replaced by its shrunk version $\hat{\boldsymbol{\Sigma}}_s$, this does not apply for the separate optimization problem. The reason for this is that the cross-covariance matrix between the assets and currencies $\hat{\boldsymbol{\Sigma}}^{ac}$ should theoretically be shrunk as well. This can be achieved by slightly modifying

[12] See Ledoit and Wolf [2004] for details on constructing this matrix.

the currency overlay step in the separate optimization problem. We do so by augmenting the currency overlay problem by one dimension and introducing the vector $\boldsymbol{\varphi}_t = (1, \phi_{2,t}, .., \phi_{M+1,t})$. Here, the first weight corresponds to the asset portfolio. Its weight is already predetermined from the asset optimization step since the asset weights sum up to one. Accordingly, we introduce parameters $\hat{\boldsymbol{\Sigma}}^{\tilde{c}} \in \mathbb{R}^{(M+1)\times(M+1)}$ and $\hat{\boldsymbol{\mu}}^{\tilde{c}} \in \mathbb{R}^{(M+1)\times 1}$, which represent the augmented covariance matrix and the augmented mean vector of the joint asset portfolio and currencies vector, as defined below. The stable separate international portfolio optimization with shrinkage of the covariance matrix is given by

$$\begin{aligned} \boldsymbol{x}_t^* &= \underset{\boldsymbol{x}_t}{\arg\min} \left\{ \frac{\gamma}{2} \boldsymbol{x}_t^T \hat{\boldsymbol{\Sigma}}_s^a \boldsymbol{x}_t - \boldsymbol{x}_t^T \hat{\boldsymbol{\mu}}^a \right\}, \\ &\text{subject to} \qquad \boldsymbol{x}_t^T \mathbf{1} = 1, \\ \boldsymbol{\varphi}_t^* \mid \boldsymbol{x}_t^* &= \underset{\boldsymbol{\varphi}_t}{\arg\min} \left\{ \frac{\gamma}{2} \boldsymbol{\varphi}_t^T \hat{\boldsymbol{\Sigma}}_s^{\tilde{c}} \boldsymbol{\varphi}_t - \boldsymbol{\varphi}_t^T \hat{\boldsymbol{\mu}}^{\tilde{c}} \mid \boldsymbol{x}_t^* \right\}, \\ &\text{subject to} \qquad \boldsymbol{A}\boldsymbol{\varphi}_t = 1, \end{aligned} \tag{3.3.17}$$

with

$$\boldsymbol{A} = \begin{pmatrix} 1 & \mathbf{0} \end{pmatrix}, \qquad \hat{\boldsymbol{\Sigma}}^{\tilde{c}} = \begin{pmatrix} \boldsymbol{x}_t^{*T} \hat{\boldsymbol{\Sigma}}^a \boldsymbol{x}_t^* & \boldsymbol{x}_t^{*T} \hat{\boldsymbol{\Sigma}}^{ac} \\ (\boldsymbol{x}_t^{*T} \hat{\boldsymbol{\Sigma}}^{ac})^T & \hat{\boldsymbol{\Sigma}}^c \end{pmatrix}, \text{ and } \qquad \hat{\boldsymbol{\mu}}^{\tilde{c}} = \begin{pmatrix} \boldsymbol{x}_t^{*T} \hat{\boldsymbol{\mu}}^a \\ \hat{\boldsymbol{\mu}}^c \end{pmatrix}.$$

Here, the linear equality constraint in the second optimization step simply enforces that the weight assigned to the asset portfolio is equal to one and the optimization is performed only with respect to the currency weights. The matrices $\hat{\boldsymbol{\Sigma}}_s^a$ and $\hat{\boldsymbol{\Sigma}}_s^{\tilde{c}}$ are calculated via Eq. (3.3.16), whereby the shrinkage target and the shrinkage intensity are computed as proposed in Ledoit and Wolf [2004].[13]

The stable joint international portfolio optimization with shrinkage of the covariance matrix is then accordingly given by

$$\begin{aligned} \boldsymbol{\theta}_t^* &= \underset{\boldsymbol{\theta}_t}{\arg\min} \left\{ \frac{\gamma}{2} \boldsymbol{\theta}_t^T \hat{\boldsymbol{\Sigma}}_s \boldsymbol{\theta}_t - \boldsymbol{\theta}_t^T \hat{\boldsymbol{\mu}} \right\}, \\ &\text{subject to} \qquad \boldsymbol{\theta}_t^T \mathbf{q}_{N,M} = 1. \end{aligned} \tag{3.3.18}$$

Another approach to induce stability in the portfolio is achieved by constraining the $\mathcal{L}^2$-norm of the portfolio weights. This approach is well-connected to the shrinkage of the covariance matrix described above. While $\mathcal{L}^1$ penalization induces extreme shrinkage (i.e., sparsity), the $\mathcal{L}^2$ penalization induces smooth shrinkage–stability. As Li [2015] shows, imposing a ridge type of constraint is equivalent to shrinking the covariance matrix towards the identity matrix. Therefore, constraining the $\mathcal{L}^2$-norm of the portfolio weights or employing a covariance matrix shrinkage estimator can be interpreted as two analogous approaches.

[13] The shrinkage target is constructed from the common constant correlation (i.e., the average of all sample correlations) and the vector of sample variances. The shrinkage intensity is optimized with respect to a quadratic measure of distance between the true and the estimated covariance matrices based on the Frobenius norm. For more information consult the original paper.

3.3.3.3 Sparse and Stable Multi-Currency Portfolio

Next, we combine the methods introduced above and construct international portfolios that are both sparse and stable. We start with the separate optimization and combine sparsity as induced by the $\mathcal{L}^1$ penalization such as in Eq. (3.3.10) with shrinkage of the covariance matrix as presented in Eq. (3.3.17). Thereby, the SAS separate international portfolio optimization with shrinkage of the covariance matrix is given by

$$\begin{aligned}
\boldsymbol{x}_t^* &= \underset{\boldsymbol{x}_t}{\arg\min}\left\{ \frac{\gamma}{2}\boldsymbol{x}_t^T\hat{\boldsymbol{\Sigma}}_s^a\boldsymbol{x}_t - \boldsymbol{x}_t^T\hat{\boldsymbol{\mu}}^a + \lambda_1^a\|\boldsymbol{x}_t\|_1 \right\}, \\
&\text{subject to} \qquad \boldsymbol{x}_t^T\mathbf{1} = 1, \\
\boldsymbol{\varphi}_t^* \mid \boldsymbol{x}_t^* &= \underset{\boldsymbol{\varphi}_t}{\arg\min}\left\{ \frac{\gamma}{2}\boldsymbol{\varphi}_t^T\hat{\boldsymbol{\Sigma}}_s^{\tilde{c}}\boldsymbol{\varphi}_t - \boldsymbol{\varphi}_t^T\hat{\boldsymbol{\mu}}^{\tilde{c}} + \lambda_1^c\|\boldsymbol{\varphi}_t\|_1 \mid \boldsymbol{x}_t^* \right\}, \\
&\text{subject to} \qquad \boldsymbol{A}\boldsymbol{\varphi}_t = 1,
\end{aligned} \tag{3.3.19}$$

with

$$\boldsymbol{A} = \begin{pmatrix} 1 & \mathbf{0} \end{pmatrix}, \qquad \hat{\boldsymbol{\Sigma}}^{\tilde{c}} = \begin{pmatrix} \boldsymbol{x}_t^{*T}\hat{\boldsymbol{\Sigma}}^a\boldsymbol{x}_t^* & \boldsymbol{x}_t^{*T}\hat{\boldsymbol{\Sigma}}^{ac} \\ (\boldsymbol{x}_t^{*T}\hat{\boldsymbol{\Sigma}}^{ac})^T & \hat{\boldsymbol{\Sigma}}^c \end{pmatrix}, \text{ and } \qquad \hat{\boldsymbol{\mu}}^{\tilde{c}} = \begin{pmatrix} \boldsymbol{x}_t^{*T}\hat{\boldsymbol{\mu}}^a \\ \hat{\boldsymbol{\mu}}^c \end{pmatrix}.$$

Here, $\hat{\boldsymbol{\Sigma}}_s^a$ and $\hat{\boldsymbol{\Sigma}}_s^c$ are the shrinkage estimators of the covariance matrix as specified in Eq. (3.3.16), with the shrinkage target and the skrinkage intensity as proposed in Ledoit and Wolf [2004].

Equivalently, the SAS joint international portfolio optimization with shrinkage of the covariance matrix is given by

$$\begin{aligned}
\boldsymbol{\theta}_t^* &= \underset{\boldsymbol{\theta}_t}{\arg\min}\left\{ \frac{\gamma}{2}\boldsymbol{\theta}_t^T\hat{\boldsymbol{\Sigma}}_s\boldsymbol{\theta}_t - \boldsymbol{\theta}_t^T\hat{\boldsymbol{\mu}} + \|\boldsymbol{S}\boldsymbol{\theta}_t\|_1 \right\}, \\
&\text{subject to} \qquad \boldsymbol{\theta}_t^T\mathbf{q}_{N,M} = 1,
\end{aligned} \tag{3.3.20}$$

with parameters as already defined before. The optimization problems from Eqs. (3.3.19) and (3.3.20) are solved analogously to the augmented quadratic problem as specified in Eqs. (3.3.14) and (3.3.15), with the appropriately shrunk covariance matrix.

Instead of employing the covariance shrinkage to construct SAS portfolios, it is also possible to induce stability into the sparse portfolio by constraining also the $\mathcal{L}^2$-norm of the portfolio weights. Such a specification that includes both the lasso and the ridge type of constraint, often referred to as the elastic net, can be implemented by adding both the $\mathcal{L}^1$ and $\mathcal{L}^2$ penalty terms to the mean–variance objective function. For details on the exact specification of the elastic net SAS international portfolio optimization see the Appendix.

3.3.3.4 Embedding Currency Exposure Constraints

The portfolio optimization rules introduced in previous subsections do not address the practical concern of the allowed minimal/maximal currency exposure in an international portfolio. Many (e.g., institutional) investors, such as pension funds, are in practice limited in the allowed foreign currency exposure of their portfolios. Such constraints are usually driven by regulators. Moreover, constraining currency weights can from another angle also be seen as a form of shrinkage. This motivates the introduction of the net currency exposure constraints. Consider the net currency exposure ψ_c and assume that

$$V_c^l \leq \psi_c = w_c - \phi_c \leq V_c^u \qquad \text{for} \quad c = 2, \dots, M+1, \tag{3.3.21}$$

where V_c^l and V_c^u are constants indicating the lower and the upper allowed exposure to currency c, expressed as a fraction of total portfolio value.[14] In what follows, we assume $V^l = V_2^l = \ldots = V_{M+1}^l$ and $V^u = V_2^u = \ldots = V_{M+1}^u$, indicating that the lower and the upper bound are equal among all foreign currencies.[15] Using Eqs. (3.3.5) and (3.3.21), the upper bound on the net currency exposure can be expressed as

$$\boldsymbol{w}_t - \boldsymbol{\phi}_t \leq V^u \mathbf{1}_M \iff \boldsymbol{C}^T \boldsymbol{x}_t - \boldsymbol{\phi}_t \leq V^u \mathbf{1}_M \iff \begin{pmatrix} \boldsymbol{C}^T & -\boldsymbol{I}_M \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_t \\ \boldsymbol{\phi}_t \end{pmatrix} \leq V^u \mathbf{1}_M. \tag{3.3.22}$$

The lower bound can be expressed equivalently.

For the separate optimization problem, the net currency exposure constraints are implemented by imposing an upper and a lower bound on the currency weights in the second optimization (i.e., currency overlay) step. The first optimization step (i.e., asset allocation) remains unchanged. Let $\boldsymbol{\phi}_t^-$ and $\boldsymbol{\phi}_t^+$ denote the lower and upper bound on the currency weight, $\boldsymbol{\phi}_t^- \leq \boldsymbol{\phi}_t \leq \boldsymbol{\phi}_t^+$. Using Eq. (3.3.22), the lower and the upper bounds of the currency weights are given by

$$\boldsymbol{\phi}_t^+ \mid \boldsymbol{x}_t^* = \boldsymbol{C}^T \boldsymbol{x}_t^* - V^l \mathbf{1}_M \quad \text{and} \quad \boldsymbol{\phi}_t^- \mid \boldsymbol{x}_t^* = \boldsymbol{C}^T \boldsymbol{x}_t^* - V^u \mathbf{1}_M, \tag{3.3.23}$$

highlighting that the currency bounds depend on the asset weights chosen in the first optimization step. A general SAS separate international portfolio optimization setting, as described before, can easily be augmented for currency exposure constraints by adding the constraint from Eq. (3.3.23) in the second optimization step.

Next, we investigate the joint optimization problem. The representation in Eq. (3.3.22) shows that an implicit currency exposure, attained through optimizing over $\boldsymbol{x}_t$, affects the possible values currency weights

[14] For a specific choice of $V_c^l = 0$ and $V_c^u = w_c$, currency c is not allowed to be over or under hedged (i.e., hedging between 0% and 100% of the exposure is allowed). For $V_c^l < 0$ and $V_c^u > w_c$, over- and under-hedging, respectfully, are permitted.

[15] Nevertheless, it is straightforward to relax this assumption and allow for constraints that vary among different currencies.

$\boldsymbol{\phi}_t$ are allowed to take. In the joint optimization problem, these two vectors are determined in a combined fashion and influence each other. Hence, the currency exposure constraints in the joint optimization have to be enforced with a linear inequality constraint specification.

Consider the definition of $\boldsymbol{\theta}_t$ and the constraints imposed by the upper and lower bounds, as obtained in Eq. (3.3.22). Then, a joint international portfolio optimization problem can be augmented for currency exposure constraints via an additional inequality constraint given by

$$\boldsymbol{Q}\boldsymbol{\theta}_t \leq \boldsymbol{D},$$

with

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{C}^T & -\boldsymbol{I}_M \\ -\boldsymbol{C}^T & \boldsymbol{I}_M \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{D} = \begin{pmatrix} V^u \mathbf{1}_M \\ -V^l \mathbf{1}_M \end{pmatrix},$$

where $\boldsymbol{Q} \in \mathbb{R}^{2M \times (N+M)}$ and $\boldsymbol{D} \in \mathbb{R}^{2M \times 1}$.

3.4 Empirical Analysis

The goal of this section is to empirically examine the regularized international portfolio optimization methods out of sample. We compare the proposed joint portfolio optimization approach with its currency overlay counterpart and the $1/N$ benchmark.

3.4.1 Data and Summary Statistics

The analysis is based on data ranging from January 1999 to December 2019. The data was obtained from Thomson Reuters Datastream and Bloomberg and consists of 21 equity broad market indices. We consider ten major developed-market currencies, the US dollar, euro, Japanese yen, Swiss franc, British pound, Canadian dollar, New Zealand dollar, Singapore dollar, Norwegian krone, and Australian dollar. All of the assets are denominated in one of these currencies and there is at least one asset denominated in every currency throughout the analysis. The data series are available at a daily frequency and the sample period ranges from January 1999, when the euro was introduced to the global financial markets, to December 2019. Table 3.1 shows the average annual returns and annual volatilities for the ten currency zones. The returns are converted to the US dollar and are benchmarked against their fully hedged counterparts.[16]

The average annual returns fall between 5 and 13 percent with the lowest average return of 5.51% being observed for Japan. On the other hand, with 12.79% New Zealand shows the highest average return over the sample period. The table further shows that fully hedging the assets yields a lower average return for

[16] Note that there is no fully hedged number reported for the US dollar since returns in the domestic currency cannot be currency hedged.

Table 3.1: Summary Statistics: 1999 to 2019.

This table presents the summary statistics for the dataset of 21 equity broad market indices for the period from 1999 to 2019. All returns are converted to the US dollar. A fully hedged asset position consists of a long position in an asset and a short position of the same size in the currency the asset is denominated in. Both average returns and return volatilities are computed with daily data and are reported as annualized percentages.

	Unhedged	Fully Hedged	Unhedged	Fully Hedged
	Annual Average Return		Annual Volatility	
Australia	10.91	7.49	20.67	12.20
Canada	10.20	8.91	18.87	12.87
Switzerland	9.70	9.28	18.99	17.18
Eurozone	6.66	7.08	19.24	14.94
United Kingdom	5.73	6.05	16.44	13.51
Japan	5.51	7.13	16.26	17.23
Norway	12.01	11.18	24.73	18.63
New Zealand	12.79	8.40	19.10	10.93
Singapore	10.35	10.02	20.98	18.17
United States	7.75		14.16	

all markets except for the Eurozone, the United Kingdom, and Japan. Furthermore, the volatilities of the fully hedged returns are lower than those of their unhedged counterparts for all markets except for Japan. This shows, as already previously observed in the international portfolio optimization literature, that full hedging has the potential to improve the risk–return spectrum by lowering the portfolio volatility.

Table 3.2 reports the full-sample correlations of currency returns (Panel A) and the cross–correlations between currencies and stocks (Panel B). Panel A shows that all currency returns are positively correlated. Moreover, geographically closely related markets such as Australia and New Zealand or Switzerland and the Eurozone exhibit particularly high currency return correlations. The cross–correlations between currencies and equities in Panel B are generally positive with only the Japanese yen, Swiss franc, and the US dollar being negatively correlated to any of the stock markets. This is similar to the findings of Campbell et al. [2010], who found that the Swiss franc and the Japanese yen move against global equity markets and are consequently attractive for risk-minimizing international investors based in the US. Panel B further shows that these three currencies exhibit very low correlations with international stock markets. This further increases their value to international investors by acting as a potential source of additional diversification. Consequently, these currencies are known as the so-called safe-haven currencies.

Note that the currency correlations, as shown in Panel A of Table 3.2, are considered in both separate and joint optimization problems, while the cross–correlations between currencies and assets are only considered in the joint optimization. The reason for this is that the currency overlay approach only takes into account

Table 3.2: Cross-Country Return Correlations: 1999 to 2019.

This table presents the cross-country correlations of currency and stock returns. Each cell of Panel A reports the correlation of the currency returns between the row and the column currencies. Panel B presents the same for the stocks. Here, each cell is a correlation between an equity market, indicated by the row, and a currency, represented by the column. The correlations are all reported as an average across all possible base currencies.

	AUD	CAD	CHF	EUR	GBP	JPY	NZD	USD	NOK	SGD
					Panel A: Currencies					
AU	1.00									
CA	0.56	1.00								
CH	0.33	0.26	1.00							
EU	0.47	0.43	0.69	1.00						
UK	0.34	0.49	0.41	0.55	1.00					
JP	0.12	0.27	0.43	0.31	0.26	1.00				
NZ	0.68	0.41	0.37	0.45	0.32	0.09	1.00			
US	0.17	0.51	0.37	0.38	0.53	0.57	0.15	1.00		
NO	0.48	0.49	0.48	0.65	0.48	0.19	0.34	0.27	1.00	
SG	0.49	0.60	0.56	0.58	0.57	0.60	0.42	0.75	0.48	1.00
					Panel B: Stocks					
AU	0.76	0.39	0.05	0.21	0.12	-0.19	0.55	-0.10	0.34	0.20
CA	0.60	0.67	-0.02	0.17	0.22	-0.10	0.40	0.03	0.36	0.24
CH	0.47	0.34	0.31	0.28	0.23	0.00	0.42	0.09	0.33	0.27
EU	0.52	0.37	0.18	0.45	0.25	-0.17	0.45	-0.04	0.44	0.22
UK	0.55	0.50	0.11	0.31	0.44	-0.09	0.43	0.10	0.44	0.32
JP	0.43	0.41	0.17	0.21	0.30	0.27	0.31	0.28	0.23	0.39
NZ	0.60	0.34	0.12	0.25	0.17	-0.10	0.76	-0.02	0.28	0.23
US	0.51	0.56	0.08	0.25	0.31	0.05	0.41	0.33	0.32	0.45
NO	0.52	0.37	0.04	0.23	0.19	-0.21	0.31	-0.13	0.56	0.11
SG	0.59	0.41	0.03	0.13	0.18	-0.07	0.46	0.02	0.25	0.35

the correlation between the return of the asset portfolio and the currency returns.[17] From Panel B, it is clear that for the period covered by the dataset, the cross–correlations are different from zero and take on values of up to 0.76 for correlations between local currencies and local markets. Therefore, the joint optimization approach proposed in this paper has the potential to improve the performance by taking into account this additional information.

The potential benefits of the joint optimization are further depicted in Fig. 3.1, which shows the in-sample efficient frontiers for both the separate and the joint international portfolio optimization problems. The frontiers were derived by solving the joint and the separate optimization problems from Section 3.3 for a range of risk-aversion parameters.

[17] This is the reason that the currency overlay approach could be understood as a form of dimensionality reduction.

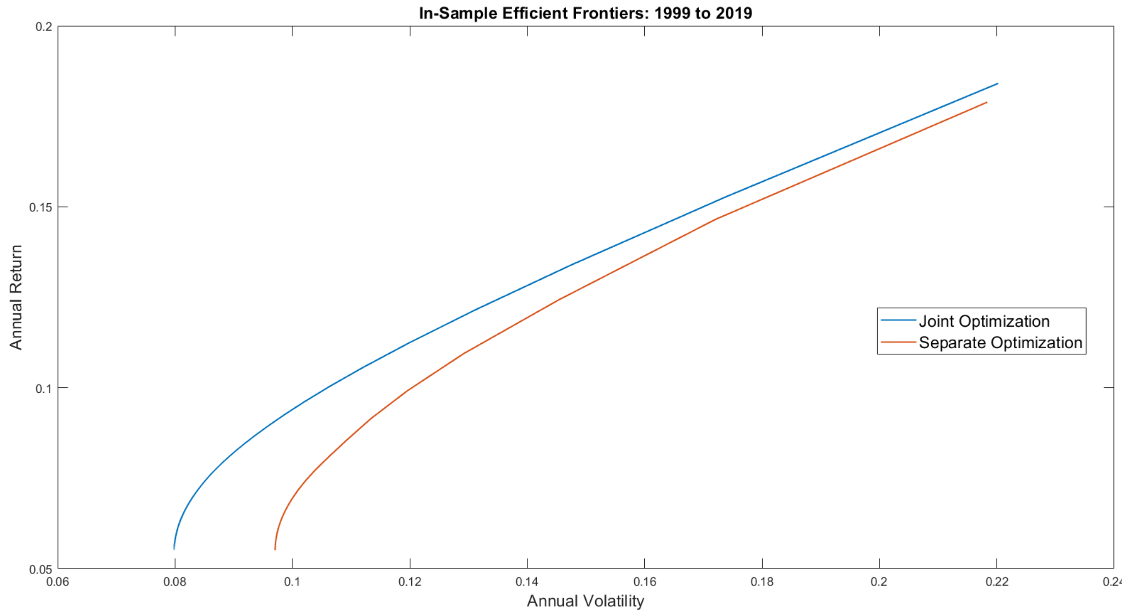

Figure 3.1: This figure depicts the in-sample efficient frontiers for the dataset of 21 equity indices and 10 currencies from 1999 to 2019. The curves were derived by varying the risk aversion parameter γ in both steps of the separate optimization as well as in the joint optimization.

As this analysis is carried out in sample (i.e., the same data is used for the estimation and the performance evaluation), it is clear that the joint optimization approach by definition has to provide improved results compared to the separate optimization approach. The separate approach is clearly suboptimal as it does not take into consideration the dependencies between the individual assets and the currencies but only between the whole asset portfolio and the currencies. The two frontiers in Fig. 3.1 would be equal in the case that assets and currencies are uncorrelated. In this case, the joint and the separate optimization would take into consideration the same amount of information.

3.4.2 Out-of-Sample Analysis

In this section, the out-of-sample performance of the regularized international portfolio optimization approaches is investigated. These approaches encompass the sparse and/or stable methods described in Section 3.3, as well as the standard mean–variance and $1/N$ portfolio benchmarks. In addition to the (currency unconstrained) SAS portfolios, we include the same optimization methods with added constraints on the allowed net currency exposures. Such constraints are unique to the international portfolio optimization problem and are driven by a practical consideration arising from potential regulatory restrictions. For these portfolio rules, the maximal absolute net currency exposure is bounded by 30%, which is equivalent to setting $-V_c^l = V_c^u = 0.3$ in Eq. (3.3.22). The list of regularized international portfolio optimization rules we analyze out of sample is:

- **MV**: The classical mean–variance portfolio (without regularization).
- **NC1**: The sparse portfolio constructed by imposing an $\mathcal{L}^1$-norm constraint.
- **NC2**: The stable portfolio constructed by imposing an $\mathcal{L}^2$-norm constraint.
- **NC1-CC**: The sparse and stable portfolio constructed by imposing an $\mathcal{L}^1$-norm constraint and shrinkage of the covariance matrix towards a given constant correlation matrix.
- **NC1-NC2**: The sparse and stable portfolio constructed by constraining both the $\mathcal{L}^1$- and the $\mathcal{L}^2$-norm.
- **RULE-CB**: A given optimization rule (RULE) as specified above with a constraint that the absolute value of the net currency exposure to any currency does not exceed 30%, i.e., a currency bound (CB).
- **1/N**: The fully hedged $1/N$ portfolio (i.e., the equally-weighted portfolio).

For each portfolio optimization rule, a separate optimization (i.e., the currency overlay) and a corresponding joint optimization over assets and currencies in a single optimization step are carried out. We aim to examine the out-of-sample performance of the regularized joint optimization approach versus the regularized separate and non-regularized optimization methods. Even though the in-sample optimality of the joint optimization is apparent, this might not be the case out of sample. The joint optimization approach requires an estimation of dependencies between single assets and currencies. Such granular estimation is alleviated in the separate optimization that can consequently be understood as an approach that by construction performs a dimensionality reduction. Therefore, to shed light on the best practice in the optimization of multi-currency portfolios, an analysis of the out-of-sample performance of both optimization approaches is required. Moreover, the $1/N$ portfolio is presented as a fully hedged portfolio that serves as a natural model-free benchmark. The reasoning for this arises from full hedging being the most extensively studied currency hedging strategy in the international asset allocation literature.

Following the existing literature, we consider the US dollar as the base currency throughout the analysis and accordingly report all the results in USD.[18] We assume that an investor rebalances her portfolio with a monthly frequency, while daily data is used for the estimation. We further assume that an investor only hedges her currency exposure one period ahead (i.e., she only enters forward contracts at time t that have their expiry at time $t + 1$). The optimization employs a rolling estimation window of two years (i.e., 500 daily observations). The covariance matrix is then computed according to the given respective portfolio optimization rule outlined above and the Bayes-Stein estimator is used for the shrinkage of the mean, see Jorion [1986].

[18] As noted above, the model is general and allows for an arbitrary base currency.

The risk-aversion parameter is set to $\gamma = 3$, which is a choice suitable for most portfolio allocation decisions, see for example Fabozzi et al. [2007]. The other model parameters, λ_1^a, λ_1^c, λ_2^a, and λ_2^c, corresponding to the $\mathcal{L}^1$ and $\mathcal{L}^2$ penalization parameters for assets and currencies, are determined via cross-validation. Here, we follow the approach of DeMiguel et al. [2009a], where leave-one-out cross-validation is carried out on the data sample of each estimation window. Using adaptive grid-search, the combination of values that achieves the lowest cross-validation portfolio variance is then chosen as the model parameter set. This set is used to estimate the optimal asset and currency weights on the given rolling window.

Table 3.3 details the results of the out-of-sample analysis. All results are reported net of transaction costs, with the transaction costs being computed as a proportion of the traded assets and currencies based on the bid-ask spread. We assume transaction costs of twenty basis points for the assets and two basis points relative to the notional of each entered currency forward contract. The derivation of the transaction costs adjusted multi-currency portfolio returns is presented in the Appendix. For each portfolio, we report the annualized Sharpe Ratio, denoted by SR, the annualized Sortino Ratio, denoted by SOR, the annualized certainty equivalent return, denoted by CEV, the annualized volatility, the turnover of the assets and currencies, denoted by $Turnover^a$ and $Turnover^c$, respectively, and the maximum drawdown, denoted by MD. The asset turnover reported in Table 3.3 is an average monthly turnover, as widely used in the existing literature, see, for example, Li [2015] or DeMiguel et al. [2009a]. As not only trading assets but also trading currencies comes at a cost, we also report the turnover of the currencies. The foreign currency exposure is hedged via the forward contracts with delivery at their maturity. Thereby, an investor at every time t makes a new decision on the amount of entered currency forward contracts (i.e., possibly long or short). Thus, the turnover for the currencies is simply the average over all absolute currency weights $|\phi_{c,t}|$, for $c = 2, \ldots, M+1$.

The results in Table 3.3 show that the portfolios derived with the joint optimization consistently achieve higher Sharpe ratios compared to their separate optimization counterparts. The comparison between portfolio rules with and without currency bounds shows that the SAS portfolio rules that include bounds on the absolute currency exposures (i.e., NC1-CC-CB and NC1-NC2-CB) outperformed the rules without such constraints (i.e., NC1-CC and NC1-NC2) in terms of Sharpe ratios. The other portfolio rules exhibit slightly worse performance when the currency bounds are introduced. This holds true for both separate and joint optimization. Additionally, both separately and jointly optimized portfolio rules outperform the fully hedged $1/N$ and MV portfolio benchmarks.

The results in terms of Sortino ratios and certainty equivalents are similar to the ones based on the Sharpe ratio. Most notably, portfolios constructed with the joint optimization approach consistently outperform their separate optimization counterparts. The norm-constrained portfolios achieve improved out-of-sample performance compared to the $1/N$ and MV benchmarks. This shows the importance of including the regularization into the international portfolio optimization process. Moreover, the performance of the SAS

Table 3.3: Out-of-Sample Portfolio Performance.

This table presents the performance measures for the international portfolio optimization rules introduced in this paper. The abbreviations of the portfolio rules are given as defined in Section 3.4.2. The employed data ranges from January 1999 to December 2019 and the first portfolio is formed in January 2001. The data consists of 21 equity indices from ten different countries and the corresponding spot and forward currency rates. All measures are computed out-of-sample and net of transaction costs.

	SR		SOR		CEV		Volatility		$Turnover^a$		$Turnover^c$		MD	
	Sep.	Joint	Sep.	Joint	Sep.	Joint	Sep.	Joint	Sep.	Joint	Sep.	Joint	Sep.	Joint
MV	0.3481	0.4625	0.3939	0.5314	0.0217	0.0322	0.2167	0.2701	0.0889	0.1002	0.3536	0.5753	0.6359	0.6808
MV-CB	0.2488	0.4119	0.2779	0.4919	0.0091	0.0390	0.1924	0.2005	0.0877	0.0960	0.2594	0.3982	0.5884	0.5778
NC1	0.4586	0.5035	0.5474	0.5702	0.0489	0.0580	0.1091	0.1425	0.0245	0.0224	0.1530	0.0953	0.5166	0.5110
NC1-CB	0.4496	0.5048	0.5193	0.6020	0.0473	0.0552	0.1046	0.1170	0.0232	0.0233	0.1637	0.1334	0.5281	0.4476
NC2	0.5107	0.5600	0.6142	0.6620	0.0582	0.0641	0.1343	0.1296	0.0204	0.0169	0.0500	0.1427	0.5942	0.4785
NC2-CB	0.5198	0.5171	0.6319	0.6246	0.0583	0.0573	0.1255	0.1212	0.0205	0.0183	0.0897	0.1884	0.5770	0.4673
NC1-CC	0.3699	0.4542	0.4307	0.5092	0.0394	0.0510	0.1150	0.1443	0.0286	0.0219	0.1276	0.0888	0.5437	0.5307
NC1-CC-CB	0.3880	0.5168	0.4397	0.6125	0.0413	0.0569	0.1110	0.1182	0.0262	0.0216	0.1375	0.1271	0.5628	0.4541
NC1-NC2	0.4962	0.6172	0.6305	0.7646	0.0488	0.0659	0.0883	0.1082	0.0165	0.0216	0.1857	0.1127	0.4504	0.4389
NC1-NC2-CB	0.5532	0.6810	0.6882	0.8314	0.0538	0.0715	0.0881	0.1046	0.0172	0.0193	0.1803	0.1557	0.4612	0.4392
1/N	0.3413		0.4071		0.0357		0.1302		0.0005		0.1054		0.5504	

portfolios is improved by adding additional bounds on the allowed absolute currency exposures. However, this is not the case for the rest of the portfolio rules.

Analyzing annualized volatilities shows that the joint optimization generally exhibits higher volatility compared to the separate optimization rules. However, as seen in the increased Sharpe and Sortino ratios, such an increase in risk is compensated by increased average returns. Moreover, the comparison between portfolio rules with and without additional currency exposure constraints shows that the introduction of the currency bounds generally leads to a volatility reduction for both joint and separate optimization rules.

On the other hand, the joint optimization tends to decrease the portfolio maximum drawdowns compared to the separate optimization for the norm-constrained portfolio rules. This shows that even though the joint optimization approach likely increases the portfolio risk measured by volatility, it tends to decrease the portfolio risk measured by the maximum drawdown. Moreover, imposing additional currency exposure constraints also tends to decrease the maximum drawdowns.

Observing the asset and currency turnovers, one can conclude that the joint optimization reduces the asset turnover compared to the separate optimization for the norm-constrained portfolios. On the other hand, less clear patterns can be observed in the currency turnovers. For example, the currency turnover for the joint optimization in comparison to the separate optimization case is reduced for the SAS portfolio rules.

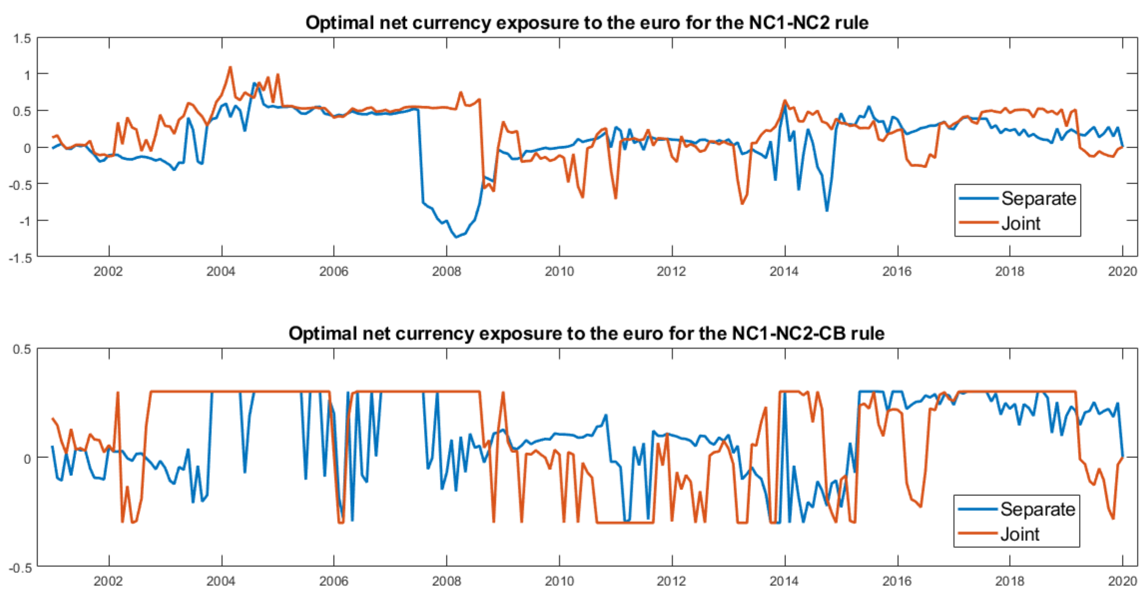

Figure 3.2: This figure illustrates the computed optimal currency exposure in EUR for the two sparse and stable portfolio rules, i.e., the NC1-NC2 portfolio (top) and the NC1-NC2-CB portfolio (bottom), for both separate and joint optimization approaches.

All in all, this analysis shows that regularization and joint optimization improve the performance of internationally diversified portfolios out of sample. Such an improvement is achieved in terms of increased risk-adjusted returns (i.e., Sharpe and Sortino ratios) as well as an increased certainty equivalent, and reduced maximum drawdown net of transaction costs. The best performing model is the SAS portfolio with the additionally imposed currency bounds (i.e., NC1-NC2-CB). Compared to the $1/N$ benchmark, this portfolio rule exhibits doubled Sharpe and Sortino ratios as well as a doubled certainty equivalent, while reducing the portfolio maximum drawdown by 20%.

Figure 3.2 depicts the computed optimal currency exposure in EUR for the SAS portfolio rules with and without currency bounds (i.e., NC1-NC2 and NC1-NC2-CB) for both separate and joint optimization. One can observe the difference between the optimal currency exposure as calculated by the separate and joint optimization. Generally, this difference is not substantial and both approaches manage to enhance the out-of-sample portfolio performance, as seen in Table 3.3. Moreover, the optimal exposures are stable over time, whereby the joint approach tends to provide increased stability compared to the separate optimization approach. Clearly, the optimal currency exposure computed by the portfolio rule with the currency bounds exhibits an increased stability over time since the currency bound can be interpreted as an additional form of shrinkage. Such enhanced stability is one of the main drivers for the out-of-sample outperformance of the NC1-NC2-CB rule.

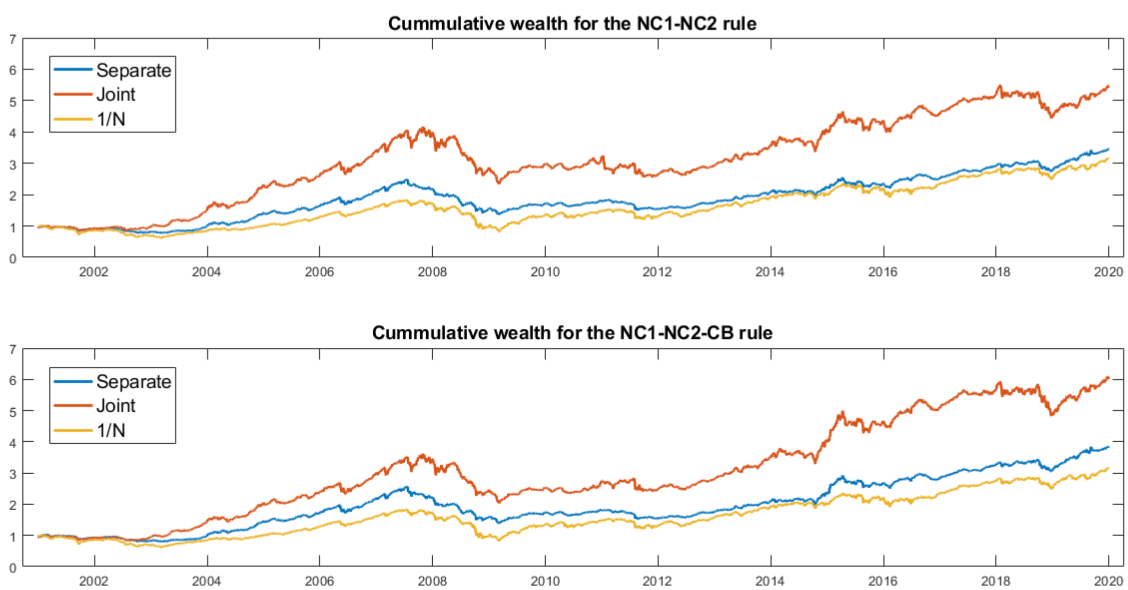

Figure 3.3: This figure portrays the cumulative wealth evolution for the two sparse and stable portfolio rules, i.e., the NC1-NC2 portfolio (top) and the NC1-NC2-CB portfolio (bottom), with an initial wealth of \$1. All reported values are adjusted for transaction costs.

Figure 3.3 illustrates the evolution of the cumulative wealth for the two SAS portfolio rules, NC1-NC2 and NC1-NC2-CB. For each rule, the separate and joint optimizations are depicted, and the fully hedged $1/N$ portfolio is shown as the benchmark. One can observe a consistent outperformance of the joint optimization approach compared to its separate counterpart as well as the $1/N$ portfolio. This outperformance is here achieved in terms of cumulative wealth and is observed over the whole backtest period. Furthermore, one can observe that the constraints on the currency exposure improve the out-of-sample performance in terms of cumulative wealth for the joint SAS portfolio rule. The other portfolio rules covered in this paper show similar cumulative wealth performance patterns, with the joint optimization consistently outperforming the separate optimization as well as the $1/N$ portfolio net of transaction costs.

3.5 Conclusions

In this paper, we develop a sparse and stable international asset allocation framework where asset and currency weights are determined in a regularized fashion. Moreover, we study if computing the assets and currencies in a joint optimization approach provides out-of-sample benefits compared to the widespread approach of currency overlay strategies employed for managing the portfolio currency risk. By performing such an analysis, we contribute to the literature on multi-currency asset allocation.

To the best of our knowledge, this is the first paper presenting a methodology for sparse and stable multi-currency asset allocation. Similar approaches, aimed at reducing parameter uncertainty in the mean–variance framework, have been proposed for the allocation of assets denominated in a single currency. We extend these approaches by proposing a framework where asset as well as currency weights are determined in a sparse and stable manner. Furthermore, we present an extensive empirical analysis on the out-of-sample performance of the international asset allocation problem and the corresponding currency risk management policy. We demonstrate that the proposed sparse and stable joint optimization outperforms the equivalent approaches based on separate (i.e., currency overlay) optimization as well as the equally-weighted and the non-regularized global portfolios net of transaction costs.

This result shows that employing currency overlay strategies is suboptimal not only in sample but also from the out-of-sample perspective. The joint multi-currency portfolio optimization takes into account the individual dependencies between assets and currencies when constructing an optimal (mean–variance) portfolio. While such an approach has a larger potential to suffer from parameter uncertainty and overfitting, we employ several regularization approaches that tackle these problems by inducing sparsity and stability of portfolio weights. We show that the out-of-sample outperformance of the joint optimization is driven by regularization that induces a bias-variance trade-off leading to improved portfolio performance. Moreover, constraining the allowed net currency exposures improves the performance of sparse and stable portfolios. Such constraint can also be interpreted as an additional form of shrinkage.

This paper allows for possible extensions of the proposed international asset optimization approach into several research directions. For example, one could generalize the introduced methodology based on the mean–variance utility to the sparse and stable tail-based (e.g., expected shortfall) risk measures. Moreover, time-series models capturing the empirical stylized facts present in financial data (e.g., volatility clustering, asymmetry, and excess kurtosis) could also be employed for modeling the asset and currency return dynamics. The performance of such models can then be benchmarked upon our model. The current framework could also be extended to the multi-period setting. Thereby, the asset and currency rebalancing would be considered within a longer time horizon. Such a model would require the development of a method based on stochastic programming that captures the multi-period asset and currency allocation decision dynamics. On the empirical side, the performance of bond portfolios could be investigated as well. Given the generally low correlations between currency and bond returns, the common practice of institutional investors is to fully hedge bond portfolios. One could explore if this strategy could potentially be improved within a joint optimization format. Furthermore, one could also study the performance of multi-currency portfolios exposed to emerging market currencies in combination with the already explored developed markets.

3.6 Appendix

3.6.1 Elastic Net Sparse and Stable Multi-Currency Portfolio Optimization

Consider the notation from Section 3.3.3.3. Then, the elastic net SAS separate international portfolio optimization is given by

$$\begin{aligned}
\boldsymbol{x}_t^* &= \underset{\boldsymbol{x}_t}{\arg\min} \left\{ \frac{\gamma}{2} \boldsymbol{x}_t^T \hat{\boldsymbol{\Sigma}}^a \boldsymbol{x}_t - \boldsymbol{x}_t^T \hat{\boldsymbol{\mu}}^a + \lambda_1^a \|\boldsymbol{x}_t\|_1 + \lambda_2^a \|\boldsymbol{x}_t\|_2^2 \right\}, \\
&\quad \text{subject to} \qquad \boldsymbol{x}_t^T \mathbf{1} = 1, \\
\boldsymbol{\phi}_t^* \mid \boldsymbol{x}_t^* &= \underset{\boldsymbol{\phi}_t}{\arg\min} \left\{ \frac{\gamma}{2} \boldsymbol{\phi}_t^T \hat{\boldsymbol{\Sigma}}^c \boldsymbol{\phi}_t - \boldsymbol{\phi}_t^T \hat{\boldsymbol{\mu}}^c + \gamma \boldsymbol{x}_t^{*T} \hat{\boldsymbol{\Sigma}}^{ac} \boldsymbol{\phi}_t + \lambda_1^c \|\boldsymbol{\phi}_t\|_1 + \lambda_2^c \|\boldsymbol{\phi}_t\|_2^2 \ \mid \boldsymbol{x}_t^* \right\},
\end{aligned} \tag{3.6.1}$$

where λ_2^a and λ_2^c represent the penalization intensity induced by the $\mathcal{L}^2$-norm of the asset and currency weights, respectively.

Equivalently, the elastic net SAS joint international portfolio optimization is given by

$$\begin{aligned}
\boldsymbol{\theta}_t^* &= \underset{\boldsymbol{\theta}_t}{\arg\min} \left\{ \frac{\gamma}{2} \boldsymbol{\theta}_t^T \hat{\boldsymbol{\Sigma}} \boldsymbol{\theta}_t - \boldsymbol{\theta}_t^T \hat{\boldsymbol{\mu}} + \|\boldsymbol{S}\boldsymbol{\theta}_t\|_1 + \|\boldsymbol{S}_2\boldsymbol{\theta}_t\|_2^2 \right\} \\
&\quad \text{subject to} \qquad \boldsymbol{\theta}_t^T \mathbf{q}_{N,M} = 1,
\end{aligned} \tag{3.6.2}$$

where $\boldsymbol{S}_2$ is defined defined analogously to $\boldsymbol{S}$ with all off-diagonal elements equal to zero and first N diagonal elements equal to λ_2^a and the remaining M diagonal elements equal to λ_2^c. When including different penalty terms for assets and currencies, the elastic net SAS international portfolio optimization model includes four penalization hyperparameters. These hyperparameters generally have to be determined via cross-validation. The implementation of the problems introduced here is straightforward since the sparse problem can be cast as a quadratic programming problem. The ridge penalization, on the other hand, can be cast as shrinkage of the sample covariance matrix towards the identity matrix.

3.6.2 Trading Costs in International Markets

Consider the notation from Section 3.3. The hedged portfolio return net of transaction costs can then be expressed as

$$R_{\mathcal{P},t+1}^{h,net} = R_{\mathcal{P},t+1}^{h} - \tau_{t+1}^{a} - \tau_{t+1}^{c}, \tag{3.6.3}$$

where τ_{t+1}^a and τ_{t+1}^c denote the total trading costs stemming from the trade of assets and currencies, respectively. Here, we assume that all transaction costs are proportional to the amount of assets traded, which means that any fixed costs are ignored. This is a reasonable assumption for large institutional investors as fixed costs are negligible in relation to the total portfolio value. Our modeling approach for the transaction costs closely resembles the one used in Li [2015] but is extended to the international portfolio setting. The transaction costs associated with the assets are defined as

$$\tau_{t+1}^{a} = \sum_{j=1}^{N} \tau_{j,t+1}^{a}(x_{j,t+1} - x_{j,t+1}^{-}), \tag{3.6.4}$$

where $x_{j,t+1}^{-}$ denotes the weight of asset i before rebalancing at time $t+1$ takes place. When an investor trades assets denominated in a foreign currency, she not only faces costs proportional to the traded assets but also to the currencies as she first needs to exchange the domestic currency into a foreign currency (or vice-versa). These trading costs are captured in the trading costs associated with the currencies τ_{t+1}^{c}, together with the trading costs stemming from the forward contracts. These costs are given as

$$\tau_{t+1}^{c} = \sum_{c=2}^{M+1} \tau_{c,t+1}^{c}\phi_{c,t+1} + \sum_{c=2}^{M+1} \tau_{c,t+1}^{c}(w_{c,t+1} - w_{c,t+1}^{-}), \tag{3.6.5}$$

where an investor only faces costs for the change in implied currency exposures and not the change of every single asset weight. This is because the investor can use the money from selling assets in currency c to buy other assets in this currency without an exposure to the currency exchange rates. Thus, only the amount of change in the implied currency exposures is subject to currency transaction costs. Here, $w_{c,t+1}^{-}$ is defined analogously to $x_{j,t+1}^{-}$ and represents the implied currency exposure at time $t+1$ before rebalancing. The transaction costs $\tau_{i,t+1}^{a}$ and $\tau_{c,t+1}^{c}$ can be expressed as a function of the bid-ask spread by

$$\begin{aligned} \tau_{j,t+1}^{a} &= \begin{cases} \log\left(\frac{1+c_{t+1}^{a}}{1-c_{t}^{a}}\right) & \text{if} \quad x_{j,t+1} > x_{j,t+1}^{-}, \\ \log\left(\frac{1-c_{t+1}^{a}}{1+c_{t}^{a}}\right) & \text{otherwise}, \end{cases} \\ \tau_{c,t+1}^{c} &= \begin{cases} \log\left(\frac{1+c_{t+1}^{c}}{1-c_{t}^{c}}\right) & \text{if} \quad \phi_{c,t+1} > 0 \quad \text{or} \quad w_{c,t+1} > w_{c,t+1}^{-}, \\ \log\left(\frac{1-c_{t+1}^{c}}{1+c_{t}^{c}}\right) & \text{otherwise}. \end{cases} \end{aligned} \tag{3.6.6}$$

Here, we define c_t^a as one half of the proportional bid-ask spread for the assets and c_t^c as the equivalent measure for the currencies.[19] The two measures are defined as

$$c_t^a = \frac{P_{j,t}^{ask} - P_{j,t}^{bid}}{2P_{j,t}} \quad \text{and} \quad c_t^c = \frac{S_{c,t}^{ask} - S_{c,t}^{bid}}{2S_{c,t}}, \tag{3.6.7}$$

where $P_{j,t}^{ask}$, $P_{j,t}^{bid}$, and $P_{j,t}$ are the ask, bid, and mid price for asset j at time t. Accordingly, $S_{c,t}^{ask}$, $S_{c,t}^{bid}$, and $S_{c,t}$ are the ask, bid, and mid price for currency c. For simplicity, we assume constant proportional bid-ask spreads for both assets and currencies. Following Jones [2002], who found that the average proportional bid-ask spread for currencies in the year 2000 was around 0.2%, we set $c_t^a = 0.001$ for all assets j and all times t. Similarly, Neely et al. [2009] report the average proportional bid-ask spread for currencies to be around 0.02%, which motivates the choice $c_t^c = 0.0001$ that is again assumed to hold for all currencies at all times.

19 If at time $t+1$ we sell an asset for $P_{i,t+1}^{bid} S_{c_i,t+1}^{bid}$ and the asset was previously bought at time t at the price $P_{i,t}^{ask} S_{c_i,t}^{ask}$, then the log return of the trade is given by

$$\begin{aligned}
\log\left(\frac{P_{i,t+1}^{bid} S_{c_i,t+1}^{bid}}{P_{i,t}^{ask} S_{c_i,t}^{ask}}\right) &= \log\left(\left(\frac{P_{i,t+1} - \frac{\delta^a}{2}}{P_{i,t} + \frac{\delta^a}{2}}\right)\left(\frac{S_{c_i,t+1} - \frac{\delta^c}{2}}{S_{c_i,t} + \frac{\delta^c}{2}}\right)\right) \\
&= \log\left(\left(\frac{P_{i,t+1}(1 - c_{t+1}^a)}{P_{i,t}(1 + c_t^a)}\right)\left(\frac{S_{c_i,t+1}(1 - c_{t+1}^c)}{S_{c_i,t}(1 + c_t^c)}\right)\right) \\
&= \log\left(\left(\frac{P_{i,t+1} S_{c_i,t+1}}{P_{i,t} S_{c_i,t}}\right)\left(\frac{1 - c_{t+1}^a}{1 + c_t^a}\right)\left(\frac{1 - c_{t+1}^c}{1 + c_t^c}\right)\right) \\
&\approx \bar{R}_{i,t+1} - \log\left(\frac{1 - c_{t+1}^a}{1 + c_t^a}\right) - \log\left(\frac{1 - c_{t+1}^c}{1 + c_t^c}\right),
\end{aligned}$$

where δ^a and δ^c denote the bid-ask spread of assets and currencies, respectfully. Clearly, it holds that $c_t^a = \frac{\delta^a}{2P_t}$ and $c_t^c = \frac{\delta^c}{2S_t}$.

Chapter 4

Accelerated American Option Pricing with Deep Neural Networks

Authors:

David Anderson, Department of Banking and Finance, University of Zurich

Urban Ulrych, Department of Banking and Finance, University of Zurich & Swiss Finance Institute

Abstract

Given the competitiveness of a market-making environment, the ability to speedily quote option prices consistent with an ever-changing market environment is essential. Thus, the smallest acceleration or improvement over traditional pricing methods is crucial to avoid arbitrage. We propose a novel method for accelerating the pricing of American options to near-instantaneous using a feed-forward neural network. This neural network is trained over the chosen (e.g., Heston) stochastic volatility specification. Such an approach facilitates parameter interpretability, as generally required by the regulators, and establishes our method in the area of eXplainable Artificial Intelligence (XAI) for finance. We show that the proposed deep explainable pricer induces a speed-accuracy trade-off compared to the typical Monte Carlo or Partial Differential Equation-based pricing methods. Moreover, the proposed approach allows for pricing derivatives with path-dependent and more complex payoffs and is, given the sufficient accuracy of computation and its tractable nature, applicable in a market-making environment.

Keywords: American Option Pricing, Deep Neural Networks, Explainable Artificial Intelligence, Speed-Accuracy Trade-Off, Market Making, Heston Model, Computational Finance.

JEL Classification: C45, C63, G13.

4.1 Introduction

Equity derivative market-making environments are known for their competitive pricing requirements in terms of both speed and accuracy. Large fluctuations of the underlying security price, together with other pricing parameters, can occur fast and without warning and are influenced by company news, as well as macroeconomic, financial, and political factors. If market makers are not able to adjust the corresponding option prices instantaneously, arbitrage opportunities can occur.

There are many reasons why a market maker could be slow in updating the price of a product. Firstly, high overall system latency can potentially cause a delay in the arrival of market data to the trading system. Secondly, the calibration process of an implied volatility surface (e.g., using a chosen parametric volatility model such as the Heston stochastic volatility model) based on new market price arrivals can represent another potential source of latency. The third step, however, is typically the most prevalent source of computational latency. Here, current market data and the corresponding fitted parametric volatility model are inputted into an algorithm to evaluate the price of a derivative. Depending on the option type, the computational performance of a pricer is affected by its underlying algorithm and the complexities in the payoff structure of the option.

In this paper, we propose a novel method that accelerates the American-style option pricing mechanism. We do not investigate the closely connected data acquisition, calibration, and price publishing parts. We assume that a trading system gathers market data and other parameters during the data acquisition and calibration phases and provides them to the pricing engine. We focus on the pricing algorithm that calculates what is known as the fair price of the option. This is the price at which no theoretical arbitrage exists – the probabilistic expectation of a profit on both sides of a transaction is equal to zero. In practice, a bid-ask spread is then generated around the computed fair price, and both bid and ask prices are subsequently published on the market platform.[1]

Since the main objective of the proposed method is its adoption in market making, a key responsibility of the proposed algorithm lies in its compliance with regulatory guidelines. Even though we use neural networks for the acceleration of American option pricing, we ensure that our algorithm is not a black box. This is achieved by generating a training set using a standard well-known pricing method (e.g., the PDE-based pricer under the Heston stochastic volatility model), as opposed to alternative neural network-based methods employing market data in the training set. Subsequently, we utilize the generated training set for training a neural network over the assumed (e.g., Heston-based) parameter space. By training the neural

[1] Depending on the current inventory (i.e., risk position) of the market-making trading desk that is pricing the option, the trading system can allocate aggressiveness to either the bid or the ask side by adjusting the spread around the fair price. Aggressiveness settings allow a trading desk to reduce its risk without hedging by attracting a specific direction of transactions.

network over a set of all possible parameter space configurations given a chosen parameter range offline, such an approach allows for: i) decreasing the latency in the pricing of American options to near-instantaneous and ii) enhancing the transparency and explainability of the American option pricing algorithm. Furthermore, by evaluating the neural network over the entire parameter space, we show that the prices resulting from the proposed deep explainable pricer are smooth with respect to the input parameters–one of the algorithmic criteria in eXplainable Artifical Intelligence.

In the pricing of path-dependent options, conventional methods rely extensively on computationally expensive numerical techniques. The most common pricing approaches are binomial models, numerical partial differential equation (PDE) methods, and Monte Carlo (MC) methods. Since the difficulties in pricing American options lie in the question of how to evaluate optimal exercise, the key differences between the methods arise from how they approach the optimal exercise decision. Cox et al. [1979] proposed the initial binomial tree method, which is fast and reasonably accurate. However, it does not work in the case of stochastic volatility. Furthermore, a trinomial tree that converges faster than the binomial tree was developed by Boyle [1986]. Tree models rely on constant volatility and determine the optimal exercise by comparing the value of immediate exercise and the discounted probability-weighted cashflows at each node.

To generalize to the non-constant volatility models, researchers developed methods based on numerical solutions to the Black-Scholes-Merton PDE, which holds in general for both European and American option styles. PDE methods for pricing American options admit the representation of the optimal exercise decision as a PDE boundary condition. The PDE with its boundary condition is solved numerically using various numerical solvers such as modified finite difference methods, finite element methods, and boundary projection methods. One of the pioneers in this area were Brennan and Schwartz [1977], who priced American put options using a boundary projection method. Later, Ikonen and Toivanen [2004] proposed an approach using the Crank-Nicolson method (i.e., a type of finite difference method, see Crank and Nicolson [1947]), which was found to be more efficient than the boundary projection method. Nevertheless, PDE methods are in general too slow for a market-making environment, especially in the presence of stochastic interest rates, stochastic volatility models, jump processes, multi-dimensional options, or derivatives with other complicated features.

Furthermore, researchers analyzed the use of statistical simulation methods known as Monte Carlo (MC) methods for option pricing. These methods are slower than the other methods mentioned above. However, they proved to be more flexible in terms of their ability to price derivatives under different volatility dynamics, more complex path-dependent payoff structures, and further departures from the basic Black et al. [1973] framework. Initial MC methods for American option pricing suffered under the curse of dimensionality since the decision process of optimal exercise relied on performing an additional MC simulation for each in-the-money path as proposed by Boyle et al. [1997]. Thus, the number of paths increased exponentially.

Famously, Longstaff and Schwartz [2001] proposed an MC method where the exercise decision relies upon the comparison of the current exercise payoff and a least-squares approximation of the option continuation value. This method circumvents the dimensionality problem of the nested MC approach and is thus highly effective.[2] However, the Longstaff-Schwartz (LS) method is too slow for a market-making environment. Recently, the area of rough volatility models has evolved, see Bayer et al. [2016]. Due to the non-Markovian dynamics of rough volatility models, pricing with Monte Carlo methods is required.

The use of artificial neural networks (ANNs) for option pricing started with Malliaris and Salchenberger [1993] and Hutchinson et al. [1994]. The authors show that the Black-Scholes-based European-style option valuation can be successfully recovered with a neural network.[3] These works serve as a valuable reference point that spurred a wide range of future research involving machine learning methods applied to option pricing and hedging.[4] Lajbcygier and Connor [1997] study the difference between hybrid neural network-based option valuation and observed market option prices and show that the pricing bias can be reduced with bootstrap methods. Anders et al. [1998] show that applying statistical inference techniques to build neural network pricing models leads to superior out-of-sample pricing performance compared to the Black-Scholes model. Garcia and Gençay [2000] incorporate financial domain knowledge in the form of a homogeneous option pricing function directly into the construction of an ANN pricer and show that such an approach reduces the out-of-sample mean squared prediction error. On a similar note, Dugas et al. [2009] show that the generalization performance of a learning algorithm can be improved by incorporating prior knowledge into its architecture. The authors propose new types of function classes that incorporate the a priori constraints based on the knowledge that the function to be learned is non-decreasing in its two arguments and convex in one of them. Pricing of S&P-500 European call options with a non-parametric modular neural network model is investigated in Gradojevic et al. [2009].

Neural networks have also been studied in the context of option hedging as well as implied volatility smile and its calibration. The European-style option hedging is investigated in Chen and Sutcliffe [2012] and Shin and Ryu [2012]. More recently, Buehler et al. [2019] present a deep hedging-based framework for hedging a portfolio of derivatives in the presence of market frictions. The implied volatility smile is, in a GARCH-neural network approach, studied in Meissner and Kawano [2001]. Computation of implied volatilities is, among others, also studied in Andreou et al. [2010] and Liu et al. [2019b].[5] More recently, Horvath et al. [2021] propose an application of neural networks in a two-step process where firstly, the European option pricing map is learned, and secondly, the European option implied volatility surfaces are

[2] More recently, Scheidegger and Treccani [2021] introduce a fast, scalable, and flexible framework for pricing American options in high-dimensional settings that substantially improves on the limitations of Monte Carlo methods.

[3] On a more recent note, Culkin and Das [2017] revisit and confirm the results from Hutchinson et al. [1994].

[4] An extensive literature review of the applications of neural networks to option pricing is provided in Ruf and Wang [2020].

[5] In a European option pricing framework, the obtained implied volatilities can be easily inverted back to the option prices by the use of the Black-Scholes formula.

calibrated to various parametric models.[6]

The literature presented up to now studies an application of ANNs to the pricing of European-style options. Additionally, applying machine learning methods to solving the American option pricing and hedging problem has already been explored. Kelly [1994] presents the first study of pricing and hedging American put options using neural networks. A similar analysis is presented also in Pires and Marwala [2004] and Morelli et al. [2004]. More recently, Gaspar et al. [2020] use neural networks to price American put options on four large U.S. companies. Generally, the American option pricing approaches fall into the categories of i) regression-enhancement for basket options as in the case of Kohler et al. [2010], ii) solving the optimal stopping problem as in the case of Becker et al. [2020], or iii) a mix of both as in Chen and Wan [2021]. Moreover, machine learning methods can be used to approximate value functions emerging in dynamic programming[7], for example arising in the American option pricing problem, see Ye and Zhang [2019].

This paper extends the current literature by using a direct application of deep neural networks to price specifically American-style options by learning not only asset price dynamics but also the dynamics of various volatility models, as well as the American option optimal exercise strategy. Building upon Horvath et al. [2021] who work in a European framework, we extend the pricing to American-style options. Moreover, our neural network learns the direct map between model parameters and price instead of the map between model parameters and an implied volatility surface. We further demonstrate that the proposed neural network pricer considerably improves upon the existing American-style option pricing models (e.g., PDE or Monte Carlo methods) in terms of the speed-accuracy trade-off.[8] As such, it is practically applicable in a market-making environment.

In the last years, methods from the field of Artificial Intelligence (AI) have been successfully applied in the field of finance. As more and more black-box approaches are being adopted, the demand for transparency from regulators, investors, developers, and managers has been increasing. This paper presents a contribution by proposing a novel method for pricing American-style options that falls within the field of eXplainable AI (XAI).[9] The neural network, often considered as a black-box algorithm, is in our approach applied solely as a function approximator that learns from a training set generated by a chosen (e.g., PDE-based) well-known pricing model. We show that by doing this, our pricer behaves in an interpretable, tractable, and trustworthy fashion. As such, our approach can be understood as a machine learning method that

[6] Deep learning calibration is studied also in Liu et al. [2019a] and Itkin [2019].

[7] On a similar note, Fecamp et al. [2019] apply ANNs to solve the pricing and hedging problem under market frictions such as transaction costs.

[8] Speed-accuracy trade-off in terms of derivative pricing, hedging, and fitting has also been confirmed by Spiegeleer et al. [2018]. Moreover, Glau et al. [2020] extend the tensorized Chebyshev interpolation of conditional expectations in the parameter and state space and show that, at comparable accuracy, the interpolation in American option prices provides a promising gain in computational efficiency, leading to a speed-accuracy trade-off.

[9] An extensive review of the literature around XAI is provided in Arrieta et al. [2020].

produces an accelerated and explainable option pricing algorithm while maintaining a high level of pricing accuracy. Model explainability is one of the crucial aspects that would ensure investors and regulators with understanding and trust to use the proposed machine learning pricing model in a market-making environment.

The remainder of the paper is organized as follows: Section 4.2 introduces the computational model and its underlying framework. In Section 4.3, the numerical results are presented and the speed-accuracy trade-off is analyzed. Finally, we conclude in Section 4.4.

4.2 Model

We propose a mixed-method where American-style option prices are generated using a standard pricing method, and the price data is subsequently utilized for training a neural network over the underlying parameter space. Since the option prices are constructed with a well-known pricing method, the neural network learns the pricing map by utilizing the generated training set that is by construction free of arbitrage. Essentially, this neural network provides the capability to learn the behavior of the American option price to changes within its parameter space. By generating our own training data and training the neural network offline (i.e., not during trading hours), the latency in the pricing of American options during trading hours can be reduced to near-instantaneous. Furthermore, the same neural network could be used for other purposes such as calculating implied volatilities or calculating the Greeks[10] of options at near-instantaneous speeds.

4.2.1 Deep Explainable Pricing Algorithm

We consider a typical parameter set prevalent in a derivative pricing framework $\Psi := (r, q, T-t, M, \psi_\sigma^{Model})$, where r denotes the interest rate, q the continuous dividend yield, $T-t$ the time-to-maturity, M the moneyness, and ψ_σ^{Model} the set of variables parameterizing a chosen volatility specification. A standard Black-Scholes-Merton setting represents a trivial case of constant volatility with $\psi_\sigma^{BS} := \sigma_0$, whereas under the Heston stochastic volatility model the parameterization is denoted by $\psi_\sigma^{Heston} := \{v_0, \sigma, \kappa, \theta, \rho\}$, where the parameter definitions are provided in the model description below.

By computing American option prices using, for example, a PDE method on a dense grid across these parameter sets, we generate a training set over which a neural network pricing functional is trained. As standard methods are used offline to generate the training data for the neural network, a major component of this paper is the accurate generation of the training data. Due to various universal approximation theorems (discussed below), it can be assumed that the neural network is able to approximate the option pricing dynamics to, at best, the precision of the underlying standard model-generated data. Thus, the error of the

[10] A set of derivatives of an option with respect to various parameters.

standard model data serves as a lower bound to the neural network error, illustrating the requirement for accurate standard model input data.

Definition 4.2.1 (Heston Stochastic Volatility Model)**.** Consider a probability space $\{\Omega, \mathcal{F}, \mathbb{Q}\}$ adapted to a filtration $\mathcal{F}_t$. Let S_t and v_t denote the asset price and instantaneous variance at time t. Then, according to the Heston model, as presented in Heston [1993], the asset price and volatility processes are, under the pricing measure $\mathbb{Q}$, defined by a system of stochastic differential equations (SDEs)

$$\begin{aligned} dS_t &= (r-q)S_t dt + \sqrt{v_t} S_t dW_t, \\ dv_t &= \kappa[\theta - v_t]dt + \sigma\sqrt{v_t} dZ_t, \\ dZ_t dW_t &= \rho dt, \end{aligned} \tag{4.2.1}$$

where r denotes the risk-free rate of return, q the expected continuous dividend yield, $\kappa > 0$ the speed of mean reversion, $\theta > 0$ the long term variance of the price process, $\sigma > 0$ the volatility of volatility, $-1 \leq \rho \leq 1$ the correlation coefficient between the geometric Brownian motion processes Z_t and W_t, and $S_0, v_0 > 0$ are the starting values of the stochastic processes S_t and v_t, respectively. In order that the instantaneous variance $v_t > 0$ is strictly positive, the parameters need to obey the Feller condition $2\kappa\theta > \sigma^2$.

The value of an American call option with maturity T and strike K, is, at time $t = 0$, given by

$$V(S_0, 0) = \sup_{\tau \in \mathcal{T}(T)} \mathbb{E}_{\mathbb{Q}}\left[e^{-r\tau}(S_\tau - K)^+\right], \tag{4.2.2}$$

where $\mathcal{T}(T)$ denotes the set of stopping times τ with values in $[0, T]$. Assuming the system of SDEs from Eq. (4.2.1), one can derive the solution to Eq. (4.2.2) in terms of a PDE.[11] The resulting Heston American call option PDE (expressed in terms of the option price) is given by

$$\frac{\partial V}{\partial t} + \frac{1}{2} v S^2 \frac{\partial^2 V}{\partial S^2} + \rho\sigma v S \frac{\partial^2 V}{\partial v \partial S} + \frac{1}{2}\sigma^2 v \frac{\partial^2 V}{\partial v^2} - rV + (r-q)S\frac{\partial V}{\partial S} + [\kappa(\theta - v) - \lambda(S, v, t)]\frac{\partial V}{\partial v} = 0, \tag{4.2.3}$$

in the region $\mathcal{D} = \{0 \leq t \leq T,\, 0 \leq S_t \leq b\,(v,t)\,,\, 0 \leq v < \infty\}$, subject to the boundary conditions

$$\begin{aligned} V\left(S_T, T\right) &= (S_T - K)^+, \\ V\left(b\,(v,t)\,, t\right) &= b\,(v,t) - K, \\ \lim_{S \to b(v,t)} \frac{\partial V}{\partial S} &= 1, \\ \lim_{S \to b(v,t)} \frac{\partial V}{\partial v} &= 0, \end{aligned}$$

[11] The corresponding self-financing replicating portfolio consists of positions in an underlying stock and two options.

where $b(v, t)$ represents the early exercise boundary at time t for the Heston volatility v, and $\lambda(S, v, t)$ denotes the corresponding market price of volatility risk. See Heston [1993] and AitSahlia et al. [2010] for a detailed proof and more information.

Algorithm 2 Neural network-based American option pricer under stochastic volatility

---

Input: Chosen parameter ranges for r, q, v_0, σ, κ, θ, ρ, M, and $T - t$.
Output: Current price of an American-style option per underlying applicable in a market-making environment.

(A) <u>Training data generation</u>: Evaluation of American option prices using a chosen standard pricing method across an entire parameter space Ψ.
(B) <u>Learning</u>: Offline training of a neural network over a parameter-price data set.
(C) <u>Pricing</u>: Application of a neural network pricer in a market-making environment.

In Algorithm 2, the proposed American option pricing methodology is summarized in a three-step process that generates a direct explainable neural network pricer. In the case when the training data is generated with a standard PDE method, Step (A) reduces to solving the Heston PDE from Eq. (4.2.3) using a finite difference method. The parameter ranges over which the learning takes place need to be chosen in this step. In our case, these are the Heston model parameters whereas the same algorithm works for any other parametric specification.[12] We detail the chosen parameter ranges (assuring that the Feller condition is satisfied) in Section 4.3.1. In Step (B), the offline training takes place. One needs to choose the structure of the neural network and the corresponding hyperparameters in this step of the algorithm. We detail the architecture of our chosen neural network in Section 4.2.2 and provide the analysis of the training process in Section 4.3.1. One can assume that the employed neural network in Step (B) inherits the underlying no-arbitrage pricing properties arising from the arbitrage-free construction of the training data set from Step (A) of the algorithm. The deep neural network is here utilized merely as a tool to learn the pricing map between the chosen standard (e.g., PDE-based) pricing method in Step (A) and the corresponding American option prices. Such an approach enhances the trust and transparency of our algorithm. Even though a neural network is applied, the algorithm itself is not a black box. As shown in the next section, our algorithm simply accelerates the computation of American option prices obtained by a well-understood pricing approach chosen in Step (A) of the algorithm. The pricing is performed in an interpretable fashion basing our approach in the area of XAI for finance. Finally, after the offline learning process from Step (B) is completed, the deep explainable (i.e., Heston-based) neural network pricer can be employed in a market-making environment in Step (C) of the algorithm. Given our aim for computational explainability, the proposed algorithm has a large potential of being understandable and trusted by humans, and as such accepted by regulators and investors.

---

[12] Note that the chosen parameter ranges largely depend on the available computational capabilities.

Since this neural network is targeted for market making, a key responsibility of any algorithm in a trading environment is its compliance with regulation such as the Markets in Financial Instruments Directive (MiFID) or Securities and Exchange Commission (SEC) guidelines. For example, MiFID specifically states that banks employing algorithms in trading should be able to "reconstruct efficiently and evaluate the strategies that algorithmic traders employ", see Directive-65/EU [2014]. This implies a requirement that black-box pricing algorithms such as neural networks need to have an explainable element. As such, we find it important to stress that our XAI-based pricing method indeed fulfills this requirement since we can show, see Section 4.3.2, that the error is bounded for any region of the underlying (i.e., Heston) parameter space.

Note that one can use any underlying data generating process as well as any American option pricing method for the training data generation since the calibration of the deep neural network works independently of the training price generation. As such, we provide a possible extension of the algorithm proposed above to pricing under the rough Bergomi volatility specification in Appendix 4.5.1.[13]

To improve upon the computationally slower standard American option pricing algorithms, we apply neural networks to serve as the functional evaluators–essentially substituting the underlying numerical pricing algorithm. This is primarily due to the ability of neural networks to model complex non-linear dynamics, differentiability, fast execution, and the ability to be evaluated using very basic low-level functionals. Furthermore, given the universal approximation properties of neural networks, it is theoretically possible to create a neural network (trained upon a chosen basic pricing method) that is able to price American-style options to any desired accuracy up to the pricing error imposed by the generated training data.

We aim to improve the process of American option pricing with respect to the speed-accuracy trade-off compared to the standard American option pricers. The motivation for improving the speed-accuracy trade-off lies in the ability of the neural network to learn the American option pricing map offline. Because of the simplicity of the neural network representation, their evaluation speed is unparalleled to the speed of other option pricing models. The boost in speed of the computation and the algorithm interpretability are the main arguments granting our approach a suitable candidate improving upon the currently used market-making pricing systems.

4.2.2 Deep Neural Network Structure

We exploit a feed-forward neural network with a rectified linear unit (ReLU) activation function merely as a tool to map from a given parameter space to the American option prices. The near-instantaneous evaluation of neural networks and their ability to approximate continuous functions provide for the above-mentioned

[13] Rough Bergomi model improves upon the classical stochastic volatility counterparts by reflecting the implied volatility surface dynamics more accurately (e.g., especially for short maturities). Under the rough Bergomi model, derivative pricing must be performed via Monte Carlo simulations. Thus, the computation time could potentially be greatly reduced using our proposed accelerated pricing method.

speed advantages over standard pricing methods. Neural networks can evaluate option prices in less than milliseconds, whereas PDE methods, binomial methods, or MC methods have significantly slower evaluation speeds. The deep neural network we employ is structured as a generic multi-layer perceptron as described in Caterini and Chang [2018].

Definition 4.2.2 (Multi-Layer Perceptron)**.** Given a data set with an input dimension of k and an output dimension of m, a multi-layer perceptron is defined as the neural network with an input layer of dimension $n_0 := k$ and an output layer of dimension $n_{L+1} := m$, where L specifies the number of hidden layers between the input and the output layer. Analogously, each layer between layers 0 and $L+1$ is of dimension n_i. For each layer i, we denote the weight and bias matrices by $W_i \in \mathbb{R}^{n_{i+1} \times n_i}$ and $b_i \in \mathbb{R}^{n_{i+1}}$, respectively. Per layer, the following function is applied to provide the input to the subsequent layer: $g_i(x_i; W_i, b_i) = \gamma_i(W_i \cdot x_i + b_i)$, where $x_i \in \mathbb{R}^{n_{i+1}}$ and γ_i denotes the corresponding activation function of the layer.

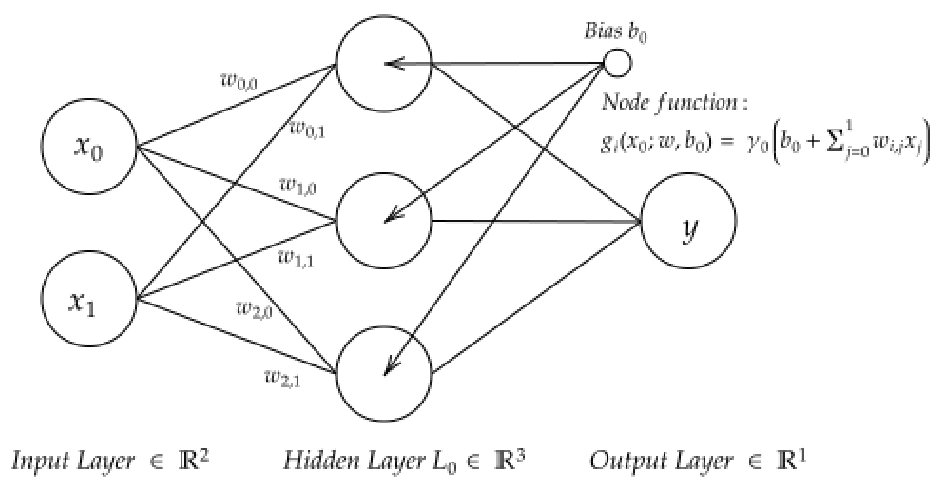

Figure 4.1: This figure depicts an example of a neural network with one hidden layer and two input variables.

Some of the foundational theorems of deep neural networks postulate that under a specific set of criteria, neural networks can approximate continuous functions to any degree of accuracy. These theorems are known as universal approximation theorems. Some of them focus on the depth (i.e., a number of hidden layers) of a neural network, and others on the width (i.e., the number of nodes per hidden layer). A widely known universal approximation theorem from Hornik et al. [1989] specifies that a neural network can approximate any continuous real-valued function on a compact set to an arbitrary accuracy $\epsilon > 0$ provided certain criteria for activation functions and depth hold. This is known as the depth-bounded universal approximation theorem. Extensions of this universal approximation theorem for ReLU activation functions are provided in Stinchcombe and White [1989]. Moreover, Lu et al. [2017] propose a specific universal approximation function that provides an upper bound to the width of a neural network with ReLU activation functions.

This result is relevant for this paper since we use the ReLU activation functions on the neural network's hidden layers. The corresponding theorem is stated below.

Theorem 4.2.1 (Width-Bounded Universal Approximation Theorem for ReLU Networks)**.** *For any Lebesgue-integrable function f:* $\mathbb{R}^n \to \mathbb{R}$ *and any* $\epsilon \geq 0$*, there exists a fully-connected ReLU network G with a width* $d_L \leq n+4$*, such that the neural network function* F_G *approximating* f *satisfies:*

$$\int_{R^n} |f(x) - F_G|\, dx < \epsilon,$$

where d_L *denotes the number of nodes in layer L.*

Our choice of the ReLU activating function relies on some of its special properties. ReLU activation functions are special since for $x < 0$, the activation function does not fire[14]—causing a sparser neural network. Furthermore, as x becomes large, the neurons do not become saturated[15] as in popular sigmoid activation functions. Lastly, ReLU functions are known to train faster compared to other activation functions, as shown by Krizhevsky et al. [2017]. One problem with ReLU activation functions noted by Caterini and Chang [2018] is that neurons can easily die, becoming what are known as "dead neurons" or neurons that always give a zero weight—a downside of allowing a sparse structure. This problem can be fixed by using a leaky ReLU activation function defined by

$$\gamma\left(x\right) \coloneqq \left\{ \begin{array}{ll} x, & x > 0 \\ 0.01x, & x \leq 0 \end{array} \right\},$$

which returns a small negative value for $x < 0$, allowing the neurons to recover.

Following standard literature, we utilize the ADAM optimizer[16] for the training with the Mean Absolute Error (MAE) cost function:

$$\mathrm{MAE} \coloneqq \frac{1}{n} \sum_{\Psi} \left| V_{NN}(\psi_k) - V_{TS}(\psi_k) \right|,$$

where n is the number of price samples, $\psi_k \in \Psi$ is a particular parameter specification, $V_{NN}(\psi_k)$ is the output of the neural network at each training iteration for a respective parameter combination ψ_k, and $V_{TS}(\psi_k)$ is the training data set for a respective parameter combination ψ_k. We decided to train with the MAE cost function because using an alternative cost function, for example, the widely-used mean squared error

[14] The so-called 'firing' of a node in a neural network is etymologically derived from the biological definition of neurons firing, i.e., sending a non-null impulse when an input is received.

[15] Saturated means that for a very large positive or negative number, the gradient of the sigmoid activation function is near zero.

[16] ADAM optimizer offers many advantages compared to the typical stochastic gradient descent, see Kingma and Ba [2014] for more details.

(MSE), would exaggerate the contribution induced by outliers by squaring their error in the cost function. This is relevant in our case since prices can vary widely within the parameter ranges (e.g., from 0.5 USD to 30 USD) and, consequently, the error varies as well.[17]

4.3 Numerical Experiments

4.3.1 Training Process

The training of the neural network consists of the determination of the hyperparameters such as the learning rate, exponential learning rate adjustment (ELRA), batch size, epochs, and depth and width of the hidden layers. The final neural network structure is set to four hidden layers with 40 nodes per layer. We follow Horvath et al. [2021] and Liu et al. [2019b] in that we decide to use a neural network with four hidden layers ensuring that the depth requirement for the universal approximation theorems is reached.

We generate the training sample (TS) parameter set Ψ_{TS} by choosing 12 equally spaced points for the parameter ranges $M \in [90, 110]$, $T - t \in [1, 255]$, $\kappa \in (0, 5]$, $\rho \in [-1, 0]$, and $\sigma \in (0, 1]$, and 12 logarithmically spaced points for $\theta \in (0, 0.41]$ and $v_0 \in (0, 0.41]$, since the long-term variance and the initial variance levels prevalently undertake lower (i.e., less than 50%) values.[18] This results in 12^7 price samples prior to the application of the Feller condition. Following the application of the Feller condition, we are left with a training set consisting of 24,634,368 parameter-price pairs.

To prepare the data for training, we standardize the input parameter data Ψ_{TS} as is a typically required practice in the training of neural networks since different input data have different absolute scales. We apply the same standardization (i.e., same mean and variance) to the parameter sets for all validation and test input data. Then, by applying the ADAM optimizer on the MAE cost function with the leaky ReLU activation function as described above, we train the neural network using an exponential learning rate adjustment. The training stops when the validation error does not improve for another 1000 iterations after reaching a running minimum. Naturally, during training, the difference between the validation and the training error is closely observed in order to prevent over-fitting to the input data.

Due to instabilities in the ADAM optimizer, one typically applies an exponential learning rate adjustment that exponentially decreases the learning rate during the training session. This allows the optimizer to initially explore the cost function surface for the minima, and subsequently, by slowly reducing the learning rate, the likelihood that the optimizer jumps out of the minimum is reduced. In Fig. 4.2, the effect of this adjustment for a sample training session is illustrated.

---

[17] We also performed training with a Vega-Weighted Absolute Error (VWAE) cost function as presented in Hamida and Cont [2005]. Thereby, the vega weight "converts" the errors in price into errors in implied volatility. The results, however, did not improve compared to the use of the conventional MAE cost function.

[18] Our training set parameter ranges are in line with the ones studied in the existing literature, see Spiegeleer et al. [2018].

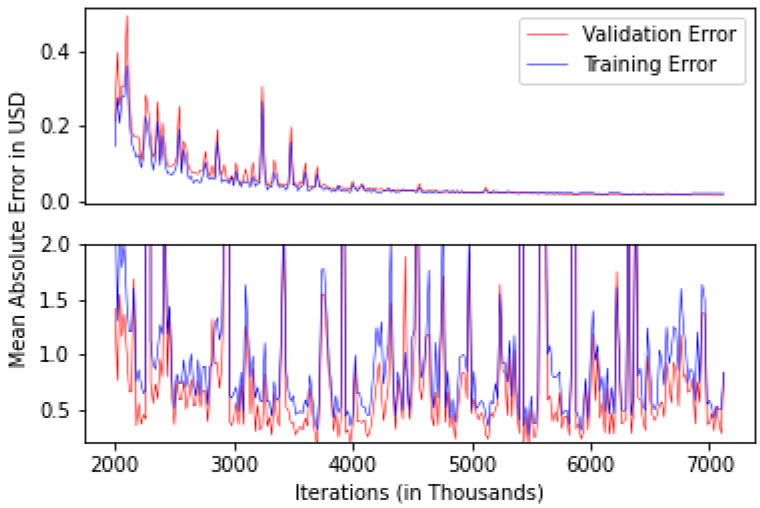

Figure 4.2: Training (in blue) and validation error (in red) per ADAM iteration with and without the exponential learning rate adjustment. The same pattern continues also after the 7000th iteration.

After 23,000 iterations over the entire training set, the stopping rule was satisfied and the training was stopped. The applied neural network structure is presented in Table 4.1.

Neural Network Structure:	
n	24,634,368
Batch Size	1,231,719
Initial Learning Rate	0.5
ELRA	0.997
Hidden Layers (Depth)	4
Nodes per Layer (Width)	40
Epoch Size	23,000
Training Time	41 hours

Table 4.1: This table presents the chosen neural network hyperparameters.

4.3.2 Pricing Results

Having trained the neural network offline, we test its out-of-sample performance in this section. We generate the out-of-sample (OOS) parameter set Ψ_{OOS} independent of the training sample parameter set Ψ_{TS} so that the error bias is avoided. Using Ψ_{OOS}, we evaluate the neural network-based prices V_{NN} and the corresponding near-exact (i.e., PDE-based) V_{PDE} prices. By “exact” we imply that a highly fine-meshed finite difference method was applied to construct the option prices. In Table 4.2, the neural network prices are compared to the near-exact PDE prices.

Heston Stochastic Volatility Case:			
OOS n	14336		
NN OOS MAE	0.02 USD	NN OOS MAPE	0.26%
NN OOS SD of Error	0.014 USD	Evaluation Time for OOS n	0.000026 sec.
NN OOS Min APE	0.0000236%	NN OOS Max APE	3.26%

Table 4.2: Model Summary under the Heston Stochastic Volatility Model.

Now that we have an overview of the average performance of the neural network pricer, we take a closer look at how it performs for each area of the parameter space Ψ. Consider a regular pricing environment with parameter values of $v_0 = 5\%$, $\theta = 15\%$, $\sigma = 0.3$, $\kappa = 0.8$, $\rho = -50\%$, $T - t = 165$ (i.e., in days), and $M = 100\%$. In Fig. 4.3, we vary one parameter at a time over its respective parameter space and plot the resulting neural network price in comparison to the near-exact price as computed by the PDE method. Such an analysis demonstrates the explainability nature of our pricing algorithm.

Notice that the neural network is able to learn the pricing function under the Heston stochastic volatility specification. Moreover, the well-studied Heston parametric structure is inherited by the deep explainable pricer with overall low relative pricing errors.[19] However, the neural network under-performs in several specific areas of the parameter space. Notice that, for example, $\rho > -0.4$, $M < 95\%$, $T - t < 20$, $v_0 < 0.05$, and σ all exhibit above-average errors. For some of these parameters (e.g., all besides the correlation and the higher vol-vol), we hypothesize that the areas of these variables where the error is particularly high would induce a lower option price. The reason for this might be that the cost function applied during the neural network training focused on reducing the high absolute errors rather than the high percentage errors. The MAE-based cost function was chosen to avoid the opposite version of this problem – high absolute errors for high price values. On the other hand, this could occur if the chosen cost function would have been based on the mean squared error (MSE). Further analysis into alternative cost functions to MSE and MAE is required (i.e., will follow in the subsequent version of the paper). With regards to the other two variables, ρ and σ, we conclude that further analysis is necessary as to why these variables exhibit clear patterns in their errors.

Figure 4.4 shows the implied volatility surface as computed by the deep explainable pricer and its PDE-based benchmark. One can observe that the deep explainable pricer reproduces the shape of the implied volatility surface by training from the data generated by the PDE-based pricer. The right subplot presents the error calculated as the value of the implied volatility as computed by the deep explainable pricer minus the value computed by the PDE-based pricer. Note that the errors are generally very low. The only area with larger deviations are the points close to maturity.

[19] Being able to provide such parametric interpretations positions our model in the field of XAI.

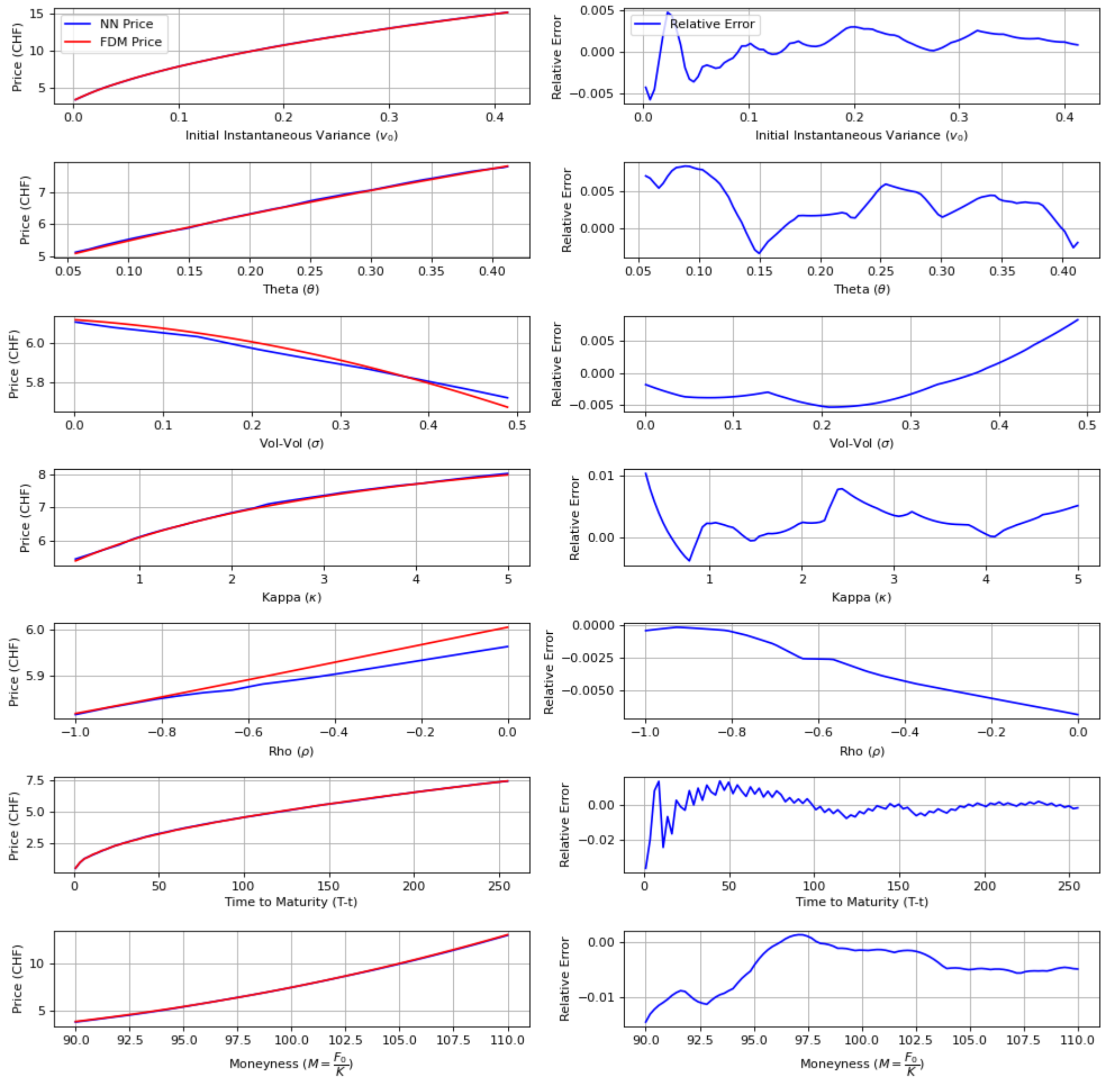

Figure 4.3: The left part of the figure depicts a comparison between the prices computed by the neural network (NN) under the Heston stochastic volatility—in blue—and the training sample computed by the finite difference method (FDM)—in red—in dependence of different model parameters. The right part of the figure equivalently shows the relative error of the NN approach in comparison to the FDM.

4.3.3 Computational Speed Comparison

The main incentive for applying neural networks in pricing American options are the benefits in terms of pricing speed. Next, we perform an analysis comparing the speed of evaluation of the proposed neural network pricer (over a set of 14,366 input parameter sets) with the speeds of standard methods such as the finite difference method, Monte Carlo simulation of Longstaff and Schwartz [2001], and binomial method of Cox et al. [1979] in Table 4.3. The computations were performed on an Intel Xeon Gold 6126 2.60GHz CPU.

The PDE method is implemented as proposed by Ikonen and Toivanen [2007] for an option of $T = 0.25$. Naturally, the methods' speed differs given different maturities, step sizes, iterations, and a number of paths. Nevertheless, no other method is able to evaluate an American option price nearly as fast as the

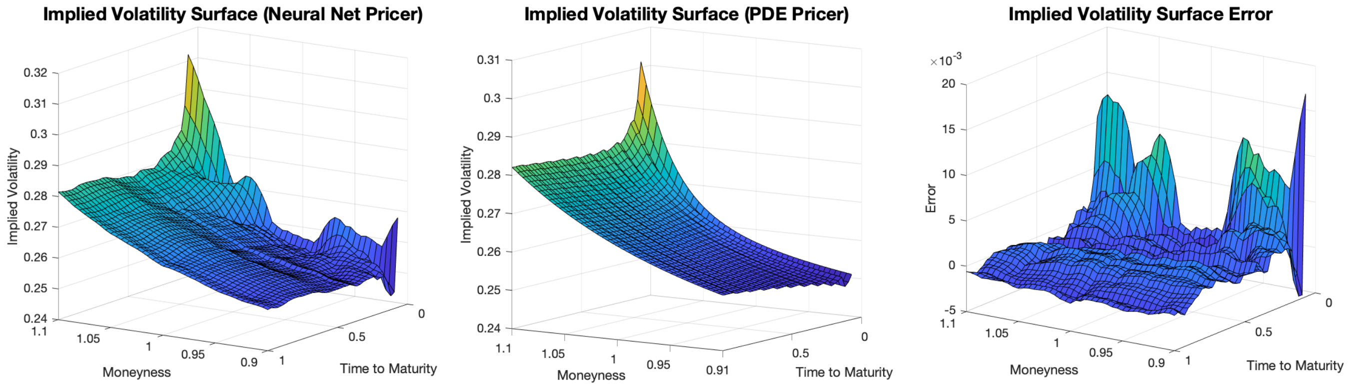

Figure 4.4: The left subplot depicts the implied volatility surface as computed by the deep explainable pricer. The middle subplot shows the implied volatility surface as computed by the benchmark PDE-based pricer. The right subplot presents the error in the implied volatility surface of the neural network pricer computed as the difference between the deep explainable and PDE-based implied volatility surfaces.

Pricing Method	Average Evaluation Speed (secs)	Compared to Neural Network Speed
Neural Network Method	0.000026	—
Finite Difference PDE Method	0.61	23,461 times slower
Binomial Method (Constant Volatility)	7.3	280,769 times slower
Longstaff-Schwartz Method	139.2	53,538,461 times slower

Table 4.3: Evaluation speed of the neural network pricer in comparison to other pricing methods for Heston stochastic volatility.

neural network pricer.

To illustrate the speed-accuracy trade-off, we choose the FDM method (i.e., the fastest competitor of the NN) and reduce the fineness of its grid such that, on average, the mean absolute error of the neural network is matched with the FDM-based error. By comparing the two methodologies at the same error level, we analyze the practicality of both methodologies in a market-making environment where both speed and accuracy are of utmost importance. Table 4.4 shows that at the same level of accuracy, the neural network pricer exhibits a considerable speed improvement over the FDM-based pricing method. In particular, the FDM method is, at the same level of accuracy, 58 times slower compared to the deep explainable American option pricer proposed in this paper.

Heston Stochastic Volatility Case:		
Model:	Error	Average Speed(secs)
Neural Network	26bps	0.000026
FDM (26x26x11)	26bps	0.0015

Table 4.4: Speed-accuracy trade-off of the neural network pricer in comparison to the FDM pricing method for Heston stochastic volatility.

Figure 4.5 depicts the overall speed-accuracy spectrum for constant volatility (CV) and Heston stochastic

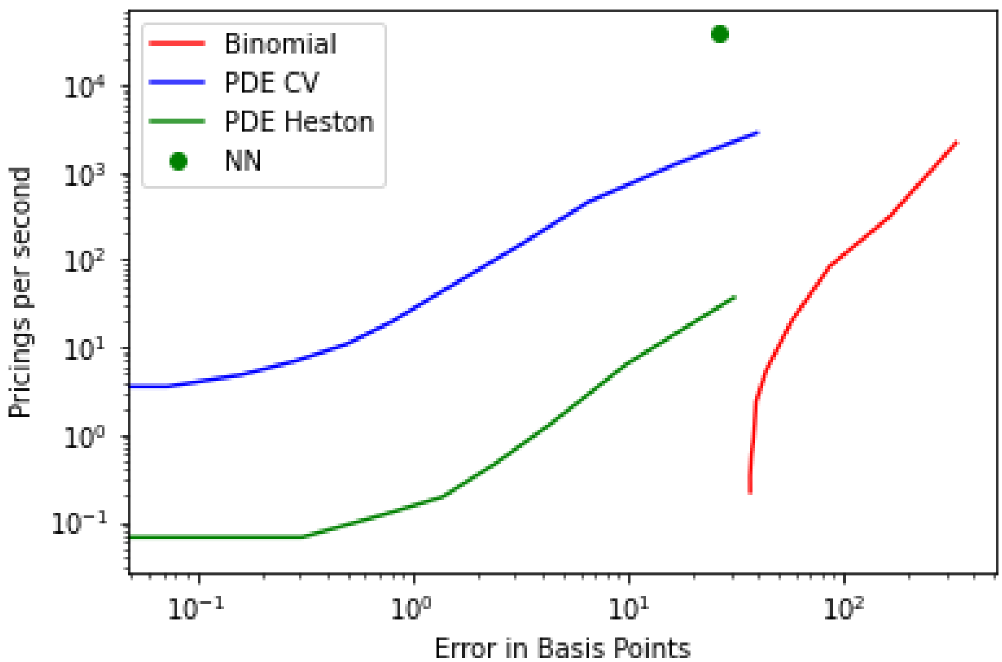

Figure 4.5: Speed-accuracy trade-off for various American option pricing methods in both constant volatility (CV) and Heston stochastic volatility environments compared to our neural network (NN) pricer.

volatility specifications compared to the deep explainable pricer. Speed is measured in terms of the number of pricings per second and accuracy is given in terms of relative error expressed in basis points against a "true" value defined as the PDE method (for CV or Heston) with a very fine mesh grid. One can observe the outperformance of the neural network with respect to the alternative pricing methods. The deep explainable pricer not only significantly outperforms the PDE Heston benchmark, but also outperforms PDE and binomial methods under constant volatility. Moreover, with an enhanced training set and longer training times, further improvements in error could be achieved–shifting the neural network point far to the left since the evaluation speed of a neural network is approximately constant. These results show that the proposed deep explainable pricer achieves a reasonable accuracy of computation while greatly reducing the speed of the computation. As such, it serves as a speedy alternative to more classical American option pricing methods currently used in a market-making environment.

4.4 Conclusions

Market-making in the equity derivatives markets requires high-precision and fast pricing algorithms. If the prices are not adjusted accurately and instantaneously, arbitrage opportunities occur. This paper presents a novel American-style option pricing method based on a deep neural network that accelerates the pricing while maintaining sufficient accuracy for market-making applicability.

The proposed model is based on generating training data by using a standard American option pricing method. We utilize a PDE-based method and price under the Heston stochastic volatility model in order to construct the training data. The obtained training data is subsequently utilized to train a deep neural network over the underlying parameter space offline. Consequently, the underlying (i.e., Heston) parameter interpretability is inherited by the neural network pricer. Such interpretability improves the trust and transparency of the proposed deep explainable pricer and associates our algorithm with the area of XAI for finance. Moreover, our pricer not only significantly increases the speed of American-style option pricing but is also able to compute implied volatilities and option Greeks at near-instantaneous speeds.

The simulation analysis shows that our approach provides a substantial improvement with respect to the speed-accuracy trade-off over standard American option pricers. The improvement primarily arises due to the ability of neural networks to learn the American option pricing map offline. Moreover, given the simplicity of neural network representations, their evaluation speed is considerably higher compared to the speed of standard option pricing models. This speed increase is naturally more pronounced for Monte Carlo methods than for PDE methods due to the complexity of simulation methods. Also, we show that this increase in speed occurs without a significant loss in accuracy when methods are compared at the same precision level. Therefore, our model exhibits the speed-accuracy and explainability features required in a challenging market-making environment.

Furthermore, potentially increased computing resources and the universal approximation properties of neural networks imply that the level of accuracy presented in this paper could be improved by i) increasing the density of the PDE finite difference method grid (or equivalently, increasing the number of paths per MC simulation), ii) experimenting with alternative neural network designs, iii) further optimizing the neural network hyperparameters, or iv) experimenting with other cost functions in order to reduce the training bias.

While these steps could potentially improve the accuracy of the neural network pricer, other interesting areas of future research could expand the framework in the sense of its modeling capabilities. Particularly, since the speed improvements are greatest when compared to Monte Carlo methods, any pricing model that relies purely on Monte Carlo methods would be a perfect candidate for neural network acceleration. One could, for example, expand the model to utilize large-deviation MC techniques for out-of-the-money and short maturity options. Moreover, one could also consider alternative dynamics for the underlying asset and volatility processes, such as the rough volatility, possibly including cash dividends and more complex payoffs that are priceable only by the Monte Carlo techniques.

4.5 Appendix

4.5.1 Deep Explainable Pricing Algorithm - Rough Volatility Extension

Definition 4.5.1 (Rough Bergomi Volatility Model)**.** Let S_t and v_t denote the asset price and instantaneous variance at time t. The rough Bergomi model, introduced by Gatheral et al. [2018], is, under a pricing measure, defined by a system of stochastic differential equations

$$\frac{dS_t}{S_t} = (r-q)dt + \sqrt{v_t}(\rho\, dW_t + \sqrt{1-\rho^2}\, dW_t^{\perp}),$$
$$v_t = \xi_0(t)\mathcal{E}(\eta W^{\alpha})_t,$$

where r is the interest rate, q the is continuous dividend yield, W_t and $W_t^{\perp}$ are independent Brownian motions, $-1 \leq \rho \leq 1$, and the instantaneous variance v_t is a product of a forward variance curve $t \mapsto \xi_0(t)$, known at time 0, and the Wick exponential $\mathcal{E}(\eta W^{\alpha})_t = \exp(\eta W_t^{\alpha} - \frac{1}{2}\mathrm{Var}(\eta W_t^{\alpha}))$ for $\eta > 0$ and a Gaussian random variable W_t^{α}. The random variable W_t^{α} is given by the Gaussian Riemann–Liouville process

$$W_t^{\alpha} = \sqrt{2\alpha+1}\int_0^t (t-s)^{\alpha} dW_s, \quad t \geq 0,$$

where the parameter $-\frac{1}{2} < \alpha < 0$ controls the roughness of paths. The corresponding paths have a Hölder regularity of $\alpha + \frac{1}{2}$ and locally behave like the paths of a fractional Brownian motion with $H = \alpha + \frac{1}{2}$. The three time-homogeneous parameters, η, ρ, and α, can be interpreted as the smile, the skew, and the near-maturity explosion of the smile and the skew, expressed in terms of the implied volatility surface.

In the case of rough Bergomi volatility, a Monte Carlo simulation is required to generate the initial training data needed for the neural network calibration. This means that the gains in speed due to training a neural network price evaluator are potentially greater than in the case of the Heston stochastic volatility model.

Algorithm 3 Neural network-based American option pricer under rough volatility

Input: ξ_0, η, α, ρ, M, $T-t$.
Output: Current price of an American-style option per underlying applicable in a market-making environment.

(A) <u>Simulation</u>: Simulation of the volatility and price processes.
(B) <u>Optimal exercise</u>: Evaluation of optimal exercise strategies for each simulated path.
(C) <u>Training data generation</u>: Evaluation of American option prices across an entire parameter space Ψ.
(D) <u>Learning</u>: Offline training of a neural network over a parameter-price data set.
(E) <u>Pricing</u>: Application of a neural network pricer in a market-making environment.

In Algorithm 3, a neural network pricer for the case of Monte Carlo generated training data set is presented. In Step (A), the Euler-Maruyama method, see Maruyama [1955], is applied to simulate paths of the volatility and price processes for a fine mesh across the parameter set Ψ. On each path, optimal exercise opportunities are evaluated, and a corresponding payoff for each path is determined.

For the evaluation of optimal exercise opportunities in Step (B), the methodology presented in Andersen et al. [2016] can be utilized. Their method of approximating the early exercise boundary allows for the computation of an early exercise boundary value for every point along the MC paths. The optimal exercise strategy is calculated by comparing the current simulated spot price level with the Andersen boundary value.[20] By discounting and averaging over the payoffs of the simulated paths for each parameter combination, a Monte Carlo estimator of the American option price is obtained. Finally, by repeating this procedure for all combinations of parameters across Ψ, one arrives at a set of American option prices across the entire parameter space – Step (C). In Step (D), the offline training occurs, and the deep neural network rough Bergomi-based pricer can potentially be employed in a market-making environment in Step (E).

[20] As mentioned in Bayer et al. [2020], under rough volatility models, the optimal exercise strategy should depend on past observations of the spot price and not only the current price. However, the authors concluded that there is no significant difference in the computed price due to including past spot prices in the exercise decision. Based on this reasoning, one could assume that the current spot price contains sufficient information to decide upon the American option exercise. However, further investigation is necessary.

Chapter 5

References

F. Ackermann, W. Pohl, and K. Schmedders. Optimal and naive diversification in currency markets. *Management Science*, 63(10):3347–3360, 2017.

F. AitSahlia, M. Goswami, and S. Guha. American option pricing under stochastic volatility: an efficient numerical approach. *Computational Management Science*, 7(6):171–187, 2010.

C. Alexander and A. Chibumba. Multivariate orthogonal factor GARCH, 1996. Working paper.

S. Álvarez-Díez, E. Alfaro-Cid, and M. O. Fernández-Blanco. Hedging foreign exchange rate risk: Multi-currency diversification. *European Journal of Management and Business Economics*, 25(1):2–7, 2016.

U. Anders, O. Korn, and C. Schmitt. Improving the pricing of options: a neural network approach. *Journal of Forecasting*, 17(5-6):369–388, 1998.

L. Andersen, M. Lake, and D. Offengenden. High performance American option pricing. *Journal of Computational Finance*, 20(1):39–87, 2016.

D. Anderson and U. Ulrych. Accelerated American Option Pricing with Deep Neural Networks. *Swiss Finance Institute Research Paper No. 22-03*, 2022.

E. W. Anderson, L. P. Hansen, and T. J. Sargent. Risk and robustness in general equilibrium. *Preprint University of Chicago*, 1998.

E. W. Anderson, L. P. Hansen, and T. J. Sargent. A quartet of semigroups for model specification, robustness, prices of risk, and model detection. *Journal of the European Economic Association*, 1(1):68–123, 2003.

R. W. Anderson and J.-P. Danthine. Cross hedging. *Journal of Political Economy*, 89(6):1182–1196, 1981.

P. C. Andreou, C. Charalambous, and S. H. Martzoukos. Generalized parameter functions for option pricing. *Journal of Banking & Finance*, 34(3):633–646, 2010.

A. Ang and G. Bekaert. International asset allocation with regime shifts. *The review of financial studies*, 15(4):1137–1187, 2002.

A. B. Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, R. Chatila, and F. Herrera. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58:82–115, 2020.

D. Avramov. Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64(3): 423–458, 2002.

D. Avramov. Stock return predictability and asset pricing models. *The Review of Financial Studies*, 17(3): 699–738, 2004.

D. Avramov and T. Chordia. Asset pricing models and financial market anomalies. *The Review of Financial Studies*, 19(3):1001–1040, 2006.

D. Avramov and G. Zhou. Bayesian portfolio analysis. *Annu. Rev. Financ. Econ.*, 2(1):25–47, 2010.

N. Barberis. Investing for the long run when returns are predictable. *The Journal of Finance*, 55(1):225–264, 2000.

G. Barone-Adesi, K. Giannopoulos, and L. Vosper. VaR without correlations for portfolios of derivative securities. *The Journal of Futures Markets*, 19(5):583–602, 1999.

C. Bayer, P. Friz, and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

C. Bayer, R. Tempone, and S. Wolfers. Pricing American options by exercise rate optimization. *Quantitative Finance*, 20(11):1749–1760, 2020.

S. Becker, P. Cheridito, and A. Jentzen. Pricing and hedging American-style options with deep learning. *Journal of Risk and Financial Management*, 13(158), 2020.

J. Bender, R. Kouzmenko, and Z. Nagy. Why currency returns and currency hedging matters. *MSCI Research Bulletin*, 2012.

L. Berger and L. Eeckhoudt. Risk, ambiguity, and the value of diversification. *Management Science*, 67(3): 1639–1647, 2021.

M. Bianchi and J.-M. Tallon. Ambiguity preferences and portfolio choices: Evidence from the field. *Management Science*, 65(4):1486–1501, 2019.

F. Black. Universal hedging: Optimizing currency risk and reward in international equity portfolios. *Financial Analysts Journal*, 45(4):16–22, 1989.

F. Black. Equilibrium exchange rate hedging. *The Journal of Finance*, 45(3):899–907, 1990.

F. Black and R. Litterman. Global portfolio optimization. *Financial analysts journal*, 48(5):28–43, 1992.

F. Black, M. Scholes, and R. Merton. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

T. Bollerslev. Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. *The Review of Economics and Statistics*, 72(3):498–505, 1990.

J. Boudoukh, M. Richardson, A. Thapar, and F. Wang. Optimal currency hedging for international equity portfolios. *Financial Analysts Journal*, 75(4):65–83, 2019.

P. Boyle. Option valuation using a three-jump process. *International Options Journal*, 3:7–12, 1986.

P. Boyle, M. Broadie, and P. Glasserman. Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control*, 21:1267–1321, 1997.

M. Brennan and E. Schwartz. The valuation of American put options. *Journal of Finance*, 32(2):449–462, 1977.

M. Britten-Jones. The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance*, 54(2):655–671, 1999.

J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris. Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences*, 106(30):12267–12272, 2009.

C. Brown, J. Dark, and W. Zhang. Dynamic currency hedging for international stock portfolios. *Review of futures markets*, 20:419–455, 2012.

V. Bruno and H. S. Shin. Currency depreciation and emerging market corporate distress. *Management Science*, 66(5):1935–1961, 2020.

H. Buehler, L. Gonon, J. Teichmann, and B. Wood. Deep hedging. *Quantitative Finance*, 19(8):1271–1291, 2019.

J. D. Burger, F. E. Warnock, and V. C. Warnock. Currency matters: Analyzing international bond portfolios. *Journal of International Economics*, 114:376–388, 2018.

R. Burkhardt and U. Ulrych. Sparse and Stable International Portfolio Optimization and Currency Risk Management. *Swiss Finance Institute Research Paper No. 22-07*, 2022.

J. Y. Campbell, L. M. Viceira, and J. S. White. Foreign currency for long-term investors. *The Economic Journal*, 113(486):C1–C25, 2003.

J. Y. Campbell, K. Serfaty-De Medeiros, and L. M. Viceira. Global currency hedging. *The Journal of Finance*, 65(1):87–121, 2010.

M. Caporin, J.-A. Jimenez-Martin, and L. González Serrano. Currency hedging strategies in strategic benchmarks and the global and Euro sovereign financial crises. *Journal of International Financial Markets, Institutions and Money*, 31:159–177, 2014.

A. Caterini and D. E. Chang. *Deep Neural Networks in a Mathematical Framework*. Springer, 2018.

K. H. Chang. Currency hedging: A free lunch? *MSCI Barra Research*, 2009.

N. Chatsanga and A. J. Parkes. International Portfolio Optimisation with Integrated Currency Overlay Costs and Constraints. *Expert Systems with Applications*, 83:333–349, 2017.

F. Chen and C. Sutcliffe. Pricing and hedging short sterling options using neural networks. *Intelligent Systems in Accounting, Finance and Management*, 19(2):128–149, 2012.

Y. Chen and J. W. L. Wan. Deep neural network framework based on backward stochastic differential equations for pricing and hedging American options in high dimensions. *Quantitative Finance*, 21(1): 45–67, 2021.

Y.-W. Cheung, M. D. Chinn, A. G. Pascual, and Y. Zhang. Exchange rate prediction redux: New models, new data, new currencies. *Journal of International Money and Finance*, 95:332–362, 2019.

J.-B. Cho, H.-G. Min, and J. A. McDonald. Volatility and dynamic currency hedging. *Journal of International Financial Markets, Institutions and Money*, 64:101163, 2020.

J.-W. Cho, J. H. Choi, T. Kim, and W. Kim. Flight-to-quality and correlation between currency and stock returns. *Journal of Banking & Finance*, 62:191–212, 2016.

B. Chowdhry. Corporate hedging of exchange risk when foreign currency cash flow is uncertain. *Management Science*, 41(6):1083–1090, 1995.

C. Conflitti, C. De Mol, and D. Giannone. Optimal combination of survey forecasts. *International Journal of Forecasting*, 31(4):1096–1103, 2015.

J. C. Cox, S. A. Ross, and M. Rubinstein. Option pricing: A simplified approach. *Journal of Financial Economics*, 7(3):229–263, 1979.

J. Crank and P. Nicolson. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Mathematical Proceedings of the Cambridge Philosophical Society*, 43(1):50–67, 1947.

R. Culkin and S. R. Das. Machine learning in finance: the case of deep learning for option pricing. *Journal of Investment Management*, 15(4):92–100, 2017.

I. Daubechies, M. Defrise, and C. De Mol. An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11):1413–1457, 2004.

J. de Boer, K. J. Bövers, and S. Meyer. Business cycle variations in exchange rate correlations: Revisiting global currency hedging. *Finance Research Letters*, 33:101195, 2020.

S. de Boer. Smart currency hedging for smart beta global equities. *The Journal of Investing*, 25(4):64–78, 2016.

F. De Roon, E. Eiling, B. Gerard, and P. Hillion. Speculative profits or hedging benefits? Currency investing in global portfolios. *Working Paper*, 2011.

F. A. De Roon, T. E. Nijman, and B. J. Werker. Currency hedging for international stock portfolios: The usefulness of mean–variance analysis. *Journal of Banking & Finance*, 27(2):327–349, 2003.

V. DeMiguel, L. Garlappi, F. J. Nogales, and R. Uppal. A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms. *Management Science*, 55(5):798–812, 2009a.

V. DeMiguel, L. Garlappi, and R. Uppal. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *The Review of Financial Studies*, 22(5):1915–1953, 2009b.

Directive-65/EU. Directive 2014/65/eu of the european parliament and of the council of 15 may 2014 on markets in financial instruments and amending directive 2002/92/ec and directive 2011/61/eu. *Official Journal of the European Union*, 2014.

W. Du, A. Tepper, and A. Verdelhan. Deviations from covered interest rate parity. *The Journal of Finance*, 73(3):915–957, 2018.

C. Dugas, Y. Bengio, F. Bélisle, C. Nadeau, and R. Garcia. Incorporating functional knowledge in neural networks. *Journal of Machine Learning Research*, 10(42):1239–1262, 2009.

M. R. Eaker and D. M. Grant. Cross-hedging foreign currency risk. *Journal of International Money and Finance*, 6(1):85–105, 1987.

B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least Angle Regression. *The Annals of Statistics*, 32 (2):407–499, 2004.

D. Ellsberg. Risk, ambiguity, and the savage axioms. *The quarterly journal of economics*, pages 643–669, 1961.

C. S. Eun and B. G. Resnick. Exchange rate uncertainty, forward contracts, and international portfolio selection. *The Journal of Finance*, 43(1):197–215, 1988.

F. J. Fabozzi, P. N. Kolm, D. A. Pachamanova, and S. M. Focardi. *Robust Portfolio Optimization and Management*. John Wiley & Sons, 2007.

S. Fecamp, J. Mikael, and X. Warin. Risk management with machine-learning-based algorithms. *Working paper*, 2019.

F. Filipozzi and K. Harkmann. Optimal currency hedge and the carry trade. *Review of Accounting and Finance*, 19(3):411–427, 2020.

R. Finlay. *The Variance Gamma (VG) Model with Long Range Dependence*. PhD thesis, School of Mathematics and Statistics, University of Sydney, 2009.

C. Francq and J.-M. Zakoian. *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons, 2019.

J. Friedman, T. Hastie, and R. Tibshirani. *The Elements of Statistical Learning*. Springer Series in Statistics New York, 2001.

K. A. Froot. Currency hedging over long horizons. Technical report, National Bureau of Economic Research, 1993.

L. Gagnon, G. J. Lypny, and T. H. McCurdy. Hedging foreign currency portfolios. *Journal of Empirical Finance*, 5(3):197–220, 1998.

R. Garcia and R. Gençay. Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics*, 94(1-2):93–115, 2000.

G. W. Gardner and T. Wuilloud. Currency risk in international portfolios: How satisfying is optimal hedging. *Journal of Portfolio Management*, 21(3):59, 1995.

L. Garlappi, R. Uppal, and T. Wang. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies*, 20(1):41–81, 2007.

R. M. Gaspar, S. D. Lopes, and B. Sequeira. Neural network pricing of American put options. *Risks*, 8(3), 2020.

G. L. Gastineau. The currency hedging decision: A search for synthesis in asset allocation. *Financial Analysts Journal*, 51(3):8–17, 1995.

J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.

I. Gilboa and D. Schmeidler. Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

K. Glau, D. Kressner, and F. Statti. Low-rank tensor approximation for Chebyshev interpolation in parametric option pricing. *SIAM Journal on Financial Mathematics*, 11(3):897–927, 2020.

J. Glen and P. Jorion. Currency hedging for international portfolios. *The Journal of Finance*, 48(5):1865–1886, 1993.

N. Gradojevic, R. Gençay, and D. Kukolj. Option pricing with modular neural networks. *IEEE Transactions on Neural Networks*, 20(4):626–637, 2009.

M. Guidolin and F. Rinaldi. Ambiguity in asset pricing and portfolio choice: A review of the literature. *Theory and Decision*, 74(2):183–217, 2013.

H. Guo and L. Ryan. Currency-hedging optimization for multi-asset portfolios. *The Journal of Portfolio Management*, 44(2):100–113, 2018.

T. Haefliger, D. Wydler, and U. Waelchli. Hedging currency risk: Does it have to be so complicated? *Available at SSRN 363080*, 2002.

S. B. Hamida and R. Cont. Recovering volatility from option prices by evolutionary optimization. *Journal of Computational Finance*, 8(4):43–76, 2005.

Y. Han. Asset allocation with a high dimensional latent factor stochastic volatility model. *The Review of Financial Studies*, 19(1):237–271, 2006.

L. Hansen and T. J. Sargent. Robust control and model uncertainty. *American Economic Review*, 91(2): 60–66, 2001.

R. D. F. Harris and J. Shen. Hedging and value at risk. *The Journal of Futures Markets*, 26(4):369–390, 2006.

T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.

K. Hornik, M. Stinchcombe, and H. White. Multilayer feed-forward networks are universal approximators. *Neural Networks*, 1989.

B. Horvath, A. Muguruza, and M. Tomas. Deep learning volatility: a deep neural network perspective on pricing and calibration in (rough) volatility models. *Quantitative Finance*, 21(1):11–27, 2021.

C.-C. Hsu, C.-P. Tseng, and Y.-H. Wang. Dynamic hedging with futures: A copula-based GARCH model. *Journal of Futures Markets*, 28(11):1095–1116, 2008.

X. Huang and X. Wang. International portfolio optimization based on uncertainty theory. *A Journal of Mathematical Programming and Operations Research*, 70(2):225–249, 2021.

J. M. Hutchinson, A. W. Lo, and T. Poggio. A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance*, 49(3):851–889, 1994.

S. Ikonen and J. Toivanen. Operator splitting methods for American option pricing. *Applied Mathematics Letters*, 17(7):104–126, 2004.

S. Ikonen and J. Toivanen. Pricing American options using lu decomposition. *Numerical Methods for Partial Differential Equations*, 24(51):104–126, 2007.

A. Itkin. Deep learning calibration of option pricing models: some pitfalls and solutions. *Working paper*, 2019.

R. Jagannathan and T. Ma. Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *The Journal of Finance*, 58(4):1651–1683, 2003.

M. R. Jahan-Parvar and F. Zikes. When do low-frequency measures really measure transaction costs? *Finance and Economics Discussion Series 2019-051*, 2019.

C. M. Jones. A Century of Stock Market Liquidity and Trading Costs. *Available at SSRN 313681*, 2002.

P. Jorion. Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, pages 279–292, 1986.

P. Jorion. Asset allocation with hedged and unhedged foreign stocks and bonds. *Journal of Portfolio Management*, 15(4):49, 1989.

P. Jorion. Mean/variance analysis of currency overlays. *Financial Analysts Journal*, 50(3):48–56, 1994.

R. Kan and G. Zhou. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, pages 621–656, 2007.

S. Kandel and R. F. Stambaugh. On the predictability of stock returns: an asset-allocation perspective. *The Journal of Finance*, 51(2):385–424, 1996.

D. L. Kelly. Valuing and hedging American put options using neural networks. *Unpublished manuscript*, 1994.

S. F. Kim and D. M. Chance. An empirical analysis of corporate currency risk management policies and practices. *Pacific-Basin Finance Journal*, 47:109–128, 2018.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *Third International Conference For Learning Representations, San Diego*, 2014.

R. W. Klein and V. S. Bawa. The effect of estimation risk on optimal portfolio choice. *Journal of financial economics*, 3(3):215–231, 1976.

P. Klibanoff, M. Marinacci, and S. Mukerji. A smooth model of decision making under ambiguity. *Econometrica*, 73(6):1849–1892, 2005.

F. H. Knight. *Risk, uncertainty and profit.* Houghton Mifflin, 1921.

M. Kohler, A. Krżyzak, and N. Todorovic. Pricing of high-dimensional American options by neural networks. *Mathematical Finance*, 20:383–410, 2010.

H. Konno and J. Li. Applications of the Integrated Approach to International Portfolio Optimization. *Asia-Pacific Financial Markets*, 7(2):121–144, 2000.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 2017.

K. F. Kroner and J. Sultan. Time-varying distributions and dynamic hedging with foreign currency futures. *Journal of Financial and Quantitative Analysis*, pages 535–551, 1993.

P. R. Lajbcygier and J. T. Connor. Improved option pricing using artificial neural networks and bootstrap methods. *International Journal of Neural Systems*, 8(4):457–471, 1997.

O. Ledoit and M. Wolf. Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection. *Journal of Empirical Finance*, 10(5):603–621, 2003.

O. Ledoit and M. Wolf. Honey, I Shrunk the Sample Covariance Matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004.

H.-T. Lee and J. K. Yoder. A bivariate Markov regime switching GARCH approach to estimate time varying minimum variance hedge ratios. *Applied Economics*, 39(10):1253–1265, 2007a.

H.-T. Lee and J. K. Yoder. Optimal hedging with a regime-switching time-varying correlation GARCH model. *Journal of Futures Markets*, 27(5):495–516, 2007b.

H. Levy and K. C. Lim. Forward exchange bias, hedging and the gains from international diversification of investment portfolios. *Journal of International Money and Finance*, 13(2):159–170, 1994.

J. Li. Sparse and stable portfolio selection with parameter uncertainty. *Journal of Business & Economic Statistics*, 33(3):381–392, 2015.

S. Liu, A. Borovykh, L. A. Grzelak, and C. W. Oosterlee. A neural network-based framework for financial model calibration. *Journal of Mathematics in Industry*, 9, 2019a.

S. Liu, C. Oosterlee, and S. Bohte. Pricing options and computing implied volatilities using neural networks. *Risks*, 7(1):1–22, 2019b.

F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: A simple least-squares approach. *The Review of Financial Studies*, 14(1):113–147, 2001.

H. Lu, Zhou; Pu, F. Wang, Z. Hu, and L. Wang. The expressive power of neural networks: A view from the width. *NIPS 2017*, 2017.

H. Lustig, N. Roussanov, and A. Verdelhan. Common risk factors in currency markets. *The Review of Financial Studies*, 24(11):3731–3777, 2011.

F. Maccheroni, M. Marinacci, and D. Ruffino. Alpha as ambiguity: Robust mean–variance portfolio analysis. *Econometrica*, 81(3):1075–1113, 2013.

D. B. Madan and E. Seneta. The variance gamma (V.G.) model for share market returns. *The Journal of Business*, 63(4):511–524, 1990.

J. Madura and W. Reiff. A hedge strategy for international portfolios. *The Journal of Portfolio Management*, 12(1):70–74, 1985.

M. Maggiori, B. Neiman, and J. Schreger. International currencies and capital allocation. *Journal of Political Economy*, 128(6):2019–2066, 2020.

M. Malliaris and L. M. Salchenberger. A neural network model for estimating option prices. *Applied Intelligence*, 3(3):193–206, 1993.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

G. Maruyama. Continuous Markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo*, 4(48):276–292, 1955.

T. Maurer, T.-D. Tô, and N.-K. Tran. Pricing risks across currency denominations. *Management Science*, 65(11):5308–5336, 2019.

A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton University Press, 2015.

G. Meissner and N. Kawano. Capturing the volatility smile of options on high-tech stocks—a combined GARCH-neural network approach. *Journal of Economics and Finance*, 25(3):276–292, 2001.

M. Mhiri and J.-L. Prigent. International Portfolio Optimization with Higher Moments. *International Journal of Economics and Finance*, 2(5):157–169, 2010.

R. O. Michaud. The Markowitz Optimization Enigma: Is 'Optimized' Optimal? *Financial Analysts Journal*, 45(1):31–42, 1989.

M. J. Morelli, G. Montagna, O. Nicrosini, M. Treccani, M. Farina, and P. Amato. Pricing financial derivatives with neural networks. *Physica A: Statistical Mechanics and its Applications*, 338(1–2):160–165, 2004.

C. J. Neely, P. A. Weller, and J. M. Ulrich. The Adaptive Markets Hypothesis: Evidence from the Foreign Exchange Market. *Journal of Financial and Quantitative Analysis*, 44(2):467–488, 2009.

S. Nestler and A. Hall. The variance gamma distribution. *Significance*, 16(5):10–11, 2019.

V.-A. Nguyen, S. Shafieezadeh-Abadeh, D. Filipović, and D. Kuhn. Mean-covariance robust risk measurement. *Swiss Finance Institute Research Paper No. 21-93*, 2021.

W. Opie and S. J. Riddiough. Global currency hedging with common risk factors. *Journal of Financial Economics*, 136(3):780–805, 2020.

M. S. Paolella and P. Polak. COMFORT: A common market factor non-Gaussian returns model. *Journal of Econometrics*, 187(2):593–605, 2015a.

M. S. Paolella and P. Polak. ALRIGHT: Asymmetric LaRge-scale (I)GARCH with Hetero-Tails. *International Review of Economics & Finance*, 40:282–297, 2015b.

M. S. Paolella and P. Polak. COBra: Copula-based portfolio optimization. In *Predictive Econometrics and Big Data, TES 2018*, volume 753 of *Studies in Computational Intelligence*, pages 36–77. Springer, Cham, 2018.

M. S. Paolella, P. Polak, and P. S. Walker. Regime switching dynamic correlations for asymmetric and fat-tailed conditional returns. *Journal of Econometrics*, 213(2):493–515, 2019.

M. S. Paolella, P. Polak, and P. S. Walker. A non-elliptical orthogonal GARCH model for portfolio selection under transaction costs. *Journal of Banking & Finance*, 125:106046, 2021.

L. Pástor. Portfolio selection and asset pricing models. *The Journal of Finance*, 55(1):179–223, 2000.

L. Pástor and R. F. Stambaugh. Comparing asset pricing models: an investment perspective. *Journal of Financial Economics*, 56(3):335–381, 2000.

L. Pástor and R. F. Stambaugh. Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics*, 63(3):315–349, 2002.

L. Pástor and R. F. Stambaugh. Predictive systems: Living with imperfect predictors. *The Journal of Finance*, 64(4):1583–1628, 2009.

A. F. Perold and E. C. Schulman. The free lunch in currency hedging: Implications for investment policy and peformance standards. *Financial Analysts Journal*, 44(3):45–50, 1988.

G. C. Pflug, A. Pichler, and D. Wozabal. The $1/n$ investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance*, 36(2):410–417, 2012.

M. M. Pires and T. Marwala. American option pricing using multi-layer perceptron and support vector machine. *In 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, 2:1279–1285, 2004.

P. Polak and U. Ulrych. Dynamic Currency Hedging with Non-Gaussianity and Ambiguity. *Swiss Finance Institute Research Paper No. 21-60*, 2021.

R. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2(3):21–41, 2000.

T. Roncalli. *Introduction to Risk Parity and Budgeting*. CRC Press, 2013.

J. Ruf and W. Wang. Neural networks for option pricing and hedging: A literature review. *Journal of Computational Finance*, 24(1):1–45, 2020.

S. Scheidegger and A. Treccani. Pricing American options under high-dimensional models with recursive adaptive sparse expectations. *Journal of Financial Econometrics*, 19(2):258–290, 2021.

D. Schmeidler. Subjective probability and expected utility without additivity. *Econometrica: Journal of the Econometric Society*, pages 571–587, 1989.

J. M. Schmittmann. Currency hedging for international portfolios. *International Monetary Fund Working Paper*, 10/151, 2010.

H. J. Shin and J. Ryu. A dynamic hedging strategy for option transaction using artificial neural networks. *International Journal of Software Engineering and its Applications*, 6(4):111–116, 2012.

J. J. Siegel. Risk, interest rates and the forward exchange. *The Quarterly Journal of Economics*, pages 303–309, 1972.

B. Solnik. Currency hedging and Siegel's paradox: On Black's universal hedging rule. *Review of International Economics*, 1(2):180–187, 1993.

B. H. Solnik. An equilibrium model of the international capital market. *Journal of economic theory*, 8(4): 500–524, 1974.

J. D. Spiegeleer, D. B. Madan, S. Reyners, and W. Schoutens. Machine learning for quantitative finance: fast derivative pricing, hedging and fitting. *Quantitative Finance*, 18(10):1635–1643, 2018.

R. F. Stambaugh. Predictive regressions. *Journal of Financial Economics*, 54(3):375–421, 1999.

M. Stinchcombe and H. White. Universal approximation using feedforward networks with non-sigmoid hidden layer activation functions. *Proceedings of the International Joint Conference on Neural Networks*, pages 613–618, 1989.

K.-I. Suzuki, H. Konno, and M. Morjiri. An International Portfolio Optimization Model Hedged with Forward Currency Contracts. *Financial Engineering and the Japanese Markets*, 4(3):275–286, 1997.

R. Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

A. Timmermann. *Chapter 4: Forecast Combinations.* Handbook of Economic Forecasting, Vol. 1. Elsevier, 2006.

W. H. Tong. An examination of dynamic hedging. *Journal of International Money and Finance*, 15(1): 19–35, 1996.

N. Topaloglou, H. Vladimirou, and S. A. Zenios. CVaR models with selective hedging for international asset allocation. *Journal of Banking & Finance*, 26(7):1535–1561, 2002.

J. Tu. Is regime switching in stock returns important in portfolio decisions? *Management Science*, 56(7): 1198–1215, 2010.

J. Tu and G. Zhou. Incorporating economic objectives into Bayesian priors: Portfolio choice under parameter uncertainty. *Journal of Financial and Quantitative Analysis*, pages 959–986, 2010.

U. Ulrych and N. Vasiljević. Ambiguity, Optimal Currency Overlay, and Home Currency Bias. *Swiss Finance Institute Research Paper No. 20-73*, 2020.

W. N. van Wieringen. Lecture notes on ridge regression, Version May 2021. Vrije Universiteit Amsterdam.

A. Verdelhan. The share of systematic variation in bilateral exchange rates. *The Journal of Finance*, 73(1): 375–418, 2018.

J. A. Wachter and M. Warusawitharana. Predictable returns and asset allocation: Should a skeptical investor time the market? *Journal of Econometrics*, 148(2):162–178, 2009.

Y. Wang, C. Wu, and L. Yang. Hedging with futures: Does anything beat the naïve hedging strategy? *Management Science*, 61(12):2870–2889, 2015.

K. P. Wong. Cross-hedging ambiguous exchange rate risk. *The Journal of Futures Markets*, 37(2):132–147, 2017.

T. Ye and L. Zhang. Derivatives pricing via machine learning. *Journal of Mathematical Finance*, 9(3): 561–589, 2019.

Part III

Curriculum Vitae

Curriculum Vitae

Personal Details:

Name	Urban Ulrych
Date of Birth	03 August 1992
Place of Birth	Celje, Slovenia

Education:

September 2017 – April 2022:	Swiss Finance Institute Ph.D. Program in Finance Department of Banking and Finance, University of Zurich Zurich, Switzerland
September 2014 – July 2017:	Master of Science in Quantitative Finance Department of Banking and Finance, University of Zurich Department of Mathematics, ETH Zurich Zurich, Switzerland
October 2011 – September 2014:	Bachelor in Financial Mathematics Faculty of Mathematics and Physics, University of Ljubljana Ljubljana, Slovenia

Professional Experience:

Since August 2021:	Quantitative Strategist Credit Suisse AG, Investment Bank Zurich, Switzerland
March 2017 – September 2017:	Quantitative Analyst ETH Juniors Zurich, Switzerland
February 2016 – November 2016:	Analytics Strategist Credit Suisse AG, Investment Bank Zurich, Switzerland